languages

Special Issue Reprint

Current Issues in Vietnamese Linguistics

Edited by
Trang Phan, Nigel Duffield and Tim Chou

mdpi.com/journal/languages

MDPI

Current Issues in Vietnamese Linguistics

Current Issues in Vietnamese Linguistics

Guest Editors

Trang Phan
Nigel Duffield
Tim Chou

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors

Trang Phan
Department of Asian and North African Studies
Ca' Foscari University of Venice
Venice
Italy

Nigel Duffield
Faculty of Letters
Konan University
Kobe
Japan

Tim Chou
Department of Foreign Languages & Literatures
National Yang Ming Chiao Tung University
Hsinchu
Taiwan

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Languages* (ISSN 2226-471X), freely accessible at: https://www.mdpi.com/journal/languages/special_issues/NW77923508.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. *Journal Name*, *Volume Number*, Page Range.

ISBN 978-3-7258-6191-0 (Hbk)
ISBN 978-3-7258-6192-7 (PDF)
https://doi.org/10.3390/books978-3-7258-6192-7

© 2026 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editors . . . . . **vii**

Preface . . . . . **ix**

Trang Phan, Nigel Duffield and Chao-Ting Tim Chou
Current Issues in Vietnamese Linguistics: An Introduction
Reprinted from: *Languages* **2025**, *10*, 293, https://doi.org/10.3390/languages10120293 . . . . . **1**

Mark Alves
An Updated Overview of the Austroasiatic Components of Vietnamese
Reprinted from: *Languages* **2024**, *9*, 377, https://doi.org/10.3390/languages9120377 . . . . . **5**

Giorgio Francesco Arcodia and Trang Phan
The Sino-Vietnamese Negative Prefixes *bất*, *vô*, *phi* and Their Coexistence with Sentential Negators: A Synchronic and Diachronic Analysis
Reprinted from: *Languages* **2025**, *10*, 146, https://doi.org/10.3390/languages10060146 . . . . . **28**

Tue Trinh
On Similative Reduplication in Vietnamese
Reprinted from: *Languages* **2025**, *10*, 150, https://doi.org/10.3390/languages10070150 . . . . . **60**

Yi-Ling Irene Liao and Tzong-Hong Jonah Lin
"Again" and "Again": A Grammatical Analysis of *lại* and *nữa* in Vietnamese
Reprinted from: *Languages* **2025**, *10*, 18, https://doi.org/10.3390/languages10020018 . . . . . **74**

Trần Phan and Chao-Ting Tim Chou
Locative Inversion in Vietnamese
Reprinted from: *Languages* **2025**, *10*, 50, https://doi.org/10.3390/languages10030050 . . . . . **92**

Nigel Duffield
Whys and Wherefores: The Aetiology of the Left Periphery (With Reference to Vietnamese)
Reprinted from: *Languages* **2025**, *10*, 116, https://doi.org/10.3390/languages10050116 . . . . . **120**

Andrew Simpson and Linh Pham
Reflexivization and Mình-Exceptional Local Binding by a Monomorphemic Anaphor?
Reprinted from: *Languages* **2025**, *10*, 44, https://doi.org/10.3390/languages10030044 . . . . . **169**

Andrew Simpson and Linh Pham
The Syntax of Null Possessors with Kinship Terms and Body Part Nouns in Vietnamese
Reprinted from: *Languages* **2025**, *10*, 158, https://doi.org/10.3390/languages10070158 . . . . . **185**

Michael Yoshitaka Erlewine and Anne Nguyen
Varieties of Polar Question Bias: Lessons from Vietnamese
Reprinted from: *Languages* **2025**, *10*, 238, https://doi.org/10.3390/languages10090238 . . . . . **208**

Tue Trinh
Partition by Exhaustification and Polar Questions in Vietnamese
Reprinted from: *Languages* **2025**, *10*, 233, https://doi.org/10.3390/languages10090233 . . . . . **221**

About the Editors

Trang Phan

Trang Phan (Department of Asian and North African Studies, Ca' Foscari University of Venice) investigates Vietnamese syntax—its cross-linguistic contrasts, diachronic development, and acquisition by learners from diverse linguistic backgrounds. She is the author of the monograph *The Syntax of Vietnamese Tense, Aspect, and Negation* (Routledge, 2024) and the primary author of three trilingual (Vietnamese–English–Italian) textbooks entitled *Vietnamese Vibes 1, 2, 3* (Cafoscarina Press, 2024). Phan co-edited several edited volumes including *Interdisciplinary Perspectives on Vietnamese Linguistics* (John Benjamins, 2019), *Vietnamese Linguistics: State of the Field* (University of Hawai'i Press, 2022), *Progress in Vietnamese Linguistics* (Taiwan Journal of Linguistics, 2024), and *Studies in Vietnamese Historical Linguistics: Southeast and East Asian Contexts* (Springer Nature, 2024).

Nigel Duffield

Nigel Duffield (Department of English and American Literature and Language, Konan University, Japan) pursues research in the areas of generative syntax, with particular reference to Vietnamese grammar, and psycholinguistics, including both acquisition and language processing. He is the author of two books, *Particles and Projections in Irish Syntax* (Kluwer 1995), *Reflections on Psycholinguistic Theories* (CUP 2018), and the co-editor (with Trang Phan and others) of a number of edited volumes on Vietnamese, including *Interdisciplinary Perspectives on Vietnamese Linguistics* (John Benjamins, 2019) and *Progress in Vietnamese Linguistics* (Taiwan Journal of Linguistics, 2024). In addition, he has published a significant number of articles on Vietnamese (*Linguistics, Lingua, Journal of East Asian Linguistics*) and psycholinguistics (*Journal of Linguistics, Language Acquisition, Second Language Research, Studies in Second Language Acquisition, Cognitive Linguistics, Language and Cognitive Processes*), plus numerous chapters in edited volumes.

Tim Chou

Tim Chou (Department of Foreign Languages and Literatures, National Yang Ming Chiao Tung University, Taiwan) investigates theoretical syntax and syntax-semantics/pragmatics interfaces. He has published papers on hyper-raising, composite probing, locative inversion, and logophoric binding in Vietnamese.

Preface

This Reprint arises from the International Symposium on Vietnamese Linguistics (ISVL), an ongoing forum dedicated to advancing the study of the Vietnamese language in its historical depth and contemporary diversity. The papers collected here were first presented at the 4th and 5th ISVL workshops, held in Hsinchu in 2024 and Venice in 2025, and have been revised for publication.

The scope of this Reprint is intentionally focused. Rather than providing a broad survey, it offers a series of targeted studies that examine inheritance and contact in Vietnamese, the structure of the clause and the left periphery, reduplication and semantic interpretation, reference and binding, and the pragmatics of particles and polar questions. The aim is to highlight new empirical findings and to situate them within broader debates in typology, historical linguistics, and formal theory.

The motivation for preparing this Reprint is twofold. First, Vietnamese continues to serve as a testing ground for key questions about language change, clause architecture, and the interface between syntax, semantics, and pragmatics. Second, the ISVL series seeks to make this work more visible to an international readership, promoting exchange between specialists in Vietnamese and the wider community of linguists.

This Reprint is addressed to researchers, teachers, and students of linguistics who are interested in Vietnamese, Southeast Asian languages, and comparative approaches to grammar and meaning. It is our hope that the contributions gathered here will inform future scholarship and stimulate further discussion across linguistic traditions.

Trang Phan, Nigel Duffield, and Tim Chou
Guest Editors

languages

MDPI

Editorial

Current Issues in Vietnamese Linguistics: An Introduction

Trang Phan [1,*], Nigel Duffield [2] and Chao-Ting Tim Chou [3,*]

[1] Department of Asian and North African Studies, Ca' Foscari University of Venice, 30123 Venice, Italy
[2] Faculty of Letters, Konan University, Kobe 658-8501, Japan; duffield@konan-u.ac.jp
[3] Department of Foreign Languages & Literatures, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
* Correspondence: thihuyentrang.phan@unive.it (T.P.); ctchou@nycu.edu.tw (C.-T.T.C.)

This Special Issue of *Languages*, part of the ISVL (International Symposium on Vietnamese Linguistics) series, presents revised papers from the 4th and 5th ISVL workshops, hosted by National Yang Ming Chiao Tung University in 2024 and Ca' Foscari University of Venice in 2025. Earlier workshops were held at the Harvard-Yenching Institute (2021), Konan University/Vietnam National University, Hanoi (2022), and Konan University (2023), with related volumes published in Phan et al. (2022), Duffield and Phan (2024), and Phan and Duffield (2025). The aim here is modest: to present focused case studies that shed light on specific aspects of Vietnamese grammar and its history. The contributions cover topics including inheritance and contact, reduplication, clause structure, reference and binding, and the semantics–pragmatics of particles. This is not intended as a comprehensive survey, but rather as a concise set of results that together point toward the next stage of ISVL research.

In assembling this issue, we also aim to demonstrate some ways in which Vietnamese evidence bears on active debates in typology, historical change, clause architecture, and the semantics–pragmatics interface, providing analyses that can be tested and compared beyond the Vietnamese domain. Vietnamese offers a compact testbed for questions central to current theory: how contact reshapes inherited structure; how a particle-rich, morphologically light system encodes scope, focus, and polarity; and how left-peripheral layering interacts with binding and reference in a language without inflectional reflexes. The papers in this Reprint contribute new data and analyses that speak directly to these issues, refining models of negation, question bias, movement vs. topicalization, and the division between anaphoric and pronominal null elements (Rizzi, 1997; Cinque, 1999; Reinhart & Reuland, 1993; von Stechow, 1996; Cole et al., 2000; Romero, 2024).

Three main question threads run through the volume:

(i) Inheritance under contact: How much of present-day Vietnamese grammar is Austroasiatic/Vietic in origin, how much is Sinitic or calqued through literacy, and how much reflects area-wide convergence?
(ii) Architecture: Where does Vietnamese project structure in the middle and left periphery? How does information structure map to those positions, and how A/A′ movement and locality should be modeled?
(iii) Interpretation: How are repetitive, restitutive, incremental, and bias meanings compositionally derived in a particle-rich, morphologically light system?

Across these themes, the papers offer a tighter account of Vietic inheritance and contact pathways; clearer predictions for the distribution of negation and polar-question forms; and parameter settings that distinguish Vietnamese from close comparators such as Mandarin while capturing internal variation.

Taken together, the contributions move from history to structure to interpretation. The diachronic studies in the Diachrony, Contact, and Inheritance subsection set the baseline for what is inherited vs. contact-driven; the semantics papers in the subsection Morphology–semantics interfaces and lexical pragmatics show how small positional changes yield distinct meanings; the structural papers in the subsection Argument structure, Movement, and the Clausal Spine locate the landing sites that make those meanings possible; the reference and polarity papers in the subsections Reference, Binding, and Null elements and Polar questions and Bias test these structures with binding and question bias diagnostics.

(i) Diachrony, contact, and inheritance

Mark Alves re-examines the Austroasiatic core of Vietnamese using enhanced Vietic reconstructions and a more precise view of Sinitic contact. He shows that native etyma and structural templates remain visible in the lexicon and in patterns of change that are not reducible to borrowing. This supports a cleaner partition between inherited structure and contact-driven remodeling, and it narrows what should count as "Sinitic" given the timing and channels of influence. This diachronic foundation constrains later analyses of negation, particles, and movement options.

Giorgio Francesco Arcodia and Trang Phan focus on the negative prefixes *bất*, *vô*, and *phi* and their interaction with clausal negators. Using a parallel set of *Analects* translations (17th, 19th, 21st centuries; *nôm* and *quốc ngữ*), they trace a non-linear trajectory: early prominence of native *chẳng*; 19th-century rise of *không* with selective prefixal use; and modern dominance of *không* alongside renewed use of *bất* and *vô* in literate and technical registers. The authors link this pattern to modern East Asian literacy and the spread of Chinese-based vocabulary, yielding a layered model of Vietnamese negation where register, position, and history matter. This model explains why derivational and sentential negation can co-exist and sets expectations for particle distribution elsewhere in the volume.

(ii) Morphology–semantics interfaces and lexical pragmatics

Tue Trinh analyzes similative reduplication with the rime *-iếc*. The construction is productive, produces "similar but not identical" readings, and tends to trigger ignorance-style inferences. Trinh models it with alternatives akin to disjunction, deriving the ignorance effect from how similarity is evaluated. The observed incompatibility with classifiers follows if *-iếc* relies on context-based similarity, while classifier constructions rely on grammatically fixed comparisons. These results show how Vietnamese encodes graded similarity without morphology, anticipating structural constraints discussed in the subsection Argument structure, movement, and the clausal spine.

Yi-Ling Irene Liao and Tzong-Hong Jonah Lin revisit *lại* and *nữa*. Preverbal *lại* can be repetitive or restitutive; postverbal *lại* is restitutive only; *nữa* supports both repetitive and incremental ("add more") readings. They derive these contrasts by adjunction height and focus association (*v*P vs. VP) and by movement that yields the surface-final position of *nữa*. The account links word order to "again" semantics and provides diagnostics for attachment height in the verbal domain. This directly interfaces with the landing sites identified in Argument structure, movement, and the clausal spine and with bias-related contrasts in Polar questions and bias.

(iii) Argument structure, movement, and the clausal spine

Trần Phan and Chao-Ting Tim Chou separate two superficially similar patterns with clause-initial locatives. In the "locative" construction, the initial phrase is an argument of a locative predicate; in the "existential" construction, it is an adjunct that can be optionally topicalized. Only the existential type allows independent topicalization. They argue for a high, hybrid A/A′ landing site in the TP domain, where a fronted locative competes with

an expletive for the same specifier. A composite featural probe captures the movement profile and supports a featural approach to the A/A′ distinction. The identified high-TP position helps locate where focus-sensitive operators can take scope.

Nigel Duffield maps the embedded left periphery, addressing the position of all pre-subject elements: subordinators, embedded topics, fronted quantifiers and adverbials, as well as the Vietnamese counterpart of English "why." Whilst the results of the study broadly align with the previous cartographic proposals, they also suggest revisions, especially with respect the "why-to" restriction, and the position of fronted quantifier expressions The Vietnamese findings urge a reconsideration of earlier claims based on Italian data, especially concerning the motivation for fronting to the left periphery. This mapping provides the structural backdrop for the polarity and bias results in Polar questions and bias.

(iv) Reference, binding, and null elements

Andrew Simpson and Linh Pham test the behavior of the anaphoric element mình. Based on controlled judgments across regions and on diagnostics with ellipsis and quantificational subjects, they show that mình in object position can be locally bound without the reflexivizer *tự*. The pattern reflects genuine binding rather than accidental coreference. This challenges approaches that tie local reflexivization to morphological complexity and implies a recalibration of anaphor typology. The findings refine predictions about where reflexivizers are required in Vietnamese and related languages.

In a second paper, Simpson and Pham analyze bare kinship terms (KNs) and body-part nouns (BPNs). KNs behave as if they project a null possessor that is anaphoric and subject to locality, while BPNs pattern with a null pronominal (pro). This split differs from previous proposals for Mandarin and supports a parameter that distinguishes anaphoric from pronominal null elements in the nominal domain. This nominal-domain contrast complements the clausal distinctions drawn in Argument structure, movement, and the clausal spine.

(v) Polar questions and bias

Michael Yoshitaka Erlewine and Anne Nguyen examine Vietnamese polar-question constructions, with particular attention to the final particle *à*. They argue for a three-way distinction: original bias (speaker beliefs), contextual bias (evidence in discourse), and projected bias (constraints encoded by the form). Some bias effects are argued to follow from pragmatic competition among question types rather than from belief states alone. This analysis is related to the structural observations in the subsection Argument structure, movement, and the clausal spine and the focus-related patterns discussed in the subsection Morphology–semantics interfaces and lexical pragmatics, placing the Vietnamese data in dialogue with broader cross-linguistic studies of question bias.

A companion note by Tue Trinh proposes a bi-clausal architecture for polar questions. Together with Partition by Exhaustification and independently motivated constraints, this analysis explains contrasts involving definite vs. quantificational subjects, plain vs. *only*-focused subjects, and the placement of *chắc chắn* 'certainly' relative to verum focus. The proposal links form, focus, and discourse control in a way that predicts observed distributions and blocks unattested ones.

Together, these ten papers provide detailed descriptions and empirically testable analyses. They make a clear contribution to an agenda: to show that how findings from Vietnamese can drive theoretical development; that its empirical particularities (inheritance under contact, particle-rich compositionality, and flexible but rule-governed clausal architecture) force revisions to received views in historical linguistics, cartography, binding theory, and the semantics–pragmatics interface. We anticipate that the volume will serve as a reference point for future work, within Vietnamese linguistics and beyond.

In terms of limitations, the Special Issue does not aim to cover phonetics and phonology, prosody, dialectology, sociolinguistic variation, acquisition/processing, or computational modeling. Some conclusions rest on specialized historical corpora and controlled judgment tasks that merit replication on larger, regionally balanced samples and across genres.

Future research would benefit from (i) a balanced diachronic and synchronic corpora spanning regions and registers; (ii) integrated prosody–syntax studies to refine left-peripheral mapping; (iii) expanded experimental work on bias, focus, and anaphora; and (iv) comparative studies across East and Mainland Southeast Asia to separate areal from language-specific effects. We also encourage open diagnostics and datasets to support replication and comparison.

Author Contributions: For academic purposes, authorship of the Introduction is as follows: T.P. is responsible for the Introduction, Diachrony, contact, and inheritance, and Morphology–semantics interfaces and lexical pragmatics, N.D. for Argument structure, movement, and the clausal spine and Reference, binding, and null elements; and C.-T.T.C. for Polar questions and bias and Conclusions. All authors have read and agreed to the published version of the manuscript.

Funding: Nigel Duffield's contribution to this project was partially supported by a KAKEN grant from the Japan Society for the Promotion of Science (21K00538 'Progress in Vietnamese Linguistics', 2021–2024). Tim Chou gratefully acknowledges the research grant MOST 110-2410-H-A49-013-MY2, provided by the National Science and Technology Council in Taiwan.

Acknowledgments: We would like to thank all contributors and reviewers for their valuable work. Trang Phan acknowledges the support of the Department of Asian and North African Studies and the Marco Polo Centre for Global Europe–Asia Connections, Ca' Foscari University of Venice, for sponsoring the 2025 workshop in Venice.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Cole, P., Hermon, G., & Huang, C. J. (Eds.). (2000). *Long distance reflexives*. Brill.

Duffield, N., & Phan, T. (Eds.). (2024). Progress in Vietnamese linguistics. *Introduction, Taiwan Journal of Linguistics*, *22*(1). [CrossRef]

Phan, T., & Duffield, N. (Eds.). (2025). Further advances in Vietnamese linguistics. *Taiwan Journal of Linguistics*, *23*(3). Available online: http://hdl.handle.net/10524/52500 (accessed on 11 November 2025).

Phan, T., Phan, J., & Alves, M. (Eds.). (2022). *Vietnamese linguistics: State of the field* (Special publication of *Journal of the Southeast Asian Linguistics Society* no. 9). University of Hawai'i Press.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, *24*(4), 657–720.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Kluwer.

Romero, M. (2024). Biased polar questions. *Annual Review of Linguistics*, *10*, 279–302. [CrossRef]

von Stechow, A. (1996). The different readings of *wieder* 'again': A structural account. *Journal of Semantics*, *13*(2), 87–138. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

An Updated Overview of the Austroasiatic Components of Vietnamese

Mark Alves

Department of ELAP, Linguistics, and Communication Studies, Montgomery College, Rockville, MD 20850, USA; mark.alves@montgomerycollege.edu

Abstract: This article presents an updated view of the language history of Vietnamese from its native Austroasiatic roots, including key historical phonological, morphological, and syntactic features and developments; a characterization of its Austroasiatic etyma; and the context of this information in Vietnamese linguistic ethnohistory. It is now possible to make better supported claims and more precise characterizations due to improved understanding of the history of Austroasiatic and Vietic and their reconstructions, the nature and effect of language contact with Chinese, and the process of typological convergence of the ancestral language of Vietnamese. This study shows that, while Vietnamese is not a typologically characteristic Austroasiatic language, the Austroasiatic components of the Vietnamese lexicon and linguistic structure are more prominent than previously supposed.

Keywords: historical linguistics; language contact; Austroasiatic; Vietic; Vietnamese

1. Introduction

This study highlights the Austroasiatic lexical and structural elements (phonology, morphology, syntax) in modern Vietnamese and characterizes Vietnamese language history in terms of its Austroasiatic roots. This can be accomplished via the use of modern digital resources and tools, substantive comparative data from both modern languages and proto-language reconstructions, and recent ethnohistorical research in the region.

The Austroasiatic origin of Vietnamese has long been recognized, but there was considerable debate in the first half of the 20th century. Uncertainty about the linguistic affiliation of Vietnamese was due to (a) the limited amount of available comparative data, (b) what is now understood to be major changes in phonological word structure from Austroasiatic to Vietnamese, (c) the complex language contact situation in Mainland Southeast Asia over past millennia, (d) the intense lexical influence of Chinese on Vietnamese and neighboring languages, and (e) the typological resemblance of Vietnamese to Chinese and Tai languages, among other factors. Even though the Austroasiatic origin of Vietnamese was widely accepted, there has been considerable focus in linguistic literature on the Chinese influence on Vietnamese and less detail about the Austroasiatic elements. It is only in the 21st century that enough digital lexical resources and proto-language reconstructions (and crucially, advances in reconstructions of Chinese, Proto-Austroasiatic, and Proto-Vietic), combined with ethnohistorical and archaeological information, have become available to create a clearer picture of both the retentions of and changes to the original Austroasiatic components of Vietnamese.

The Chinese typological and lexical features of Vietnamese are a highly visible aspect of Vietnamese. Vietnamese has significant typological similarity with Chinese in its tone system, isolating morphology, vocabulary, and disyllabic lexical compounding. Also, Chinese loanwords in Vietnamese, which number in the thousands, are an essential part of Vietnamese etymological study and are seen in its literary tradition and cultural matters. However, despite the effect of two thousand years of language contact with Sinitic, the similarities should not be overstated. First, the typological convergence leading to the

modern template happened over a period stretching from the beginning of the 1st millennium CE to the 1800s: this was not a sudden event. Next, many native Austroasiatic and Vietic features have been retained, while conversely, many features seen among Chinese languages did not become part of Vietnamese typology (e.g., the A-not-A question pattern, pre-nominal adjectival modifiers, post-nominal possessive markers, etc.). Also, though the lexical impact of Chinese is statistically significant, Vietnamese still has a robust core native Austroasiatic and Vietic component. However, the latter aspect has been difficult to assess, as previously, knowledge of Austroasiatic language history was less developed, and no reliable Proto-Austroasiatic (hereafter pAA) reconstructions were available.

The primary goal of this article is to characterize the historical changes from Austroasiatic to Vietnamese at primary linguistic levels: phonological, morphological, and syntactic. An attempt is made to identify features that are considered retentions of Austroasiatic, while putting aside those which may arguably be considered the result of areal contact. Despite the Chinese-like features of Vietnamese, Vietnamese still has core Austroasiatic elements at all linguistic levels, some of which can only be seen through reconstructions, but others which are much like pAA or any other modern Austroasiatic language.

2. Overview of Vietnamese Language History

To provide context for the historical linguistic hypotheses in this study, this section summarizes past stages of the language situations of Vietnamese supported by ethnohistorical and archaeological information. Table 1 presents the posited historical linguistic periods, their approximate timing, and hypothesized stages of typological characterizations from Austroasiatic to modern Vietnamese.

Table 1. Tentative stages from Austroasiatic to modern Vietnamese.

Period	Chronological Points of Reference	Hypothesized Typological Traits
Proto-Austroasiatic	c. 2000 BCE, pre-Metal Age, dispersal of Austroasiatic peoples in Mainland Southeast Asia	Disyllabic, nontonal, CCVC syllables
Proto-Vietic	c. 1000 BCE, Metal Age, settled Austroasiatic groups	Disyllabic, nontonal, CCVC syllables, emerging phonation[1] (?)
Heavy Sinitic–Vietic contact	First half of 1st mill. CE, Chinese administration	Disyllabic, nontonal, CCVC syllables, emerging north–south speciation in Vietic (?)
Proto-Viet-Muong	By 1000 CE, end of Chinese administration, start of local administrations	Beginning of reduction in disyllabicity (?), CCVC syllables, tonal (pre-register)
Archaic Vietnamese	Early 2nd mill. CE, use of nativized Nôm script	Reduced disyllabicity, CCVC syllables, tonal (post-register (?))
Middle Vietnamese	Mid-2nd mill. CE to the early 1800s, increased regional trade and contact	Monosyllabic, CCVC with progressive loss of onset clusters, tonal
Modern Vietnamese	Late 2nd mill. CE, Late Colonial period	Monosyllabic, CGVC, tonal

Note: (?) indicates that a feature cannot be demonstrated with high certainty based on comparative data but is feasible based on the larger context.

Table 1 begins with the Austroasiatic dispersal,[2] over 4000 years ago, at which point the ancestral language of Vietnamese would have been a typical Austroasiatic language, with sesquisyllabic word structure (i.e., disyllabic words with an iambic stress pattern) and onset clusters with medial -l- and -r-, but no tones. At some point after the dispersal, the Vietic group speciated but kept the Austroasiatic morphophonological template, like other modern conservative Vietic languages. The beginning of the Metal Age in the region in c. 1000 BCE is noted as a chronological point of reference, as it was long after the dispersal, enough for linguistic differentiation, but at a time when presumably there was contact with other groups and potential sociocultural developments and corresponding linguistic changes. After contact with the incoming Sinitic ethnolinguistic community around the beginning of the Common Era, there may have been increasing linguistic distinctions between the northeast and southern Vietic speech communities, as suggested by the many early Chinese cultural loanwords present in Vietnamese but not Vietic groups to the south (Sidwell and Alves 2021, p. 184).

The hypothesized lexical differentiation of Vietic sub-branches, with associated cultural implications, precedes further speciation of the northeast group consisting today of Vietnamese and a few dozen lects collectively called Muong.[3] The Viet–Muong languages share lexical and phonological features, including tone system patterns. The timing of tonogenesis in this group is unknown, but Alves (2018, p. 17) has posited that it happened by the end of the 1st millennium CE in Vietnamese, though whether this is also the case in Muong lects has not yet been determined. Also, no analysis has shown a precise phylogenetic tree of this sub-branch, and it has been posited that, considering the lack of shared phonological innovations, no precise tree is even possible (Phan 2012).

Next, at the start of the 2nd millennium, a variety of Chinese of a hypothesized Chinese ethnolinguistic group merged with the Viet–Muong language(s) (Phan 2013). The speech community after that merging can be said to have spoken Archaic Vietnamese. At this point, thousands of Late Middle Chinese words—in both literary and vernacular Vietnamese—became part of Vietnamese and Muong, which in Vietnamese is typically called *từ Hán-Việt* or 'Sino-Vietnamese vocabulary'.

As for the phonological word at that time, it is difficult to reliably determine the Proto-Viet–Muong template given the transitional nature of the language, but there was likely retention of presyllables into the early 2nd millennium. Two sets of evidence for this are (a) lenited onsets in Vietnamese that softened in intervocalic position (discussed in Section 3 before Table 5) and (b) textual evidence from the 1200s suggesting that Archaic Vietnamese still retained presyllables.[4] The Chinese-based Nôm orthography at that time offers evidence of a lingering number of disyllabic words (Shimizu 2015). Regarding the timing of the process of monosyllabification, we can only speculate that there may have been decreasing percentages of disyllabic words over the centuries, but there are insufficient data to offer details. However, by the Middle Vietnamese stage, as exemplified by the 1651 Romanized Vietnamese–Portuguese–Latin dictionary (Rhodes [1651](1991), the process was complete, as Vietnamese had no prosodic disyllabic words. However, it retained some onset clusters with -l- at that time. It is only by the mid-1800s that mainstream Vietnamese had its complete modern template with only medial [-w-] (in CGVC, G refers to a 'glide') (though some rural Vietnamese dialects in north–central Vietnam have been reported to have/tl/onset clusters (Nguyễn 1999)).

Thus, based on available evidence, in this hypothesized scenario, the Vietic ancestor of Vietnamese retained its Austroasiatic template into the late 1st millennium CE, though it had developed tones by that time. Then, at some point in the first half of the 2nd millennium, it completely lost presyllables. Finally, in the 19th century, Vietnamese largely reached its modern largely CVC syllable canonical shape, which resembles the typology that Chinese had developed by the period of Early Middle Chinese in the first half of the 1st millennium CE. Thus, it can be said that only in the past half millennium did Vietnamese take only a truly Chinese-like appearance, and that prior to that, it was still very much an Austroasiatic-like language.

The next sections review the phonological, morphological, and syntactic matters in more detail and connect the Austroasiatic origins to modern Vietnamese features.

3. Historical Phonological Aspects

Thus, Vietnamese has been substantially restructured from the original Austroasiatic template, and today, the phonological typology of Vietnamese resembles that of Chinese in many ways. However, as noted, this situation was only fully reached in the last several hundred years. The Vietnamese syllable is most often a simple CV(C) structure, with only [-w-] as a medial in at most 10 percent of Vietnamese syllables (Kirby and Alves 2022), or as little as 5.2 percent (Trần and Vallée 2009, p. 232). Vietnamese has a complex tone system (i.e., six contour tones with phonation features, while two tones exhibit phonetic distinctions in syllables with p/t/k codas) that maps directly onto the system of the Chinese historical phonological tradition (i.e., the four ping/shang/qu/ru tone categories divided into eight tones by height according to onset voicing). In early 20th century research, this situation created uncertainty regarding its linguistic affiliation, with speculation of Tai or Chinese origins (see Gage 1985 and Alves 2006 for a summary). However, language contact and typological convergence extends beyond Chinese. Vietnamese is in a larger sphere of language contact, sharing the typology of Kra–Dai and Hmong–Mien languages of southern China and Mainland Southeast Asia. These languages are also largely monosyllabic and have typologically comparable tone systems.

Despite linguistic influence due to language contact, we can still identify features that are retentions of Austroasiatic features. Among the pAA etyma in Vietnamese, pAA vowels *i, *a, *u and common pAA consonants *m/n/ɲ/ŋ, * k, and *l/r in onsets of main syllables are readily found retained in modern Vietnamese reflexes. Indeed, some modern Vietnamese words closely or completely resemble their original proto-language source forms (main syllables in cases of pAA disyllabic words), including both pAA and Proto-Vietic (pV hereafter), as in Table 2. Basic vocabulary items such as 'dog', 'bird', 'fish', and others are clearly connected to their pAA counterparts, which have 4000-plus-year histories.

Table 2. Vietnamese words with retentions of pAA main syllable forms.

N. Vietnamese	pV	pAA	Gloss
cɔː^{5} (chó)	*cɔʔ	*cɔʔ (AA030)	dog
ciːm^{1} (chim)	*ciːm	*ciːm (AA016)	bird
kaː^{5} (cá)	*kaːʔ	*kaʔ (AA009)	fish (n)
kɔːn^{1} (con)	*kɔːn	*koan (AA213)	child
mat^{6} (mặt)	*mat	*matʔ (AA024)	eye
muːn^{6} (mụn)	*muːnʔ	*muːnʔ (AA070)	pimple
ɲaː^{2} (nhà)	*ɲaː	*ɲaːʔ (AA110)	house
laː^{5} (lá)	*səˑlaːʔ	*səˑ'laʔ (AA010)	leaf
ɽuːŋ^{6} (rụng)	*ruŋʔ	*ruŋʔ (AA444)	fall, drop (v)
ɽuː^{5} (rú)	*bəˑruːʔ	*bəˑ'ruʔ (AA322)	hill, mountain
ɽuəj^{2} (ruồi)	*Cəˑrɔːj	*roaj (AA028)	fly (n)

Following are notes on data presentation in this article. Subsequent tables with data contain (a) neutral IPA transcriptions of Vietnamese words (i.e., not matching a dialect but rather based on the Vietnamese orthography, following the system of phonetization of (Kirby 2008), followed by Vietnamese Quốc Ngữ romanized orthographic spelling in parentheses, (b) pV reconstructions based on those of Ferlus as in the SEALang Mon-Khmer Etymological Dictionary but modified to parallel the format of the pAA reconstructions, and (c) the pAA reconstructions of Sidwell's (2024) 500-word list with their specified indica-

tors (e.g., AA030) after them. Some potential pAA reconstructions (mostly by Sidwell in the spreadsheet assembled by Sidwell and Alves) that currently lack identifying numbers are marked 'AAP', while others are from Shorto's (2006) reconstructions (marked with # and the reference number) when they are attested in at least several Austroasiatic branches, indicating deep history in Austroasiatic but not necessarily back to the pAA stage.

Naturally, many other pAA segments have changed over time in Vietnamese, but the phonological slots have been retained. The main-syllable onset system has undergone restructuring not only in voicing but also manner of articulation, with some becoming the implosives /ɓ/ and /ɗ/, as in Table 3, or fricatives such as /v/, /z/, or/ ɣ/, as in Table 5, as discussed below.

Table 3. Onset voicing of pAA and corresponding tone height in Vietnamese.

Vietnamese	pV	pAA	AA Gloss
ɓɛː^{3} (bẻ)	*pɛh	*pɛːh (AA188)	break in two
ɗaːn^{1} (đan)	*taːɲ	*taːɲ (AA005)	weave (v)
ciːn^{5} (chín)	*ciːnʔ	*ciːnʔ (AA017)	cooked, ripe
ʔɔːj^{5} (ói)	*ʔoːlʔ	*ʔoːlʲ, *ʔuəl (AA213)	choke, vomit (v.i.)
ɓuːŋ^{6} (bụng)	*buŋʔ	*buŋ, *buuŋʔ (Shorto #624a.b)	belly
ɗoːj^{2} (đồi)	*doːl	*doːl, *duəl (AA423)	hill, mound
caj^{6} (chạy)	*ɟalʔ	*ɟarʔ (AAP)	run (v)
kam^{2} (cằm) 'chin'	*gam (tent.)	*-gaːm (AAP)	molar tooth

Changes in onset voicing, which have occurred in many languages in the region, is seen in pAA and pV etyma in Vietnamese, though the details of this change in Vietnamese are not well understood. Onset voicing is also related to tone height in Vietnamese, as in other neighboring tonal languages: original pAA voiceless stops *p/t/c/k occur in syllables with upper register tones in Vietnamese (1 = the *ngang* tone, 3 = the *hỏi* tone, 5 = the *sắc* tone), while proto-language voiced onsets *b/d/g occur in those with lower register tones (2 = the *huyền* tone, 4 = the *ngã* tone, 6 = the *nặng* tone). For example, pAA *ciːnʔ—'cooked, ripe'—is the source of a modern Vietnamese word with the upper register tone 5, while *ɟarʔ—'run (v)'—is the source of a modern word with the lower register tone 6. While counterexamples exist, and not all details are accounted for, the pattern is reasonably consistent.

Some changes in the onset system have resulted in some original sounds being lost. For instance, pAA implosive onsets have become Vietnamese nasals instead (Ferlus 1992, p. 117), as in Table 4. And yet, though not retained from a proto-language stage, the implosives /ɓ/and/ɗ/ have been revived in the modern Vietnamese phonological inventory by developing from voiceless plosives, as seen in Table 3. Implosives are common in Austroasiatic languages, but the WALS database (Maddieson 2013) shows that such sounds are not as common in languages in other language families in Southern China and Southeast Asia. Thus, we can speculate that there has been a regional typological pressure towards retaining or redeveloping these. In Vietnamese, implosive onsets are distinct from plosive onsets in that (a) they tend to correspond to upper register tones 1, 3, and 5, unlike plosive voiced onsets, which correspond to lower register tones 2, 4, and 6, and (b) they do not undergo lenition to become voiced fricatives in intervocalic position, unlike plosive consonants, as in Table 4.

Table 4. Nasal onsets from pAA implosive *ɓ/ɗ.

Vietnamese	pV	pAA	Gloss
maŋ^{1} (măng)	*tə.ɓaŋ	*tə.ʼɓaŋ (AA023)	bamboo shoots
mɔ:^{3} (mỏ)	*kə.ɓɔh	*cə.ʼɓɔh, *cə.ʼɓuːʔ (AA442)	beak
nɨək^{5} (nước)	*ɗaːk	*ɗaːkʔ (AA039)	water
nuːŋ^{1} (nung)	*ɗuŋ	*ɗɨŋ, *ɗuŋ (AA492)	cook (v)
nɔːŋ1 (nong)	*ɗoːŋ; *ɗoːŋʔ	*kə.ʼɗoːŋ (AA214)	winnowing basket
mɔːk^{5} (móc)	NA	*ɓɔkʔ, *ɓuəkʔ (AA392)	hook (v)
muːk^{5} (múc)	*ɓuːk	*ɓəkʔ (Shorto #367)	draw water

One particularly notable change that makes Vietnamese typologically less like other Austroasiatic or even other Vietic languages is its series of voiced onset fricatives: /v/, /z/, and /ɣ/ (represented in Vietnamese orthography by 'v', 'd' and 'gi' for /z/, and 'g/gh' respectively).[5] These developed from main syllable onsets in intervocalic position (see Ferlus 1992, p. 113 and updates by Alves 2024), as in the examples of Table 5, in which all the reconstructed forms are disyllabic.

Table 5. Voiced fricative onsets in intervocalic position in pAA presyllables.

Vietnamese	pV	pAA	Gloss
vɨən^{6} (vượn)	*kə.waŋʔ	*kwaaŋʔ (Shorto #934)	gibbon
voːj^{1} (vôi)	*kə.puːr	*knpur (Shorto #1636)	lime (mineral)
zaːm^{1} (dam) 'field crab'	*kə.taːm	*kə.ʼtaːm (AA004)	crab
zɔː^{5} (gió)	*kə.jɔːʔ	*kə.ʼjaːl; *kə.ʼjɔʔ (AA075)	wind, air (n)
ɣəw^{5} (gấu)	*cə.guːʔ; cə.kuːʔ	*cguːʔ; ckuːʔ (Shorto #1817)	bear (n)
ɣoːj^{5} (gối)	*tə.kuːlʔ	*Cə.ʼkuːl (AA326)	knee
ɣaːw^{6} (gạo)	*rə.koːʔ	*rəŋ.ʼkoːʔ (AA066)	husked rice

While many instances retroflex onsets in Vietnamese words are retentions of retroflex onsets in Middle Chinese loanwords (Ferlus 1992), in native Vietnamese words, such sounds often correspond to earlier onset clusters, as in Table 6. In a development distinct from the just noted lenition, some of these onset clusters appear to have come from previous presyllables that at some point lost syllable status. There is a regional process in Austroasiatic of presyllable reduction due to the iambic stress pattern weakening presyllables, leading to a variety of manifestations of the so-called sesquisyllabic typology (Pittayaporn 2015).

Table 6. Retroflex onsets from onset clusters.

Vietnamese	pV	pAA	Gloss
ʂaːŋ^{5} 'sáng'	*plaːŋʔ	*plaaŋʔ (NA)	shine (v)
ʂɔːk^{5} 'sóc'	*prɔːk	*pə.ʼrɔːkʔ (AA091)	squirrel
ʂaː^{5} 'sá'	*kraːʔ	*kraʔ (AA162)	path, road
ʈaʲŋ^{1} 'tranh'	*plaŋ	*pə.ʼlaŋ (AA060)	thatch-grass
ʈaːj^{5} 'trái'	*pə.leːʔ > *pleːʔ	*pə.ʼleʔ (AA043)	fruit

The Vietnamese tone system has served as a key example of the historical process of tonogenesis. Haudricourt's (1953) seminal study of the origins of Vietnamese tones as

a seminal study, and that theoretical model has been applied to the study of the emergence of tones in Chinese, Kra–Dai, Hmong–Mien, and Tibeto-Burman languages, as well as other Vietic languages. Among cognates in other Austroasiatic languages, there are correspondences between Austroasiatic coda fricatives *-h/-ɛ and Vietnamese tones 3 or 4 and between final voiceless stops *-p/-t/-k/-ʔ and tones 5 or 6, as exemplified in Table 7. Crucially, Vietnamese has lost final *-h/ɛ and *-ʔ, while reflexes of these segments are found among many Austroasiatic languages. The rephonologization of previous codas as tones in Proto-Viet–Muong was likely stimulated by contact with Chinese. However, tonogenesis in Viet–Muong may also be seen as the result of a larger trajectory predating Sinitic–Vietic language contact. Pharyngeal and registral features and phonation have been seen as stimulants of suprasegmental features in Austroasiatic languages (e.g., Gehrmann 2022), such that tonogenesis could have occurred regardless of contact with Chinese. Nevertheless, the borrowing of numerous Chinese words with matching tones allows for the influence of the Chinese tone system on the Vietnamese one.

Table 7. Tone categories corresponding to pAA coda types.

Vietnamese	pV	pAA	Gloss
taj^{1} (tay)	*si:	*si:ʔ (AAP)	hand, arm
siəŋ^{1} (xương)	*tʃ.ʔa:ŋ	*cə.ʼʔa:ŋ (AA012)	bone
ɗo:j^{2} (đồi)	*do:l	*do:l, *duəl (AA423)	hill, mound
ruəj^{2} (ruồi)	*mə.rɔ:j	*roaj (AA028)	fly (n)
ɗo:^{3} (đổ)	*toh	*toh (AAP)	pour
ɓaj^{3} (bảy)	*pəs	*pɔh, *pəɛ (AA209)	seven
mu:j^{4} (mũi)	*mu:s	*mu:h, *mu:ɛ (AA003)	nose
ɣɔ:^{4} (gõ)	*C.kɔ:h ;*gɔ:h	*guəh ‘knock (v)’ (Shorto #1978)	knock (v)
caw^{5} (cháu)	*cu:ʔ	*cuʔ (AA096)	grandchild
mu:n^{6} (mụn)	*mu:nʔ	*mu:nʔ (AA070)	pimple
kat^{5} (cắt)	*kac	*katʔ (AA186)	cut (v)
ʂɔ:k^{5} (sóc)	*pə.rɔ:k	*pə.ʼrɔ:kʔ (AA091)	squirrel

Overall, while various pAA consonants and vowels have changed in Vietnamese, many have still been retained in their original forms. The later stage of pV phonology also shows examples of changes from pAA phonology, but in many cases, the reconstructions are very close to or the same as the pAA reconstructions, so they also help demonstrate the ultimate pAA origin of these speech sounds and words.

Regarding the influence of Chinese, despite the evidence of its impact on Vietnamese phonology, the native elements of Vietnamese phonology are still evident. In a study of some 8000 Vietnamese phonemically distinct monosyllabic words (with the number including distinctions due to tones), Alves (2022, pp. 122–23) determined that, while native words in Vietnamese have statistically prominent phonological features identifying them, Sino-Vietnamese loanwords (i.e., Late Middle Chinese and the system of readings of Chinese characters) mostly do not. For instance, the onset /ɣ/ ‘g/gh’ occurs only in native Vietnamese words (and very early Chinese loanwords, at least several centuries before the Sino-Vietnamese layer), not in Sino-Vietnamese Chinese character pronunciations. Another example is the diphthong /uə/ ‘uô/ua’, which is found in only nine Sino-Vietnamese words but in 279 words of native Vietnamese (again, the latter including some early Chinese loanwords). Regarding other phonemes of Vietnamese, there are consistently higher ratios of those sounds in native words than among Sino-Vietnamese vocabulary such that no phonological features specifically mark syllables as Chinese loanwords.

4. Native Lexicon: The Vietic and Austroasiatic Components

We now consider the native portion of the Vietnamese lexicon in terms of quantitative data and semantic/cultural domains of words. The Vietnamese lexicon is well known for its Chinese component, though in the past, inflated numbers (e.g., 70 percent) of the portion of Chinese loanwords have been mentioned without reference to a specific study or research method that determined such numbers. More recently, numbers based on focused counts of entries in dictionaries (Hùng Việt et al. 2018, pp. 348–51) and selected lexical sets (Alves 2009, p. 619) show a range of 25 to 35 percent.[6] Thus, according to these recent studies, the bulk of the Vietnamese lexicon—some two-thirds to three-quarters—is native, not borrowed. Based on data assembled for this study, of the native portion of the Vietnamese lexicon, (a) nearly two hundred words are pAA etyma or at least early Austroasiatic items, (b) at least several hundred are likely pV items,[7] (c) some can be reconstructed only to the Proto-Viet–Muong stage, and (d) an uncertain number but likely thousands of words can only be found in Vietnamese.[8]

Which strata words belong to can be determined by collating lexical data for Vietic languages, determining which are widespread in Vietic and which are restricted to Vietnamese and Muong lects, identifying those which are likely Austroasiatic etyma, and excluding those of likely external origin (primarily Chinese, other Austroasiatic languages, Tai, and Indic).[9] Ferlus's Vietic reconstructions contained in the Mon–Khmer Etymological Dictionary, combined with a dozen other sources of Vietic languages, constitute a spreadsheet with over 1200 reconstructions that I have assembled and to which I have added reconstructions. Vietic reconstructions have then been considered for their potential relationship to other Austroasiatic etyma, though this has previously been a problem as a developed set of pAA reconstructions was lacking.

An early statistical study of the lexical relationship of Vietnamese to other Austroasiatic languages is that of Thomas and Headley (1970), who used 100-word lists of basic vocabulary. Their study showed that inter-branch lexical relatedness of the languages mostly ranged from 20 to 30 percent, and such was the case for Vietnamese and Muong, making them as close to Austroasiatic languages as any others. As for studies that propose pAA etyma of Vietnamese, two are by Hô (1992) and Alves (2020a). Hồ's work, listing about 200 words, does not explicitly refer to pAA reconstructions, making those claims tentative. Alves (Ibid.) similarly identified about 200 possible Austroasiatic items (many of which are not the same as Hồ's) and another 500 pV etyma in Vietnamese. For the Austroasiatic words, he referred to Shorto's (2006) 2200 Mon–Khmer reconstructions, but as Shorto's work should be regarded as a work in progress, Alves also considered the geography of Shorto's reconstructions and utilized only items with distribution widespread enough to validate a claim of proto-language status.

However, only recently has Sidwell (2024) provided a list of pAA etyma reconstructions based on criteria to focus on words which have demonstrable historical depth. Applying Sidwell and Alves's (2023) approach in determining which words are reconstructable to the time of the pAA dispersal over 4000 years ago, Sidwell's reconstructions consist only of items with attestations in Austroasiatic branches both inside and outside of Mainland Southeast Asia, the latter consisting of Munda, Nicobarese, Aslian, and Khasian. This geographic distribution is strong evidence for the posited deep-history time-depth. In Sidwell & Alves's working spreadsheet, besides the 500 pAA reconstructions, a few hundred more have been marked potential pAA items. Another few hundred items are found only in Mainland Southeast Asia, which means they do not meet Sidwell & Alves's criterion to be high-confidence pAA etyma, but they still may have substantial time-depth in Austroasiatic language history. Thus, both of these latter categories are also considered in the current study.

The results of checking these data are that 146 Vietnamese words are modern reflexes of pAA etyma (Appendix B), and an additional 46 items can be regarded as potential pAA etyma or at least words early in Austroasiatic history (Appendix C), a total of 192 words. While the number of Austroasiatic words in Vietnamese is modest, those words include

a wide range of basic vocabulary (e.g., body parts, common animals, natural phenomena, basic actions, etc.). Moreover, other words provide somewhat more specific information about early Austroasiatic activities (e.g., to cook, to dream, to meet, to gather/associate, to comb, to hunt, etc.) and families (e.g., grandchild, child, mother). Crucially, many of the words match archaeological portrayals of the daily life of settled agriculturalists in this region (e.g., Higham 2021). Some of these words are related to farming practices in the Neolithic period of Austroasiatic history (e.g., rice, bran/husk, mortar, pestle, winnowing basket, to plant seed, etc.). The 13 function words in the lists include number terms, units of time, and deictic terms. It is notable that, while neighboring Kra–Dai languages have mostly lost their native numeral word systems, Vietnamese has retained an Austroasiatic set, in spite of other significant lexical borrowing from Chinese. Table 8 lists the various categories and provides examples taken from Appendices A and B.

Table 8. Semantic domains of Austroasiatic vocabulary in Vietnamese.

Semantic Domains	No. of Words	Examples of Senses (Complete Data in Appendices B and C)
Actions of general nature	36	to comb, to hunt, to open, to cover, to scatter/sprinkle, etc.
Body terms (nouns and verbs)	32	eye, gall/bile, to open one's mouth, hair, to choke/vomit
Animals	28	dog, crab, fish, python, horn, etc.
Actions for motion	16	to carry on one's back, to pour, to fall/drip, to wade/swim, to run, etc.
Agricultural and vegetation	14	husked rice, bran/husk, taro, fruit, to plant seed, etc.
Food and related actions	7	cooked/ripe, to suck(le), to cook, etc.
Function, time, and numeral words	13	this, that, you, year, day, four, nine, etc.
Natural environment	11	hill, mountain, rain, cloud, gravel, etc.
Items and related actions	11	house, pestle, mortar, to weave, to scoop/bale, etc.
Cultural aspects (cognition, sense, kinship, society, and speech)	10	grandchild, mother, to dream, to meet/associate, to ask, to call out, etc.
Spatial	11	distant, deep, supine, side, etc.
Sense	8	sour, bitter, to rub/stroke, warm (adj), etc.

Altogether, despite the seeming modest number, these nearly 200 words form a core of Austroasiatic etyma in Vietnamese that can hardly be considered trivial. Indeed, these words highlight the Austroasiatic essence of Vietnamese.

5. Morphology and Syntax

Identifying the Austroasiatic elements of Vietnamese morphology and syntax requires comparison with both modern languages and earlier stages of Austroasiatic and Vietic. While there are more grammatical descriptions of Austroasiatic languages than ever, many of them are short treatments, with only some more thorough descriptions in articles and dissertations. Regarding pAA syntax, recent advances in understanding have been made, but the claims in this area are still general and somewhat speculative. As for morphology, some affixes are well represented among Austroasiatic languages, such that categories of derivational morphology can be reconstructed. Austroasiatic-style reduplicative can be assumed to have been an operational word formation strategy, though no reduplicative words have been reconstructed. The connections of these matters to Vietnamese are less direct than for the lexicon and speech sounds but still demonstrable.

5.1. Syntactic Structures

One impediment to developing hypotheses of Vietnamese diachronic syntax is that some features are shared by many neighboring languages in the region regardless of language family, such as SVO/AVP clause structure and post-nominal modifiers in noun phrases. Thus, it is difficult to distinguish retention from contact-induced typological convergence. Still, according to recent work, we can posit that pAA had head-initial phrase structure in both clauses and noun phrases, while modern Vietnamese has a blend of this due to later contact-induced factors. These issues are discussed below.

In the past decade, pAA clause structure has been hypothesized to be verb-initial (Jenny et al. 2015; Jenny 2020). Jenny (2020) provides supporting evidence for this claim from multiple sources: (a) main clauses in some Khasian, Nicobarese, and Aslian languages, (b) embedded clauses of some Palaungic languages, and (c) noun compounds of Munda languages. If this hypothesis is valid, one distinct possibility is that a shift to verb-medial main clause structures happened among many Austroasiatic languages in the region due to language contact, though this cannot as yet be determined. As for pAA noun-phrase structure, no specific studies have presented a complete argument for a reconstruction, but a head-initial type is strongly indicated. Noun phrases with strictly post-nominal modifiers occur in Old Khmer texts from the pre-Angkorian period of the early 1st millennium CE, and head-medial order in Vietic appears to be later a development due to the spread of classifier phrases at least partly due to language contact with Chinese (Alves 2020c, p. 63).

For Vietic clause structure, Alves (2020c, p. 69) reconstructs topic–comment structures (with subjects/agents morphologically indistinguishable from topics), verb-medial clauses, middle-passive constructions (lacking any explicit lexical means of marking the passive), and a right-branching VP with objects/patients, other complements, and adjuncts after verbs. These hypothesized proto-language features are feasible considering the most general properties of Vietic languages and such features in early Vietnamese texts.

Alves (2020c, p. 75) posits the following syntactic structures in pV, in which clauses are verb-medial and noun phrases are head-initial. There is no language-internal evidence to hypothesize an earlier verb-initial stage like that of Jenny's claim for pAA. In modern Vietnamese, clauses are similarly verb-medial.

The Early Vietic Clause	**Modern Vietnamese**
*[topic-verb-complement]	[topic-verb-complement]
	nó bắt cá
	3s catch fish
	'He/She/It... catches fish.'

Similarly, Modern Vietnamese noun phrase structure has a head-medial structure with elements both before and after head nouns. In Vietnamese, quantification (i.e., phrases consisting of quantity words followed by measure words or classifiers) precedes head nouns, while all other elements (i.e., modifiers, determiners, and possessives) follow them. In a reconstruction of the pV noun phrase, Alves (2020b, p. 67) posits that it matched the modern Khmer structure of head-initial, with all modifiers following a head noun. Additional support comes from Vietnamese textual evidence from the early 2nd millennium CE in which there was substantial flexibility of the position of quantification. Quantification commonly occurred varyingly before and after head nouns. The following early text sample from the mid-2nd millennium shows a quantification phrase (e.g., 'one unit') following the head noun (e.g., 'letter').

情	書	蔑	幅	封	群	謹
tình	thư	một	bức	phong	còn	kín
love	letter	One	CLS	seal	still	tight

'A love letter still sealed tight.' (From the poem "Ba Tiêu" in "Quốc Âm Thi Tập")

This was hypothetically a transitional stage leading from the earlier post-nominal position to the modern stage with only pre-noun position quantification. While the adoption of Chinese classifiers and measure words may have influenced the position of quantification in the Vietnamese noun phrase, the position of other modifying elements, including adjectival elements, possessive nouns, and determiners, was unaffected.

The Early Vietic Noun Phrase	**Modern Vietnamese**
*[noun–modifier–quantity]	[quantity–noun–modifier]
	một con **cá** lớn ấy
	one CLS **fish** big distal
	'That one large fish.'

If the pAA verb-initial hypothesis holds, at best, we can say that Vietic participated in the early regional change to topic-comment clauses and that Vietnamese shares with pAA a post-verbal object. As for noun phrases, the post-nominal position of modifiers is a retention from pAA. A summary of clause and noun-phrase structures at different stages is provided in Table 9.

Table 9. Clause and noun-phrase structures in pAA, pV, and Vietnamese.

Structural Aspect	Stage	Pattern
Clauses	pAA	verb-initial
	pV	verb-medial
	Viet.	verb-medial
Noun phrases	pAA	head-initial
	pV	head-initial
	Viet.	head-medial

5.2. Word Formation Strategies

As for morphology, pAA has been reconstructed with derivational prefixes and infixes (Sidwell 2008), and lexical reduplication involving alliteration, riming, and ablaut is widespread enough to consider a word formation strategy in pAA. A more challenging matter is to determine how much phonesthemes played a part in pAA word formation, but some evidence allows for that possibility. The Austroasiatic-style reduplication is robustly attested in Vietnamese, and some evidence of pAA phonesthemes is evident. As for Austroasiatic derivational morphology, there is some evidence of this in the remaining monosyllabic roots, though most of them have been completely lost with no means of recovering them.

5.2.1. Reconstructing Lost Affixation

The complete loss of presyllables in Vietnamese has made recovery of past prefixes and infixes challenging at best and impossible in many instances. Reconstructing the presence of presyllables alone is difficult, let alone their precise phonological material. However, one type of retention for which there is some persuasive evidence is pAA nominalizing nasal infixes, a widespread word formation strategy among Austroasiatic branches. Table 10 shows examples of source verbs and derived nouns with vowel-plus-nasal sequences in several branches of Austroasiatic. This type of derivational infix is even seen in Munda and Nicobaric, demonstrating that these date to the pAA stage.

Table 10. Nominalizing infixes in Austroasiatic languages.

Language (Branch)	Source Verbs	Derived Nouns
May (Vietic)	pi^3 'to carry on one's back'	pani^3 'basket'
Maleng Bro (Vietic)	tęʔ 'to urinate'	trnęʔ 'urine'
Buxing (Mangic)	tain 'to weave'	tɣm.lain 'a braid'
Car (Nicobaric)	kahúl 'to cook'	kamhũ̀l 'a cook'
Mlabri (Khumic)	tɛk 'to hit'	trnɛk 'hammer'
Santali (Munda)	ʤɔk 'to sweep'	ʤɔnɔk 'broom'
Stieng (Bahnaric)	puːs 'to sweep'	pənus 'broom'

While derivational morphology is mostly fossilized even in conservative Vietic languages, some examples can be provided, as in the first two instances in Table 10. The possibility then exists for Vietnamese in a previous stage to have lost the presyllables of these derived nouns, leaving main syllables with nasal onsets.

Ferlus (1977, pp. 53–54) provides such instances in Vietnamese of pairs of verbs and nouns that share rhymes (i.e., vowel, coda, and tone), and the actions apply to the related nouns, as in Table 11. While Ferlus notes that a couple of these pairs have cognates in Khmu in the Khmuic branch of Austroasiatic, he provides no instances in other Vietic languages.[10]

Table 11. Possible instances of retained nominalizing infixes in Vietnamese.

Source Verbs	Derived Nouns
ɗaːn^1 (đan) 'weave'	naːn^1 (nan) 'bamboo slat'
ɗap^5 (đắp) 'cover'	nap^5 (nắp) 'lid, stopper'
ɗəː^6 (đợ) 'pawn; pledge'	nəː^6 (nợ) 'debt'
ɗuːt^5 (đút) 'insert'	nuːt^5 (nút) 'stopper, plug (n)'
ceːm^1 (chêm) 'wedge'	neːm^1 (nêm) 'wedge (n)'
ceːp^5 seːp^5 (chếp xếp) 'fold'	neːp^5 (nếp) 'wrinkle (in nếp nhăn)'
cɔːk^6 (chọc) 'thrust, poke'	nɔːk^6 (nọc) 'stake, picket'
ɽiːt^6 (rịt) 'tie up'	niːt^6 (nịt) 'belt'
taːj^3 (tải) 'bag (n)'	naːj^3 (nải) 'sack, bag'
ciə^4 (chĩa) 'pitchfork; fish with a fish lance'	niə^4 (nĩa) 'fork'

While nominalizing infixes and causative prefixes are common among Austroasiatic languages, affixes with other functions have developed among Austroasiatic languages. We can thus speculate that other affixes may have existed as well in earlier stages of Vietic. As noted, onset clusters *kl in pAA and pV have become [ʈ] 'tr' in modern Vietnamese, and in some cases, such phonological material may have come from presyllables. Table 12 contains instances of potential pairs of verbs and either stative or active verbs in which the posited source words have [l] onsets while the derived verbs have [ʈ] onsets. Also in Table 12 are the words spelled in early romanized script from de Rhodes' 1651 dictionary, in which these words had [tl] cluster onsets instead of the possible [ʈ], which also existed at that time. These must be regarded as tentative until additional comparative evidence can confirm or refute them (e.g., similar evidence in other Vietic languages, early Vietnamese textual data showing evidence of presyllables, etc.), but they at least show early onset clusters and thus possible previous presyllables that represent earlier stages of Vietic with a more Austroasiatic-like template.

Table 12. Vietnamese source verbs and derived verbs that may have had prefixes.

Source Verbs	Derived Verbs	17th Cent.
leːn^{1} (lên) 'ascend'	ʈeːn^{1} (trên) 'atop'	tlên
lɛːw^{1} (leo) 'climb'	ʈɛːw^{1} (treo) 'hang, suspend' (cf. ʈɛːw^{2} (trèo) 'climb')	tleo
loːj^{1} (lôi) 'pull, drag'	ʈoːj^{1} (trôi) 'drift, float'	tlôi
liət^{5} (lướt) 'glide'	ʈiət^{6} (trượt) 'slip, skid'	tlợt
NA (cf. pAA *lun, *lɔːn 'round, roll (v)')	ʈɔːn^{2} (tròn) 'round'	tlòn
ʈaːj^{3} (rải) 'spread, scatter'	ʈaːj^{3} (trải) 'spread, lay'	klải

5.2.2. Austroasiatic-Style Reduplication

Another Austroasiatic-like feature of Vietnamese is reduplication. Austroasiatic languages exhibit a type of reduplication in which templates of monosyllabic or disyllabic words are copied, but typically one or more segment alternates from the root. This results in patterns of alliteration, rhyming, or ablaut, as exemplified in Table 13. While precise reduplicative forms cannot be reconstructed to pAA or even pV, they are widespread enough among Austroasiatic languages (e.g., note the instance in Mundari in India, as shown in Table 13) to regard as a word formation strategy at the proto-language stage.

Table 13. Examples of reduplication with alternating segments in Austroasiatic languages.

Types	Examples
Alliteration	Vietnamese: lɔːŋ^{1}-laʲŋ^{1} (long lanh) 'sparkling' Bru (Katuic): tamaj-tamiːŋ 'new' Sre (Bahnaric): crih-craːj 'strange, wonderful'
Ablauting	Vietnamese: muːm^{4}-miːm^{4} (mũm mĩm) 'chubby, plump' Cuoi (Vietic): həŋ^{44ʔ} haŋ^{44ʔ} 'empty' Pacoh (Katuic): puːc-paːc 'to flutter'
Rhyming	Vietnamese: boːj^{5}-roːj^{5} (bối rối) 'embarrassed; puzzled' Mundari (Munda): cali-bali 'dressing up nicely' Phong (Khmuic): kup-plup 'butterfly'

While Vietnamese has instances of full reduplication (e.g., *th*ɯə*ŋ*^{2}-*th*ɯə*ŋ*^{2} (thường) 'sometimes'), a type of reduplication involving alternating segments is much more common, with thousands of words formed in this way, as in the Từ Điển Từ Láy Tiếng Việt [Dictionary of Vietnamese Reduplicative Words] (Viện Ngôn Ngữ Học 1995). This type of reduplication is distinct from the prominent types of reduplication in Chinese languages (e.g., full reduplication with pragmatic functions, partial reduplication such as the A-not-A questions, an ABB pattern of copied second syllables, etc.).[11] As expected, it is also a feature of conservative Vietic languages, such as the instance of Cuoi in Table 13.

5.2.3. Phonestheme Onsets

While Vietnamese lacks presyllables—and thus lacks prefixes or infixes—it does exhibit some phonesthemic word formation patterns with ideophonic features related to oral actions or senses. In some instances, these patterns are shared by other Austroasiatic languages, including some that can be seen in pAA etyma. One can cast an excessively wide net and find chance similarities or even more common typological patterns connected to sound symbolism. Nevertheless, the ideophonic sound symbolism was undoubtedly a factor in the original creation of the words and patterns, and their widespread distribution suggests a deep history in the language family, especially when neighboring language families (e.g., Kra–Dai, Sinitic and Tibeto-Burman, etc.) show different patterns that cannot be

considered part of a real phenomenon. The proposed phonestheme onsets in Vietnamese and other Austroasiatic languages include (a) nasal onsets for words related to oral activities (snoring, choking, sucking) and taste (sweetness), (b) /h/ onsets in words referring to inhaling, singing, sniffing, or singing, and (c) /l/ onsets for words expressing licking or sticking out a tongue.

Nasal Onsets

pAA etyma in this category include *ɲɔːkʔ 'suck' (AA483), *ɲɔːtʔ 'drink, suck (v)' (AAP), and *ŋaːm 'sweet' (AA113), but Vietnamese does not have cognates of these. Instead, several pV etyma in Vietnamese with senses related to oral and nasal activities similarly have nasal onsets, mostly engma, as shown in Table 14. It is not clear what the purpose is of the presyllables on these items.

Table 14. Vietnamese words with nasal onsets expressing oral and nasal activities.

Vietnamese	Gloss	PV
ɲaːj^{1} (nhai)	chew	*təˌɲaːj
ŋəm^{6} (ngậm 'hold in mouth')	suck	*ŋəmʔ
ŋɛːn^{6} (nghɛːn)	choke (while eating)	*ŋɛːnʔ
ŋɔːt^{6} (ngọt)	sweet	*təˌŋɔːc
ŋaj^{5} (ngáy)	snore	*təˌŋarʔ
ŋɨːj^{4} (ngửi)	sniff, smell (v)	*təˌŋəs

/h/ Onsets

Another set of words has/h/onsets and semantics of verbs to denote inhaling or breathing. The number of words with this feature suggests that this onomatopoetic feature became part of a modestly productive word formation pattern. pAA has only *huːtʔ 'drink, sip (v)' (AA233), but pV and Vietnamese have additional relevant instances, as in Table 15.

Table 15. Vietnamese words with /h/ onsets expressing oral activities.

Gloss	AA	pV	Viet.
suck/inhale	*huːtʔ	*huːc	huːt^{5} (hút)
inhale, sniff, snuffle	*hɨːtʔ (AA402a)	*hiːt (tent.)	hiːt^{5} (hít)
open (mouth) (v)	*haʔ (AAP)	*haːʔ	haː^{3} (hả)
kiss/sniff (v)	NA (Proto-Katuic *huuɲ, *huoɲ smell, sniff')	*huːɲ	hoːn^{1} (hôn)
sing	NA	*haːt	haːt^{5} (hát)
cough (v)	NA	*hɔː	hɔː^{1} (ho)
suck in	NA	NA	huːp^{5} (húp)

/l/ Onsets

Words with /l/ onsets that are related to the mouth can be readily found in languages outside of Austroasiatic, for example, Proto-Indo-European *leigh 'lick', Proto-Sino-Tibetan *liek 'lick, tongue', and Proto-Tai (Pittayaporn 2009) *C̬lwiəA 'to lick' and *liːnC 'tongue'. Thus, the ideophonic status of such words seems likely. Regardless, in pAA, similar categories of words with /l/ onsets have been reconstructed, such as *lən.'taːkʔ 'tongue' (AA054) and *liəmʔ, *liːmʔ 'lick (v)' (AA1h19) (and also *liətʔ and *liəpʔ 'lick (v)' (AAP)). Of these, only *liəmʔ, *liːmʔ has a reflex in Vietnamese, but Vietnamese has other such words with an /l/ onset, which are pV etyma, and in other Austroasiatic languages, there are also instances of words (both cognates and non-cognates) with /l/ onsets having senses, as in Table 16.

Table 16. Vietnamese and cognate or non-cognate Austroasiatic words with /l/ onsets and senses related to the mouth.

Viet	pV	Other Austroasiatic languages
liəm^{5} (liếm) 'lick (v)'	*-lɛːmʔ	Proto-Katuic *laas, *l̥ias 'tongue, lick' Proto-Khmuic *-lɛːl 'to lick, taste'
lɯəj̀^{4} (lưỡi) 'tongue'	*laːs	Bolyu (Pakanic) lɔ^{31} l̥jiːm^{53} 'tongue' Bolyu (Pakanic) l̥jim^{31} 'to lick' Munda (not reconstructed but seen in a widespread form) laːŋ 'tongue'
lɛː^{1} (le) 'stick out tongue'	*t.lɛːl	Proto-Bahnaric *liər ~ *lɛːr 'to stick out tongue' Kensiu (Aslian) lʌhleh 'to wiggle/stick out tongue'

Overall, while words in this section can be considered part of an ideophonic feature in other languages, the number of Vietnamese words with cognates in Austroasiatic proto-language and modern Austroasiatic languages altogether suggest an ideophone phones-theme pattern of word formation in Austroasiatic that has been retained in Vietnamese.

6. Implications in Vietnamese Linguistic Ethnohistory

Despite the impact of Chinese on Vietnamese, especially in its lexicon, but also its phonology (e.g., syllable structure and tones) and noun phrase structure (i.e., movement of quantification to before head nouns), there are many retentions of Austroasiatic features at all linguistic levels. In addition, Vietnamese maintained its Austroasiatic typology for many centuries after the start of Sinitic–Vietic language contact; the linguistic restructuring was not an instance of rapid transformation. While tones in Vietnamese may have emerged in the later part of the 1st millennium, disyllabic words were retained into the early centuries of the 2nd millennium (Shimizu 2015), and onset clusters lasted into the early 1800s (Vu 2019). Also, while hundreds of early Chinese loanwords were borrowed in the 1st millennium, the large-scale borrowing of Chinese words did not take place until the beginning of the 2nd millennium. This was the time of the hypothesized shift of a local Chinese community to Archaic Vietnamese and the contemporaneous spread of Chinese rime dictionaries and a tradition of Chinese-character reading.

Altogether, despite the Chinese administration in northern Vietnam for a millennium and significant impact on the Vietnamese lexicon and some elements of its phonology, these changes did not result in a language that was unrecognizable as an Austroasiatic language. Thus, we can speculate that the Vietic community in northern Vietnam in that period held considerable sociocultural status. There was even retention of a native numeral system (in contrast with replacement of the numeral systems of Tai languages with the Chinese one), despite an influential resident Chinese community in which early Vietic speakers intermarried, while borrowing many other categories of words. While details of the sociocultural situation in northern Vietnam in the 1st millennium CE are not available, the sociolinguistic status of Viet–Muong and, later, Vietnamese, was apparently significant enough to stave off many of the potential effects of intense language contact.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: See Appendix A for a list of the online lexical data resources consulted for this study. Other publications are noted throughout this article and are publicly available.

Acknowledgments: I must thank Paul Sidwell, Hilario de Sousa, and two anonymous readers for providing helpful helpful suggestions in details of the text.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Online Resources Consulted

- Blust, Robert and Stephen Trussel. 2010. Austronesian Comparative Dictionary. Updated 21 December 2014. Web. http://www.trussel2.com/ACD/acd-ak_a.htm (accessed on 1 September 2024)
- Dictionary of Old Khmer. SEACLASSICS Khmer. Bangkok: Center for Research in Computational Linguistics. Web. http://sealang.net/ok/ (accessed on 12 August 2024.)
- Digital Dictionaries of South Asia. University of Chicago. Web. https://dsal.uchicago.edu/dictionaries/list.html (accessed on 15 July 2024)
- Old Javanese Dictionary. Bangkok: Center for Research in Computational Linguistics. Web. http://sealang.net/ojed/ (accessed on 15 July 2024)
- Proto-Tai-o-matic. Bangkok: Center for Research in Computational Linguistics. Web. http://sealang.net/crcl/proto/ (accessed on 1 August 2024)
- SEALang Library (Digital dictionaries of Burmese, Indonesian, Karen, Khmer, Lao, Malay, Mon, Shan, Thai, Vietnamese, and various other major languages in Southeast Asia). Bangkok: Center for Research in Computational Linguistics. Web. http://sealang.net/library/ (accessed on 5 September 2024)
- SEALang Mon-Khmer Etymological Dictionary. Bangkok: Center for Research in Computational Linguistics. Web. http://www.sealang.net/monkhmer/dictionary/ (accessed on 30 July 2024)
- SEALang Munda Etymological Dictionary. Bangkok: Center for Research in Computational Linguistics. Web. http://www.sealang.net/munda/dictionary/ (accessed on 15 July 2024)
- The Sino-Tibetan Etymological Dictionary and Thesaurus. The University of California, Berkeley. Web. http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl (accessed on 20 August 2024)

Appendix B. Proto-Austroasiatic

VIET	GLOSS	SEM	PV	PAA	ID
ɓaːw^{1} (bao)	dream (n, v)	cognition	*səˑpoː	*ʔəmˑ’poːˀ	AA029
ɓaj^{1} (bay)	fly (v)	motion	*pər	*par, *pər	AA042
ɓaj^{3} (bảy)	seven	numeral	*pəs	*pɔh, *pəɛ	AA209
ɓan^{5} (bắn)	shoot (v)	basic	*pəɲʔ	*paɲˀ	AA057
ɓɛː^{3} (bẻ)	break off (v.t.)	basic	*pɛh	*pɛːh	AA188
ɓɨək^{5} (bước)	step (v)	motion	NA	*təˑ’paːkˀ	AA465
ɓoːn^{5} (bốn)	four	numeral	*poːn	*puənˀ	AA018
ɓuː^{5} (búː)	suck(le) (v)	food	*buːʔ	*ɓuːˀ	AA033
caːj^{3} (chải)	comb (v)	basic	*caːs	*caːɛ	AA170
caj^{2} (chày)	pestle	item	*tʃəˑreː	*ɟənˑ’reːˀ	AA044
caj^{6} (chạy)	run (v)	motion	*ɟalʔ	*ɟarˀ	AAP
caw^{5} (cháu)	grandchild	kinship	*cuːʔ	*cuʔ	AA096
cɔː^{5} (chó)	dog	animal	*ʔaəˑcɔːʔ	*cɔʔ	AA030
cɔːk^{6} (chọc)	prick, pierce (v)	basic	NA	*cɔkˀ	AA396
cəj^{5} (chấy)	head louse	animal	*ciːʔ	*ciʔ	AA001
cɛː^{3} (chẻ)	tear, split open (v)	basic	*cɛh	*cɛːh	AAP

VIET	GLOSS	SEM	PV	PAA	ID
ci:m^{1} (chim)	bird	animal	*ci:m	*ci:m	AA016
ci:n^{5} (chín)	cooked, ripe	food	*ci:nʔ	*ci:nˀ	AA017
cu:t^{6} (cụt)	cut short (v)	basic	NA	*gutˀ	AA461
cuə^{1} (chua)	sour	sense	*ɟɔ:ʔ	*ɟuʔ, *ɟɔʔ	AA068
za:m^{1} (dam 'field crab')	crab	animal	*kə.ta:m	*kə.'ta:m	AA004
ɗa:n^{1} (đan)	weave (v)	item	*ta:ɲ	*ta:ɲ	AA005
ɗam^{1} (đăm (archaic))	right (side)	spatial	*dam, *tam	*tam	AA117
ɗaŋ^{5} (đắng)	bitter	sense	*taŋʔ	*ʘaŋˀ	AA041
ɗap^{5} (đắp 'embank')	layer (v)	basic	NA	*tapˀ	AA265
ɗɔ:^{5} (đó)	that, there	function	NA	*tu:ˀ, *tɔ:ˀ	AAP
ɗəm^{1} (đâm 'to pound (rice)')	pound (v)	food	*təm	*təm	AAP
ɗəp^{6} (đập	dam, block (v)	basic	*tap	*dapˀ	AA223
ɗi:t^{5} (đứt)	break, sever (v)	basic	*tVc (tent.)	*tacˀ	AA252
ɗo:^{3} (đổ)	pour (v)	motion	*toh	*toh	AAP
ɗo:j^{2} (đồi)	hill, mound	nat. env.	*do:l	*do:l, *duəl	AA423
ɣa:j^{4} (gãi 'to scratch due to itching')	scratch (v)	basic	*ka:s	*ka:cˀ (also cf. *ka:ɛ 'shave, scrape (v)')	AA162
ɣa:w^{6} (gạo)	husked rice	agr. and veg.	*rə.ko:ʔ	*rəŋ.'ko:ˀ	AA066
ɣap^{6} (gặp)	meet, associate (v)	social	NA	*gapˀ	AAP
ɣɔ:m^{1} (gom 'gather, collect')	gather, associate (v)	social	NA	*kom	AA317
ɣəj^{2} (gầy)	thin, lean	spatial	NA	*rə.'gɨj	AA315
ɣəp^{5} (gấp)	cover (v)	basic	*kəp	*kɨpˀ, *kɔpˀ	AA067
ɣo:j^{5} (gối)	knee	body	*tə.ku:lʔ	*Cə.'ku:l	AA326
ha:^{3} (hả)	open mouth (v)	body	*ha:ʔ	*haʔ	AAP
hi:t^{5} (hít)	inhale, sniff (v)	sense	*hi:t (tent.)	*hɨ:tˀ	AA402a
hwi:t^{5} (huýt)	whistle (v)	speech	NA	*huəcˀ	AA138
ka:m^{5} (cám)	bran, husk	agr and veg	*tə.ka:mʔ	*sə.'ka:mˀ	AA076
ka:t^{5} (cát)	sand	nat. env.	*tə.ka:c	*sə.'ka:cˀ	AA238
ka:^{5} (cá)	fish (n)	animal	*ʔə.ka:ʔ	*kaʔ	AA009
kam^{2} (cằm)	molar tooth	body	*gam (tent.)	*-ga:m	AAP
kap^{6} (cặp)	pinch, grip (v)	basic	*gap	*gapˀ	AA361
kɔ:n^{1} (con)	child	kinship	*kɔ:n	*koan	AA006
kɔ:ŋ^{1} (cong)	bend (v)	motion	*kɔ:ŋ (tent.)	*koŋ	AA409
ko:j^{5} (cối)	mortar	item	*tə.ko:lʔ	*go:lˀ	AAP
ko:t^{6} (cột 'tie, bind (v)')	knot (v)	item	NA	*kuətˀ	AA139
kwat^{6} (quặt)	turn, bend (v)	motion	NA	*kə.'watˀ	AAP
la:^{5} (lá)	leaf	agr and veg	*sə.la:ʔ	*sə.'laʔ	AA010
liem^{5} (liếm)	lick (v)	body	*C.lɛ:mʔ	*liəmˀ, *li:mˀ	AA119
liət^{6} (lượt 'to glide')	slippery	sense	*gə.le:t 'slip (v)'	*li:tˀ, *le:tˀ	AAP

VIET	GLOSS	SEM	PV	PAA	ID
lo:^{4} (lõ)	hole, perforate (v)	basic	NA	*luh	AA050
lo:j^{6} (lội ‘wade’)	wade, swim (v)	motion	*lo:jʔ	*lo:jʔ, *lɔ:jʔ	AA135
lo:t^{6} (lột ‘slough’)	slip (off, down) (v)	motion	NA	*lo:cʔ, *lu:cʔ	AA303
maj^{2} (mày)	you (singular)	function	*mi:	*mi:ʔ, *me:ʔ	AA047
maŋ^{1} (măng)	bamboo shoots	agr. and veg.	*tə.ɓaŋ	*tə.’ɓaŋ	AA023
mat^{5} (mắt)	eye	body	*mat	*matʔ	AA024
mɔ:^{3} (mỏ)	beak	animal	*kə.ɓɔh	*cə.’ɓɔh, *cə.’ɓu:ʔ	AA442
mɔ:j^{5} (mói ‘a tool used when transplanting’)	plant seed (v)	agr. and veg.	*cə.mɔ:lʔ ‘dibble stick’	*ɟə.’mɔ:lʔ	AA131
mɔ:k^{5} (móc)	hook (v)	item	NA	*ɓɔkʔ, *ɓuəkʔ	AA392
mɔ:k^{6} (mọc ‘grow, come out’)	emerge (v)	motion	NA	*muəkʔ, *mɔ:kʔ	AA219
mə:^{3} (mở)	open (v)	basic	*pəh	*pəh	AA398
məj^{1} (mây)	cloud	nat. env.	*kə.məl	*rə.’mɨl	AA431
mət^{6} (mật)	gall, bile	body	*mɨ̂t	*kə.’matʔ, *kə.’mɨtʔ	AA079
mɛ:^{6} (mẹ)	mother	kinship	*me:ʔ ~ mɛ:ʔ	*meʔ	AA061
mɨə^{1} (mưa)	rain (n,v)	nat. env.	*kə.ma:	*gə.’ma:ʔ	AA036
mɨən^{6} (mượn ‘borrow’)	ask (v)	speech	*ma:ɲʔ	*sə.’ma:ɲ	AA335
miəŋ^{6} (miệng)	mouth, jaw	body	*mɛ:ŋʔ	*miəŋʔ	AA273
mo:t^{6} (một)	one	numeral	*mo:c	*muəjʔ, *mo:jʔ	AA049
mu:j^{4} (mũi)	nose	body	*mu:s	*mu:h, *mu:ɐ	AA003
mu:n^{6} (mụn)	pimple	body	*mu:nʔ	*mu:nʔ	AA070
mu:t^{5} (mút)	suck (v)	food	*cə.ɓu:c	*ɓu:cʔ	AAP
muəj^{4} (muỗi)	mosquito	animal	*mɔ:s	*mɔ:s	AA133
ɲa:^{2} (nhà)	house	item	*ɲa:	*ɲa:ʔ	AA110
naj^{2} (này)	this	function	*Cə.ni:	*ni:ʔ	AA065
nam^{1} (năm)	year	time	*cə.nəm	*cə.’nam, *cə.’nɨm	AA058
nap^{5} (nắp)	cover (v)	basic	*Cə.nap	*ɗapʔ	AA360
nɔ:^{6}/nɔ:^{5} (nọ/nó)	this	function	NA	*nɔh	AA391
nɔ:ŋ^{1} (nong)	winnowing basket	item	*ɗo:ŋ; *ɗo:ŋʔ	*kə.’ɗo:ŋ	AA214
nəp^{5} (nắp ‘to hide’)	cover (v)	basic	*ɗəp (tent.)	*ɗipʔ	AA224
ɲəw^{6} (nhậu ‘drink (v)’)	drink, suck (v)	food	*ɲu:ʔ	*ɲuʔ	AA218
ni:t^{5} (nít ‘small (of children)’)	small	spatial	*ɗi:t	*ɗi:tʔ	AAP
nɨ:t^{5} (nứt)	crack open/apart (v)	basic	*ɗac (tent.)	*dacʔ, *ɗacʔ	AA349
nɨək^{5} (nước)	water	nat. env.	*ɗa:k	*ɗa:kʔ	AA039
nu:ŋ^{1} (nung)	cook (v)	food	*ɗuŋ	*ɗɨŋ, *ɗuŋ	AA492
nu:p^{5} (núp)	cover (v)	basic	(Arem n◎tɯp)	*ɗu:pʔ	AA231
ŋa:j^{5} (ngái)	distant	spatial	*sə.ŋa:jʔ	*cə.’ŋa:jʔ	AA038
ŋa:p^{5} (ngáp)	yawn (v)	body	*səŋ.ʔa:p	*sə.’ŋa:pʔ, *səŋ.’ʔa:pʔ	AA022
ŋaj^{2} (ngày)	day	time	*C.ŋi:	*tə.’ŋi:ʔ	AA027

VIET	GLOSS	SEM	PV	PAA	ID
ɽa:j^{5} (rái)	otter	animal	*pə.se:ʔ	*bə.ɕeʔ	AA080
ɽa:ŋ^{6} (rạng ‘break (of daylight)’)	daylight (n., v.)	basic	NA	*ra:ŋˀ	AA167
ɽa:w^{1} (rao)	call out (v)	speech	NA	*kə.ʼra:w	AA250
ɽam^{5} (rắm)	fart (v)	body	*kə.səmʔ	*pə.ɕu:m	AA101
ɽan^{5} (rắn)	snake	animal	*pə.səɲʔ	*bə.ʼsaɲ	AA115
ɽe:^{4} (rễ)	root	agr. and veg.	*kə.riɛs	*ʔə.ʼriəɕ	AA026
ɽə:j^{1} (rơi)	scatter, sprinkle (v)	basic	NA	*rə:j	AA421
ɽən^{6} (rận)	body louse	animal	*mə.rəɲʔ	*ʔəm.ʼrəɲˀ	AA089
ɽiət^{5} (riết)	tighten, squeeze (v)	basic	NA	*riətˀ	AA381
ɽo:^{6} (rộ (‘noisily’ in reduplicative compounds))	speak, make sound (v)	speech	NA	*roʔ	AA090
ɽu:^{5} (rú)	hill, mountain	nat. env.	*bə.ru:ʔ	*bə.ʼruʔ	AA322
ɽu:ŋ^{6} (rụng)	fall, drop (v)	motion	*ruŋʔ	*ruŋˀ	AA444
ɽuəj^{2} (ruồi)	fly (n)	animal	*mə.rɔ:j	*roaj	AA028
ʂa:^{5} (sá)	path	motion	*kra:ʔ	*kə.ʼraʔ	AA073
ʂa:ŋ^{5} (sáng)	shine (v)	sense	*pə.la:ŋʔ	pə.ʼla:ŋˀ	AAP
ʂaw^{1} (sau)	back, behind	body	*kraw	*kə.ʼrɔ(◉)ʔ, *kə.ʼraw	AAP
ʂaw^{5} (sáu)	six	numeral	*pə.ru:ʔ	*tə.ʼruʔ, *pə.ʼruʔ	AA137
ʂɔ:^{6} (sọ)	taro	agr. and veg.	*sə.ro:ʔ	*sə.ʼroʔ	AA084
ʂɔ:j^{3} (sỏi)	gravel	nat. env.	*kə.rɔ:s	*kə.ʼruəs	AAP
ʂɔ:k^{5} (sóc)	squirrel	animal	*pə.rɔ:k	*pə.ʼrɔ:kˀ	AA091
ʂəw^{1} (sâu)	deep	spatial	*cə.ru: > *cru:	*ɟə.ʼru:ˀ	AA019
ʂi:^{1} (si)	banyan, ficus	agr. and veg.	*ɟə.ri:	*ɟə.ʼri:ˀ	AA064
ʂi:ŋ^{2} (sừng)	horn	animal	*kə.rəŋ	*kə.ʼriŋ	AA313
siəŋ^{1} (xương)	bone	body	(*tʃə.ʔa:ŋ)	*cə.ʼʔa:ŋ	AA012
ʈa:j^{5} (trái)	fruit	agr. and veg.	*pə.le:ʔ	*pə.ʼleʔ	AA043
ta:t^{5} (tát)	scoop, bale (v)	item	*ʔaə.sa:c	*sa:cˀ	AA111
taj^{1} (tay)	hand, arm	body	*si:	*si:ˀ	AAP
tam^{5} (tám)	eight	numeral	*sa:mʔ (or *tha:mʔ)	*təN.ɕa:m	AA338
ʈan^{1} (trăn)	python	animal	*kə.lən	*tə.ʼlan	AA040
ʈaɲ^{1} (tranh)	thatching grass	agr. and veg.	*plaŋ	*pə.ʼlaŋ	AA060
tɔ:k^{5} (tóc)	hair	body	*C.suk	*sukˀ, *sɔkˀ	AA032
ʈɔ:ŋ^{1} (trong)	inside; belly	body	(Ruc klɔ:ŋ^{1}, pPong *kluŋ^{4})	*kə.ʼlu:ŋ, *kə.ʼluəŋ	AA106
ʈe:c^{6} (trệch)	mistake, miss (target) (v)	basic	NA	*lacˀ, *pə.ʼlacˀ	AA172
ʈi:ŋ^{5} (trứng)	egg	food	*kə.lə:ŋʔ	*pən.ʼlə:ŋˀ	AA485
tiək^{5} (tước)	skin (n.,v.)	body	NA	*sa:kˀ	AA074
ʈo:n^{2} (tròn)	round, roll (v)	motion	*gə.lɔ:n > *klɔ:n	*lun, *lɔ:n	AA225

VIET	GLOSS	SEM	PV	PAA	ID
va:^{3} (vả)	fig tree	agr. and veg.	NA	*lə.ʼwa(◉)ʔ	AA160
va:k^{5} (vác)	carry on back/slung (v)	motion	*ɓa:k	*ɓakˀ	AA176
van^{6} (vặn)	twist (v)	basic	*vaɲʔ	*waɲˀ	AA269
ve:^{2} (về)	go around, return (v.i.)	motion	*ve:r	*wi:r	AA123
vən^{1} (vần ‘roll (v)’)	twist (v)	basic	NA	*wəɲ	AA405
vo:^{4} (võ)	hit, slap (v)	basic	*pəh	*pəh	AA205
vu:n^{1} (vun)	heap up (v)	basic	*Cə.pun (tent.)	*ɓu:n	AAP
vuət^{5} (vuốt)	rub, stroke (v)	sense	*pɔ:cˀ (tent.)	*po:tˀ, *pɔ:tˀ	AA217
xa:k^{6} (khạc)	hawk up phlegm (v)	body	*krə.ha:k	*kə.ʼha:kˀ	AA164
ʑɔ:^{5} (gió)	wind, air	nat. env.	*k◉jɔ:ʔ	*kə.ʼja:l; *kə.ʼjɔʔ	AA075
ʑɔ:t^{6} (giọt)	fall, drip (v)	motion	NA	*ɟɔ:tˀ	AA416
zəp^{6} (dập)	bury (v)	motion	*Cə.təp ‘to bury’ (tent.)	*tɨpˀ	AA143
ʔa:k^{5} (ác)	crow (n)	animal	*krə.ʔa:k	*kə.ʼʔa:kˀ	AA011
ʔɔ:j^{1} (oi ‘sultry’)	warm	sense	NA	*ʔu:r, *ʔo:r	AA328
ʔɔ:j^{5} (ói)	choke, vomit (?) (v.i.)	body	*ʔo:lʔ	*ʔo:lˀ, *ʔuəl	AA213
ʔɔ:ŋ^{1} (ong)	wasp, hornet	animal	*ʔɔ:ŋ	*ʔɔ:ŋ	AA132
ʔəm^{5} (ấm)	warm	sense	*sə.ʔəmʔ	*cə.ʼʔɨmˀ	AA433

Appendix C. Austroasiatic Words in Vietnamese Found Only in Mainland Southeast Asia

VIET	GLOSS	SEM	PV	Shorto’s Recon.	Shorto ID
ɓɔ:^{3} (bỏ)	throw away (v)	items	NA	*pooh	#2025
ɓi:^{5} (bí)	squash/pumpkin	agr. and veg.	*cə.pi:r	*cpiir	#1637
ɓi:t^{5} (bít)	cover (v)	spatial	NA	*cɓiit, *cɓiət	#1039
ɓɨt^{5} (bứt)	pluck, pick	basic	*pəc	*pic, *piic, *piəc, *pəc	#823
ɓu:ŋ^{6} (bụng)	belly	body	*buŋʔ	*buŋ, *buuŋʔ	#624a.b
ci:n^{5} (chín)	nine	quantity	*ci:nʔ	*dciinʔ	#1144
cu:m^{6} (chụm)	gather, group (n)	spatial	NA	*bɟum, *bɟuum, *bɟəm	#1338a,b
ɗɛ:^{3} (đẻ)	born, give birth	body	*tɛh	*ɗeh	#2013
ɗi:t^{5} (đít ‘buttocks’)	anus	body	NA	*ktiit	#1007
ɗiək^{5} (điếc)	deaf	body	*tɛ:k	*tik, *tiək, *tək	#303
ɗu:j^{1} (đui)	blind	body	*du:l, *tu:l	*tuəl	#1734b
ɣɔ:^{4} (gõ)	knock (v)	basic	*Cə.kɔ:h, *gɔ:h	*guəh	#1978
ɣəw^{5} (gấu)	bear (n)	animal	*cə.gu:ʔ; cə.ku:ʔ	*ɟkaw	#1817
ku:t^{5} (cút)	quail	animal	NA	*tgut, *tguut	#969
la:^{5} la̰k^{5} (lá lách)	spleen	body	*la:ʔ, la:	*slaʔ	#232
ma:^{6} (mạ)	seed	agr. and veg.	*sə.ma:ʔ	*maʔ	#133
miə^{5} (mía)	sugarcane	agr. and veg.	*kə.mɛ:ʔ	*klmiəʔ	#137

VIET	GLOSS	SEM	PV	Shorto's Recon.	Shorto ID
mo:^{1} (mô (dialect))	what, where	function	*Cmo:	*moʔ	#136.a
mo:j^{5} (mối)	kind of lizard	animal	*bolʔ	*ɓulʔ, *ɓuəl	#1767
mo:k^{5} (mốc 'mould, mouldy')	rotten	basic	*ɓuk	*ɓuk, *kɓuk	#368
mu:k^{5} (múc)	draw water	basic	*ɓu:k	*ɓəkʔ	#367
na:ŋ^{1} (nang 'eye tooth, tusk')	tooth	body	*kə.nɛ:ŋ	*gnaiŋ, *gniəŋ	#597
nam^{1} (năm)	five	quantity	*ɗam	*pɗam	#1363
nɔ:n^{5} (nón)	wide-brim hat	items	*ɗɔ:nʔ	*.ɗuən	#1159
no:^{3} (nổ)	burst, explode (v.i.)	basic	*ɗoh 'to explode'	*pɗuh, *pɗuəh	#2015a.a,b
no:k^{5} (nốc)	boat	items	NA	*ɗuuk, *ɗuk	#336
ŋa:^{3}, ŋiə^{3} (ngả, ngửa)	face upward, supine	spatial	*C.ŋah	*lŋaarh	#1590
ŋəm^{1} (ngâm)	soak, steep	basic	NA	*.ŋəm	#1321
ŋiə^{5} (ngứa)	itch (v.i.)	body	*Cə.ŋa:ʔ	*lŋaaʔ	#35
ŋiəj^{1} (ngươi)	pupil of eye	body	*Cə.ŋa:j	*ŋaaj	#1453
ŋiək^{5} (ngước)	bend back head (v)	body	NA	*lŋaak	#288.b
ŋiəŋ^{5} (nghiêng 'inclined, on the side')	side	spatial	*sə.gɛ:ŋ	*ɟkiiŋ, *ɟkiəŋ *ɟkaiŋ	#504
ɽiə^{3} (rửa)	wash (v)	basic	*ʔə.ra:wh	*raawh	#1841
ɽu:^{4} (rũ)	wash, rinse (clothes)	basic	NA	*ruuh	#2053
ɽuət^{6} (ruột)	intestine	body	*rɔ:c	*ruuc, *ruəc	#844
ʂa:j^{3} (sải)	fathom (n)	spatial	*pə.la:s	*lais	#1944
sa:k^{5} (xác)	body, carcass	body	NA	*cak	#290
ʂap^{5} (sắp)	ready, prepared	cognition	*srap	*srap	#1282
ʂəm^{5} (sấm)	thunder	nat. env.	*kə.rəmʔ	*grəmʔ, *gram	#1392
ʂo:ŋ^{1} (sông)	river	nat. env.	*kro:ŋ	*ruŋ, *ruuŋ, *ruəŋ	#668
ʈam^{1} (trăm)	hundred	quantity	*klam	*klam	#1405
va:^{3} (vả)	hit, slap (v)	basic	*tə.pah	*pah	#2022.a
vat^{5} (vắt)	squeeze, wring (v)	basic	*pat	*pat	#1021.a
viən^{6} (vượn)	gibbon	animal	*kə.vaŋ	*kwaŋ, *kwaaŋʔ	#934
vo:j^{1} (vôi)	lime (mineral)	nat. env.	*kə.pu:r	*knpur	#1636
ʑiət^{5} (giết)	die (v)	body	*kə.ce:t	*kcət	#987

Notes

1 Questions of the development of phonation features—and thus potential suprasegmental developments early in Vietic phonological history (e.g., Tạ 2021)—cannot yet be adequately addressed.

2 Early Austroasiatic speakers are posited to be among the groups of original hunter–gatherers and incoming agriculturalists in northern Vietnam at the Man Bac archaeological site dated to nearly 4000 years before the present (Higham 2017). A continuous chain of archaeological periods from that time led to the Dong Son culture (c. 700 BCE to 200 CE) of that region. Vietnamese has a significant number of Han Dynasty Chinese loanwords, providing support for a Vietic presence in the region at that time and also connecting Vietnamese to the end point of that chronological chain (Alves 2022).

[3] The exact number of Muong lects is not known, but there is considerable variety. Nguyễn (2005) provides data on 30 varieties of Muong.

[4] The past situation of varieties of Muong with respect to this development is less clear. Many do not have lenited onsets, so they lack the evidence Vietnamese has. However, some varieties do have them, but they lack history textual evidence, so it is not yet possible to provide chronological information with certainty that these were not borrowed from Vietnamese with those sounds.

[5] Vietnamese dialectal variety is more complex. The IPA transcriptions here are for standard northern Vietnamese. Dialects in other regions have other realizations, such as /j/ for 'v', 'gi', and 'd' and unsoftened / / for 'g/gh' in southern and central Vietnamese, and other variants in small areas of north–central Vietnamese. These variants still highlight the phonetic effect of their past intervocalic position.

[6] Claims of loanwords in Vietnamese from other languages are beyond the scope of this study, but broadly, possible early Tai loanwords (perhaps two dozen (Alves 2022)) are extremely small in number compared to the thousands of Chinese loanwords. In the Late Colonial period, hundreds of words were borrowed from French (Schlovin 2018), though only some of those words are highly integrated into Vietnamese, while many of these are not in common usage, as they are old-fashioned, highly technical, or not a common part of Vietnamese culture (e.g., *a-nốt* 'anode', *flăng* 'flan', etc.).

[7] Cognates of many native words can also be found in other Vietic languages, including several hundred that have been reconstructed to the Proto-Vietic level.

[8] Some words in minority Vietic languages are likely recent loanwords from Vietnamese, not retentions from an earlier common proto-language.

[9] Determining loanwords neighboring languages is done by searching through various digitized lexical databases, proto-language reconstructions, and various digital dictionaries. A list of these resources is provided in Appendix A.

[10] Such derivationally related cognates are not easily located in available data, but the Vietic language Ruc has *tùnú t* 'stopper', corresponding to Vietnamese *nu:t*^{5} (*nút*), though the data does not contain a source verb. Whether this represents an earlier stage in Vietic, borrowing, or a chance shared innovation cannot be known.

[11] There are some instances in Chinese of alliteration, as in Cantonese 咧啡 *le*^{5} *fe*^{5} 'messy', 撈捎 *laau*^{4} *saau*^{4} 'messy', 立亂 *lap*^{6} *lyun*^{6} 'chaotic', and others. English has such words as well (e.g., 'hocus-pocus', 'flim-flam', etc.), but such word formation -patterns are not as productive in English or varieties of Chinese as in most Austroasiatic languages.

References

Alves, Mark. 2006. Linguistic research on the origins of the Vietnamese language: An overview. *Journal of Vietnamese Studies* 1: 104–30. [CrossRef]

Alves, Mark. 2009. Loanwords in Vietnamese. In *Loanwords in the World's Languages*. Edited by Martin Haspelmath and Uri Tadmor. New York: De Gruyter Mouton, pp. 617–37. [CrossRef]

Alves, Mark. 2018. Early Sino-Vietnamese lexical data and the relative chronology of tonogenesis in Chinese and Vietnamese. *Bulletin of Chinese Linguistics* 11: 3–33. [CrossRef]

Alves, Mark. 2020a. Báo cáo tóm tắt về từ Proto-Nam Á và Proto-Vietic trong tiếng Việt [A Brief Report on Proto-Austroasiatic and Proto-Vietic Vocabulary in Vietnamese]. Paper presented at Hội Thảo Quốc Tế Ngôn Ngữ Học Lần Thứ IV, Vietnam, December 19–20, unpublished paper for keynote talk.

Alves, Mark. 2020b. Historical ethnolinguistic notes on Proto-Austroasiatic and Proto-Vietic vocabulary in Vietnamese. *Journal of the Southeast Asian Linguistics Society* 13: xiii–xlv. Available online: http://hdl.handle.net/10524/52472 (accessed on 5 December 2024).

Alves, Mark. 2020c. Initial steps in reconstructing Proto-Vietic syntax. In *Austroasiatic Syntax in Areal and Diachronic Perspective*. Edited by Mathias Jenny, Paul Sidwell and Mark Alves. Boston: Brill, pp. 46–81.

Alves, Mark. 2022. The Đông Sơn speech community: Evidence for Vietic. *Crossroads* 19: 138–74. [CrossRef]

Alves, Mark. 2024. From Vietic presyllables to Vietnamese simplex onsets. *Taiwan Journal of Linguistics* 22: 1–46. [CrossRef]

Ferlus, Michael. 1977. L'infixe instrumental <rn> en Khamou et sa trace en Vietnamien. *Cahiers de Linguistique, Asie Orientale* 2: 51–55.

Ferlus, Michel. 1992. Histoire abrégée de l'évolution des consonnes initiales du Vietnamien et du Sino-Vietnamien. *Mon-Khmer Studies Journal* 20: 111–25.

Gage, William W. 1985. Vietnamese in Mon-Khmer perspective. In *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt*. Edited by Suriya Ratanakul, David D. Thomas and Suwilai Premsrirat. Salaya: Institute of Language and Culture for Rural Development, Mahidol University, pp. 493–524.

Gehrmann, Ryan. 2022. Desegmentalization: Towards a Common Framework for the Modeling of Tonogenesis and Registrogenesis in Mainland Southeast Asia with Case Studies from Austroasiatic. Doctoral dissertation, the University of Edinburgh, Edinburgh, UK.

Haudricourt, André-Georges. 1953. La place du vietnamien dans les langues austroasiatiques. *Bulletin de la Société de linguistique de Paris* 49: 122–28.

Higham, Charles F. W. 2017. First farmers in mainland Southeast Asia. *Journal of Indo-Pacific Archaeology* 41: 13–21. [CrossRef]

Higham, Charles. 2021. The Neolithic occupation of Southeast Asia. In *The Languages and Linguistics of Mainland Southeast Asia: A Comprehensive Guide*. Edited by Paul Sidwell and Mathias Jenny. Boston: De Gruyter Mouton, pp. 21–32.

Hô, Lê. 1992. Từ Nam Á trong tiếng Việt [Austroasiatic words in Vietnamese]. In *Tiếng Việt va Các Ngôn Ngữ Dân Tộc Phía Nam [Vietnamese and Languages of Ethnic Groups in the South]*. Hanoi: Nhà Xuất Bản Khoa Học Xã Hội, pp. 65–110.

Hùng Việt, Lê Xuân Thại, Lý Toàn Thắng, Nguyễn Hoàng Anh, Trịnh Thị Hà, Nguyễn Thị Tân, and Dương Thị Thu Hà. 2018. *Từ Ngữ Hán Việt: Tiếp Nhận và Sáng Tạo [Vietnamese Vocabulary: Receiving and Creating]*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.

Jenny, Mathias. 2020. Verb-initial structures in Austroasiatic languages. In *Austroasiatic Syntax in Areal and Diachronic Perspective*. Edited by Mathias Jenny, Paul Sidwell and Mark Alves. Boston: Brill.

Jenny, Mathias, Tobias Weber, and Rachel Weymuth. 2015. The Austroasiatic languages: A typological overview. In *The Handbook of Austroasiatic Languages*. Edited by Mathias Jenny and Paul Sidwell. Boston: Brill, vol. 1, pp. 13–43.

Kirby, James. 2008. vPhon: A Vietnamese Phonetizer. Available online: https://github.com/kirbyj/vPhon (accessed on 5 August 2024).

Kirby, James, and Mark Alves. 2022. Exploring Statistical Regularities in the Syllable Canon of Sino-Vietnamese Loanmorph Phonology. In *Vietnamese Linguistics: State of the Field*. Edited by Trang Phan, John Phan and Mark Alves. JSEALS Special Publication No. 9. Honolulu: University of Hawai'i Press, pp. 104–26.

Maddieson, Ian. 2013. *Glottalized Consonants. In WALS Online (v2020.3) [Data Set]*. Edited by Matthew S. Dryer and Martin Haspelmath. Available online: http://wals.info/chapter/7 (accessed on 5 September 2024).

Nguyễn, Nhã Bản (chủ biên). 1999. *Từ Điển Tiếng Địa Phương Nghệ Tĩnh [Dictionary of the Nghe Tinh Dialect]*. Hanoi: Văn hoá thông tin.

Nguyễn, Văn Tài. 2005. *Ngữ âm tiếng Mường qua các phương ngôn [The Phonetics of the Mường Language Across Its Various Dialects]*. Hanoi: Nhà Xuất bản Từ điển Bách khoa.

Phan, John D. 2012. Mường is not a subgroup: Phonological evidence for a paraphyletic taxon in the Viet-Muong sub-family. *Mon-Khmer Studies Journal* 40: 1–18.

Phan, John D. 2013. Lacquered Words: The Evolution of Vietnamese under Sinitic Influences from the 1st Century BCE through the 17th Century CE. Doctoral dissertation, Cornell University, Ithica, NY, USA.

Pittayaporn, Pittayawat. 2009. The Phonology of Proto-Tai. Doctoral dissertation, Cornell University, Ithica, NY, USA.

Pittayaporn, Pittayawat. 2015. Typologizing sesquisyllabicity: The role of structural analysis in the study of linguistic diversity in Mainland Southeast Asia. In *Languages of Mainland Southeast Asia: The State of the Art*. Edited by Nick. J. Enfield and Bernard Comrie. Volume 649 in the series Pacific Linguistics. Berlin, München and Boston: De Gruyter Mouton, pp. 500–31. [CrossRef]

Rhodes, Alexandre de. 1991. *Từ Điển An Nam-Lusitan—La Tinh (Thường Gọi là Từ Điển Việt—Bồ-La) [An Annamese-Portuguese-Latin Dictionary]*. Translated by Thanh Lãng, Hoàng Xuân Việt, and Đỗ Quang Chính. Viện Khoa Học Xã Hội Tại Thành Phố Hồ Chí Minh. Ho Chi Minh: Nhà Xuất Bản Khoa Học Xã Hội. First published 1651.

Schlovin, Vera. 2018. A Comprehensive Corpus of French Loanwords into Vietnamese. Unpublished File. Available online: https://refubium.fu-berlin.de/handle/fub188/23231 (accessed on 15 June 2024).

Shimizu, Masaaki. 2015. A Reconstruction of Ancient Vietnamese Initials Using Chữ Nôm Materials. *NINJAL Research Papers* 9: 135–58.

Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Edited by Paul Sidwell, Doug Couper and Christian Bauer. Canberra: Pacific Linguistics.

Sidwell, Paul. 2008. Issues in the morphological reconstruction of Proto-Mon–Khmer. In *Morphology and Language History: In Honour of Harold Koch*. Edited by Claire Bowern. Philadelphia: John Benjamins.

Sidwell, Paul. 2024. 500 Proto Austroasiatic Etyma: Version 1.0. *Journal of the Southeast Asian Linguistics Society*, 17, i–xxiii.

Sidwell, Paul, and Mark Alves. 2021. The Vietic languages: A phylogenetic analysis. *Journal of Language Relationship* 19: 166–94.

Sidwell, Paul, and Mark Alves. 2023. Re-evaluating Shorto's MKCD reconstructions. In *Papers from ICAAL 9 & 10*. Edited by Mark Alves and Paul Sidwell. JSEALS Special Publication No. 12. Honolulu: University of Hawai'i Press, pp. 98–126.

Tạ, Thành Tấn. 2021. The Tone System of Rục and Tonogenesis in Vietic Languages. *Tạp Chí Phát Triển Khoa Học và Công Nghệ—Khoa Học Xã Hội và Nhân Văn* 5: 955–76. [CrossRef]

Thomas, David D., and Robert K. Headley, Jr. 1970. More on Mon-Khmer subgroupings. *Lingua* 25: 398–418. [CrossRef]

Trần, Thị Thúy Hiền, and Nathalie Vallée. 2009. An acoustic study of interword consonant sequences in Vietnamese. *Journal of the Southeast Asian Linguistics Society* 1: 231–50.

Viện Ngôn Ngữ Học [The Institute of Linguistics]. 1995. *Từ Điển Từ Láy Tiếng Việt [Dictionary of Vietnamese Reduplicative Words]*. Hanoi: Trung Tâm Khoa Học Xã Hội và Nhân Quốc Gia.

Vu, Duc Nghieu. 2019. Vietnamese Initial Consonant Clusters in Quốc Ngữ Documents from the 17th to Early 19th Centuries. *Journal of the Southeast Asian Linguistics Society* 12: 143–62.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

The Sino-Vietnamese Negative Prefixes *bất*, *vô*, *phi* and Their Coexistence with Sentential Negators: A Synchronic and Diachronic Analysis

Giorgio Francesco Arcodia and Trang Phan *

Department of Asian and North African Studies, Ca' Foscari University of Venice, Dorsoduro 1686, Campo San Sebastiano, 30123 Venice, Italy; giorgio.arcodia@unive.it
* Correspondence: thihuyentrang.phan@unive.it

Abstract: This paper presents a comprehensive synchronic and diachronic analysis of the Sino-Vietnamese negative prefixes *bất* (Chinese 不 *bù*), *vô* (無 *wú*), and *phi* (非 *fēi*), examining their historical development and modern usage in Vietnamese, with a comparative perspective on their Chinese equivalents. By investigating the interaction between these prefixes and Vietnamese sentential negators—such as the native *chẳng* and the Chinese-derived *không*—the study explores the evolution of negation in Vietnamese over several centuries. The research draws on a corpus of three bilingual Classical Chinese–Vietnamese translations of Confucius's *Analects* from the 17th, 19th, and 21st centuries, two written in traditional *Nôm* script and one in the modern *Quốc ngữ* alphabet. This corpus provides valuable insights into linguistic shifts driven by language contact in Vietnam. The findings reveal that in the 17th century, the Sino-Vietnamese prefixes *bất*, *vô*, and *phi* were largely absent, with native *chẳng* dominating. By the 19th century, *chẳng* persisted, but *không* emerged as a sentential negator, and *bất* appeared, both reflecting Chinese forms and demonstrating innovative uses. In the 21st century, *không* became the dominant negator, with *bất* and *vô* seeing increased usage, reflecting broader trends of linguistic modernization. This study situates these changes within the broader context of 20th-century East Asian literacy expansion, where Japan played a pivotal role in disseminating modernized Chinese-based vocabulary. By examining the selective adaptation and integration of Sino-Vietnamese elements, this paper contributes to a deeper understanding of language contact, syntactic influence, and lexical innovation in the evolving Vietnamese lexicon.

Keywords: negative prefix; sentential negator; Sino-Vietnamese

1. Introduction

This paper investigates the Sino-Vietnamese negative prefixes *bất*, *vô*, and *phi*, tracing their historical development and modern usage in Vietnamese, while comparing them with their Chinese counterparts 不 *bù*, 無 *wú*, and 非 *fēi*. Although significant work exists on Sino-Vietnamese linguistic interaction (Haudricourt, 2010; Hashimoto, 1978; T. C. Nguyễn, 1979; Alves, 2007a, 2007b, 2017; Shimizu, 2017, a.o.), less attention has been paid to the specific evolution of these negative prefixes and their interaction with native Vietnamese sentential negators like *chẳng* and the Chinese-derived *không*. This study seeks to address this gap by focusing on both the synchronic and diachronic changes in the Vietnamese negation system.

1.1. The Vietnamese Negation System

Vietnamese negation operates on two levels: sentential and lexical. At the sentential level, contemporary Vietnamese primarily uses ***không*** as the standard negator[1], which originates from the Chinese word 空 *kōng* 'empty, void' (P. P. Nguyễn, 1994; Trinh, 2005; Phan, 2023). However, ***không*** only became a sentential negator in the late 17th century (Phan et al., 2024), whereas *chẳng*, a native negator, was dominant before the mid-19th century. By the 19th and 20th centuries, ***không*** overtook *chẳng* as the primary negator in Vietnamese (Trinh et al., 2024).

At the lexical level, Sino-Vietnamese prefixes like *bất*, *vô*, and *phi* have been productive in forming negative adjectives, nouns, and verbs (Alves, 2007a, 2007b). However, little research has explored when these prefixes became productive in Vietnamese or how they interact with sentential negators such as ***không*** and *chẳng*. This paper provides a detailed analysis of how these negative prefixes evolved and how they coexist and compete with native negators in the Vietnamese language.

1.2. Why Focus on Vietnamese Translations of the Analects?

A key contribution of this study is its analysis of three bilingual Classical Chinese–Vietnamese translations of Confucius's *Analects* from the 17th, 19th, and 21st centuries. The *Analects* (論語 *Lúnyǔ*), a cornerstone of Confucian philosophy, had a profound cultural and linguistic influence in both China and Vietnam (Lau, 1979; Ames & Rosemont, 1998; Gardner, 2003). These translations, two written in the traditional character-based ***Nôm*** script and one in the modern *Quốc ngữ* alphabet, offer a unique opportunity to observe how Sino-Vietnamese prefixes (*bất*, *vô*, *phi*) were integrated into Vietnamese alongside native negators like *chẳng* and ***không***. By analyzing this corpus, we can observe patterns of linguistic coexistence, competition, and innovation, revealing shifts in negation usage over time.

1.3. Chinese Negators and Their Vietnamese Adaptations

In the history of Chinese, 不 ***bù***, 無 *wú*, and 非 *fēi* have been used as core negators at both the sentential and lexical levels (Pulleyblank, 1995; Li & Thompson, 1981). When these elements were borrowed into Vietnamese, they were restricted to word formation, functioning only as prefixes to form negative adjectives, nouns, and verbs. In contrast, Vietnamese relied on native negators such as *chẳng*, with ***không*** later incorporated from Chinese. As a result, in Vietnamese translations of Chinese texts, translators utilized both Sino-Vietnamese prefixes and sentential negators to accurately maintain the original meaning.

1.4. Classical Chinese and the Evolution of Literary Sinitic in East Asia

'Classical Chinese' is used to refer to the form of written Chinese used in classical texts from the late Spring and Autumn period to the Han Dynasty (roughly corresponding to Peyraube's (1996) 'Late Archaic Chinese'), including Confucius's *Analects*. These texts, particularly the Confucian classics, shaped the intellectual history of East Asia. In English-language Sinology, the terms 'Classical Chinese' and 'Literary Chinese' have been traditionally used to describe this variety and the written variety modeled on the classics which was used in later epochs (i.e., since the first century CE, known in Chinese as 文言 *wényán*), respectively. Mair (2004) introduced the term 'Literary Sinitic' to replace 'Classical Chinese' or 'Literary Chinese'. In a recently published work, King (2021) proposed using the new term 'Literary Sinitic' to both refer to the term of wényán used in modern Chinese, and also to refer to the concept of 漢文 *hànwén* in other East Asian countries (kanbun in Japan, hammun in Korea, and Hán văn in Vietnam) because Literary Sinitic is the most important linguistic feature of classical texts in East Asian countries.

In Vietnam, Chinese characters were introduced during the Nam Việt 南越 kingdom period (203–111 BCE) and became entrenched during the Chinese domination of 交州 *Jiāozhōu*/Giao Châu (modern-day northern Vietnam) from 179 BCE to 938 CE (T. C. Nguyễn & Bùi, 2020). By the 8th and 9th centuries, under the influence of the Chinese language in Tang dynasty, the Hán-Việt (Sino-Vietnamese) reading system emerged, a system that remains in use today (T. C. Nguyễn, 1979; Phan, 2013, 2025)[2]. After Vietnam's independence, Chinese characters (*chữ Hán*) remained the official script for administrative documents. The Confucian civil service examination system (1075–1919) ensured the continued mastery of Literary Sinitic, allowing scholars to compose official documents and literary works in this written variety. Surveys show that the nine most important Confucian classics, including the *Analects*, were all translated into Vietnamese and recorded in *chữ Nôm* during the monarchal period (T. C. Nguyễn, 2016). These translations, along with later versions in *Quốc ngữ*, cemented *wényán*'s long-standing impact on Vietnamese linguistic and intellectual history.

Choosing the *Analects* for this study offers several advantages. First, its mix of dialogue, narrative, and didactic passages provides a rich array of syntactic environments in which negation appears. Second, the availability of translations of the *Analects* from the 17th, 19th, and 21st centuries provides a well-spaced corpus for examining linguistic change over time. Third, its broad dissemination—first via woodblock printing and later through modern editions—means these versions were read by diverse audiences, helping us gauge both scholarly and popular language use.

1.5. Aims and Main Findings

The primary objective of this study is to offer a synchronic and diachronic analysis of the Sino-Vietnamese negative prefixes *bất*, *vô*, and *phi*, comparing their usage with their Chinese equivalents. It also explores the interaction between these prefixes and sentential negators, revealing patterns of linguistic coexistence and innovation.

Our key findings are as follows:

(i) 17th-Century Native Dominance: In the 17th-century translation, *chẳng* dominates the negation system, with no occurrences of the prefixes *bất*, *vô*, or *phi*, reflecting the dominance of indigenous negators during this period.

(ii) 19th-Century Transition: By the 19th century, while *chẳng* remains prevalent, *không* emerges as a sentential negator. This period also sees the appearance of *bất*, which corresponds directly to Chinese forms (e.g., *bất nhân* 不仁) or appears in innovative formations (e.g., *bất tài* 不才), signaling the increasing influence of Chinese on Vietnamese negation patterns.

(iii) 21st-Century 'Modernization': In the 21st-century translation, *không* becomes the dominant negator, and there is a notable increase in the use of *bất*, *vô*, and *phi*, reflecting broader innovative trends in Vietnamese. This shift is part of the broader linguistic and cultural 'modernization' of East Asia, influenced by Japan's role in transmitting modernized Chinese characters through reformist writings (Sinh, 1993).

Structure of the Paper

This paper is organized as follows. Section 2 provides an overview of the Chinese negative prefixes 不 *bù*, 無 *wú*, and 非 *fēi* in Classical Chinese and their development in (Modern Standard) Mandarin. Section 3 focuses on the modern usage of *bất*, *vô*, and *phi* in Vietnamese, particularly on their interaction with native negators like *chẳng* and *không*. Section 4 introduces the historical corpus, and Section 5 presents a diachronic analysis of the three bilingual *Analects* translations. Section 6 discusses the broader implications of these findings and suggests future research directions.

2. 不 *bù*, 無 *wú*, and 非 *fēi* in Chinese

Chinese negation involves a nuanced system of items that have undergone significant changes throughout the history of the language. Among the most central negators are 不 *bù*, 無 *wú*, and 非 *fēi*, which have played key roles at both the sentential and lexical levels. This section focuses on their use and evolution from Classical Chinese to Modern Standard Mandarin, providing crucial context for understanding their adaptation into Vietnamese.

2.1. Classical Chinese

Given that the translations of the *Analects* chosen for this study are bilingual—featuring both Vietnamese and Classical Chinese—understanding the role of these negators in Classical Chinese is essential.

2.1.1. Verbless Nominal Predication: 非 *fēi*

In *Classical Chinese*, 非 *fēi* was commonly used to negate nominal predicates in verbless sentences. Unlike copular verbs in other languages, *fēi* is not considered a verb but rather a particle that negates the predicate directly (see Vogelsang, 2021). This structure, which was well established in the Classical period, became a defining feature of nominal negation throughout Premodern Chinese.

An example of *fēi* in a nominal predication may be found in the *Mencius*:

(1)

非	我	也，	兵	也
fēi	*wǒ*	*yě*,	*bīng*	*yě*
NEG	1SG	PTC,	weapon	PTC

'It was not I, it was the weapon'. (*Mencius*; Pulleyblank, 1995, p. 16)

In this example, *fēi* negates the subject 我 *wǒ* ('I'), contrasting it with 兵 *bīng* ('weapon'). The particle 也 *yě* is often omitted in negative clauses, in the Classical language.

In another example from the *Zhuangzi*, the nominal negation is simpler, and *fēi* serves again as the main negator:

(2)

子	非	魚
zǐ	*fēi*	*yú*
2SG	NEG	fish

'You are not a fish'. (Vogelsang, 2021, p. 34)

Here, *fēi* negates the subject–predicate relationship directly, contrasting 子*zǐ* ('you') with 魚*yú* ('fish'). The use of *fēi* in this manner highlights its role in creating binary contrasts in Classical Chinese.

Historically, 非 *fēi* evolved from the phrase 不唯 *bù wéi*, where *bù* functioned as a negator and 唯 *wéi* acted as an emphatic particle. Over time, 不唯 *bù wéi* ceased to be used, and 非 *fēi* became the standard form of nominal negation (Vogelsang, 2021, p. 32). Grammarians differ in their classification of 非 *fēi*: some treat it as a particle rather than an adverb (e.g., Guo et al., 1999), which makes it unique among negators in Chinese (Vogelsang, 2021).

2.1.2. Verbal Predication: 不 bù

Unlike 非 *fēi*, which negates nominal predicates, 不 *bù* is primarily used to negate verbal predicates. In Classical Chinese, and up until the stage of Modern Chinese, *bù* played a central role in negating actions and states. It could negate verbs and adjectives.

Consider the following example from the *Mencius*, where 不 *bù* negates the verb 見 *jiàn* ('see'):

(3)

就	之	而	不	見	所	畏	焉
jiù	*zhī*	*ér*	***bù***	*jiàn*	*suǒ*	*wèi*	*yān*
approach	3SG	and	NEG	see	NMLZ	fear	PTC

'Going up to him, I did not see anything to fear in him'.
(*Mencius*; Pulleyblank, 1995, p. 29)

Here, ***bù*** directly negates the act of seeing, marking the absence of the action. This example illustrates the broad applicability of ***bù*** to verbal negation in Chinese.

Additionally, 不 ***bù*** is used to negate adjectives, which are sometimes classified as stative/property verbs in Chinese. The following example shows the negation of an adjective:

(4)

日	月	不	高
rì	*yuè*	***bù***	*gāo*
sun	moon	NEG	high

'The sun and moon are not high'. (Vogelsang, 2021, p. 45)

In this example, 不 ***bù*** negates the adjective *gāo* (高, 'high'), treating it as a stative verb. This reflects the flexibility of 不 ***bù*** in Classical Chinese, as it can negate both actions and states.

2.1.3. Possessive and Existential Negation: 無 wú

The third negator we considered in the present study is 無 *wú*, which negates possession or existence. In Classical Chinese, 無*wú* was used to negate existential or possessive predicates, distinguishing it from 不 ***bù***, which negates actions or states.

For example, in the following sentence, *wú* negates the existence of 'gentlemen':

(5)

無	君子，	莫	治	野人
wú	*jūnzǐ*,	*mò*	*zhì*	*yěrén*
NEG	gentleman,	no one	rule	rustic-person

'If there were no gentlemen, there would be no one to rule the rustics'.
(*Mencius*; Pulleyblank, 1995, p. 29)

Here, *wú* negates the existence of 君子 *jūnzǐ* 'gentleman', building a conditional clause about governance.

2.2. Modern Standard Mandarin

In Modern Standard Mandarin, the negators 不 ***bù***, 沒 *méi*, and 無 *wú* have developed distinct syntactic functions, each associated with specific contexts of negation. This section explores how these negators operate in the modern language and compares their roles to their counterparts in Classical Chinese. This comparison is essential for understanding the Sino-Vietnamese prefixes *bất*, *vô*, and *phi* and their integration into Vietnamese linguistic structures.

2.2.1. Irrealis Negation: 不 *bù*

In Modern Standard Mandarin, 不 ***bù*** functions as the general negator for irrealis predicates, negating both verbs and adjectives. This is similar to its role in Classical Chinese, where ***bù*** negates verbs and adjectives directly, without the need for additional auxiliaries. Just as in Classical Chinese, predicative adjectives are sometimes conflated with verbs in the literature (see, e.g., Li & Thompson, 1981), but the criteria for distinguishing them from verbs are somewhat clearer in the modern language (see, e.g., Arcodia, 2012).

Example (6) illustrates the negation of an adjectival predicate:

(6)

張三	不	高
Zhāngsān	*bù*	*gāo*
Zhangsan	NEG	tall

'Zhangsan isn't tall'.

bù is also used to negate the copula 是 *shì*, as in example (7):

(7)

張三	不	是	日本人
Zhāngsān	*bù*	*shì*	*Rìběn-rén*
Zhangsan	NEG	COP	Japan-person

'Zhangsan is not Japanese'.

Here, *bù* negates the copula 是 *shì*, used in nominal predication. In Classical Chinese, *fēi* would have been used in such cases, without a copula (see Exx. 1–2). In Modern Standard Mandarin, 非 *fēi* is used only in lexical items and conventionalized constructions, while *bù* serves as the general negator for the copula.

bù is also the general irrealis negator for verbs, as may be seen in example (8):

(8)

我	不	去
wǒ	*bù*	*qù*
1SG	NEG	go

'I'm not going'.

In this sentence, the verb 去 *qù* ('go') is negated by *bù*. The pattern is similar to that found in Classical Chinese verbal negation, such as in example (3) from the *Mencius*.

2.2.2. Realis Negation: 没(有) *méi*(*yǒu*)

In contrast to 不 *bù*, 沒(有) *méi(yǒu)* is used to negate realis predicates: while *bù* covers irrealis contexts, 没 *méi* is mostly used to negate the occurrence of past episodic events.

For example, in (9):

(9)

我	没	去
wǒ	*méi*	*qù*
1SG	NEG	go

'I didn't go'.

Here, 没 *méi* negates the occurrence of the action of going. This differs from 不 *bù*, which negates a habitual or future event (compare Ex. 8).

Méi is also used as the existential negator: in (10), it negates the existential predicate 空 *kòng* ('free time'):

(10)

今天	我	没	空
jīntiān	*wǒ*	*méi*	*kòng*
today	1SG	NEG	free time

'I don't have time today'.

In this case, 沒 *méi* negates the existence of free time, much like 無 *wú* in Classical Chinese negates existence or possession, as example (5) above. However, in Modern Standard Mandarin, 沒 *méi* is also the primary negator for realis situations, especially when discussing past episodic events (9).

2.2.3. Word Formation

In contemporary standard Mandarin Chinese, the negators 不 *bù*, 無 *wú*, and 非 *fēi* play crucial roles not only in sentence-level negation but also in word formation, particularly in creating derived adjectives and nouns.

For instance, 不 *bù* is frequently used to form adjectives by negating the base adjective or noun. In example (11), 不 *bù* combines with 安 *ān* ('peace') to form 不安 *bù'ān* ('uneasy'), a predicative adjective that describes an emotional state:

(11)

他	很	不安
tā	*hěn*	*bù'ān*
3SG.M	very	uneasy

'He's feeling (very) uneasy'.

This pattern, where 不 ***bù*** is prefixed to an adjective or noun, demonstrates how 不 ***bù*** functions as a productive prefix in forming descriptive adjectives. However, 不 ***bù*** very often creates non-predicative adjectives, e.g., 不敗 *bùbài* ('invincible'); as shown in (12), such a non-predicative item cannot be used in the same predicative construction seen just above (11).

(12)

*他	很	不敗
tā	*hěn*	*bùbài*
3SG.M	very	invincible

(Meaning:) 'He's (very) invincible'.

Similarly, 無*wú* serves as a negator in morphologically complex words, typically indicating the absence of possession or attributes. In (13), for example, *wú* combines with 糖 *táng* ('sugar') to form 無糖 *wútáng* ('sugar-free'), a non-predicative adjective designating a food product lacking sugar:

(13)

無糖
wútáng
NEG sugar
'sugar-free'

In this case, 無 *wú* negates the noun 糖 *táng*, creating morphologically complex words that describe what something lacks, which is common in terms that describe attributes or states. Similarly to 不 *bù*, it can be sometimes used also to form predicative adjectives (e.g., 無力 *wúlì* 'powerless'), and we can find it also in nouns, again describing something lacking a property (e.g., 無產者 *wúchǎnzhě* 'proletarian', lit. 'one who lacks properties').

非 *fēi*, on the other hand, is used to form words that imply negation of a property, such as 非會員 *fēihuìyuán* ('non-member'). This prefix can form nouns, as 非會員 *fēihuìyuán*, but also non-predicative adjectives, both by combining with nouns (e.g., 非貿易 *fēimàoyì* 'non-commercial') and with adjectives (非官方 fēiguānfāng 'unofficial'; see Arcodia, 2012).

These examples of word formation highlight how negators in Mandarin Chinese are used not only to negate verbal or adjectival predicates but also to create new lexical items. In both Classical and Modern Chinese, 不 ***bù***, 非 *fēi*, and 無 *wú* serve distinct but related functions in negating actions, states, and existence. While 不 ***bù*** continues to dominate verbal and adjectival negation in modern Mandarin, and is also used in word formation, *wú* and *méi* have evolved specific functions, with 無 *wú* primarily used in lexical negation and *méi* in syntactic negation; 非 *fēi* is generally confined to word formation, in the modern language. These negators provide a foundation for understanding how Sino-Vietnamese prefixes like *bất*, *vô*, and *phi* were integrated into the Vietnamese lexicon.

In the following section, we examine how these elements, after being introduced into Vietnamese, interact with sentential negators like *chẳng* and *không*. By analyzing the Vietnamese translations of Confucius's *Analects*, we explore how Sino-Vietnamese negative prefixes coexist with sentential negative markers, setting the stage for the development of Vietnamese negation systems.

3. *Bất*, *vô*, and *phi* in Present-Day Vietnamese

A crucial distinction between the Chinese negators 不 ***bù***, 無 *wú*, and 非 *fēi* and their Vietnamese counterparts *bất*, *vô*, and *phi* lies in their syntactic scope. In Chinese, 不 ***bù***, 無 *wú*, and 非 *fēi* function at both the sentential and lexical levels, whereas in contemporary

Vietnamese, *bất*, *vô*, and *phi* are restricted to word formation (i.e., to the lexical level) and co-exist with independent sentential negators. This distinction is central to understanding how negation operates in Vietnamese, and how it diverges from its Chinese origins.

In present-day Vietnamese, negation is expressed through both sentential negators and negative prefixes. There are two primary sentential negators: *không*, which is derived from Chinese 空*kōng* and functions as the standard negator, and *chẳng*, a native negator that is considered non-standard and often adds an emphatic or non-neutral tone.

(16)	a.	Nó	**không**	đi	thư viện
		3SG	NEG	go	library
		'He didn't go to the library'.			
	b.	Nó	**chẳng**	đi	thư viện
		3SG	NEG	go	library
		'He didn't go to the library'.			

Both (16a) and (1tb) convey the meaning 'He didn't go to the library,' but the latter carries a stronger, more emphatic connotation due to the use of *chẳng*. While the use of sentential negators like *không* and *chẳng* has been widely investigated (P. P. Nguyễn, 1994; Phan, 2023; Trinh et al., 2024), there has been significantly less attention paid to the negative prefixes *bất*, *vô*, and *phi* in modern Vietnamese. These Sino-Vietnamese prefixes, which originate from their Chinese counterparts, are primarily used in the formation of negative adjectives, nouns, and verbs. This section, therefore, is devoted to examining the role and usage of *bất*, *vô*, and *phi* in present-day Vietnamese, shedding light on how they function within the broader system of Vietnamese negation.

Drawing on data from the *Great Dictionary of Vietnamese* (N. Y. Nguyễn, 1999), we compiled a list of 125 frequently used words containing either *bất*, *vô*, or *phi* in the prefixal position. From this list, 51 occurrences of *bất*, 67 occurrences of *vô*, and 7 occurrences of *phi* were identified, excluding archaic or idiomatic usages, as summarized in Table 1.

Table 1. The number of words containing bất, vô, phi in the *Great Dictionary of Vietnamese*.

	***Great Dictionary of Vietnamese* (1999)**
number of words containing *bất*	51
number of words containing *vô*	67
number of words containing *phi*	7

While these prefixes retain core semantic functions related to negation, their usage patterns in Vietnamese show significant divergence from Chinese, particularly in terms of allowable combinations and lexical productivity.

3.1. Bất in Present-Day Vietnamese

The prefix *bất* is used to create negative adjectives, verbs, and nouns in Vietnamese, often expressing negative qualities or states. In our data, we identified 51 frequently used words with *bất* such as:

(17)	*bất*	*an*	'insecure'
	bất	*biến*	'invariant'
	bất	*cẩn*	'careless'
	bất	*chính*	'unrighteous'
	bất	*công*	'unjust'

These examples illustrate how *bất* functions to negate qualities and states, forming compound words that convey negative meanings. This pattern aligns with its use in Modern Standard Chinese, where, as shown earlier (§2.2.3), 不 *bù* similarly negates adjectives

and nouns to express states of negation or absence, as in 不安 *bù'ān* ('uneasy') or 不正 *bùzhèng* ('unrighteous'). However, Vietnamese demonstrates selective adaptation, as some combinations commonly used in Chinese are not productive or acceptable in Vietnamese. For instance, while *bất tài* ('incompetent') is a viable term in Vietnamese, neither *vô tài* nor *phi tài* are considered acceptable combinations. This contrasts with Chinese, where both 不才 *bùcái* ('incompetent') and 無才 *wúcái* ('talentless') are permissible. This divergence highlights how Vietnamese has selectively integrated *bất* while maintaining constraints on its combinations. In this sense, *bất* has retained its core semantic function of negating qualities but has been limited in terms of lexical flexibility when compared to Chinese.

3.2. Vô in Present-Day Vietnamese

The prefix *vô* is more frequently used in Vietnamese than *bất*, with 67 occurrences identified in the data. *vô* generally negates qualities and states, often conveying the absence or lack of something. Common examples include:

(18)	*vô*	*tâm*	'unthoughtful'
	vô	*tận*	'endless'
	vô	*thần*	'atheist'
	vô	*tội*	'innocent/guilt-free'
	vô	*trách nhiệm*	'irresponsible'

These compounds highlight how *vô* in Vietnamese is used to negate the existence or presence of a quality, much like 無 *wú* in Chinese. However, Vietnamese demonstrates unique constraints on how *vô* can be used.

3.3. Phi in Present-Day Vietnamese

The prefix *phi* is the least frequently used of the three, with only seven occurrences identified in the data. However, *phi* plays an important role in forming compound words related to non-standard or non-normative categories. Some examples include:

(19) *phi đạo đức* 'unethical'
phi lý 'irrational'
phi pháp 'illegal'
phi quân sự 'non-military'
phi thường 'extraordinary'
phi nghĩa 'unrighteous'
phi phàm 'extraordinary'

The prefix *phi* typically negates normative categories, indicating something that deviates from social or legal standards. This usage parallels that of 非 *fēi* in Mandarin, where it negates similar concepts, such as in 非貿易 *fēimàoyì* ('non-commercial') or 非會員 *fēihuìyuán* ('non-member'). However, as seen with *bất*, Vietnamese imposes restrictions on how *phi* can combine with certain base words. While *phi phàm* ('extraordinary') is commonly used, *bất phàm* is not, even though the equivalent forms in Chinese are both acceptable.

Across all three prefixes—*bất*, *vô*, and *phi*—Vietnamese demonstrates a strong parallel with Chinese. For example, combinations like *bất an* (insecure) in Vietnamese closely align with the Chinese 不安 *bù'ān*, and *vô biên* (boundless) mirrors 無邊 *wúbiān* in Chinese. This alignment suggests that Vietnamese has maintained much of the original form and function of these prefixes as they were borrowed from Chinese. Particularly in formal and philosophical domains, the overlap between the two languages is significant, as seen in terms like *bất khả thi* (impossible), which directly corresponds to 不可能 *bù kěnéng* in Chinese. This reflects the deep and enduring influence of Chinese on the Vietnamese lexicon.

However, a crucial difference emerges in how these prefixes function within each language. In the history of Chinese, 不 ***bù***, 無 *wú*, and 非 *fēi* have operated flexibly at the sentential and lexical levels, although only 不 ***bù*** is productively used as a syntactic negator in the modern language. In contrast, their Vietnamese counterparts—*bất*, *vô*, and *phi*—are strictly confined to lexical negation, forming negative compounds such as adjectives, nouns, and verbs. This aligns with typological studies showing that core grammatical elements like negation are more resistant to borrowing than lexical items (Heine & Kuteva, 2005; Sakel, 2007). Sentential negation in Vietnamese is handled separately by the native negators *không* and *chẳng*, which coexist alongside the Sino-Vietnamese prefixes. This compartmentalization of negation in Vietnamese, where sentential and lexical negators are separated, marks a significant structural divergence from Chinese. It illustrates how Vietnamese, while adopting Chinese lexical elements, has tailored its negation system to its own syntactic norms. Vietnamese also demonstrates a degree of innovation with these borrowed prefixes. In some cases, Vietnamese has created terms that do not have exact counterparts in Chinese, reflecting creative linguistic adaptations. This dynamic process of borrowing and adaptation exemplifies how languages can integrate foreign elements while still evolving in unique ways.

In the next section, we explore how these prefixes function in translations of Confucius's *Analects*, examining their appearance across the 17th, 19th, and 21st centuries. This analysis provides further insights into the interaction between Sino-Vietnamese prefixes and sentential negators, and how these elements have coexisted and evolved in Vietnamese over time. Specifically, we focus on the timing of their appearance and their frequency in translated texts, shedding light on broader patterns of language contact and adaptation.

4. Data Collection

To trace the evolution of Sino-Vietnamese negators from Classical Chinese to modern Vietnamese, this study examines three key bilingual texts spanning the 17th to 21st centuries: *Luận ngữ ước giải* (17th century), *Luận ngữ thích nghĩa ca* (19th century), and *Khổng phu tử và Luận ngữ* (21st century). These texts, blending Classical Chinese with Vietnamese translations, offer a unique corpus for analyzing the integration of the prefixes *bất*, *vô*, and *phi* alongside sentential negators such as *chẳng* and *không*.[3]

4.1. Luận ngữ ước giải (Concise Interpretation to the Analects) (17th Century)

Luận ngữ ước giải 論語約解, known as the *Concise Interpretation to the Analects*, is the earliest extant Vietnamese translation of Confucius's *Analects*, composed in the 17th century. Its full title is '*Tứ thư ước giải Luận ngữ giải quốc âm nghĩa tăng bổ Đại toàn bị chỉ*' 四書約解論語解國音義增補大全備旨 (Explanation of the *Analects* using national pronunciation and meaning, supplemented with important annotations from the *Daquan* 大全, part of the *Concise Interpretation to the Four Books*). In this paper, it is referred to as *Luận ngữ ước giải*. It forms part of the *Tứ thư ước giải* 四書約解 (*Concise Interpretation to the Four Books*), which translates and annotates the *Four Books of Confucianism* into Vietnamese using the Nôm script, accompanied by Literary Sinitic annotations. The translator remains anonymous, and the text is housed at the Institute of Sino-Nom Studies (Hanoi, Vietnam), with call number AB.270/1–5.

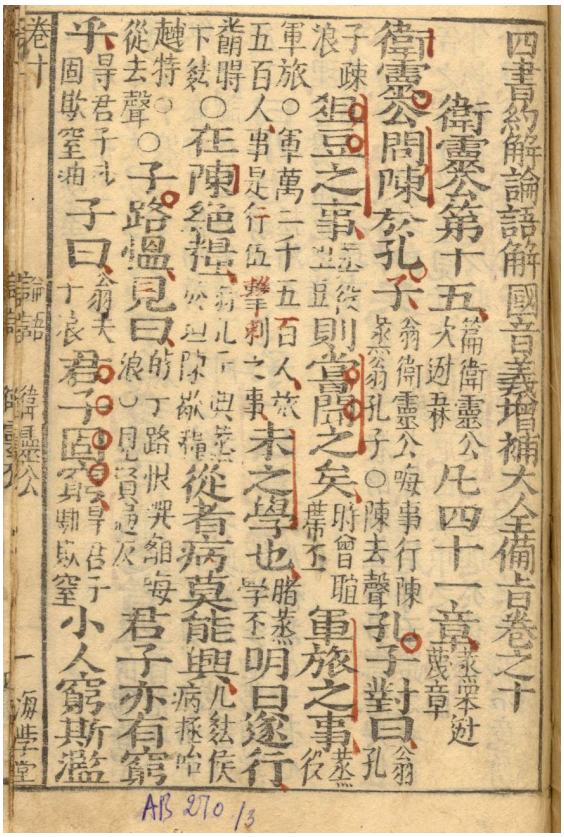
Although the book cover dates the text to the 20th year of Emperor Minh Mạng's reign (1839), subsequent research by T. C. Nguyễn (2009), T. C. Nguyễn (2014) revealed that the composition was completed in the 17th century. The wood blocks used for printing were re-carved multiple times from the 17th to the late 19th century. Despite the cover date of 1839, the text itself, written in Nôm script and Vietnamese language, retains the characteristics of 17th-century language and script.

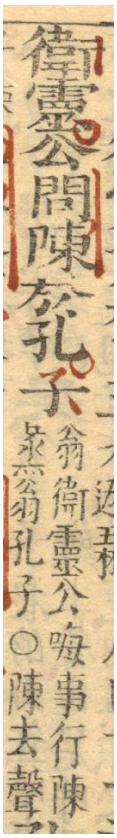
The *Luận ngữ ước giải* originally consisted of 19 volumes, but only 12 volumes remain, with volumes 3–9 missing. The remaining portion includes the last six chapters of the *Analects*, from chapter 15 'Weiling Gong' 衛靈公 to chapter 20 'Yao yue' 堯曰, spanning 98 pages. To maintain a strict comparative framework, our study focuses exclusively on these final six chapters, comparing them across the 17th, 19th, and 21st-century Vietnamese translations alongside the original text in Literary Sinitic.

The format of *Luận ngữ ước giải* is bilingual, with large traditional Sinographs used for the original Literary Sinitic text and smaller Nôm characters for the Vietnamese translation. The structure of each paragraph in the *Luận ngữ ước giải* is divided into three parts, as illustrated in Table 2:

(i) Original Text in Literary Sinitic: The main text is written in large traditional Sinographs at the beginning of each paragraph.
(ii) Nôm Translation: Below the original text, small-sized Nôm characters provide the Vietnamese translation.
(iii) Annotation Section: Small Sinographs give additional annotations, typically sourced from the *Great Complete Works of the Four Books* (四書大全), compiled during the Ming dynasty by a group led by Hu Guang (胡廣, 1369–1418). A circular bullet point separates the Nôm translation from the annotation, where present as they both use small font sizes.

Table 2. The first page of the remaining portion of the *Luận ngữ ước giải*, within the *Tứ thư ước giải*, call number AB.270/3, page 1a.





1. Large traditional Sinographs, original text: 衛靈公問陳於孔子 *Vệ Linh công vấn trận ư Khổng tử* (Sino-Vietnamese reading).
Weiling Gong inquired of Confucius regarding the matter of warfare.
2. Small-sized Nôm characters, Vietnamese translation: 翁衛靈公嗨事行陣蒸翁孔子 *ông Vệ Linh công hỏi sự hành trận chưng ông Khổng tử* (Vietnamese phonetic transcription).
The elder Weiling Gong asked about the conduct of warfare to Master Confucius.
3. Circular bullet point and small-sized Sinographs annotation: ¡陳去聲 *Trận, khứ thanh. (Sino-Vietnamese reading).*
The word 'Trận' has a departing tone.[4]

The first page of the remaining portion of the *Luận ngữ ước giải*, within the *Tứ thư ước giải*, call number AB.270/3, page 1a

Exposition of the translation of the opening passage of chapter 15 'Weiling Gong.'

This format, with its bilingual structure and detailed annotations, allows for in-depth analysis of how negators in Literary Sinitic were translated into Vietnamese, particularly the use of 不 *bất*, 無 *vô*, and 非 *phi*.

4.2. Luận ngữ thích nghĩa ca (Song of Interpreting the Meaning of the Analects) (19th Century)

Luận ngữ thích nghĩa ca 論語釋義歌, or *Song of Interpreting the Meaning of the Analects*, is a 19th-century Vietnamese translation of the *Analects* with 2970 lines of verse. Its full title is *Tự Đức thánh chế Luận ngữ thích nghĩa ca* 嗣德聖制論語釋義歌 (The Song of Interpreting the Meaning of the Analects Composed by Emperor Tự Đức), hereafter referred to as the *Luận ngữ thích nghĩa* ca. This translation was compiled by Emperor Tự Đức (1828–1883), a monarch of the Nguyễn dynasty renowned for his poetic compositions, during his reign from 1848 to 1883. It was authorized for engraving and printing by the court around the years 1896–1907 after the author's death. The text exhibits the linguistic and script characteristics (Vietnamese language and Nôm script) of the second half of the 19th century, reflecting the time of the author's completion of the work. The composition incorporates many expressions from the Huế dialect, the capital of the Nguyễn dynasty (1802–1945), such as the words *mần răng* 嗼凌 (how), *mần ri* 嗼夷 (like this), *chừ* 除 (now), and *ni* 尼 (this).

At the Institute of Sino-Nom Studies (Hanoi, Vietnam), there are currently two sets of books with call numbers A.186/1-2 and HVv.709/3-6, with the latter being incomplete. Therefore, this article references the first set. Call number A.186/1-2 comprises seven volumes, divided into two booklets, totaling 1,104 pages. Corresponding to the last six chapters of the remaining portion of the *Luận ngữ ước giải* presented above, from chapter 'Weiling Gong' to chapter 'Yao yue,' the *Luận ngữ thích nghĩa ca* cover from the latter part of volume 5 to the end of volume 7, totaling 760 lines of verse (26% of the total length of 2970 lines).

Luận ngữ thích nghĩa ca is written in the poetic *lục bát* (six-eight, i.e., the number of syllables in every two lines) verse form and uses Nôm script, a reflection of the linguistic and cultural environment of the Nguyễn dynasty. The text demonstrates the highly formalized nature of Vietnamese poetic composition in this era.

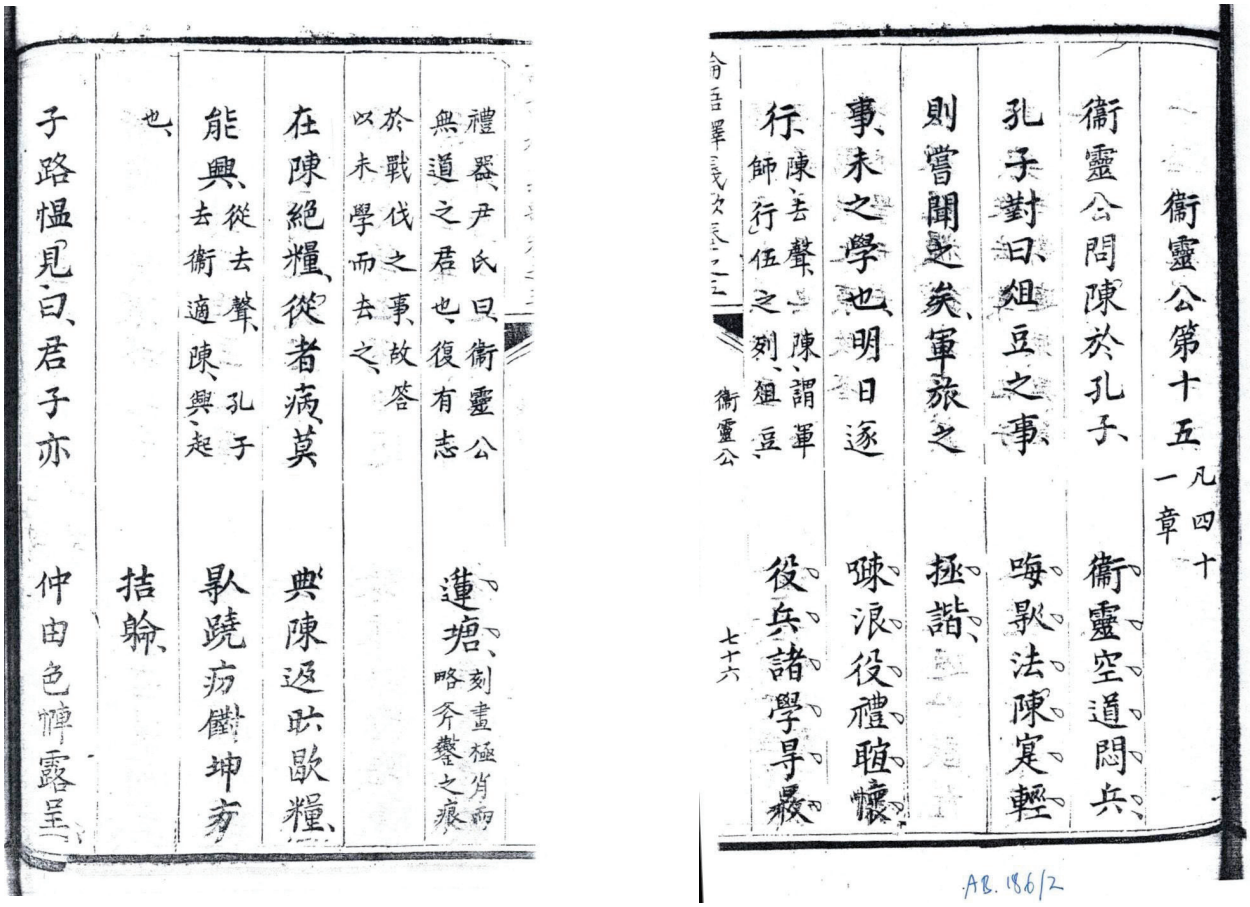
The presentation format of *Luận ngữ thích nghĩa ca* is structured into two tiers, as shown in Table 3:

(i) Top Tier: This contains (i) the original text of the *Analects* in large traditional Sinographs, accompanied by (ii) small Sinographs for annotations. These annotations are mainly drawn from Zhu Xi's (朱熹, 1130–1200) *Collected Annotations on the Analects* 論語集注.

(ii) Bottom Tier: (iii) The large Nôm characters in the bottom tier provide the Vietnamese translation in *lục bát* verse, followed by (iv) commentary in small Sinographs written by Nguyễn Phúc Miên Trinh 阮福綿寘 (1820–1897), an uncle of Emperor Tự Đức.

This text exemplifies how Vietnamese translators engaged with Classical Chinese texts using the *lục bát* form, while still incorporating complex philosophical concepts. The use of *bất*, *vô*, and *phi* in this poetic context provides insight into how Classical Chinese negators were translated into vernacular Vietnamese during the 19th century, when both Chinese influence and native linguistic forms coexisted.

Table 3. The *Luận ngữ thích nghĩa ca* (*Song of Interpreting the Meaning of the Analects*), call number AB.186/2, volume 5, pages 76a (right side) and 76b (left side).



The *Luận ngữ thích nghĩa ca* (*Song of Interpreting the Meaning of the Analects*), call number AB.186/2, volume 5, pages 76a (right side) and 76b (left side).

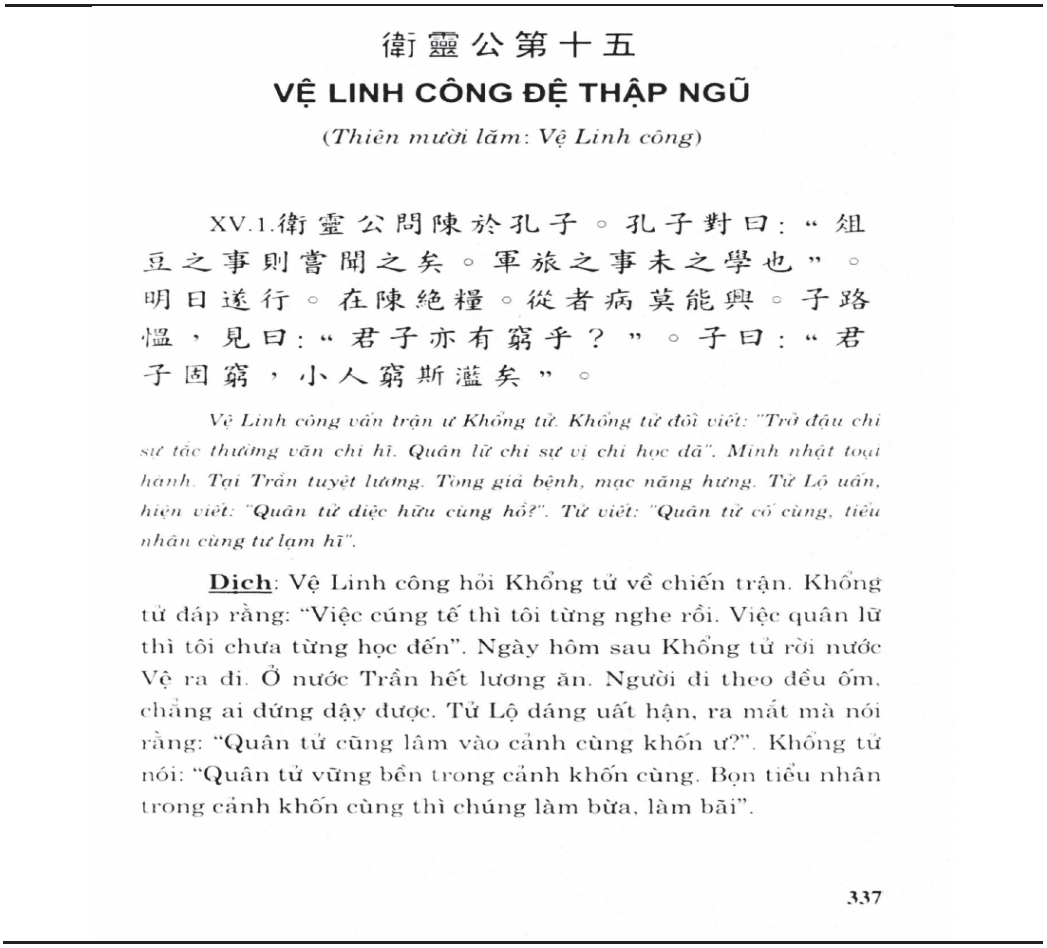
4.3. Khổng phu tử và Luận ngữ (Confucius and the Analects) (21st Century)

Khổng phu tử và Luận ngữ (*Confucius and the Analects*) by Phạm (2004) is a modern translation of the *Analects* into Vietnamese using the Latin-based *Quốc ngữ* script. This work reflects the modernization of the Vietnamese writing system, marking the transition from Nôm script to Quốc ngữ, which became the dominant script from the 20th century. The text was first published by the National Political Publisher in Hanoi in 2004. The translated portion from chapter 15 'Weiling Gong' to chapter 20 'Yao yue' spans from page 337 to page 391.

This translation adheres closely to the segmentation of the *Analects* as outlined in Zhu Xi's *Collected Annotations on the Analects*, using a systematic approach to label each section. For example, XV.1 refers to the 15th chapter and the 1st section. The text is organized into four main components, as seen in Figure 1:

(i) Original Text in Literary Sinitic: The source material is presented in traditional Sinographs.
(ii) Sino-Vietnamese Phonetic Transcription: This provides a phonetic transcription using Sino-Vietnamese readings.
(iii) Vietnamese Translation: The translation is rendered in modern Vietnamese prose using Quốc ngữ.

(iv) Footnote-Style Annotations: These provide additional commentary, although they are less extensive than the annotations found in the earlier Nôm-script texts.



衛靈公第十五

VỆ LINH CÔNG ĐỆ THẬP NGŨ

(Thiên mười lăm: Vệ Linh công)

XV.1.衛靈公問陳於孔子。孔子對曰：“俎豆之事則嘗聞之矣。軍旅之事未之學也”。明日遂行。在陳絕糧。從者病莫能興。子路慍，見曰：“君子亦有窮乎？”。子曰：“君子固窮，小人窮斯濫矣”。

Vệ Linh công vấn trận ư Khổng tử. Khổng tử đối viết: "Trở đậu chi sự tắc thường văn chi hĩ. Quân lữ chi sự vị chi học dã". Minh nhật toại hành. Tại Trần tuyệt lương. Tòng giả bệnh, mạc năng hưng. Tử Lộ uẩn, hiện viết: "Quân tử diệc hữu cùng hồ?". Tử viết: "Quân tử cố cùng, tiểu nhân cùng tư lạm hĩ".

Dịch: Vệ Linh công hỏi Khổng tử về chiến trận. Khổng tử đáp rằng: "Việc cúng tế thì tôi từng nghe rồi. Việc quân lữ thì tôi chưa từng học đến". Ngày hôm sau Khổng tử rời nước Vệ ra đi. Ở nước Trần hết lương ăn. Người đi theo đều ốm, chẳng ai đứng dậy được. Tử Lộ đáng uất hận, ra mắt mà nói rằng: "Quân tử cũng lâm vào cảnh cùng khốn ư?". Khổng tử nói: "Quân tử vững bền trong cảnh khốn cùng. Bọn tiểu nhân trong cảnh khốn cùng thì chúng làm bừa, làm bãi".

337

Figure 1. Khổng phu tử và Luận ngữ, page 337.

The modern prose translation allows for a clear understanding of how Sino -Vietnamese negators *bất*, *vô*, and *phi* have been fully integrated into contemporary Vietnamese. The text reflects the culmination of centuries of linguistic adaptation and demonstrates how these prefixes have been normalized within modern Vietnamese vocabulary.

4.4. Comparison of the Three Bilingual Texts (the Last Six Chapters)

The three texts—*Luận ngữ ước giải*, *Luận ngữ thích nghĩa ca*, and *Khổng phu tử và Luận ngữ*—offer a unique corpus that spans over 400 years of Vietnamese linguistic history, as summarized in Table 4. Together, they allow us to trace how Sino-Vietnamese negators have evolved across different periods, genres, and writing systems.

In terms of script and language, the first two texts are written in Nôm script, while the 21st-century text is written in Quốc ngữ. This transition from Nôm to Quốc ngữ reflects broader cultural and educational reforms in Vietnam.

In terms of translation style, the 17th-century *Luận ngữ ước giải* uses prose for translation, while the 19th-century *Luận ngữ thích nghĩa ca* uses the poetic *lục bát* form. We acknowledge that the lục bát form's fixed syllable count and tonal pattern may preferentially accommodate shorter negators (e.g., *chẳng*, *không*) and limit the inclusion of longer, multi-morphemic prefixes such as *bất-* and *vô*—potentially influencing their observed frequencies. The 21st-century translation returns to prose, offering a more straightforward ap-

proach. These shifts in genre and style reflect changes in how the Vietnamese approached the translation of Classical Chinese texts.

Table 4. Comparison of the 3 bilingual texts.

Book Title	***Luận ngữ ước giải***	***Luận ngữ thích nghĩa ca***	***Khổng phu tử và Luận ngữ***
Date of Completion	17th century	Late 19th century	Early 21st century
Language of Translation	Vietnamese	Vietnamese	Vietnamese
Script of Translation	Nôm script	Nôm script	Latin-based Quốc ngữ script
Genre of Translation	Prose	Lục bát (six-eight) verse	Prose
Segmentation Method of Each Translation	The segmentation is much smaller compared to Zhu Xi's segmentation, with most segments not forming complete sentences but rather individual phrases; when concatenated, they form complete chapters and sections.	Segmentation is smaller or equal to Zhu Xi's chapters, but each segment is either a complete sentence or a concise paragraph.	Follow Zhu Xi's *Collected Annotations on the Analects*, each chapter is divided into sections.
	When concatenating the Nôm translation segments of the 2 Nôm translations together according to Zhu Xi's chapter units, it will form the corresponding Vietnamese translated text for each chapter, following the segmentation structure of the Quốc ngữ script translation.		
Literary Sinitic Annotations.	Available, based on the *Great Complete Works of the Analects* 論語大全 Hu Guang group	Available, based on Zhu Xi's *Collected Annotations on the Analects*	Not available

In terms of annotation and commentary, the 17th- and 19th-century texts include extensive annotations, often explaining the translation choices and providing additional context for the use of Sino-Vietnamese negators. The 21st-century text includes fewer annotations, relying instead on a direct translation of the *Analects* into modern prose. These annotations are crucial for understanding the nuanced use of negation in earlier translations and how they were understood in different periods.

By analyzing these texts, we can observe the gradual integration of Sino-Vietnamese negators into the Vietnamese lexicon, from the early stages in the 17th century, through their poetic adaptation in the 19th century, to their full incorporation into modern prose in the 21st century. These shifts reflect broader cultural and linguistic changes in Vietnam, offering a comprehensive view of the evolution of Sino-Vietnamese negation.

5. *Bất*, *vô*, *phi* and Their Coexistence with Sentential Negators in the Three Translated Texts

This section examines the use of the Sino-Vietnamese prefixes *bất*, *vô*, and *phi* in the three translations analyzed in this study. By analyzing these texts, we can trace the integration of these Sino-Vietnamese prefixes into the Vietnamese language and their interaction with native negators like *chẳng* and *không*. In particular, we explore when these prefixes first appear, how frequently they are used, and how they function alongside sentential negators in each translation.

In the original text of the last six chapters of the *Analects*, 153 instances of 不 *bù*, 38 of 無 *wú*, and 4 of 非 *fēi* were identified, totaling 195 sentences containing these negators,

as seen in Table 5. Our bilingual corpus includes all these Chinese sentences along with their equivalents in the three Vietnamese translations, resulting in 585 sentences across the 17th-, 19th-, and 21st-century translations. This brings the total number of sentences in the corpus to 780, providing a robust foundation for comparative analysis of the usage and adaptation of negation in both languages. In particular, the 153 instances of 不 *bù* make it the most frequent negator, reflecting its prominence in both sentential and lexical negation. The 38 instances of 無 *wú* and the 4 instances of 非 *fēi* show that these negators, though less frequent, are still essential for marking lexical negation, particularly in abstract or moral contexts. These frequency patterns in Literary Sinitic offer a point of comparison when analyzing how the corresponding Vietnamese prefixes (*bất*, *vô*, and *phi*) were adapted across the three translations.

Table 5. The number of 不 *bù*, 無 *wú*, and 非 *fēi* in the last 6 chapters of the *Analects*.

	The *Analects* (Last 6 Chapters)
不 *bù*	153
無 *wú*	38
非 *fēi*	4

5.1. Distribution of Negators in the 17th-Century Translation

Our analysis shows that in the 17th-century Vietnamese translation of the *Analects*, most instances of 不 *bù*, 無 *wú*, and 非 *fēi* were translated using *chẳng*, the native Vietnamese negator. Specifically, *chẳng* was used in 152 out of 153 cases of *bù*, 32 out of 38 cases of *wú*, and all 4 instances of *fēi*, as seen in Table 6. This overwhelming reliance on *chẳng* reflects a clear preference for native sentential negation over the adoption of Sino-Vietnamese lexical prefixes like *bất*, *vô*, or *phi*, which do not appear in the translation at all.

Table 6. 不 *bù*, 無 *wú*, and 非 *fēi* and their equivalents in the 17th-century Vietnamese translation of the *Analects*.

Literary Sinitic	**Vietnamese Translation (17th Century)**
不 *bù* (153)	*chẳng* (152), others (1)
無 *wú* (38)	*chẳng* (32, in which *chẳng có* (12)), others (6)
非 *fēi* (4)	*chẳng* (4, in which *chẳng phải* (3))

In the original Literary Sinitic text, 不 *bù* is used 153 times; 152 of these instances were translated using the native Vietnamese negator *chẳng*.

(20) *bù* (不) => *chẳng*

a. Literary Sinitic (Chapter 15.21)

君子	矜	而	不	爭、
Quân tử	căng	nhi	**bất**	tranh.
gentleman	proud	but	NEG	contentious
群	而	不	黨、	
Quần	nhi	**bất**	đảng”.	
join	but	NEG	cliquish	

‘The gentleman is proud but not contentious. He joins with others but is not cliquish.’

b. Vietnamese translation (17th century)

导	君子	嚴	麻	拯	爭		
Người	quân tử	nghiêm	mà	**chẳng**	tranh		
CLF	gentleman	proud	but	NEG	contentious		

'The gentleman is proud but not contentious'.

固	排	麻	拯	覓	夕	笈	丕
Có	bầy	mà	**chẳng**	mếch	làm	bè	vậy
has	group	but	NEG	wrongly	make	ally	PRT

'He joins with others but is not cliquish'.

無 *wú* is used 38 times in the original Chinese text, and it is typically translated as *chẳng* or *chẳng có* 'not have' in the Vietnamese translation.

(21) 無*wú* => *chẳng có* 'not have'

a. Literary Sinitic (Chapter 17.23)

君子	有	勇	而	無	義
Quân tử	hữu	dũng	nhi	**vô**	nghĩa
gentleman	possess	courage	but	lack	rightness

'A gentleman who possesses courage but lacks rightness'.

b. Vietnamese translation (17th century)

导	君子	固	孟	麻	拯	固	義
Người	quân tử	có	mạnh	mà	**chẳng**	**có**	nghĩa
gentleman		have	courage	but	not	have	rightness

'A gentleman who possesses courage but lacks rightness'.

Here, 無 *vô* is rendered as *chẳng có*, indicating the absence of a quality (rightness). This structure mirrors the negation pattern found in Chinese but uses the native Vietnamese negator *chẳng có* instead of *vô*.

非 *fēi*, which appears four times in the original Chinese text, is mostly translated as *chẳng phải* 'not right' in the Vietnamese translation, a construction that aligns with (verbless) nominal negation in Vietnamese.

(22) 非 *fēi* => *chẳng phải* 'not right'

a. Literary Sinitic (Chapter 15.28)

人	能	弘	道、	非	道	弘	人、
Nhân	năng	hoằng	đạo,	phi	đạo	hoằng	nhân
Person	can	enlarge	Way,	NEG	Way	enlarge	person

'A person can enlarge the Way; it is not the case that the Way can enlarge a person'.

b. Vietnamese translation (17th century)

㑒	咍	夕	贕	道,	拯	沛	道	咍	夕
Người	hay	làm	rộng	đạo,	**chẳng**	**phải**	đạo	hay	làm
person	can	do	large	Way	not	right	Way	can	do

贕	导
rộng	người
large	person

'A person can enlarge the Way; it is not the case that the Way can enlarge a person'.

The 17th-century Vietnamese translation of the *Analects* preserved the semantic distinctions of the Literary Sinitic negators through native syntactic mean; we do not see examples of prefixation (lexical negation) yet in the Vietnamese text. For example, the complex form *chẳng có* 'not have' (a negator and an existential verb) was used for existential nega-

tion and *chẳng phải* 'not right' (a negator and a predicate) was used for (verbless) nominal negation. Chinese syntactic negators, therefore, are consistently rendered using native Vietnamese syntactic negators—a significant observation, as later translations show the gradual emergence of Sino-Vietnamese negators.

This dominance of native Vietnamese negators during the 17th century points to a period where Vietnamese largely retained its own negation system, even as it borrowed Chinese vocabulary in other lexical domains. The findings illustrate the selective nature of linguistic adaptation, where core grammatical elements like negation were slower to absorb Sino-Vietnamese influences. In the following section, we turn to the 19th-century translation of the *Analects*, examining how the prefixes *bất*, *vô*, and *phi* began to be integrated into the Vietnamese language while continuing to coexist with native negators such as *chẳng*.

5.2. Distribution of Negators in the 19th-Century Translation

The data for the 19th-century translation reveal a mix of native and borrowed negators, as illustrated in Table 7:

Table 7. 不 *bù*, 無 *wú*, and 非 *fēi* and their equivalents in the 19th-century Vietnamese translation of the *Analects*.

Literary Sinitic	Vietnamese Translation (19th Century)
不 ***bù*** (153)	*chẳng* (90), ***không*** (35), *bất* (2), *others* (26)
無 *wú* (38)	*chẳng* (11), ***không*** (14), *others* (13)
非 *fēi* (4)	*chưa* (1), *lầm* (1), *nào có* (1), no translation (1)

In the 19th-century Vietnamese translation of the *Analects*, we observe a more diverse distribution of negators compared to the 17th-century text. Native negators like *chẳng* continue to be dominant, appearing 90 times to translate 不 ***bù*** and 11 times to translate 無 *wú*, but we also see the increasing use of ***không***, which appears 35 times to translate 不 ***bù*** and 14 times to translate 無 *wú*. This shift aligns with the historical trend of ***không***, which is derived from Chinese 空 *kōng*, becoming more integrated into the Vietnamese language, particularly as a sentential negator. Recent studies have noted that ***không*** began to grammaticalize into a general negator in the late 17th century (Phan et al., 2024), and by the late 19th century, it started to rival or even replace *chẳng* as the dominant negator (Trinh et al., 2024).

Additionally, we observe that, among the three prefixes, *bất* is the first to appear in the Vietnamese translation, albeit in just two instances. Its early inclusion marks the initial stage of the integration of Sino-Vietnamese prefixes into Vietnamese negation. This reflects a gradual linguistic shift, where native negators like *chẳng* still dominated but were slowly accompanied by borrowed Sino-Vietnamese elements. The appearance of *bất* signals the beginning of increased lexical borrowing from Chinese, which would become more prominent in subsequent periods.

In the 19th-century translation, 不 ***bù*** is predominantly translated as sentential negators *chẳng* or ***không***.

(23) 不 *bù* => *chẳng/không*

a. Literary Sinitic (Chapter 15.21)

君子	矜	而	不	爭、
Quân tử	căng	nhi	**bất**	tranh.
gentleman	proud	but	NEG	contentious
群	而	不	黨、	
Quần	nhi	**bất**	đảng".	
join	but	NEG	cliquish	

'The gentleman is proud but not contentious. He joins with others but is not cliquish'.

b. Vietnamese translation (19th century)

𠊛	賢	嚴	吏	空	爭、		
Người	hiền	nghiêm	lại	**không**	tranh,		
person	gentle	proud	but	NEG	contentious		
和	𠊓	悧	眾	拯	唌	誽	隊く、
Hoà	nên	ưa	chúng	**chẳng**	bênh	dua	đòi,
join	then	like	group	NEG	incline to	cliquish	

'The gentleman is proud but not contentious. He joins with others but is not cliquish'.

Having compared example (20) and example (23), we see that in the same Literary Sinitic sentence, 不 *bù* appeared twice and was translated as *chẳng* in both instances in the 17th-century translation, whereas in the 19th-century translation, one instance was rendered as *chẳng* and the other as *không*. The shift from the native sentential negator *chẳng* to the Chinese-derived sentential negator *không* suggests an increased adoption of Sino-Vietnamese lexical elements during the 19th century, although the chosen item does not match Chinese usage.

Another distinctive feature of the 19th-century translation, among the three bilingual texts under investigation, is the first appearance of one of the three Sino-Vietnamese prefixes. This is the earliest text where 不 *bù* is translated as *bất*, marking the initial stage of integrating these items into Vietnamese negation.

(24) 不 *bù* => bất

a. Literary Sinitic (Chapter 19.3)

嘉	善	而	矜	不	能
gia	thiện	nhi	căng	**bất**	**năng**
applaud	good	but	sympathize	NEG	ability

'He applauds good men and sympathizes with those who lack ability'.

b. Vietnamese translation (19th century)

嗐	𠊛	唯く	卒	傷	台	不	才
Khen	người	giỏi	tốt	thương	thay	**bất**	**tài**
applaud	person	good	nice	sympathize	PRT	unable	

'He applauds good men and sympathizes with those who lack ability'.

(25) 不 *bù* => bất

a. Literary Sinitic (Chapter 17.21)

子	曰、	予	之	不	仁	也
Tử	viết:	"Dư	chi	**bất**	**nhân**	dã
Confucius	say	Yu	PRT	no	human	PRT

'The Master said, Yu (Zai Wo) has no humaneness!'

b. Vietnamese translation (19th century)

吠	浪	予	意く	罕	時	不	仁
Dạy	rằng	Dư	ấy	hẳn	thì	**bất**	**nhân**
teach	that	Dư	that	clearly	TOP	unhuman	

'The Master said, Yu (Zai Wo) has no humaneness!'

Interestingly, these two examples where 不 ***bù*** is translated as *bất* reveal distinct translation strategies: one slightly more innovative, and the other more conservative. In the first example (24), 不能 *bù néng*, which means 'lacking ability' in Literary Sinitic, is translated as *bất tài* in Vietnamese. While *bất tài* can still be understood as 'lacking ability,' it typically carries a stronger connotation of 'incompetence' or 'lack of talent.' This translation reflects a minor shift toward a more specific interpretation. In the second example (25), 不仁 *bù rén*, meaning 'inhumane' or 'lacking benevolence,' is translated as *bất nhân*, which is a direct and faithful reflection of the original. Here, the translator adheres closely to the source text, preserving the core meaning of the Confucian concept of 仁 *rén*, or 'benevolence' (Ivanhoe, 2000). Unlike the slight shift seen in the previous example, this translation remains conservative, ensuring that key philosophical terms retain their original sense and integrity. These two examples illustrate the early stages of the integration of *bất* into Vietnamese negation. In one case, as with *bất tài*, the translator makes subtle adjustments to fit Vietnamese usage, while in another, such as *bất nhân*, the original term is maintained. Together, these cases suggest that the appearance of *bất* in the 19th-century translation marks a gradual yet careful adoption of Sino-Vietnamese prefixes into the Vietnamese lexicon.

Similar to 不 ***bù***, the negator 無 *wú* continues to be predominantly translated as *chẳng* (11 times) or *không* (14 times). However, we observe a more frequent use of *không* as the preferred translation for *vô*, reinforcing the trend of increased Sino-Vietnamese lexical borrowing in this period. This is consistent with recent studies on the grammaticalization path of *không* (Phan et al., 2024), which suggest that, before *không* assumed the role of a general negator in Vietnamese, it initially functioned as an existential negator meaning 'not have'. In the following example from Chapter 17.23 of the *Analects*, 無 *wú* is translated as *không*, reinforcing this trend:

(26) 無 *wú* => không

a. Literary Sinitic (Chapter 17.23)

君子	有	勇	而	無	義	爲亂
quân tử	hữu	dũng	nhi	**vô**	nghĩa.	vi loạn
gentleman	have	courage	but	NEG	rightness	dorebellious

'A gentleman who possesses courage but lacks rightness will become rebellious'.

b. Vietnamese translation (19th century)

勇	麻	空	義	亂	醝	苦く而	之
Dũng	mà	**không**	nghĩa	loàn	gây	khó	gì
courage	but	NEG	rightness	rebellious	cause	trouble	thing

'A gentleman who possesses courage but lacks rightness will become rebellious'.

In the case of 非 *fēi*, the 19th-century translation shows a variety of renderings, including *chưa* 'not yet', *lâm* 'wrong', and *nào có* 'not have', depending on the context.

(27) 非 *fei* => *nào có* 'not have'

a. Literary Sinitic (Chapter 15.28)

人	能	弘	道、	非	道	弘	人、
Nhân	năng	hoằng	đạo,	phi	đạo	hoằng	nhân
Person	can	enlarge	Way,	NEG	Way	enlarge	person

'A person can enlarge the Way; it is not the case that the Way can enlarge a person'.

b. Vietnamese translation (19th century)

𠊛	哈	𢷾	穬	道	常、		
Người	hay	mở	rộng	đạo	thường,		
person	can	open	large	Way	often		
道	箕	芾	固	𢷾	芒	鄧	𠊛、
Đạo	kia	**nào**	**có**	mở	mang	đặng	người,
Way	DEM	not	have	open	PRT	can	person

'A person can enlarge the Way; but the Way cannot enlarge a person'.

This variability may be influenced by the poetic nature of the text, which often allows for broader interpretation. The translator's approach to 非 *fēi* seems more flexible, moving away from a strict adherence to uniform translation norms. This flexibility may reflect their efforts to convey the nuances of meaning and context, using different forms of negation depending on the specific situation.

Compared to the 17th-century translation, where the native sentential negator *chẳng* was overwhelmingly dominant and there was no sign of *bất*, *vô*, or *phi*, the 19th-century translation marks a transitional phase. A notable difference is the increased use of the Chinese-derived sentential negator *không*, especially in translating both 不 *bù* and 無 *wú*. This shift aligns with recent research on the development of *không*, which transitioned from an existential negator to a general negator (Phan et al., 2024). Additionally, the 19th-century translation introduces *bất* for the first time, signaling the early stages of integrating Sino-Vietnamese prefixes into the negation system.

Overall, the 19th-century translation reflects a period of transition in Vietnamese negation. Native negators like *chẳng* remained prominent, but the increasing use of *không* and the introduction of *bất* are indicative of a gradual integration of Sino-Vietnamese elements. In the following section, we explore the 21st-century translation of the *Analects* and how *bất*, *vô*, and *phi* have become more integrated into modern Vietnamese.

5.3. Distribution of Negators in the 21st-Century Translation

In contrast to earlier translations, the 21st-century version reflects a linguistic shift toward the frequent use of the Chinese derived *không* as the standard negator over the native Vietnamese *chẳng*.

As seen in Table 8, *không* has become the overwhelmingly preferred negator, translating 127 out of 153 instances of 不 *bù*, 27 out of 38 instances of 無 *wú*, and all 4 instances of 非 *fēi*. In comparison, *chẳng* plays a more limited role in this period, appearing in only 12 instances for 不*bù* and 2 for 無 *wú*, a stark contrast with its prominence in earlier translations. This decline reflects a shift in usage patterns where *không* takes precedence, pushing *chẳng* into a more emphatic or non-standard role. This supports previous research on the grammaticalization of *không*, which became a standard negator by the late 19th century (Trinh et al., 2024). Below are examples in which 不 *bù*, 無 *wú*, and 非 *fēi* are all translated as *không*.

(28) 不 *bù* => *không*

a. Literary Sinitic (Chapter 15.21)

君子	矜	而	不	爭、
Quân tử	căng	nhi	**bất**	tranh.
gentleman	proud	but	NEG	contentious
群	而	不	黨、	
Quần	nhi	**bất**	đảng".	
join	but	NEG	cliquish	

Table 8. 不*bù*, 無 *wú*, and 非 *fēi* and their equivalents in the 21st-century Vietnamese translation of the *Analects*.

Literary Sinitic	Vietnamese Translation (21st Century)
不*bù* (153)	*không* (127), *chẳng* (12), *bất* (5), others (9)
無 *wú* (38)	*không* (27) (in which *không có* (9)), *chẳng* (2), *vô* (5), others (4)
非 *fēi* (4)	*không* (4) (in which *không phải* (3))

'The gentleman is proud but not contentious. He joins with others but is not cliquish'.

b. Vietnamese translation (21th century)

Người quân tử	tự chủ bản thân,	giữ	quan điểm	nhưng
CLF gentleman	control self	hold	opinion	but
không	tranh giành,	cãi cọ.		
not	fight	argue		

Hợp quần	với	mọi người	nhưng	**không**	bè	phái
join	with	people	but	not	ally	

'The gentleman is proud but not contentious. He joins with others but is not cliquish'.

(29) 無 *wú* => *không* or *không có* 'not have':

a. Literary Sinitic (Chapter 17.23)

君子	有	勇	而	無	義	爲	亂
quân tử	hữu	dũng	nhi	**vô**	nghĩa	vi	loạn
gentleman	have	courage	but	NEG	rightness	do	rebellious

'A gentleman who possesses courage but lacks rightness will become rebellious'.

b. Vietnamese translation (21th century)

Quân tử	có	dũng	mà	**không**	**có**	nghĩa
Gentleman	has	courage	but	NEG	have	rightness
thì	làm	loạn				
TOP	make	rebel				

'A gentleman who possesses courage but lacks rightness will become rebellious'.

(30) 非 *fēi* => *không phải* 'not right':

a. Literary Sinitic (Chapter 15.28)

人	能	弘	道、	非	道	弘	人、
Nhân	năng	hoằng	đạo,	phi	đạo	hoằng	nhân
Person	can	enlarge	Way,	NEG	Way	enlarge	person

'A person can enlarge the Way; it is not the case that the Way can enlarge a person'.

b. Vietnamese translation (21th century)

Con	người	có thể	hoằng dương	đạo lý,	chứ	**không**
CLF	person	CAN	enlarge	Way	but	**NOT**
phải	đạo lý	hoằng dương	cho	con	người	
right	Way	enlarge	for	CLF	person	

'A person can enlarge the Way; it is not the case that the Way can enlarge a person'.

The 21st-century translation of the *Analects* exhibits a similar pattern to that of the 17th century, where Vietnamese mostly employed syntactic strategies rather than relying on lexical prefixation for negation. The key distinction, however, lies in the choice of the standard negator: while the 17th-century translation predominantly used *chẳng* to maintain the semantic distinctions of the Literary Sinitic negators, the 21st-century text continues this syntactic approach with *không* as the main negator. For example, in both periods, existential negation is expressed through the combination of a negator and an existential verb. In the 17th-century translation, this was typically rendered as *chẳng có* ('not have'), as seen in example (21b), whereas in the 21st-century version, the same structure is realized as *không có*, as seen in example (29b). Similarly, for (verbless) nominal negation, the earlier translation used *chẳng phải* ('not right'), as in (22b), which in the 21st-century translation becomes *không phải*, as in (30b). This suggests that, despite increased lexical borrowing from Chinese, Vietnamese uses the core grammatical strategies with compound structures like *không có* and *không phải*, allowing for precise differentiation between types of negation.

The increased appearance of Sino-Vietnamese prefixes, such as *bất* and *vô*, is also notable in this period. *bất* appears in five instances, and *vô* in five instances as well, marking their integration into the Vietnamese lexical negation system. Just as in the 19th-century translation, *bất* continues to appear, showcasing its role as a formal and fixed negator. However, while two of these instances were already present in the 19th-century translation, the 21st-century version introduces three more cases of *bất*. Two of these three retain the original lexical form of Literary Sinitic, such as *bất thiện* and *bất đồng*, while the third offers a more innovative translation: the Literary Sinitic *bất hiền* is rendered as *bất tiếu* in Vietnamese. For instance, in Chapter 16.11 of the Literary Sinitic text, 不善 *bù shàn* is translated as *bất thiện*:

(31) 不 *bù* => bất

a. Literary Sinitic (Chapter 16.11)

見	不	善	如	探	湯	
kiến	**bất**	**thiện**	như	thám	thang	
see	NEG	good	as	explore	scalding	water

'sees what is not good and acts as though he had put his hand in scalding water'.

b. Vietnamese translation (21th century)

thấy	điều	**bất**	**thiện**	chạy	như	tránh	nước	sôi
see	thing	NEG	good	run	as	avoid	water	boiling

'sees what is not good and acts as though he had put his hand in scalding water'.

Similarly, in Chapter 15.39, 不 *bù* is translated as *bất đồng*:

(32) 不 *bù* => bất

a. Literary Sinitic (Chapter 15.39)

道	不	同
Đạo	**bất**	**đồng**
way	NEG	same

'your Way is not the same'.

b. Vietnamese translation (21th century)

Đạo lý	đã	**bất**	**đồng**
way	PST	NEG	same

'your Way is not the same'

However, the translation becomes more innovative in Chapter 19.22. Instead of directly translating 不賢 *bù xián* 'unworthy' as *bất hiền*, the 21st-century translation uses *bất tiếu* (not worthy):

(33) 不 *bù* => bất

a. Literary Sinitic (Chapter 19.22)

賢	者	識	其	大	者、	
Hiền	giả	chí	kì	đại	giả,	
worthy	man	remember	PRT	essential	thing	
不	賢	者	識	其	小	者
bất	**hiền**	giả,	chí	kì	tiểu	giả,
NEG	worthy	man	remember	PRT	minor	thing

'worthy men remember the essentials and those of little worth remember the minor points'.

b. Vietnamese translation (21th century)

Bậc	hiền	giả	nhớ	được	cái	lớn,	kẻ	**bất**
CLF	worthy	man	remember	CAN	CLF	big	man	NEG
tiếu	nhớ	được	cái	nhỏ.				
worthy	remember	CAN	CLF	small				

'worthy men remember the essentials and those of little worth remember the minor points'.

This innovation highlights the growing flexibility of *bất* in the modern Vietnamese lexicon, reflecting both the preservation of classical forms and the creative use of new ones to adapt to contemporary linguistic norms.

In addition to the prefix *bất*, *vô* makes its debut in the 21st-century translation, appearing in five instances that underscore its role in formalizing lexical negation. Notably, *vô* did not appear in the 17th- and 19th-century translations, marking its emergence solely in the 21st-century version. Unlike *bất*, which shows some innovative usages in Vietnamese, *vô* remains closely aligned with its original Literary Sinitic form, without deviations from the classical meanings.

In the following examples, *vô* consistently retains its original sense of existential negation without creative reinterpretation or novel usage:

(34) 無 *wú* => vô

a. Literary Sinitic (Chapter 15.4)

無	爲	而	治	者	其	舜	也	與
vô	**vi**	nhi	trị	giả	kỳ	Thuấn	dã	dư?
NEG	action	but	rule	person	that	Shun	PRT	PRT

'Of those who ruled through inaction, surely Shun was one?'

b. Vietnamese translation (21th century)

Người	thực hiện	nền	chính trị	**vô vi**	mà
person	carry out	CLF	politics	inaction	but
đất nước	thịnh trị	có lẽ	là	Thuấn	chăng
country	peaceful	perhaps	is	Shun	PRT

'Of those who ruled through inaction but the country was peaceful, was Shun the one?'

In this instance, *vô vi* is directly retained in the Vietnamese translation, preserving the original term from Literary Sinitic.

(35) 無 *wú* => vô

a. Literary Sinitic (Chapter 15.6)

邦	無	道	如	矢
Bang	**vô**	**đạo**	như	thỉ.
state	NEG	way	as	arrow

'The state was without the Way, he was straight as an arrow'.

b. Vietnamese translation (21th century)

Nước	**vô**	**đạo**	ông	cũng	thẳng	như	tên
state	lack	way	PRN	still	straight	as	arrow

'The state was without the Way, he was straight as an arrow'.
Again, the phrase *vô đạo* is preserved directly from the classical text.

(36) 無 *wú* => vô

a. Literary Sinitic (Chapter 15.3)

終	夜	不	寢	以	思
chung	dạ	bất	tẩm	dĩ	tư
all	night	NEG	sleep	in order to	think
無	益				
vô	**ích**				
no	use				

'and all night without sleeping in order to think. It was no use'.

b. Vietnamese translation (21th century)

Trọn	đêm	không	ngủ	để	nghĩ
all	night	NEG	sleep	in order to	think
nhưng	đều	**vô ích**			
but	all	no use			

'and all night without sleeping in order to think. It was no use'
In this case, *vô ích* maintains its classical meaning of 'no use'.

(37) 無 *wú* => vô

a. Literary Sinitic (Chapter 16.2)

天下	無	道
Thiên hạ	**vô**	**đạo**
world	NEG	way

'When the Way no longer prevails in the world'.

b. Vietnamese translation (21th century)

Thiên hạ	**vô đạo**
world	lack way

'When the Way no longer prevails in the world'.
Here, *vô đạo* once again aligns perfectly with its Literary Sinitic form.

(38) 無 *wú* => vô

a. Literary Sinitic (Chapter 17.24)

惡	勇	而	無	禮	者
ố	dũng	nhi	**vô**	**lễ**	giả
hate	courage	but	no	ritual	person

'He hates courage that ignores ritual decorum'.

b. Vietnamese translation (21st century)

Ghét	kẻ	có	dũng	mà	**vô lễ**
hate	person	have	courage	but	no ritual

'He hates courage that forgets ritual decorum'.

In this final example, *vô lễ* is faithfully translated as 'no ritual' in Vietnamese, closely following the classical meaning. The consistent appearance of *vô* in these examples demonstrates its seamless integration into the Vietnamese lexicon, with no notable departures from its original Literary Sinitic usage. Unlike *bất*, which sometimes exhibits innovative translation choices, *vô* adheres closely to its classical roots, highlighting its stability as a negation prefix in modern Vietnamese.

Unlike *bất* and *vô*, *phi* does not appear in any of the three Vietnamese translations analyzed in this study. In the 21th-century translation, 非 *fēi* is consistently translated as *không*, with *không phải*, a common construction in modern Vietnamese that negates the truth of a statement or identity being the most commonly used form. This indicates a clear preference in Vietnamese for syntactic negation over lexical negation when considering (verbless) nominal negation constructions. Rather than employing a Sino-Vietnamese prefix equivalent to 非 *fēi*, such as *phi*, Vietnamese translators favor the use of *không phải*, a syntactic structure that aligns with the native patterns of negation, particularly in (verbless) nominal negative contexts. This aligns with a trend in Vietnamese to adapt and integrate negation strategies through syntactic combinations rather than adopting all lexical prefixes from Literary Sinitic.

(39) 非 *fēi* => *không phải* 'not right':

a. Literary Sinitic (Chapter 15.1)

對	曰	然、	非	與、
Đối	viết:	"Nhiên.	**Phi**	dư?"[5]
reply	say	yes	NEG	Q

'Zigong replied, Yes, isn't that so?'

b. Vietnamese translation (21th century)

Đáp	rằng:	"Vâng,	chẳng lẽ	**không**	**phải**	ư?"
reply	COMP	Yes	perhaps	NEG	right	Q

'Zigong replied, Yes, isn't that so?'

The 21st-century translation of the *Analects* marks a significant shift in the Vietnamese negation system, with *không* firmly establishing itself as the dominant negator. Additionally, there is a notable increase in the use of *bất* and *vô* reflecting broader 'linguistic modernization' trends in Vietnam. This shift aligns with the larger cultural and linguistic modernization of East Asia, heavily influenced by Japan's transmission of neologisms made of Chinese morphemes through reformist writings during the 20th century (Sinh, 1993).

In this context, the integration of *bất* and *vô* into formal expressions, alongside the creative flexibility seen in the translation of *bất tiếu 'not worthy'*, highlights the evolving role of Sino-Vietnamese prefixes in modern Vietnamese.

This 21st-century shift contrasts with earlier periods, such as the 17th-century translation, where *chẳng* overwhelmingly dominated, and the 19th century, where *không* and *chẳng* coexisted more equally. The gradual rise of *không* as the standard negator underscores the ongoing influence of Sino-Vietnamese elements in the language. The increased presence of *bất* and *vô* in the 21st century demonstrates the culmination of these linguistic trends, showing how Vietnamese has balanced the integration of Chinese lexical features with its own syntactic structures.

In summary, the 21st-century translation exemplifies the historical evolution of Vietnamese negation, with *không* taking precedence while *bất* and *vô* gain prominence in specific contexts. This evolution is part of a broader cultural shift in East Asia, driven by modernization and the transmission of modernized Chinese linguistic features (Sinh, 1993; Alves, 2017).

5.4. Evidence from Earlier Texts

As highlighted earlier, the 17th-century Vietnamese translations of the *Analects* relied almost exclusively on native Vietnamese negators like *chẳng*, with no evidence of *bất*, *vô*, or *phi*. This pattern is consistent across earlier bilingual texts, suggesting a strong reliance on native Vietnamese negators in formal translations. For instance, in the *Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú* 新編傳奇漫錄增補解音集註 (*Interpretation of Legendary Tales: Newly Edited Collection with Supplementary Explanations and Pronunciation Annotations*), (late 16th to 17th century), the translation of 不 *bù*, 無 *wú*, and 非 *fēi* similarly utilized native negators such as *chẳng*, *chăng*, and *chớ*.

As shown in Table 9, the 16th- to 17th-century Vietnamese translations predominantly employed native sentential negators without integrating Sino-Vietnamese lexical prefixes. This reliance on native structures, especially for negation, persisted in earlier texts, indicating a gradual and selective integration of Sino-Vietnamese elements into the language. The transition to using prefixes such as *bất* and *vô* would not occur until the 19th and 20th centuries, suggesting that Vietnamese language and translation practices initially remained insulated from external lexical influences in the domain of negation.

Table 9. 不*bù*, 無 *wú*, and 非 *fēi* and their equivalents in the late 16th-century–17th-century Vietnamese.

Literary Sinitic	**Vietnamese Translation (Late 16th–17th Century)**
不 *bù* (563)	*chẳng* (558), *chớ* (5)
無 *wú* (95)	*chẳng* (92), *chăng* (3)
非 *fēi* (69)	*chẳng* (44), *chớ* (25)

(Washizawa, 2019a, pp. 151–157)

By the 20th century, the rise of the sentential negator *không* and the emergence of the negative prefixes *bất* and *vô* reflect broader shifts in the intellectual and linguistic landscape of Vietnam. These shifts align with the country's 'modernization' efforts, which were significantly influenced by Chinese and Japanese intellectual movements, as well as by European colonialism. As noted by Sinh (1993), the reforms in Japan during the Meiji period played a crucial role in disseminating modernized Chinese-character-based vocabulary across East Asia, facilitating the eventual integration of Sino-Vietnamese elements, including *bất* and *vô*, into the Vietnamese lexicon. The literacy reforms and educational movements of the 20th century, particularly the promotion of the *Quốc ngữ* script, further accelerated the standardization of the Vietnamese language. This standardization allowed for greater integration of Sino-Vietnamese prefixes into formal, academic, and philosophical texts. These reforms not only improved access to education but also reinforced the use of prefixes like *bất* and *vô* in domains where precision and formality were required.

By the 21st century, *bất* and *vô* had become firmly established in formal Vietnamese, particularly in academic and philosophical discourse. This is evident in modern translations of the *Analects*, where these prefixes appear alongside *không*, which had by then become the dominant negator for sentential negation. The growing prominence of *bất*

and *vô* underscores the increasing formalization and modernization of Vietnamese in intellectual contexts, influenced by East Asian linguistic reforms in the early 20th century (Hashimoto, 1978; Alves, 2017; Takahashi, 2022). The integration of Sino-Vietnamese prefixes into the modern lexicon reflects both Vietnam's efforts to 'modernize' its language and educational system and the external influences from Chinese and Japanese reformists. Despite the increasing presence of *bất* and *vô*, however, *không* remains the dominant negator, highlighting a balance between retaining native linguistic structures and incorporating new Sino-Vietnamese elements, and a division of labor between syntactic negation and lexical negation. This gradual yet dynamic evolution demonstrates the adaptability of the Vietnamese language in response to both internal and external changes.

6. Discussion

This study has examined the development and integration of the Sino-Vietnamese prefixes *bất*, *vô*, and *phi* within the Vietnamese language, using a unique corpus of bilingual translations of the *Analects* from the 17th, 19th, and 21st centuries. By analyzing these texts, we have traced how these prefixes entered the Vietnamese language and interacted with sentential negators like *chẳng* and *không*. The shifts observed over these three centuries highlight both the resilience of native Vietnamese syntactic structures and the gradual incorporation of Sino-Vietnamese lexical elements.

6.1. Comparison of Negation Across Three Periods

A comparative analysis of the three translations spanning over three centuries reveals a dynamic evolution in how negation has been handled in Vietnamese. Each period represents a distinct phase in the adaptation and integration of Sino-Vietnamese elements into the negation system of the language.

6.1.1. The 17th Century: Dominance of Native Negators

In the 17th century, the Vietnamese translation of the *Analects* overwhelmingly relied on native negators, particularly *chẳng*. The absence of the Sino-Vietnamese prefixes *bất*, *vô*, or *phi* during this period indicates that Vietnamese was largely resistant to lexical borrowing in the domain of negation. Native constructions like *chẳng có* 'not have' for existential negation and *chẳng phải* 'not right' for (verbless) nominal negation maintained the semantic distinctions of Literary Sinitic negators through syntactic, rather than lexical, means. The reliance on these native forms suggests a period where the Vietnamese language favored its own grammatical structures over borrowed ones, even though other areas of the lexicon had already incorporated Chinese vocabulary.

6.1.2. The 19th Century: Transitional Phase

The 19th-century translation of the *Analects* marks a clear transition in the Vietnamese negation system. While native negators such as *chẳng* continued to dominate, *không*—a Chinese-derived sentential negator—began to appear more frequently, particularly for translating 不 *bù* and 無 *wú*. This shift aligns with historical linguistic studies that identify the grammaticalization of *không* as a general negator (Washizawa, 2019b; Phan et al., 2024; Trinh et al., 2024). The 19th century also saw the first instances of the Sino-Vietnamese prefix *bất*, signaling the early stages of its integration into the Vietnamese negation system. However, *bất* only appeared in two instances, one of which maintained a conservative translation (e.g., *bất nhân*), while the other reflected a more innovative use (e.g., *bất tài*). The usage of *bất* remained limited to formal or philosophical contexts, reflecting a cautious and gradual borrowing process. This transitional phase marked a shift in Vietnamese negation, where native structures were slowly complemented by borrowed Sino-Vietnamese elements, leading to a more hybrid system that combined both native and borrowed forms.

6.1.3. The 21st Century: Modernization and Integration

The 21st century marks a significant shift in the Vietnamese negation system, with *không* emerging as the dominant sentential negator across most contexts. Unlike the 17th century, where *chẳng* dominated, and the 19th century, where *không* and *chẳng* coexisted, the modern translation of the *Analects* shows a clear preference for *không*. In this period, *bất* and *vô* also appear more frequently, reflecting their growing role in formal and academic discourse. While the innovation seen in the earlier usage of *bất* continued, *vô* maintained a more conservative usage, remaining closely aligned with its original Literary Sinitic meaning. Interestingly, *phi* is still absent in the 21st-century translation. Instead, 非 *fēi* is consistently translated as *không phải* ('not right'), reflecting Vietnamese's continued preference for syntactic negation over the adoption of a lexical negator such as *phi*. This suggests that despite increased borrowing of Sino-Vietnamese prefixes in other areas, Vietnamese retains a syntactic strategy for negation in copular constructions, demonstrating the resilience of native grammatical patterns.

6.2. Linguistic Modernization and Broader Cultural Context

The increasing integration of Sino-Vietnamese prefixes like *bất* and *vô* in the 21st century reflects broader cultural and linguistic modernization in East Asia. As Sinh (1993) noted, Japan's Meiji-era reforms played a pivotal role in transmitting modernized Chinese vocabulary across East Asia, influencing countries like Vietnam. The rise of Sino-Vietnamese prefixes in Vietnamese texts, particularly in the 20th and 21st centuries, mirrors these larger regional shifts toward linguistic and intellectual modernization.

The promotion of the *Quốc ngữ* script during the 20th century also played a critical role in the standardization of the Vietnamese language. This increased literacy and educational reform not only made the language more accessible but also facilitated the integration of Sino-Vietnamese elements like *bất* and *vô* into formal contexts. As Vietnam engaged more deeply with modern Chinese and Japanese intellectual traditions, these prefixes became more embedded in the lexicon, particularly in legal, academic, and philosophical discourse.

Despite this growing presence, the dominance of *không* as the general sentential negator suggests that the Vietnamese language has balanced external influences with a preservation of native structures. While prefixes like *bất* and *vô* have found their place in formal contexts, *không* remains the most versatile and widely used negator, illustrating the selective nature of Vietnamese in the adoption of Chinese items.

6.3. The Role of our Unique Corpus and Future Research Directions

The unique bilingual corpus of the *Analects* translations, spanning three centuries, offers invaluable insights into the evolution of negation in Vietnamese. It provides a rare opportunity to trace how Vietnamese has integrated Sino-Vietnamese prefixes into its language, alongside native syntactic strategies. This corpus allows for a deeper understanding of the interplay between borrowed lexical elements and native grammatical structures, shedding light on the complexities of linguistic adaptation in Vietnam.

However, more research is needed to address key questions that this study has raised:

(i) When did the first compounds involving *bất*, *vô*, and *phi* appear in Vietnamese? While we have identified their early usage in the 19th century, understanding the precise origins of these compounds and their spread across different text types could further illuminate the process of Sino-Vietnamese lexical integration.

(ii) What are the possible word classes for these compounds, and do they behave syntactically like their counterparts in Chinese? Further research should explore how these

compounds function across various syntactic contexts in Vietnamese and whether they mirror Chinese structures or diverge in significant ways.

Expanding the corpus to include non-philosophical genres from the same periods would provide a broader picture of how Sino-Vietnamese prefixes were adopted across different domains. Additionally, further exploration of literacy movements and educational reforms, particularly the influence of Japan and China in transmitting modernized vocabulary, could offer deeper insights into the cultural factors driving these linguistic shifts.

7. Conclusions

This study provides a comprehensive analysis of the evolution of negation in Vietnamese across three centuries, tracing the gradual integration of Sino-Vietnamese lexical prefixes *bất*, *vô*, and *phi* alongside sentential syntactic negators such as *chẳng* and *không*. The findings highlight a slow but steady shift toward incorporating Sino-Vietnamese elements into the language, particularly in formal and academic contexts. However, the preservation of native syntactic structures underscores the balance Vietnam has struck between borrowing external linguistic elements in the lexicon, while maintaining its own native constructions in syntax.

The unique corpus of the *Analects* translations has offered a rare glimpse into this linguistic evolution, providing a valuable resource for future research. By examining how negation has shifted over time in Vietnamese, we gain a deeper understanding of the complex processes of language contact, adaptation, and modernization in Sinospheric East Asia. The selective integration of Sino-Vietnamese elements like *bất*, *vô*, and *phi* into the modern Vietnamese lexicon underscores the dynamic nature of language contact and the complex processes through which languages evolve over time.

Documents investigated in the paper

Luận ngữ ước giải 論語約解 [*Concise Interpretation to the Analects*], 17th century, anonymous, written in Nôm, archived of Institute of Sino-Nom Studies, call number AB.270/3, Romanized transliterated by T. C. Nguyễn (2009, manuscript).

Tự Đức thánh chế Luận ngữ thích nghĩa ca 嗣德聖制論語釋義歌 [*Tự Đức imperical explanation of the Analects*], 19th century, by Emperor Tự Đức (1828–1883), written in Nôm, archived at the Institute of Sino-Nom Studies, call number A.186/1-2.

Khổng phu tử và Luận ngữ [*Confucius and the Analects*], 21st century, by Phạm (2004), written in Romanized Quốc ngữ, Hanoi: National Political Publisher.

Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú 新編傳奇漫錄增補解音集註 (*Interpretation of Legendary Tales: Newly Edited Collection with Supplementary Explanations and Pronunciation Annotations*), 16th–17th century, anonymous, written in Nôm, archived of Institute of Sino-Nom Studies, call number HN.257 and HN.258, Romanized transliterated by Q. H. Nguyễn (2001), *Truyền kỳ mạn lục giải âm*, Hanoi: Social Sciences Publishing house.

Đại từ điển tiếng Việt [*Great Dictionary of Vietnamese*], complied by N. Y. Nguyễn (1999). Ho Chi Minh City: Culture and Information Publisher.

Author Contributions: Conceptualization, G.F.A. and T.P.; methodology, G.F.A. and T.P.; validation, G.F.A. and T.P.; formal analysis, G.F.A. and T.P.; investigation, G.F.A. and T.P.; resources, G.F.A. and T.P.; data curation, G.F.A. and T.P.; writing—original draft preparation, T.P.; writing—review and editing, G.F.A. and T.P.; supervision, G.F.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We gratefully thank Nguyễn Tuấn Cường for providing the data and engaging in invaluable discussions—this paper is as much his as ours. We also appreciate the feedback from attendees at the 4th Workshop on Vietnamese Linguistics (ISVL-4), 15–17 March 2024, at National Yang Ming Chiao Tung University, Hsinchu, Taiwan, and the VIII Study Days on Chinese Linguistics, 19–20 June 2024, at the University of Verona, Italy. We are indebted to our two reviewers for their constructive suggestions, which greatly enhanced the quality of this paper. For academic clarity, Giorgio Francesco Arcodia is responsible for Sections 1, 2, 5.1, 5.4 and 6, while Trang Phan is responsible for Section 3, Section 4, Section 5.2, Section 5.3, and Section 7.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 In perfect contexts, *chưa* serves as the negator. See Trinh (2005) and Phan (2023) for the distinction between *không* and *chưa*.

2 T. C. Nguyễn (1979) posits that during the 8th and 9th centuries, the Chang'an dialect of Chinese was taught in *Jiaozhou*/Giao Châu (modern northern Vietnam) as a foreign language. Phan (2013) further suggests that during this period, in addition to the local language spoken by the indigenous population (Vietnamese), there existed a contact language referred to as *Annamese Middle Chinese*. Rather than a mere contact variety, *Annamese Middle Chinese* was the home dialect of urban Sinitic communities in the Red River Delta, linked to other Chinese dialect islands in Guangxi and Hunan. After Vietnam's independence, these Sinitic settlers gradually shifted to speaking Annamese natively—imbuing it with heavy Sinicisms (Phan & de Sousa, 2022; Phan, 2025). This linguistic variety was part of a broader Chinese dialectal continuum, stretching from Chang'an to Giao Châu. Despite differences in their hypotheses, both scholars agree on the influence of the phonological system of the Chang'an dialect of the Tang dynasty in shaping the Sino-Vietnamese reading system (*Hán-Việt*) in Vietnam.

3 Translated texts such as our *Analects* versions offer a valuable, diachronic parallel between source and target, enabling us to observe negator choices in uniform contexts and tie shifts to specific periods. Yet we recognize that translators' aims—balancing literal fidelity against target-language fluency—may influence negator distribution. To separate translation conventions from genuine linguistic change, future work should draw on non-translated materials (official edicts, literary prose, folk tales, modern journalism) as well as both Sino-Vietnamese and purely native texts. A broad, multi-genre diachronic investigation will provide a fuller understanding of how bất-, vô-, and phi- have developed in Vietnamese. We leave this extension to future research. We are grateful to a reviewer for drawing attention to this important consideration.

4 The Chinese character 陳 has two pronunciations: *trận*/*zhèn* (departing tone 去), meaning battle, and *trần*/*chén* (level tone 平), meaning the Trần/Chén family.

5 In (39a), 非 behaves as a sentential negator—essentially a verb meaning 'not'—able to stand alone and take the interrogative particle 與. Unlike the prefixal use of 非 in compounds such as phi-pháp (<非法), where 非 is bound to the noun 法, this verbal use was never adopted into Vietnamese. We thank a reviewer for discussing this point.

References

Alves, M. (2007a). Categories of grammatical Sino-Vietnamese vocabulary. *The Mon-Khmer Studies*, *37*, 217–229.

Alves, M. (2007b). Sino-Vietnamese grammatical borrowing: An overview. In Y. Matras, & J. Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 343–361). De Gruyter Mouton.

Alves, M. (2017). Chinese loanwords in Vietnamese. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics* (Vol. 1, pp. 585–592). Brill.

Ames, R., & Rosemont, H. (1998). *The analects of confucius: A philosophical translation*. Ballantine Books.

Arcodia, G. F. (2012). Constructions and headedness in derivation and compounding. *Morphology*, *22*, 365–397. [CrossRef]

Gardner, D. (2003). *Zhu Xi's reading of the Analects: Canon, commentary, and the classical tradition*. Columbia University Press.

Guo, X., Zhang, W., & Li, M. (1999). *Gudai Hanyu—Shang* 古代漢語—上 [Old Chinese—Volume I]. Shangwu Yinshuguan.

Hashimoto, M. J. (1978). Current developments in Sino-Vietnamese studies. *Journal of Chinese Linguistics*, *6*, 1–6.

Haudricourt, A.-G. (2010). The origin of the peculiarities of the Vietnamese alphabet. *Mon-Khmer Studies*, *39*, 89–104.

Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge University Press.

Ivanhoe, P. J. (2000). *Confucian moral self cultivation*. Hackett Publishing.

King, R. (2021). Vernacular reading in the sinographic cosmopolis and beyond. In K. Bunkyō (Ed.), *Literary Sinitic and East Asia: A cultural sphere of vernacular reading* (pp. ix–xl). R. King, M. Burge, S. N. Park, A. Lushchenko, & M. Hattori (Trans.). Brill.

Lau, D. C. (1979). *Confucius: The analects*. Penguin Books.

Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Mair, V. H. (2004). Review of *Sinitic Grammar: Synchronic and Diachronic Perspectives*, edited by H. Chappell. *Sino-Platonic Papers*, *14*, 8–14.

Nguyễn, N. Y. (1999). *Đại từ điển Tiếng Việt* [Great dictionary of Vietnamese]. Culture & Information Publisher.

Nguyễn, P. P. (1994). Quelques aspects de la négation en vietnamien. *Cahiers de Linguistique—Asie Orientale*, *23*, 231–239. [CrossRef]

Nguyễn, Q. H. (2001). *Truyền kỳ mạn lục giải âm* [A transcribed and annotated version of Truyền kỳ mạn lục]. Social Sciences Publishing house.

Nguyễn, T. C. (1979). *Nguồn gốc và quá trình hình thành cách đọc Hán Việt* [The origin and formation of Sino-Vietnamese pronunciation]. Social Science Publisher.

Nguyễn, T. C. (2009). *Transliteration of Luận ngữ ước giải* 論語約解 [Concise interpretation to the Analects]. (Call number AB.270/3). Unpublished manuscript. Archives of Institute of Sino-Nom Studies.

Nguyễn, T. C. (2014). Tiếp cận văn bản học với Tứ thư ước giải [A textual approach to concise interpretation to The Four Books]. *Tạp chí Hán Nôm [Journal of Sino-Nom Studies]*, *2*, 27–45.

Nguyễn, T. C. (2016). Dịch Nôm kinh điển Nho gia tại Việt Nam nhìn từ khía cạnh tư liệu, phiên dịch, và thông diễn kinh điển [The translation of classical Confucian texts into Nôm in Vietnam: Perspectives on source materials, translation, and interpretation of classical texts]. *Tạp chí Hán Nôm [Journal of Hán Nôm Studies]*, *2*, 12–29.

Nguyễn, T. C., & Bùi, A. C. (2020). The Chinese script, Confucian script, and Nôm script: Some reflections on writing and politics in monarchical Vietnam. *Journal of Chinese Writing Systems*, *4*, 149–160. [CrossRef]

Peyraube, A. (1996). Recent issues in Chinese historical syntax. In C.-T. Huang, & Y.-H. A. Li (Eds.), *New horizons in Chinese linguistics* (pp. 161–213). Kluwer Academic Publishers.

Phan, J. (2013). *Lacquered words: The evolution of Vietnamese under Sinitic influences from the 1st century BCE through the 17th century CE* [Ph.D. dissertation, Cornell University].

Phan, J. (2025). *Lost tongues of the Red River: Annamese Middle Chinese & the origins of the Vietnamese language*. Harvard University Asia Center.

Phan, J., & de Sousa, H. (2022). Southwestern Middle Chinese: Preliminary evidence from Hunan, Guangxi, and Sino-Vietnamese. In T. Phan, J. Phan, & M. Alves (Eds.), *Vietnamese linguistics: State of the field* (pp. 59–79). (Journal of the Southeast Asian Linguistics Society Special Publication No. 9). University of Hawai◎i Press.

Phan, T. (2023). *The syntax of Vietnamese tense, aspect, and negation*. Routledge.

Phan, T., Nguyen, T. C., & Shimizu, M. (2024). The rise of negative markers in Vietnamese: Evidence from primary Nôm texts. In T. Phan, N. T. Cuong, & M. Shimizu (Eds.), *Studies in Vietnamese historical linguistics: Southeast and East Asian contexts* (pp. 29–50). Springer Nature.

Phạm, V. K. (2004). *Khổng Phu Tử và Luận ngữ* [Confucius and the Analects]. National Political Publisher.

Pulleyblank, E. G. (1995). *Outline of classical Chinese grammar*. University of British Columbia Press.

Sakel, J. (2007). Types of loan: Matter and pattern. In Y. Matras, & J. Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 15–30). De Gruyter Mouton.

Shimizu, M. (2017). Chinese and Vietnamese. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics* (Vol. 1, pp. 389–392). Brill.

Sinh, V. (1993). Chinese characters as the medium for transmitting the vocabulary of modernization from Japan to Vietnam in Early 20th century. *Asian Pacific Quarterly*, *25*(1), 1–16.

Takahashi, Y. (2022). The grammatical nature of Sino-Vietnamese verb-object compounds. *JSEALS*, *8*, 232–243.

Trinh, T. (2005). *Aspect of clause structure in Vietnamese* [Master's thesis, Humboldt University].

Trinh, T., Phan, T., & Vu, D. N. (2024). Negation and polar questions in Vietnamese: Present and past. *Taiwan Journal of Linguistics*, *22*(1), 67–88.

Tự Đức. (1828–1883). *Tự Đức thánh chế Luận ngữ thích nghĩa ca* 嗣德聖制論語釋義歌 [Tự Đức imperial explanation of the Analect]. call number A.186/1-2]. Institute of Sino-Nom Studies.

Vogelsang, K. (2021). *Introduction to classical Chinese*. Oxford University Press.

Washizawa, T. (2019a). *Nghiên cứu đối chiếu các hư từ Hán văn dịch sang các hư từ văn Nôm trong tác phẩm song ngữ Hán Nôm Truyền Kỳ Mạn Lục* [A contrastive study of Chinese functional words translated into Vietnamese in the Sino-Vietnamese bilingual work Truyền Kỳ Mạn Lục] [Ph.D. dissertation, Vietnam National University Hanoi].

Washizawa, T. (2019b). Process of grammaticalization of function words của, sự, không, and bị in Vietnamese. *Tokyo University Linguistic Papers (TULIP)*, *41*, 367–392.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

On Similative Reduplication in Vietnamese

Tue Trinh

Center for Cognitive Science of Language, University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia; tue.trinh@ung.si

Abstract: Vietnamese has a productive reduplication strategy where the reduplicant appears to the right of the base and is segmentally identical to the base except that its last rime is *iếc* (pronounced [ɪək]). In this note, I attempt to account for some observations about *iếc*, including the fact that it gives rise to ignorance inferences and is incompatible with classifiers. I propose that the semantics of *iếc* parallel the pragmatics of disjunctions and that the notion of similarity underlying the interpretation of *iếc* is contextual while that underlying the interpretation of classifiers is grammatical. The proposal makes use of some tools of predicate logic and formal semantics.

Keywords: reduplication; similarity; ignorance; classifiers; Vietnamese

1. Introduction

Vietnamese has a productive reduplication strategy where the reduplicant appears to the right of the base and is segmentally identical to the base except that its last rime is *iếc* [ɪək] (Maspero, 1912, p. 109; Emeneau, 1951, p. 186; Thompson, 1987, p. 173; Nguyễn Đình Hoà, 1997, pp. 53–59; Vu, 1998, pp. 172–174; Đoàn Thiện Thuật, 2007, p. 71; Ngo, 2021, p. 135). Some examples are given in (1):[1]

(1)

sách 'book'	→	sách siếc
[sɛk]		[sɛk.sɪək]
cam 'orange'	→	cam kiếc
[kam]		[kam.kɪək]
thuyền 'boat'	→	thuyền thiếc
[tʰwɪən]		[tʰwɪən.tʰɪək]
sinh viên 'student'	→	sinh viên sinh viếc
[sɪŋ.vɪən]		[sɪŋ.vɪən.sɪŋ.vɪək]
mỳ Ý 'spaghetti'	→	mỳ Ý mỳ iếc
[mi.i]		[mi.i.mi.ɪək]
sinh viên 'student'	→	sinh viên sinh viếc
[sɪŋ.vɪən]		[sɪŋ.vɪən.sɪŋ.vɪək]
ca-mê-ra 'camera'	→	ca-mê-ra ca-mê-riếc
[ka.me.ra]		[ka.me.ra ka.me.rɪək]
philosophy	→	philosophy philosophiếc
[fi.lo.so.fi]		[fi.lo.so.fi.fi.lo.so.fɪək]

For present purposes, I will assume that the Vietnamese lexicon contains a morpheme, and I will call it RED, which is mnemonic for "reduplication". The rule in (2) captures how this morpheme is phonologically realized:

(2) Morpho-phonemic rule
Let [$_w$ ___X] be a word whose last rime is X
[$_w$ ___X]+RED → $\underbrace{[_w \text{ ___X}]}_{base}$+$\underbrace{[_w \text{ ___ɪək}]}_{reduplicant}$

The application of this rule is exemplified in (3) for *cam* 'orange' and *sinh viên* 'student'. The rimes targeted by the rule are underlined:

(3) a. [k<u>am</u>]+RED → [k<u>am</u>]+[k<u>ɪək</u>]
b. [sɪŋ.v<u>ɪən</u>]+RED → [sɪŋ.v<u>ɪən</u>]+[sɪŋ.v<u>ɪək</u>]

As we can see, there is no constraint on the number of syllables in the base: *sách* 'book' and *cam* 'orange' have one syllable, *sinh viên* 'student' has two, *ca-mê-ra* 'camera' has three, and *philosophy* has four syllables. Note, also, that (2) does not distinguish between words of different origins. Among the examples in (1), *sách* 'book' and *cam* 'orange' are "native" Vietnamese words. With the possible exception of historical linguists, no-one can say where these words came from. In contrast, most speakers will be able to recognize that *sinh viên* 'student' is "Sino-Vietnamese" ("Hán Việt"), i.e., that it was "imported" from Chinese. However, it has undergone complete Vietnamization: both of its syllables occur independently as syllables in native Vietnamese words, and there is no native counterpart which expresses the same meaning. The tri-syllabic *ca-mê-ra* is transparently a loan word. It is quite often written with hyphens between the syllables. This fact is instructive. Vietnamese is a "monosyllabic" language, in the sense that every morpheme is a syllable.[2] Empty spaces are inserted into Vietnamese texts following the same rule that holds for European languages like German: namely, between the smallest linguistic units that are guaranteed to be a sequence of syllables (Trinh, 2014). In German, these are words, but in Vietnamese these are morphemes.[3] The fact that Vietnamese speakers (often) write <ca-mê-ra>, not <ca mê ra>, indicates their intuition that these three syllables are not morphemes, i.e., that this word is not a "regular" word in their language.[4] Also, the last syllable, [ra], is perceived by speakers of the standard dialect as "exotic", as [r] is not a possible onset in this dialect. As for *philosophy*, it is clearly a foreign word. The Vietnamese word for 'philosophy' is *triết học*. However, if *philosophy* is used as a foreign word in conversation then the option of reduplicating it as *philosophy philosophiếc* is available, together with all of its concomitant semantic effects.

What are these effects? The most prominent one is that *iếc*-reduplication licenses the inference of "uncertainty" on the part of the speaker. Consider the conversation below:[5]

(4) A: John đang ăn gì thế?
John is eating what?
B: Nó đang ăn cam+kiếc gì đó
He is eating an orange+RED DEM$_{wh}$

B's answer in (4) conveys the message that B is not sure whether John is eating an orange or just something "similar" to an orange. In other words, B does not rule out the possibility that John is in fact eating an orange, but nor does she rule out the possibility that John is eating, say, a tangerine. This intuition is corroborated by the fact that it is odd for someone to contest B's answer with the statement that John is actually eating a tangerine. In the same way, it is odd to contest it with the statement that John is eating an orange:[6]

(6) A: John đang ăn gì thế?
John is eating what?
B: Nó đang ăn cam+kiếc gì đó
He is eating an orange+RED DEM$_{wh}$

C: #Không đúng. Nó đang ăn quýt / cam
That's not true. He is eating a tangerine / orange.

I will call *iếc*-reduplication in Vietnamese "similative reduplication", borrowing a term from works on phenomena in other languages that share some family resemblance to what we are now describing.[7] Note that if B had used the plain and not the reduplicated form of *cam* 'orange', such a follow-up as C's would be totally natural. This shows that RED does make an interpretive difference:

(7) A: John đang ăn gì thế?
John is eating what?
B: Nó đang ăn cam.
He is eating an orange.
C: Không đúng. Nó đang ăn quýt.
That's not true. He is eating a tangerine.

Before we go on, a remark about the bounds of this note should be made. The reader will have seen that in the examples above, the reduplicated form of *cam*, namely, *cam+kiếc*, co-occurs with the expression *gì đó*, which I gloss as DEM_{wh}. The reason for this gloss is that *gì* has a life of its own as a question word, and *đó* is also the demonstrative article.[8]

(8) a. John đang ăn gì?
John is eating what?
'What is Nam eating?'
b. John đã mua quyển sách đó.
John bought that CL book
'Nam bought that book.'

Intuitively, the reduplication form is more natural with than without *gì đó*. I will put aside the question as to what underlies the preference for this co-occurrence. The semantic effects of RED together with DEM_{wh} will be analyzed as arising from RED alone. A more detailed account of the construction that derives such effects compositionally from its subconstituents—i.e., RED, *gì*, and *đó*—will have to be left for future research.[9]

I will end this introduction by stressing that *iếc*-reduplication is just one among many reduplication strategies of Vietnamese. My gloss of it as RED should not mislead the reader into thinking that this is *the* reduplication process in Vietnamese. Given the bound of this note, I will not present an overview of reduplication in Vietnamese. I refer the reader to Hoàng Văn Hành (2008) for such an overview.

Let us now analyze the uncertainty inference associated with RED.

2. Uncertainty

I propose that the reduplication morpheme RED weakens the meaning of the base word. Specifically, RED maps a predicate P to another predicate $sim(P)$, where $sim(P)$ is defined as in (9):

(9) $sim(P) =_{\text{def}} \{x \mid x \text{ is something similar to a } P\}$

Assuming that identity is the limiting case of similarity, i.e., that everything is maximally similar to itself, applying the *sim* function to a predicate P amounts to disjoining P with predicates similar to P.[10] Thus, $sim(P)$ means $P \vee P' \vee P'' \ldots$, where P', P'', ... are similar to P.[11] Let us consider an example, supposing that *quýt* 'tangerine' and *bưởi* 'grapefruit' are the predicates similar to *cam* 'orange':[12]

(10) a. $[\![\text{cam}]\!] = [\![\text{orange}]\!] = \{x \mid x \text{ is an orange}\}$;

b. ⟦cam hoặc quýt hoặc bưởi⟧
= ⟦orange or tangerine or grapefruit⟧
= $\{x \mid x \text{ is an orange}\} \cup \{x \mid x \text{ is a tangerine}\} \cup \{x \mid x \text{ is a grapefruit}\}$;
c. ⟦cam kiếc⟧ = ⟦cam+RED⟧ = sim(⟦cam⟧) = sim(⟦orange⟧)
= $\{x \mid x \text{ is similar to an orange}\}$
= $\{x \mid x \text{ is an orange}\} \cup \{x \mid x \text{ is a tangerine}\} \cup \{x \mid x \text{ is a grapefruit}\}$.

We see that *orange*+RED ends up meaning the same as *orange or tangerine or grapefruit*. Now, a disjunction is logically weaker than any individual disjunct: if it is true that John is eating an orange then it is also true that John is eating an orange *or* a tangerine. Given that $sim(P)$ is $P \vee Q$, we have (11-a) and (11-b):[13]

(11) a. $P \Rightarrow sim(P)$;
b. $\underbrace{\text{John is eating an orange}}_{p} \Rightarrow \underbrace{\text{John is eating an orange+RED}}_{p \vee q}$

We are now in a position to make a first attempt at explaining the uncertainty inference associated with RED. Consider the disjunctive statement in (12):[14]

(12) John ate an orange or a tangerine
$\rightsquigarrow \neg K_S(\textit{John ate an orange})$

A clear intuition about (12) is that it conveys the message that the speaker is not certain that John ate an orange.[15] This fact, as we know, can be accounted for in terms of Gricean reasoning. Grice (1967) proposes the following "maxims of conversations", i.e., rules that discourse participants are mutually assumed to be following when they talk:

(13) a. Quality: Say only what you believe to be true!
b. Quantity: Be informative!

Two things should be noted about these maxims. Firstly, they are ranked. Specifically, Quality takes precedence over Quantity, which means that one should never sacrifice truthfulness for informativity: one should try to convey as much information as possible, provided one is not lying. It follows, then, that what is more informative than what one says is also what one does not believe to be true. Secondly, informativity respects logical strength: if p is logically stronger than q then p is more informative than q.[16] As *John ate an orange* is logically stronger than *John ate an orange or a tangerine* it is also more informative, and the fact that the speaker asserts the latter proposition but not the former means, given the maxims, that she does not believe the former, i.e., that she is not certain that John ate an orange.[17]

Are we done? Well, no. It turns out that there is a problem in my account of the facts. It is based on the claim that assertion of a proposition p licenses the inference that $\neg K_S(q)$, where q is stronger than p. This claim, I argued, can be derived from the Gricean maxims of Quality and Quantity. However, there is good reason to believe that the claim is false. Consider the following conversation:

(14) A: Where does John live?
B: He lives in France.
$\not\rightsquigarrow \neg K_S(\textit{John lives in Paris})$

Suppose we happen to hear this exchange out of the blue, say between two people on the subway, we would *not* conclude that B is not sure John lives in Paris. B may very well know the particular street address in France at which John lives, but simply does not feel the need

to be that specific. Given that *John lives in Paris* is more informative than *John lives in France*, this intuition we have about (14) conflicts with what I said above, namely, that utterance of a weaker proposition p gives rise to the inference that the speaker is not sure about a stronger proposition q. Here is another example:

(15) A: Who did John talk to?
B: He talked to a student.
$\not\rightsquigarrow \neg K_S(\textit{John talked to a male student})$

Given that every male student is a student but not vice versa, a more informative statement that B could have made is that John talked to a male student. However, B's answer in (15), out of the blue, clearly does not imply that she is not sure that John talked to a male student.

Here is, then, the dialectical situation. I have proposed that *John is eating an orange*+RED implies that the speaker is not certain *John is eating an orange*, because the former is less informative than the latter, and the Gricean maxims, Quality and Quantity, together tell us that the speaker asserts the most informative proposition of which she is certain. We then saw examples that challenge this analysis, i.e., the conversations in (14) and (15), where the assertion of a less informative proposition does *not* imply the speaker's uncertainty about a more informative proposition. How do we resolve this conflict? Quality and Quantity seem to be virtual truisms of language use, so we do not want to tinker with these principles. And we cannot tinker with logic either: if John talked to a male student then he talked to a student. There is no way out of that. What we could do is either (i) revise our semantics of RED, withdrawing the claim that *orange*+RED is weaker than *orange*, or (ii) add auxiliary hypotheses to the account. I propose we take the latter course.

One addition we need turns out to be another Gricean maxim, namely, that of Relation. What the maxim of Relation tells us is that we should be "relevant" (Lewis, 1988), i.e., answer the "question under discussion" (Roberts, 1996). Recall that Quality and Quantity together derive the claim that the speaker asserts the most informative proposition of which she is certain. With Relation added to the mix, we derive something weaker: the speaker asserts the most informative *relevant* proposition of which she is certain. This means that the assertion of a weaker proposition p would imply the speaker's uncertainty about a stronger proposition q only if q is relevant, i.e., only if there is interest in knowing whether q is true. That this result is empirically correct can be seen from reflecting on our intuition about a slight modification of the scenarios in (14) and (15). Consider (16):

(16) A: Does John live in Paris?
B: He lives in France.
$\rightsquigarrow \neg K_S(\textit{John lives in Paris})$

Changing A's question from *where does John live?* to *does John live in Paris?*, we made it clear that the proposition *John lives in Paris* is relevant, i.e., that we are interested in knowing whether it is true. Given this set up, B's assertion of the proposition that John lives in France, which is less informative than the proposition that he lives in Paris, is predicted to license the inference that the speaker is not certain that John lives in Paris. This prediction is confirmed by our intuition. The same holds for the modified version of (15), which is (17):

(17) A: Did John talk to a male student?
B: He talked to a student.
$\rightsquigarrow \neg K_S(\textit{John talked to a male student})$

A's question makes sure that the proposition that John talked to a male student is relevant. Given this background, assertion of a less informative proposition—in this case, *he talked to a student*—leads to the inference that the speaker is not certain John talked to a male student.

Is the addition of the maxim of Relation enough to derive the fact that RED implies the speaker's uncertainty? No. What I have said about the semantics of RED is only that it weakens the meaning of the base. In order to generate speaker uncertainty about this meaning, I also need to say that RED makes the base relevant. Thus, there are two components to the interpretation of *John is eating an orange*+RED. Firstly, it is weaker than *John is eating an orange*. Secondly, it makes *John is eating an orange* relevant.[18]

(18) Interpretation of RED (first version, to be revised):
   a. $P \Rightarrow P$+RED;
   b. P+RED makes P relevant.

Let us recap. Consider, again, the sentence in (19):

(19) John đang ăn cam+kiếc gì đó.
   John is eating an orange+RED DEM_{wh}
   $\rightsquigarrow \neg K_S(\textit{John is eating an orange})$

Given (18), this sentence is weaker than *John is eating an orange* and, at the same time, makes *John is eating an orange* relevant. From Quality, Quantity, and Relation we conclude that the speaker is not certain John is eating an orange.

Are we done? As it turns out, no. Let us see why not.

3. Ignorance

There is a crucial difference between the speaker's uncertainty in (16) and the speaker's uncertainty in (19) that we have not discussed.

(20) A: Does John live in Paris?
   B: He lives in France.
   Possible inference: $K_S\neg(\textit{John lives in Paris})$

It is, intuitively, possible to understand B's answer in (20) as conveying the message that B knows that John does *not* live in Paris. Specifically, B knows that John lives in some city in France, and while she is not sure which city it is, she knows that it is not Paris. Note that this inference is stronger than but consistent with the uncertainty inference that we claim is associated with B's answer in (20), namely, $\neg K_S(\textit{John lives in Paris})$. Now, consider (19), reproduced below as (21):

(21) John đang ăn cam+kiếc gì đó.
   John is eating an orange+RED DEM_{wh}
   Not a possible inference: $K_S\neg(\textit{John is eating an orange})$

A clear intuition about (21) is that it does *not* convey the message that the speaker knows John is not eating an orange. The speaker does not rule out the possibility that John is eating, say, a tangerine instead of an orange, but neither does she rule out the possibility that John is, in fact, eating an orange. Thus, the uncertainty inference we have associated with (21), namely, $\neg K_S(\textit{John is eating an orange})$, is too weak. It must be strengthened into a claim that is inconsistent with $K_S\neg(\textit{John is eating an orange})$. One obvious way to do this is to conjoin the uncertainty about John eating an orange with the uncertainty about John *not* eating an orange. Here is our empirical claim:

(22) John đang ăn cam+kiếc gì đó.
John is eating an orange+RED DEM_{wh}
$\rightsquigarrow \neg K_S(\textit{John is eating an orange}) \wedge \neg K_S \neg (\textit{John is eating an orange})$

Inferences of the form $\neg K_S p \wedge \neg K_S \neg p$ are "ignorance inferences". We say that speaker S is "ignorant" about proposition p if S is not certain that p and is not certain that $\neg p$. Ignorance about p is stronger than uncertainty about p: it is uncertainty about p conjoined with uncertainty about $\neg p$.

How do we explain the ignorance inference associated with RED? Specifically, what do we need to say about the meaning of *orange*+RED so that the sentence *John is eating an orange*+RED conveys the message that the speaker is not certain that John is eating an orange and is not certain that John is not eating an orange? A good starting point, again, is to look to disjunction. It is well-known that disjunction licenses ignorance inferences: the speaker is understood to be uncommitted to the truth *and* uncommitted to the falsehood of the individual disjuncts. This makes a disjunction different from an apparently equivalent vague term. Compare (20) with (23):

(23) A: Does John lives in Paris?
B: He lives in Paris or Nice or Toulouse or Lyon or Marseille ...
Not a possible inference: $K_S \neg$ John lives in Paris

Suppose B listed all locations in France. In that case, what B says in (23) would be logically equivalent to what B says in (20). But, as we have seen, B's answer in (20) allows for the possibility that B knows John does *not* live in Paris. B's answer in (23), however, clearly implies that B does not know whether John lives in Paris or not. The crucial difference between a vague term and a disjunction of specific terms is this: the disjunction necessarily makes all the disjuncts relevant (Chemla, 2008; Sauerland, 2004), while a vague term does not necessarily make the more specific terms relevant. B's utterance in (23), i.e., *Paris* $\vee$ *Nice* $\vee$ *Toulouse* ..., is weaker than *Paris*, and at the same time makes *Paris* relevant. Therefore, it implies that B is ignorant about *Paris*, i.e., that she is committed neither to the claim that John lives in Paris nor to the claim that he does not live in Paris. B's utterance in (20), on the other hand, does not (necessarily) make *Paris* relevant, so no conclusion about *Paris* is drawn, and the utterance is consistent with B's knowing that John does not live in Paris.

What we can say to explain the ignorance inferences associated with P+RED, then, is that P+RED makes P relevant and all other alternatives relevant. In other words, $sim(P)$ is not only interpreted as $P \vee P' \vee P'' \ldots$, i.e., as a logically more inclusive concept, but also as *P or P′ or P″* ..., i.e., as an *expression* of this more inclusive concept in the form of a syntactic disjunction. Another way to say this is that P+RED has not only the semantics but also the pragmatics of a disjunction:

(24) A: What is John eating?
B: He is eating an orange+RED
$\not\approx$ 'he is eating something similar to an orange'
$\approx$ 'he is eating an orange or a tangerine or a grapefruit ...'

(25) Interpretation of RED (final version)
a. P+RED is interpreted as $P \vee P' \vee P'' \ldots$, where P', $P'' \ldots \Rightarrow sim(P)$;
b. P+RED makes P, P', P'', ... relevant.

4. Classifiers

Vietnamese is a "classifier language". The combination of nouns and numerals must be mediated by a function word, i.e., the so-called "classifier":

(26) a. John ăn một *(quả) cam
John ate one CL orange
b. John mua hai *(quyển) sách
John ate two CL book

For present purposes, I will assume the account proposed by Trinh (2011) of this fact.[19] Specifically, I will assume that nouns in classifier languages such as Chinese and Vietnamese differ from their counterparts in non-classifier languages such as English and German in being "number neutral": whereas English *book* denotes the set of singular (i.e., atomic) books, Vietnamese *sách* 'book' denotes the set containing both singular books and pluralities of books. I will write "a+b" for the plurality consisting of a and b, etc.

(27) a. $[\![\text{book}]\!] = \{x \mid x \text{ is a singular book}\} = \{a, b, c, \dots\}$
b. $[\![\text{sách}]\!] = \{x \mid x \text{ is a singular book or a plurality of books}\}$
$= \{a, b, c, \dots, a{+}b, a{+}c, b{+}c, a{+}b{+}c, \dots\}$

As counting books means counting atomic books,[20] the noun *sách* must be "atomized" before it can compose with a numeral. This task of atomization is carried out by the classifier. Specifically, the classifier maps a predicate P to the set of entities that are singular Ps, i.e., atoms of P.

(28) Interpretation of CL:
$[\![\text{CL}]\!] = \lambda P.\ \lambda x.\ x$ is a singular P.

Applying CL to *sách* 'book' yields the following set:

(29) $[\![\text{quyển sách}]\!] = \{x \mid x \text{ is a singular book}\} = \{a, b, c, \dots\}$

I have been speaking of "the classifier", as if there is a unique classifier for all nouns in Vietnamese. The reader will have seen from the example sentences in (26) that this is not the case. The classifier for *cam* 'orange' is *quả*, while that for *sách* 'book' is *quyển*. Exchanging the classifiers will result in severe deviance:

(30) a. *John ăn một quyển cam;
b. *John mua hai quả sách.

A selectional relationship, then, obtains between the classifier and the noun. The relationship is not purely syntactic. One can say that the classifier imposes a semantic requirement on the noun it combines with. In fact, the requirement seems to be that the noun must belong to a class of predicates that are "similar" in some sense. Take *quyển*, for example. It selects not only *sách* 'book' but also *sổ* 'logbook', *vở* 'notebook', *lịch* 'calendar', *tạp chí* 'magazine', etc. There is certainly some similarity between these concepts: they all denote things that have pages. As for *quả*, it selects not only *cam* 'orange' but also other fruits, such as *quýt* 'tangerine', *bưởi* 'grapefruit', *táo* 'apple', etc. The similarity in this case is quite obvious. This means we should revise our view on classifier semantics. Firstly, we should not speak of "the classifier". There are many classifiers, each selecting nouns of a certain "similarity class". Secondly, (28) is not the interpretation of any particular expression but an interpretation *scheme* for many expressions: it is what underlies the semantics of all classifiers. The interpretation of each individual classifier will have to come with presuppositions about the class of predicates that it selects. We will consider *quyển* as a representative case. Here is the entry for it:

(31) Interpretation of *quyển* (first version, to be revised):
$[\![\text{quyển}]\!] = \lambda P : P \Rightarrow sim(book).\ \lambda x.\ x$ is a singular P.

The reader can see that for the semantics of *quyển* I am making use of the same *sim* function that I used for the semantics of RED. This move is motivated by the hope that the same notion of similarity is operative in both cases. What (37) is intended to say is that *quyển* selects for predicates that are "similar" to *book* in the same way that the predicates falling under *book*+RED are, e.g., *notebook*, *logbook*, *magazine*, etc. It would, of course, be an interesting result if the hope turns out to be actualized, as that would show that we have found independent evidence for *sim* from another corner of grammar.

The hope, unfortunately, is *not* actualized. Consider the following data point:

(32) A: John đang đọc gì?
John is reading what?
'What is John reading?'
B: Nó đang đọc sách+siếc gì đó.
he is reading a book+RED DEM_{wh}
'He's reading a book+RED'
C: #Không đúng. Nó đang đọc báo.
not true he is reading a newspaper
'That's not true. He's reading a newspaper.'

The crucial observation here is that C's utterance is infelicitous. The intuition is that C objects to what B has said but then goes on to say something that does not contradict what B has said. What this suggests is that newspapers are considered similar to books as far as RED is concerned. In other words, the claim that John is reading something that falls under *book*+RED is consistent with the claim that John is reading a newspaper. Now, consider the following data point:

(33) a. John đang đọc một quyển sách / *báo
John is reading one CL_{quyen} book / newspaper;
b. John đang đọc một tờ *sách / báo
John is reading one CL_{to} book / newspaper.

As we can see, the classifier for *sách* 'book' is *quyển* and the classifier for *báo* 'newspaper' is *tờ*. Switching the classifiers gives rise to deviance. This means that, as far as the classifiers are concerned, books and newspapers are *not* similar.

What about the other classifier that I have mentioned, *quả*, which, as I have said, combines with nouns for fruits? Can we also argue that it refers to a different notion of similarity as compared to RED? The answer is yes. Consider (34):

(34) A: John đang mua cam+kiếc gì đó.
John is buying orange+RED DEM_{wh}
'John is buying orange+RED';
B: Không đúng. Nó đang mua bóng.
not true. He is buying balloons
'That's not true. He's buying balloons.'

The fact that B's response to A in (34) is perfectly felicitous suggests that as far as RED is concerned balloons are *not* similar to oranges: the claim that John is buying balloons can be used to negate the claim that John is buying *orange*+RED. Now, as it turns out, the classifier for *bóng* 'balloon' is also *quả*, i.e., the same one as for *cam* 'orange':

(35) John đang mua một quả cam / bóng.
John is buying one CL_{qua} orange / balloon.

This means that as far as the classifier *quả* is concerned, oranges and balloons *are* similar. The conclusion, then, is that *quả* and RED, and more generally classifiers and RED, refer

to two different notions of similarity. The question now is how we are to characterize this difference? I propose that we appeal to the semantics–pragmatics divide. Specifically, I propose that the notion of similarity underlying the interpretation of RED is *contextual*, notationally represented by the function sim_c, while that underlying the interpretation of classifiers is *grammatical*, notationally represented by the function sim_g. What I mean by this is simply that the selectional property of a classifier such as *quyển* is lexically specified: we can, in principle, list all nouns that can combine with *quyển*, i.e., all predicates in $sim_g(book)$, and we know that this list does not contain *báo* 'newspaper'. On the other hand, *book*-RED is a predicate whose extension varies from context to context: it may include things that fall under *báo* 'newspapers' in some but not all situations:[21]

(36) Interpretation of RED (final version)
   a. P+RED is interpreted as $P \vee P' \vee P'' \ldots$, where P', $P'' \Rightarrow sim_c(P)$;
   b. P+RED makes P, P', P'', ... relevant.

(37) Interpretation of *quyển* (final version):
$[\![\text{quyển}]\!] = \lambda P : P \Rightarrow sim_g(book).\ \lambda x.\ x$ is a singular P.

Note that, given (36) and (37), we predict that RED makes the noun it combines with incompatible with classifiers. This prediction is borne out, as it is certainly not the case that $book$+RED $\Rightarrow sim_g(book)$, (38) is a presupposition failure.

(38) *John đang đọc một quyển sách-siếc gì đó
John is reading one CL_{quyen} book+RED DEM_{wh}

Now, there is an interesting prediction that we make. We have considered the case where N+RED combines with a classifier. What about the case where RED combines with the classifier+noun complex, i.e., the case of [CL+N]-RED. The account just proposed actually predicts that such construction should be acceptable. Let us assume, following several works on the semantics of classifiers, that the classifier maps a noun's meaning, which in languages like Vietnamese would be a cumulative predicate, to an atomic predicate. Thus, whereas *sách* denotes a set that contains both singular books and pluralities of books, *quyển sách* denotes a set that contains only singular books. This means that *quyển sách*-RED would just denote the set containing things that are similar to singular books. This set is practically just the set of things similar to books. What is different now is that there should be no presupposition failure due to the selectional requirement of *quyển*, because what it combines with is just *sách*, not *sách*-RED. This prediction is borne out, as evidenced by the acceptability of (39).[22]

(39) Nam đang đọc quyển sách-quyển siếc gì đó
Nam is reading a singular book-RED DEM_{wh}
'Nam is reading a (singular) book or something like that'

Similarly, (40) is acceptable:

(40) Nam đang ăn quả cam-quả kiếc gì đó
Nam is eating a singular orange-RED DEM_{wh}
'Nam is eating a (singular) orange or something like that'

Excursus on encapsulation—What I have just said amounts to claiming that grammar is "blind" to the fact that newspapers are similar to books (in the sense that they are both reading materials which have pages) or, in the same way, blind to the fact that balloons are not similar to oranges in the same way that tangerines are. More generally, my description of similative reduplication in Vietnamese subscribes to the view that grammatical knowledge

is, to some extent, "encapsulated" from contextual, i.e., encyclopedic, knowledge. This view has been defended elsewhere (cf. Fox, 2000; Fox & Hackl, 2006; Magri, 2009). However, it is not clear to what extent it has become consensus in the field, so I will end this section with a brief illustration of encapsulation, using two examples from the literature.

The first is taken from (Fox & Hackl, 2006). These authors address the puzzle that *John has three children* implies John does not have four children, while *John has more than three children* does not imply John does not have more than four children.[23] Their explanation is based on the hypothesis that the scale underlying measurement in natural language is dense, not discreet. This means that, as far as grammar is concerned, counting is carried out in terms of rational, not natural, numbers, and that the quantity implicatures that are potentially associated with (41-a) and (42-a) are (41-b) and (42-b), respectively:

(41) a. John has three children.
b. ¬John has n children, for each *rational* number n greater than 3.

(42) a. John has more than three children.
b. ¬John has more than n children, for each *rational* number n greater than 3.

A sentence ϕ is associated with an implicature ψ only if $\phi \wedge \psi$ is consistent, i.e., non-contradictory. Since (41-a) $\wedge$ (41-b) is consistent, (41-a) ends up meaning John has three but not four children. In contrast, (42-a) $\wedge$ (42-b) is contradictory: if (42-a) is true, then it must be the case that John has $3 + \epsilon$ children, which means John has more than $3 + \epsilon/2$ children, which means there is a rational number n greater than 3, namely, $3 + \epsilon/2$, such that John has more than n children, which means (42-b) is false. Thus, the implicature in (42-b) is not associated with the sentence (42-a), and *more than three* just means more than three. The claim needed for (Fox & Hackl, 2006)'s account, which is relevant to our discussion, is this: grammar does not know that one cannot count children with non-natural numbers. The computation of quantity implicatures, which (Fox & Hackl, 2006) claim takes place in grammar, is encapsulated from the encyclopedic knowledge that John cannot have 3.5 children.

The second example is taken from Magri (2009), and it has to do with the fact that (43) is an odd sentence:

(43) #Some Italians come from a warm country.

Magri (2009) attributes the oddness of (43) to the fact that it licenses the quantity implicature that not all Italians come from a warm country, which (in conjunction with the literal meaning of the sentence) conflicts with our common knowledge that all Italians come from the same country. Now, in order for the implicature *not all* to be associated with *some*, *all* has to be more informative than *some*, but *all* is not more informative than *some* in this case. If we know, and we do, that all Italians come from the same country, saying that some of them come from a warm country should convey the same amount of information as saying that all of them do. This means that for the grammar to associate the implicature *not all Italians come from a warm country* with the sentence *some Italians come from a warm country*, it must *not* know that all Italians come from the same country. Again, we see that a computation that is carried out in grammar is blind to contextual knowledge. **End of excursus.**

5. Open Questions and Conclusions

The discussion above has concerned *iec*-reduplication of nouns. An anonymous reviewer points out that the same process can apply to verbs also. The example they provide is (44):

(44) Giờ này mà còn tắm tiếc gì nữa

The sentence, spoken by A to B, means something like 'why would you do something like washing at this time of day?'.[24] The word to pay attention to here is *tắm tiếc*, which is the reduplicated form of *tắm*, which means 'wash'. My intuition is that *tắm tiếc* describes activities that are similar to washing. If this intuition is correct, an extension of the theory I proposed above that can deal with verbs seems to be in sight. I leave the task of working out its details for future research.

Another open question is the following puzzle: when there is no ignorance inference, RED conveys dismissiveness on the part of the speaker:

(45) A: Nam làm gì?
Nam do what
B: Giáo sư+giáo siếc gì đó
professor+RED DEM_{wh}

B's response to A's question in (45) suggests that B does not think very highly of professors, i.e., that she is "dismissive" towards them. The account I gave does not explain this fact. An issue to ponder here is how much of this interpretation is to be derived from 'purely pragmatic' considerations.[25] Again, I leave this problem to future research.

To conclude, I have argued for the following claims about similative reduplication in Vietnamese: (i) it gives rise to ignorance inferences in the same way disjunctions do; (ii) it refers to a notion of similarity that is different from the one referred to by classifiers. The second claim implies that (aspects of) grammatical knowledge can be encapsulated from encyclopedic knowledge.

Funding: This work has been financially supported by the Slovenian Research Agency (ARIS) project no. J6-4615.

Institutional Review Board Statement: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest.

Notes

1 Strings of phonemes will be written between square brackets. I should mention that Vietnamese is a tonal language but I do not indicate tones in my transcriptions here. In fact, there are other features of Vietnamese phonetics that are not represented in this note, e.g., the fact that the coda of *sách* is palatalized. I chose a level of detail that I think is sufficient for the question under discussion, which pertains to the semantic interpretation of *iếc*. A note should also be made about the third example in (1), which is the reduplication of *thuyền* as *thuyền thiếc*. I assume that the rime of *thuyền* is *uyền*, i.e., [wiən]. Under this assumption, the reduplication of *thuyền* as *thuyền thiếc* accords with the rule in (2). However, an anonymous reviewer remarks that the rime of *thuyền* is actually [iən], without the [w], and the rule in (2) should capture the claim that sometimes more than the rime is replaced. While I do not agree with the reviewer's view on the rime of *thuyền*, I do agree that if it is just [iən] then the rule in (2) has to be made more complicated.

2 For this discussion, I will use the term "morphemes" to mean phonologically realized morphemes.

3 German is not a monosyllabic language. Consequently, there are morphemes in German that are not syllables. Thus, the smallest linguistic units that are guaranteed to be a sequence of syllables are the words.

4 Strings of graphemes will be written between angled brackets.

5 I will represent a reduplicated word with a plus sign connecting the base and the reduplicant, and I will gloss the reduplicant as RED.

6 Intuitively, the oddness of C's correction in (6) is the same as that of B's in (5) below:

(5) A: John talked to Mary or Sue.
B: That's not true. He talked to Mary.

We come back to the similarity between *iếc*-reduplication and disjunction below.

7 An example is *bulut-mulut* 'cloud-RED' (Armoskaite & Kutlu, 2014, p. 271) in Turkish or *ketab-metab* 'book-RED' (Smith, 2020, p. 3) and *chai-mai* 'tea-RED' (Krifka & Modarresi, 2023, p. 323) in Persian.

8 I gloss the 'classifier' *quyển* as CL. We will discuss classifiers below.

9 Note that a variant of *đó* is *đấy*. Thus, *gì đó* in the examples here can be exchanged with *gì đấy*. I thank an anonymous reviewer for drawing my attention to this fact.

10 I will say that a predicate P is "similar" to another predicate Q to mean that P denotes the set of things that are similar to things in the set denoted by Q.

11 I assume type flexible $\vee$, i.e., $p \vee q = 0$ if $p = q = 0$ for any p, q of type t, and $f \vee g = f(x) \vee g(x)$ for any f, g of type $\langle \sigma, \tau \rangle$ and x of type σ.

12 The word *hoặc* means 'or'.

13 As with $\vee$, I assume type flexibility for $\Rightarrow$: $p \Rightarrow q = 0$ if $p = 1$ and $q = 0$ for any p, q of type t, and $f \Rightarrow g = f(x) \Rightarrow g(x)$ for any f, g of type $\langle \sigma, \tau \rangle$ and x of type σ. Note, also, that I use "$\Rightarrow$" to mean *asymmetric* entailment: $P \Rightarrow Q$ means that P is strictly stronger than Q, i.e., that P entails Q but Q does not entail P.

14 Henceforth, I will write "$\rightsquigarrow$" to mean 'licenses the inference that', as in (12), and I will write "$\not\rightsquigarrow$" to mean 'does not license the inference that', as in (14) below.

15 And, also, that she is not certain that John ate a tangerine. I will, henceforth, write "$K_S(p)$" to mean that the speaker S believes that p (Meyer, 2013), i.e., that p is true in all of S's epistemically accessible worlds (Hintikka, 1969), i.e., that S is "certain" that p (A possible world w is epistemically accessible to a person x if x considers w a candidate for the actual world). Consequently, "$\neg K_S(p)$" means that it is not the case that p is true in all of S's epistemically accessible worlds, i.e., that there are worlds compatible with what S believes where p is false, i.e., that S is not certain that p.

16 This is, of course, a simplication. Informativity should take the common ground into account: p is more informative than q if $p \cap c$ is logically stronger than $q \cap c$, where c is the "context set", i.e., the conjunction of all propositions which the discourse participants take for granted, i.e., believe to be true and believe each other to believe to be true (Stalnaker, 1978, 2002). For present purposes, I assume that the common ground does not affect the logical relationship between the propositions involved, i.e., that $p \cap c \Rightarrow q \cap c$ if $p \Rightarrow q$.

17 Note that the phrase *does not believe* here must be understood in the non-neg-raising reading of *believe*. In other words, it means $\neg K_S(p)$, not $K_S(\neg p)$.

18 Note that this makes the difference between P and P+RED somewhat similar to the difference between SOME and ANY. See (Crnič, 2019; Kadmon & Landman, 1993; Krifka, 1995; Linebarger, 1980), among others.

19 Trinh's account is developed based on the theory advanced in Chierchia (1998) regarding the typological divide between "number marking" and "classifier" languages. For a similar description of Chinese, see (Cheng & Sybesma, 1999).

20 If *two books* can refer to two members of (27-b), the sentence *John bought two books* would be true if John bought *a*, *b*, and *c*; obviously, an unwelcome result.

21 An anonymous reviewer raises the question whether *quyển* is not just as 'contextual' as RED. They point out that one can say *một quyển gì đó* in Vietnamese, where the meaning of *gì đó* is roughly 'something', suggesting ignorance of the speaker as to what that something is, hence leaving it to the 'context' to determine exactly what is being referred to. My response has two parts. Firstly, the meaning *gì đó* is not 'contextual' in the relevant sense: *gì đó* means 'something' in all contexts! The fact that the sentence *John is reading something* is true in both a context where he is reading *Anna Karenina* and in a context where he is reading *War and Peace* does not mean the word *something* varies its meaning between these two contexts. Secondly, *một quyển gì đó* does imply that whatever the relevant object is, it is assumed to fall under a predicate that *can* combine with *quyển*. In other words, it is either a book (*sách*), a magazine (*tạp chí*), a journal (*nhật ký*), or even a calendar (*lịch*), but it is not, say, a letter (*thư*). To the extent that these considerations are valid, the reviewer's extremely interesting observation does not conflict with my proposal but, in fact, corroborates it.

22 I thank an anonymous reviewer for drawing my attention to this fact.

23 If *more than three* implies *not more than four* then *more than three* would mean *exactly four*, which is equivalent to *more than three but not more than four*. This is obviously wrong: *John has more than three children* certainly does not convey the message that John has exactly four children.

24 Note that 'washing' here means 'washing oneself', i.e., bathing or showering. I refrain from word-for-word glossing in this case, simply because I have no idea how to gloss the arguably functional items *mà*, *còn*, *gì*, and *nữa*, as they are used in this manner.

25 I thank an anonymous reviewer for raising this issue.

References

Armoskaite, S., & Kutlu, E. (2014). Turkish m-reduplication: A case of similative number. *Turkic Languages*, *18*, 271–288.

Chemla, E. (2008). *Similarity: Towards a unified account of scalar implicatures, free choice permission and presupposition projection*. Manuscript. ENS and MIT.

Cheng, L. L. S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, *30*(4), 509–542. [CrossRef]

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, *6*, 339–405. [CrossRef]

Crnič, L. (2019). Any: Logic, likelihood, and context. *Language and Linguistic Compass*, *13*, 1–20.

Đoàn Thiện Thuật. (2007). *Ngữ âm tiếng Việt*. Nhà xuất bản Đại học Quốc gia.

Emeneau, M. B. (1951). *Studies in Vietnamese (annamese) grammar*. University of California Press.

Fox, D. (2000). *Economy and semantic interpretation*. MIT Press.

Fox, D., & Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy*, *29*, 537–586. [CrossRef]

Grice, P. (Ed.). (1967). Logic and conversation. In *Studies in the way of words* (pp. 41–58). Harvard University Press.

Hintikka, J. (1969). Semantics for propositional attitudes. In J. W. Davis, D. J. Hockney, & W. K. Wilson (Eds.), *Philosophical logic* (pp. 21–45). Reidel.

Hoàng Văn Hành. (2008). *Từ láy trong tiếng Việt*. Nhà xuất bản Khoa học Xã hội.

Kadmon, N., & Landman, F. (1993). Any. *Linguistics and Philosophy*, *16*(4), 353–422. [CrossRef]

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis*, *25*(3–4), 209–257.

Krifka, M., & Modarresi, F. (2023, May 5–7). *Similative plurals in Persian: Alternatives, homogeneity, or strongest meanings?* 41st West Coast Conference on Formal Linguistics (Vol. 41, pp. 232–332), Santa Cruz, CA, USA.

Lewis, D. (1988). Relevant implication. *Theoria*, *54*(3), 161–174. [CrossRef]

Linebarger, M. C. (1980). *The grammar of negative polarity* [Ph.D. thesis, Massachusetts Institute of Technology].

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*(3), 245–297. [CrossRef]

Maspero, H. (1912). Etudes sur la phonétique historique de la langue annamite. Les initiales. *Bulletin de l'École Française d'Extrême-Orient*, *12*, 1–124. [CrossRef]

Meyer, M.-C. (2013). *Ignorance and grammar* [Ph.D. thesis, Massachusetts Institute of Technology].

Ngo, B. (2021). *Vietnamese: An essential grammar*. Routledge.

Nguyễn Đình Hoà. (1997). *Tiếng việt không son phấn*. John Benjamins.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J.-H. Yoon, & A. Kathol (Eds.), *OSU working papers in linguistics* (Vol. 49, pp. 91–136). The Ohio State University.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*(3), 367–391. [CrossRef]

Smith, R. W. (2020). Similative plurality and the nature of alternatives. *Semantics and Pragmatics*, *13*(15), 1–44. [CrossRef]

Stalnaker, R. (1978). Assertion. *Syntax and Semantics*, *9*, 315–332.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*, 701–721. [CrossRef]

Thompson, L. C. (1987). *A vietnamese reference grammar*. University of Hawai'i Press.

Trinh, T. (2011). Nominal reference in two classifier languages. *Proceedings of Sinn und Bedeutung*, *15*, 629–644.

Trinh, T. (2014). Das Vietnamesische und das Chinesische. In M. Krifka, A. Błaszczak, Leßmölmann, A. Meinunger, B. Stiebels, R. Tracy, & H. Truckenbrodt (Eds.), *Das mehrsprachige klassenzimmer* (pp. 247–270). Springer.

Vu, S. (1998). A unified analysis of some Vietnamese reduplication forms. *Proceedings of SEALS*, *8*, 165–191.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

Article

"Again" and "Again": A Grammatical Analysis of *lại* and *nữa* in Vietnamese

Yi-Ling Irene Liao * and Tzong-Hong Jonah Lin

Graduate Institute of Linguistics, National Tsing Hua University, Hsinchu 300044, Taiwan; jonahlin@mx.nthu.edu.tw
* Correspondence: irene0202@gapp.nthu.edu.tw

Abstract: This work examines the grammatical properties of *lại* and *nữa* in Vietnamese, both of which can express the repetition of an event. It has been observed that different syntactic positions of *lại* result in different readings, as noted in previous studies. When *lại* precedes a verb, it may assume either the repetitive reading or restitutive reading. When *lại* follows a verb, it can only assume the restitutive reading. *Nữa* can be used for the repetitive reading and the incremental reading as well, in the sense that an activity is incremented by adding subevents measured along some dimension, as discussed by Tovena & Donazzan (2008). We adopt Stechow's (1996) structural analysis and the theory of focus semantics and propose that the preverbal *lại* is adjoined to vP, which can be focus-associated with an element within its c-command domain, i.e., vP or VP. This is the origin of the ambiguous readings of the preverbal *lại*. The postverbal *lại* is adjoined to VP, and this is the reason why it does not yield ambiguous readings. We also propose that *nữa* is adjoined to vP, along with the movement of vP to a higher functional projection. This results in the surface final position of *nữa*.

Keywords: repetition; repetitive; restitutive; again; Vietnamese

1. Introduction

This work investigates the adverbs of repetition in Vietnamese, *lại* and *nữa*, and proposes syntactic accounts for them. The adverb *lại* can be used alone or with the particle *nữa* to express the meaning of "again" (Nguyễn, 1997), as shown in (1).

(1)	Trời	lại	mưa	(nữa)	rồi.[1]
	sky	again	rain	more	PERF

'It is raining again.' (Nguyễn, 1997, p. 147)

The word *lại* in Vietnamese can be used as a verb meaning "to come", an adverb, a modal particle conveying the speaker's attitude, or a sub-element in a sentence connector (Thompson, 1987; Nguyễn, 1997; Trần, 2023, and others). See (2a)–(2d).

(2) a. Em hãy **lại** đây với anh!
you IMP come here with me
'Come here to me, please!' (Trần, 2023, p. 252)

b. Tôi **lại** yêu anh ấy như ngày nào.
I again love him as day which
'I fell in love with him again like before.' (Trần, 2023, p. 255)

c. Vì sao khi con kéo đàn, bà
why when I play violin grandma
lại khóc vậy mẹ?
LAI cry such mother
'Why does grandma cry when I play the violin, mom?' (Trần, 2023, p. 261)

d. **Ngược lại** / **Trái lại**, anh ta rất chăm chỉ.
however however he very hard-working
'However, he is very hard-working.' (Trần, 2023, p. 263)

In this work, we focus on the repetitive use of *lại*. It has been pointed out that different syntactic positions of *lại* result in different readings (Thompson, 1987; Nguyễn, 1997; Phan, 2013, etc.); see (3a,b). When the adverb *lại* precedes a verb, it has a repetitive reading. When the adverb *lại* follows a verb, it can only yield a restitutive reading. According to Phan (2013), in (3a), the entire event of the subject writing a letter is repeated, while, in (3b), only the result state of the event (i.e., the letter having been written) re-occurs.[2,3]

(3) a. Ông **lại** viết thư.
he again write letter
'He wrote another letter.' (Phan, 2013, p. 98)

b. Ông viết **lại** thư.
he write again letter
'He revised the letter.' (Phan, 2013, p. 98)

Stechow (1996) postulates a structural analysis for the ambiguous readings of the adverb *wieder*, "again", in German and argues that the ambiguity of *wieder* arises from different modifying scopes. Beck and Johnson (2004) further apply this analysis to the ambiguity of the adverb *again* in English (see also Beck & Snyder, 2001; Beck, 2006). Inspired by these analyses, we propose that the preverbal *lại* in Vietnamese adjoins to vP. The modifying scope of the preverbal *lại* can be the entire event or a result state of the event. This gives rise to the two readings, namely the repetitive and restitutive readings. In addition, we argue that the postverbal *lại* adjoins to VP, and it can only modify the result state of the predicate vP. Therefore, it only yields the restitutive reading. We also show that syntactic tests support the proposed analyses of the preverbal *lại* and the postverbal *lại*.

The word *nữa*, which occurs after the predicate, expresses the meaning "more, in addition, also", as shown in (4). It can also denote the repetition of an event, as in (5).

(4) Ông dùng cơm **nữa** thôi?
you have rice more SFP
'Are you going to eat more rice?' (Thompson, 1987, p. 271)

(5) Hôm nay trời mưa **nữa** rồi.
today sky rain more PERF
'It is raining again today.'

We propose that *nữa* adjoins to vP, and, furthermore, it triggers the merger of a FocusP on the phrase structure, to which vP moves. This results in the predicate-final position of *nữa*.

This paper is structured as follows. Section 2 presents a review (Stechow, 1996). Section 3 examines the preverbal *lại* and the postverbal *lại*. Section 4 discusses *nữa* and its syntactic position. Section 5 outlines our proposal. Section 6 contains the summary.

2. The Structural Analysis of the Repetitive/Restitutive Ambiguity

Stechow (1995) points out that *wieder*, "again", in German, exhibits semantic ambiguity, as shown in (6). The adverb *wieder*, "again", may assume the repetitive reading (the reading in (6a)) or the restitutive reading (the reading in (6b)).[4] In (6a), the whole event "Randi had caught Bockhirsch" is repeated. In (6b), the result state of the event, "Bockhirsch is in the state of being a prisoner," is repeated.[5]

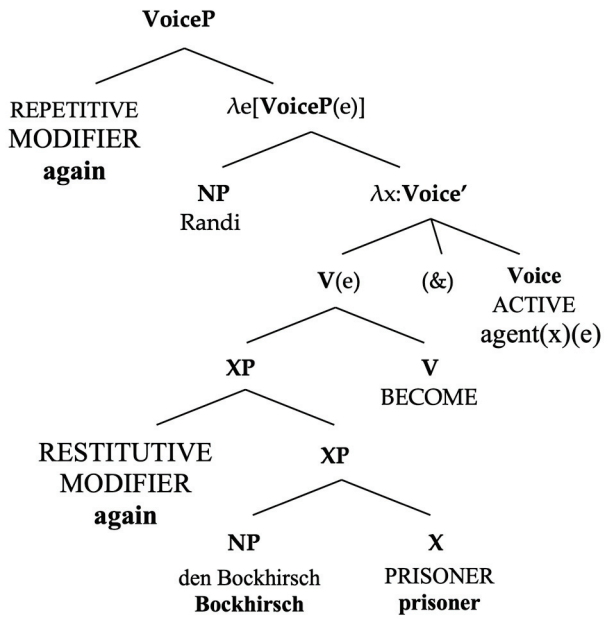
(6) Randi den Bockhirsch **wieder** fing. (Stechow, 1996, pp. 94–95)
Randi Bockhirsch again caught
a. 'Randi had caught Bockhirsch before, and she caught him again.'
b. 'Bockhirsch had been a prisoner before, and Randi caused him to be a prisoner again.'

This repetitive/restitutive ambiguity is also observed in English. See (7). When the whole agentive event "Sally opened the door" is repeated, a repetitive reading is obtained. When the state of the door's being open is repeated, the restitutive reading results instead. In other words, if the repetition includes the agent of the action, the repetitive reading is yielded.

(7) Sally opened the door again. (Beck & Johnson, 2004, p. 106)
a. 'Sally opened the door, and that had happened before.' (repetitive)
b. 'Sally opened the door, and the door had been open before.' (restitutive)

Adopting the theory of Voice projection by Kratzer (1994), Stechow (1995, 1996) postulates a structural analysis for the ambiguity of *wieder*, "again", as shown in (8) (from Stechow, 1996, p. 97).

(8)



There are two possible positions for *wieder* in the D-structure of the sentence in (6). When *wieder* is generated in a higher position, it scopes over the VoiceP, resulting in a repetitive reading. When it is in a lower position, it only scopes over the main predicate (XP here), resulting in a restitutive reading.

Words corresponding to *wieder* in other languages are also claimed to exhibit such semantic ambiguity. Some authors (Beck & Johnson, 2004; Beck, 2005) apply Stechow's (1996) theory to *again* in English for its ambiguous readings in the double-object construction.

3. The Syntactic Position of the Adverb *lại*

In this section, we discuss the syntactic structures of the preverbal *lại* and the postverbal *lại* and how their syntactic positions are correlated with their potential readings.

3.1. The Preverbal lại

To begin with, we examine the syntactic position of the preverbal *lại*. The strategy that we use consists of comparing the syntactic occurrence of *lại* with different types of elements of different syntactic positions in the same sentence, thereby determining the structural position of the element *lại*.

First, the evaluative adverb *quả nhiên*, "indeed", and the epistemic adverb *có lẽ*, "possibly", must precede the preverbal *lại*.[6] Evaluative and epistemic adverbs are usually assumed to be CP-level adverbials (see Cinque, 1999; Giorgi, 2010).

(9)	a.	Ông	quả nhiên	**lại**	viết	thư.
		he	indeed	again	write	letter
		'He indeed wrote another letter again.'				
	b.	*Ông	**lại**	quả nhiên	viết	thư.
		he	again	indeed	write	letter
(10)	a.	Ông	có lẽ	**lại**	viết	thư.
		he	possibly	again	write	letter
		'He possibly wrote another letter.'				
	b.	*Ông	**lại**	có lẽ	viết	thư.
		he	again	possibly	write	letter

Second, the preverbal *lại* must precede the subject-oriented adverb *vô tình*, "unintentionally", and the vP-level adverb *bình tĩnh*, "calmly", as shown in (11) and (12). Presumably, these adverbials are adjoined to vP.

(11)	a.	Nam	**lại**	vô tình	làm vỡ	chiếc	bình.
		Nam	again	unintentionally	break	CL	vase
		'Nam unintentionally broke the vase again.'					
	b.	*Nam	vô tình	**lại**	làm vỡ	chiếc	bình.
		Nam	unintentionally	again	break	CL	vase
(12)	a.	Nam	**lại**	bình tĩnh	trả lời	câu hỏi.	
		Nam	again	calmly	answer	question	
		'Nam calmly answered the question again.'					
	b.	*Nam	bình tĩnh	**lại**	trả lời	câu hỏi.	
		Nam	calmly	again	answer	question	

Third, the preverbal *lại* must precede the vP-level aspectual adverb *liên tục*, "continually" (Cinque, 1999), but it follows the generic-modal adverb *thông thường*, "normally", which presumably is a TP-level adverb.

(13) a. Gần đây Nam **lại** liên tục mắng Hoa.
recently Nam again continually scold Hoa
'Nam has been scolding Hoa continuously recently.'

b. *Gần đây Nam liên tục **lại** mắng Hoa.
recently Nam continually again scold Hoa

(14) a. Nam thông thường **lại** đến phòng tập
Nam normally again come room do
thể dục vào buổi chiều.
exercise in afternoon
'He normally goes to the gym again in the afternoon.'

b. *Nam **lại** thông thường đến phòng tập
Nam again normally come room do
thể dục vào buổi chiều.
exercise in afternoon

Fourth, the preverbal *lại* must precede the passive maker *bị*, which, according to (Bruening & Tran, 2015), is the head of a predicate, equivalent to v in most current syntactic analyses.

(15) a. Nam **lại** bị Hoa đánh.
Nam again PASS Hoa hit
'Nam was beaten by Hoa again.'

b. *Nam bị **lại** Hoa đánh.
Nam PASS again Hoa hit

Fifth, the preverbal *lại* can take a dynamic or a stative predicate, but it requires the perfect aspect marker *rồi* when it occurs with a stative predicate.[7] This phenomenon is similar to the adverb *you*, "again", in Mandarin. Lin and Liu (2009) observe that the adverb *you*, "again", in Mandarin requires the presence of the perfect aspect marker *le* when the main verb is stative. It appears that the same requirement is effective for the Vietnamese adverb *lại* as well. Compare the examples in (16) and (17).

(16) a. Nam **lại** hôn Hoa. [dynamic]
Nam again kiss Hoa
'Nam kissed Hoa again.'

b. Nam **lại** đánh Hoa. [dynamic]
Nam again hit Hoa
'Nam hit Hoa again.'

(17) a. Nam **lại** thích Hoa *(rồi). [stative]
Nam again like Hoa PERF
'Nam likes Hoa again.'

b. Nam **lại** béo *(rồi). [stative]
Nam again fat PERF
'Nam gets fat again.'

To summarize, the hierarchical distribution of the adverbials discussed so far is shown in (18). Based on this observation, we propose that *lại* adjoins to vP as its outer-most layer.

(18) { *quả nhiên* 'indeed' (CP) / *có lẽ* 'probably' (TP) / *thông thường* 'normally' (TP) } > ***lại*** > { *vô tình* 'unintentionally' (vP) / *bị* (v) / *liên tục* 'continually' (vP) }

3.2. The Postverbal lại

In this subsection, we turn to the grammatical properties of the postverbal *lại*. Remember that the postverbal *lại* only yields a restitutive reading. First, the postverbal *lại* is only compatible with dynamic predicates, as shown in (19). Ungrammaticality results when it co-occurs with a stative predicate, regardless of whether the perfect marker *rồi* occurs or not, as shown in (20).

(19) a. Nam sơn **lại** nhà. [dynamic]
Nam paint again house
'Nam repainted the house.'

b. Nam làm **lại** một cái bánh. [dynamic]
Nam do again one CL cake
'Nam made another cake.'

(20) a. *Nam thích **lại** Hoa (rồi). [stative]
Nam like again Hoa PERF
Intended reading: 'Nam likes Hoa again.'

b. *Nam quên **lại** chuyện này (rồi). [stative]
Nam forget again story this PERF
Intended reading: 'Nam forgot this story again.'

Second, the postverbal *lại* cannot occur after the object, as in (21b). It must precede the object, as in (21a).

(21) a. Ông viết **lại** thư.
he write again letter
'He revised the letter.'

b. *Ông viết thư **lại**.
he write letter again

Third, the postverbal *lại* must precede the postverbal manner adverb *từ từ*, "slowly". The reverse order is ungrammatical.

(22) a. Làm **lại** từ từ!
do again slowly
'Do it slowly again!'

b. ***Làm** từ từ lại!
do slowly again

Fourth, the postverbal *lại* must precede the completive particle *xông*, "finish". Phan (2013) argues that *xông* takes the "base predicate" (roughly equivalent to VP) as the complement.

(23) a. Nam vừa sơn **lại** xông nhà.
Nam just paint again finish house
'Nam has just finished the repainting of the house.'

b. *Nam vừa sơn xông **lại** nhà.
Nam just paint finish again house

Fifth, the postverbal *lại* must precede the dynamic modal *được*, "be able to", as shown in (24). Ungrammaticality results if *lại* follows the postverbal modal *được*.

(24) a. Nam viết **lại** được lá thư.
Nam write again be able to CL letter
'Nam was able to revise the letter.'

b. *Nam viết được **lại** lá thư
Nam write be able to again CL letter

Sixth, the postverbal *lại* can co-occur with the preverbal *lại*, as shown in (25).

(25) Ông **lại** viết **lại** thư.
he again write again letter
'He revised the letter again.'

To summarize, the hierarchical distribution of the postverbal *lại* is shown in (26). The postverbal *lại* must precede postverbal manner adverbs, the completive marker, the dynamic modal, and the object in a linear order.

(26) Lexical V > Postverbal *lại* > { *từ từ* 'slowly' / *xông* 'finish' / *được* 'be able to' / Object }

4. The Adverb *nữa* and Its Syntactic Position

In this section, we discuss the grammatical properties of the adverb *nữa*, "more, in addition, also" in Vietnamese. Before delving into the discussion of *nữa*, let us introduce the concept of "incremental reading" first. Tovena and Donazzan (2008, p. 91) point out that the repetitive adverb *ancora* in Italian can give rise to an incremental reading, which means that an activity is incremented by adding subevents measured along a specific dimension. See (27).

(27) a. Maria sta **ancora** leggendo.
'Maria is *still* reading.'
b. Maria sta leggendo **ancora** un libro.
'Maria is reading one *more* book.'

The word *ancora* in (27a) denotes a meaning comparable to "still," and the one in (27b) denotes a meaning similar to "more." The incremental interpretation can be thought of either as a repetition of events or as a continuation of an activity by adding more object units. The adverb *nữa* in Vietnamese is quite similar to *ancora*, as it yields an incremental reading, shown in (28) and (29).

(28) Hôm nay trời mưa **nữa** rồi.
today sky rain more PERF
'It is raining again today.'

(29) Nam có thể ăn một bát cơm **nữa**.
Nam can eat one bowl rice more
'Nam can eat one more bowl of rice.'

Now, we turn to the syntactic position of *nữa*. First, similar to the preverbal *lại*, the adverb *nữa* can occur with a stative or dynamic predicate; see (30) and (31). When it occurs with a stative predicate, the perfect aspect particle *rồi* is required, as shown in (31).

(30) a. Nam hôn Hoa **nữa**. [dynamic]
Nam kiss Hoa more
'Nam kissed Hoa again.'

b. Nam đánh Hoa **nữa**. [dynamic]
Nam hit Hoa more
'Nam hit Hoa again.'

(31) a. Nam thích Hoa **nữa** *(rồi). [stative]
Nam like Hoa more PERF
'Nam likes Hoa again.'

b. Nam lại béo **nữa** *(rồi). [stative]
Nam again fat more PERF
'He gets fat again.'

Second, *nữa* must follow postverbal manner adverbs, such as, for instance, *từ từ*, "slowly", in (32). The reverse order is ungrammatical.

(32) a. Nam ăn từ từ **nữa**.
Nam eat slowly more
'Nam ate slowly too.'

b. *Nam **nữa** ăn từ từ.
Nam eat more slowly

Third, *nữa* must follow the completive marker *xông*, "finish", as shown in (33).

(33) a. Nam (còn) ăn bánh mì xông **nữa**.
Nam even eat bread finish more
'Nam even finished the bread.'

b. *Nam (còn) ăn bánh mì **nữa** xông.
Nam even eat bread more finish

Fourth, *nữa* must precede the perfect aspect marker *rồi.* See (34). When the perfect aspect marker precedes *nữa*, the sentence is ungrammatical.

(34) a. Hôm nay trời mưa **nữa** rồi.
today sky rain more PERF
'It is raining again today.'

b. *Hôm nay trời mưa rồi **nữa**.
today sky rain PERF more

Fifth, *nữa* cannot scope over the epistemic modal adverb *chắc chắn*, "surely". For the sentence in (35), only reading 1 is possible, where *nữa* is within the scope of *chắc chắn*, "surely". In reading 2, *nữa* is intended to scope over the epistemic modal adverb *chắc chắn*. This reading is not available.

(35) Nam chắc chắn ăn cơm **nữa**.
Nam surely eat rice more
1. 'Nam surely ate rice again.'
2. *'Again, Nam surely ate rice.'

Sixth, *nữa* cannot scope over the negator *không*, "not", as shown in (36). In this sentence, the negator scopes over the particle *nữa* in readings 1 and 2, and both readings are acceptable. In reading 3, the negator is intended to fall within the scope of *nữa*. However, this reading is unacceptable.

(36) Nam sẽ không đi Mỹ **nữa**.
Nam will NEG go America more
1. 'Nam doesn't have any intention to go to the US now.' ("will" > NEG > *nữa*)
2. 'Nam will not go to the US anymore.' (NEG > "will" > *nữa*)
3. *'Nam again doesn't have any intention to go to the US.'
(**nữa* >NEG/"will")

Seventh, *nữa* can scope over the quantificational subject of a sentence, as shown in (37). When the quantificational subject *không ai*, "nobody", in (37) scopes over the particle *nữa*, reading 1 is obtained. When the particle *nữa* scopes over the quantificational subject, reading 2 is obtained. Both readings are acceptable.

(37) Không ai đến **nữa**.
NEG who come more
1. 'Nobody came again.' (NP > *nữa*)
2. 'Again, nobody came.' (*nữa* > NP)

Eighth, *nữa* can co-occur with the preverbal *lai* and the postverbal *lai*. When these three adverbial elements appear in the same sentence, as in (38), the sentence is fully grammatical.

(38) Nam **lại** sơn **lại** nhà **nữa**.
Nam again paint again house more
'Nam repainted the house again. (He also repainted other stuff)'

In summary, the following two sets of properties are observed with *nữa*. See (39) and (40). In linear order, the particle *nữa* must follow the postverbal manner adverb *từ từ*, "slowly", and the completive marker *xông*, "finish", and precede the perfect aspect marker *rồi*. Regarding the scope property of *nữa*, it must fall within the scope of the epistemic modal adverb *chắc chắn*, "surely", and the negator *không*, "not". In addition, *nữa* may scope over the quantificational subject of a sentence.

(39) Linear order

Verb > { *từ từ* 'slowly' / *xông* 'finish' } > ***nữa*** > *rồi*

(40) Scope
a. *chắc chắn*, "surely" > *nữa*
b. *không*, "not" > *nữa*
c. Quantificational subject > *nữa*
nữa > Quantificational subject

5. The Proposal

Inspired by the structural analysis of Stechow (1996), we propose an analysis for the preverbal *lại*, the postverbal *lại*, and *nữa* that is partially structural and partially focus-semantic (see Beck, 2006; Ippolito, 2007; and Csirmaz & Slade, 2020, for more focus-based accounts of repetitive adverbs). We agree with Stechow's proposal that the different readings of a repetitive adverb result from its syntactic position, rather than lexical ambiguity (see Dowty, 1979; and Fabricius-Hansen, 1983, among others). However, the approach we adopt is more flexible, since syntactic adjacency does not completely determine the reading of the adverb in question (for details, see below). We argue that the repetitive adverb can target an element within its c-commanding domain and not only the constituent to which it is directly adjoined.

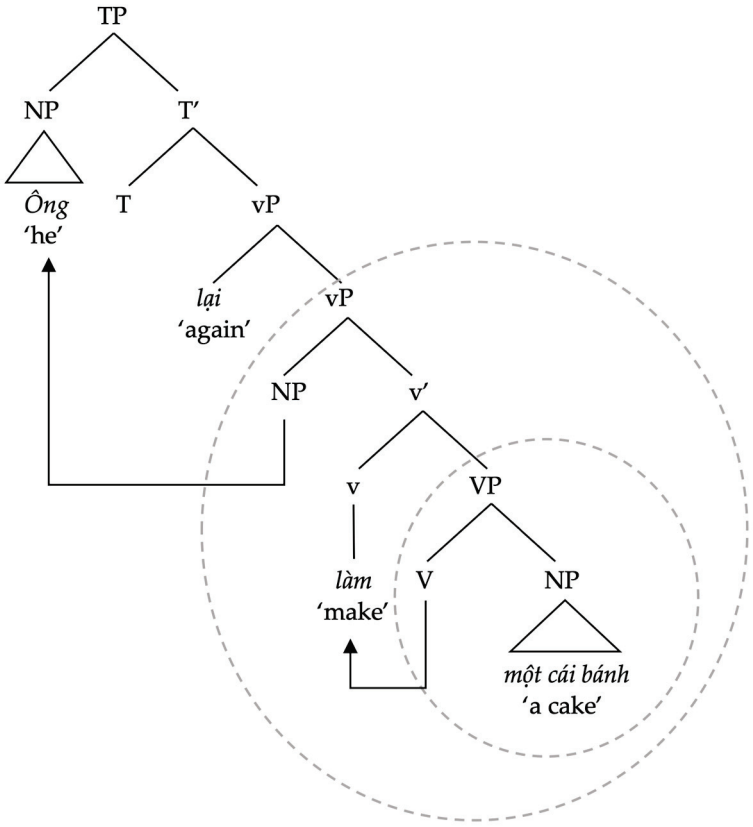
First, based on our observation of the preverbal *lại* (summarized in (18)), we propose that the preverbal *lại* adjoins to vP. Semantically, as a focus particle, it can be associated with the entire event, namely vP, or only with the result state of the event, namely VP. This results in two possible readings, i.e., the repetitive reading and the restitutive reading. See the example in (41) and its syntactic structure in (42). Note that, in Stechow's theory, *wieder* (and also the English adverb *again*; see Beck & Johnson, 2004) can only modify the syntactic domain to which it is directly adjoined. However, in the case of Vietnamese (and Mandarin, too; see Lin & Liu, 2009), non-adjacent focalization is possible. In other words, when *lại* is adjoined to vP, focalization of the complement of vP, namely VP, is possible in Vietnamese (and Mandarin). This state of affairs is actually a normal case rather than an exception. An example is the focus adverb *only* in English. In the English sentence "*John only bought books*", *only* may focalize the verb *bought* or the object NP *books*. For example, we can have the following two contrasts in mind: "*John only bought books and did not buy other things*", in which case *books* is focus-marked; or alternatively, "*John only bought books and did not borrow them*", in which case the verb *bought* is focus-marked. Having the English adverb *only* as a paradigm example, we claim that the preverbal *lại* in Vietnamese is such a focus particle. It can focalize a constituent that is within its c-command domain, immediately adjacent to it or otherwise. We assume that the focalization function of the preverbal *lại* is carried out by the probe–goal relation of current syntactic theory, which does not require adjacency.

(41)

Ông	**lại**	làm	một	cái	bánh.
he	again	make	one	CL	cake

'He made a cake again.' (repetitive or restitutive)

(42) Preverbal *lại* (circled areas = possible focus targets)



The proposed analysis explains the position of *lại* when it co-occurs with other elements, as shown in (16) and repeated in (43).

(43)
$$\left\{\begin{array}{l}\textit{quả nhiên}\text{ 'indeed' (CP)}\\ \textit{có lẽ}\text{ 'probably' (TP)}\\ \textit{thông thường}\text{ 'normally' (TP)}\end{array}\right\} > \textbf{\textit{lại}} > \left\{\begin{array}{l}\textit{vô tình}\text{ 'unintentionally' (vP)}\\ \textit{bị}\text{ (v)}\\ \textit{liên tục}\text{ 'continually' (vP)}\end{array}\right\}$$

The analysis in (42) can account for the structural properties of the preverbal *lại* shown in (43). If the preverbal *lại* adjoins to vP, it will necessarily be lower than CP-level adverbs and TP-level adverbs, such as *quả nhiên*, "indeed", *có lẽ*, "probably", and *thông thường*, "normally". In addition, since it is on vP and thus precedes the head v, it naturally precedes the passive maker *bị*, which is assumed to be the head of a predicate (equivalent to *v*) (see Bruening & Tran, 2015). And, since we assume that the preverbal *lại* adjoins to the outer-most layer of vP, it is higher than and, hence, precedes the vP-level adverbs *vô tình*, "unintentionally" (vP), and *liên tục*, "continually" (vP).

We also noted that the preverbal *lại* may occur with dynamic and stative predicates, but the perfect aspect marker *rồi* is required when it occurs with a stative predicate, as in (16) and (17), repeated below.

(44) a. Nam **lại** hôn Hoa. [dynamic]
Nam again kiss Hoa
'Nam kissed Hoa again.'

b. Nam **lại** đánh Hoa. [dynamic]
Nam again hit Hoa
'Nam hit Hoa again.'

(45) a. Nam **lại** thích Hoa *(rồi). [stative]
Nam again like Hoa PERF
'Nam likes Hoa again.'

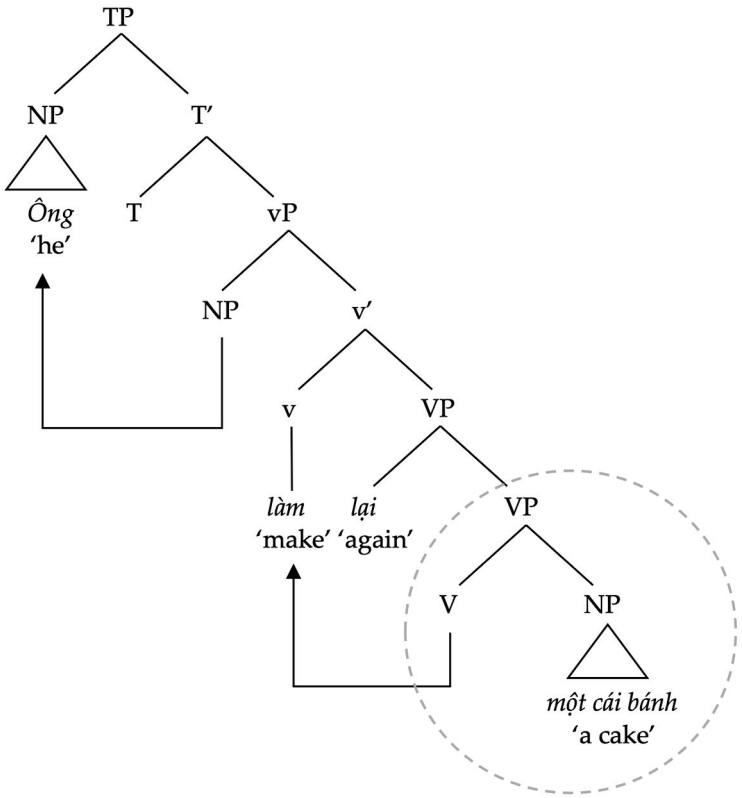
b. Nam **lại** béo *(rồi). [stative]
Nam again fat PERF
'Nam gets fat again.'

This phenomenon is similar to the adverb *you*, "again", in Mandarin. When *you*, "again", occurs with a stative predicate, it also needs the presence of the perfect aspect marker *le*. Lin and Liu (2009, p. 1188) argue that, when *you*, "again", adjoins to a static predicate, it turns the predicate into a dynamic one, namely a change-of-state predicate. We propose that the preverbal *lại* exhibits the same function. When it occurs with a static predicate, it turns the predicate dynamic. The presence of the perfect aspect marker *rồi* is therefore required to indicate that the predicate is now a dynamic one.[8]

As to the postverbal *lại*, we propose that it adjoins to the lexical VP as its outer-most layer. See (46) and (47) as a demonstration. In sentence (46), the verb *làm*, "make", raises to the head of vP, resulting in the postverbal position of *lại*. Since the postverbal *lại* only c-commands the VP, its focus domain is limited to the result state of the predicate. Consequently, only the restitutive reading is available for the postverbal *lai*. This analysis is also compatible with Phan's (2013) proposal, which posits a projection lower than the vP which denotes the result state of the asserted event, ResultativeP. For simplicity, we use the term VP here.

(46) Ông làm **lại** một cái bánh.
he make again one CL cake
'He made a cake again.' (only restitutive)

(47) Postverbal *lại* (circled area: possible focus target)



This analysis accounts for the structural properties of the postverbal *lại* shown in (24), repeated as in (48).

(48)

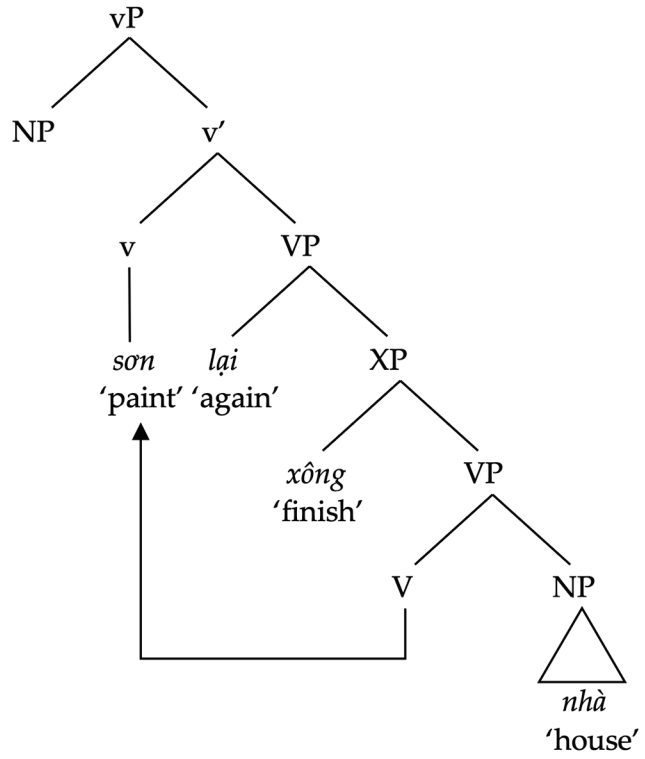
Lexical V > Postverbal *lại* > { *từ từ* 'slowly' / *xông* 'finish' / *được* 'be able to' / Object }

First, if the postverbal *lại* adjoins to VP, it should precede the object. This accounts for the fact that the postverbal *lại* must precede the object in a sentence. Second, if the postverbal *lại* adjoins to VP as its outer-most layer, other VP-level elements should be lower than it. This explains the fact that the postverbal *lại* precedes the postverbal VP-level manner adverb *từ từ*, "slowly." It also explains the fact that *lại* precedes the completive particle *xông*, "finish", and the dynamic modal *được*, "be able to." Phan (2013) proposes that the particle *xông* and the modal *được* are heads of the projection CompletiveP and ResultativeP, respectively, both of which occur between a higher VP (roughly equivalent to vP) and a lower VP of a sentence. In our framework, the postverbal *lại* is the outer-most layer of the (generalized) VP; thus, it must be higher than these two elements and precede them. We take *xông* as an illustration.[9]

(49)	Nam	vừa	[sơn	lại	xông	nhà].
	Nam	just	paint	again	finish	house

'Nam just finished the house painting.'

(50) Co-occurrence of the postverbal *lại* with the other postverbal element



In Section 3.2, we observed that the postverbal *lại* can only be used with a dynamic predicate; it cannot be used with a stative predicate, not even with the perfect aspect marker *rồi*. See (19) and (20) above. This can be explained by the proposal that the postverbal *lại* adjoins to the projection denoting the result state of a verbal predicate, namely VP. Only a dynamic verbal predicate has a result state. Furthermore, we may assume that the result state that *lại* modifies must be the result of an agentive action. Since *rồi* only introduces a change-of-state meaning to a stative predicate and no agency is brought in, the addition of *rồi* cannot save the sentence from ungrammaticality. A piece of evidence for this proposal is that sentences (17a,b), where the presence of *rồi* makes a stative sentence with the preverbal *lại* grammatical, can only have the repetitive reading. They do not have the restitutive reading. The lack of an acceptable restitutive reading for (17a,b) clearly originates from the fact that stative predicates do not yield a result state, even when they become dynamic by the function of *rồi*.

Lastly, we turn to the particle *nữa*, which occurs in the predicate-final position. The particle *nữa*, in its lexical–semantic nature, is an additive particle that yields an incremental reading. The structural properties of *nữa* are shown in (39) and (40), now repeated in (51) and (52).

(51) Linear order

Verb > { *từ từ* 'slowly' / *xông* 'finish' } > ***nữa*** > *rồi*

(52) Scope

a. *chắc chắn*, "surely" > *nữa*

b. *không*, "not" > *nữa*

c. Quantificational subject > *nữa*
nữa > Quantificational subject

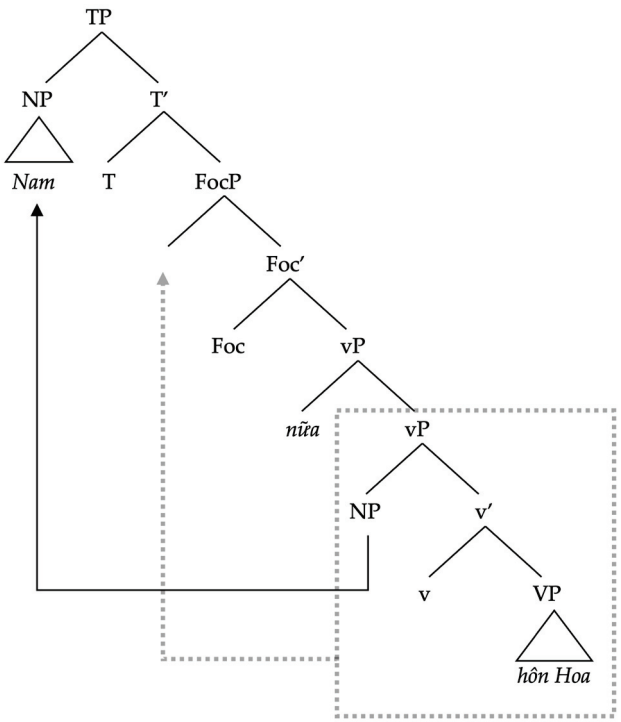
We propose that *nữa* adjoins to vP. We further propose that *nữa* triggers movement of the vP to a higher functional projection, Spec, FocP, resulting in its being stranded behind the predicate.[10] Following the analysis of light predicate raising of Simpson (2001), we assume that the motivation for such movement is to defocus the constituent in question. Examples of such movement include the sentence-final particle *kong*, "to speak", in Taiwanese (Simpson & Wu, 2002) and the sentence-final deontic modal in Cantonese and a number of Southeast Asian languages (Simpson, 2001). Let us use the sentence in (30a) as a demonstration, repeated as (53), with (54) as its structural analysis.

(53)

Nam	hôn	Hoa	nữa.
Nam	kiss	Hoa	more

'Nam kissed Hoa again.'

(54)



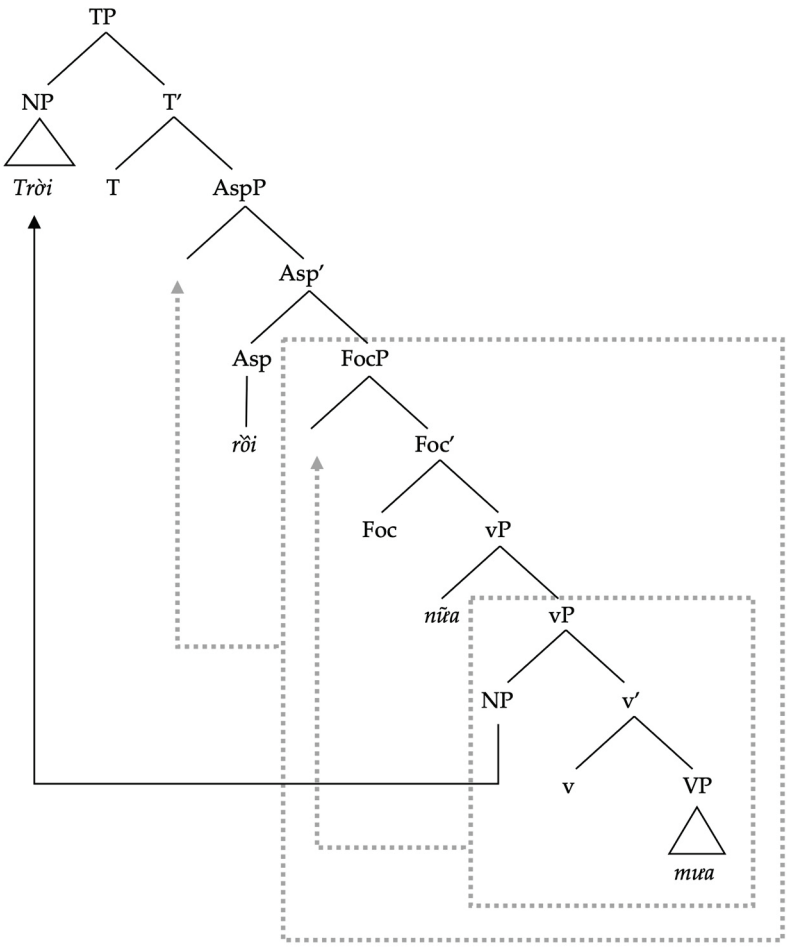
This analysis can account for the structural properties of *nữa* shown in (51) and (52). First, if *nữa* adjoins to vP, VP-level or VP-internal elements such as *từ từ*, "slowly", and *xông*, "finish", should precede it because the whole VP moves to Spec, FocP, and becomes higher than vP. Second, *nữa* precedes *rồi* because *rồi* is the head of the projection AspP, which we assume triggers the raising of its complement (FocP in this case) to its specifier. See (55) and (56) for illustration.

(55)

Trời	mưa	nữa	rồi.
sky	rain	more	PERF

'It is raining again.'

(56)



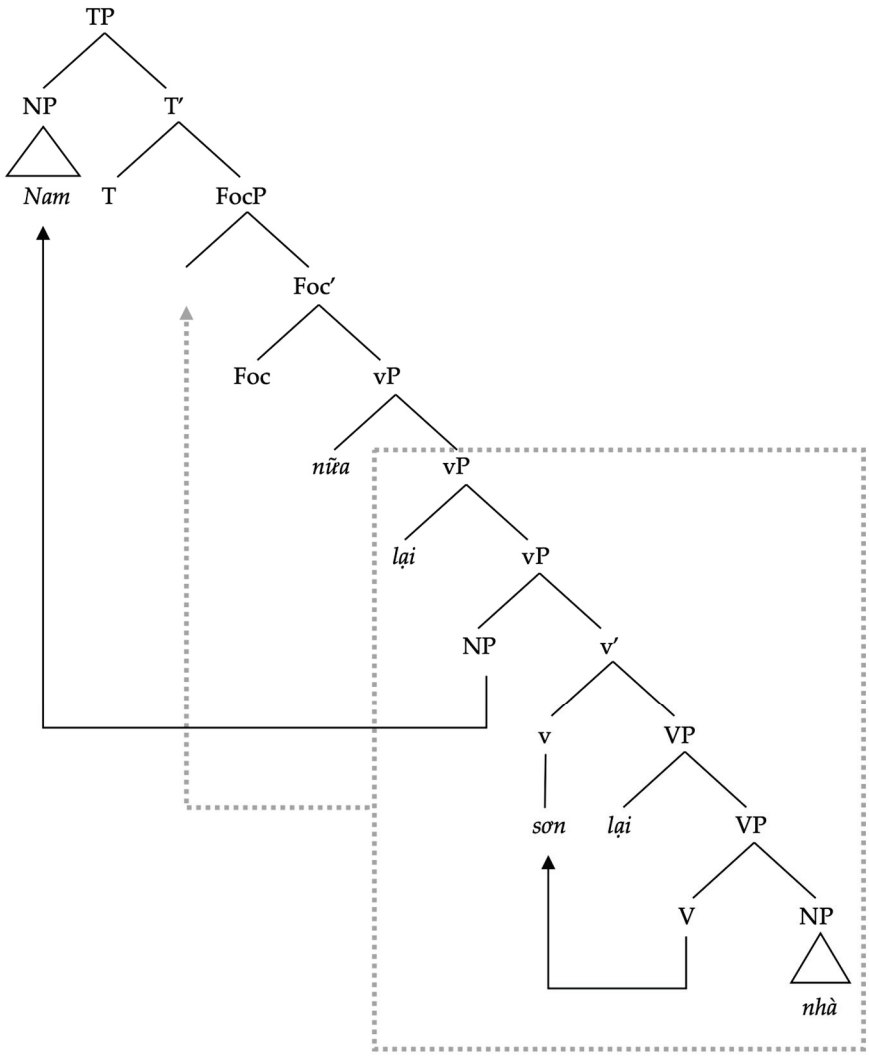
Third, if the particle *nữa* adjoins to vP, it cannot scope over TP-level elements such as the epistemic modal adverb *chắc chắn*, "surely", and the negator *không*, "not". In addition, as *nữa* adjoins to vP, it should fall within the scope of a subject quantifier. So, the scope relation "quantificational subject > *nữa*" in (52c) is obtained. On the other hand, a quantificational subject may undergo quantifier lowering (May, 1985) and assumes its scope position in Spec, vP. In that position, it falls within the scope of *nữa* since *nữa* adjoins to vP. In this way, the scope relation "*nữa* > quantificational subject" is obtained.

Fourth, our proposal can also account for the co-occurrence of *nữa* with the preverbal *lại* and the postverbal *lại*, as shown in (57) and (58). Sentence (38) is repeated in (57).

(57)	Nam	lại	sơn	lại	nhà	nữa.
	Nam	again	paint	again	house	more

'Nam repainted the house again. (He also repainted other stuff)'

(58)



6. Conclusions

Inspired by previous works (Stechow, 1996; Beck & Johnson, 2004), we propose that the ambiguity of *lại* in Vietnamese results from its adjunction to vP and VP. Additionally, the adverb *nữa* in Vietnamese adjoins to vP and yields an incremental reading. It also triggers vP movement to Spec, FocP, resulting in its predicate-final position.

Author Contributions: Conceptualization, Y.-L.I.L. and T.-H.J.L.; Methodology, Y.-L.I.L. and T.-H.J.L.; Software, Y.-L.I.L. and T.-H.J.L.; Validation, Y.-L.I.L. and T.-H.J.L.; Formal analysis, T.-H.J.L.; Investigation, Y.-L.I.L.; Resources, Y.-L.I.L. and T.-H.J.L.; Data curation, Y.-L.I.L.; Writing—original draft preparation, Y.-L.I.L.; Writing—review & editing, Y.-L.I.L. and T.-H.J.L.; Supervision, T.-H.J.L.; Project administration, Y.-L.I.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the authors.

Acknowledgments: We would like to thank our informants, Trần Phan and Viet Cao, for their assistance with grammatical judgments on Vietnamese data. We also thank the audience at ISVL-

4 (National Yang Ming Chiao Tung University, Hsinchu). We are grateful to the two anonymous reviewers for their valuable advice and comments.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 The abbreviations used in the glosses are as follows: PERF = perfect aspect marker; IMP = imperative particle; CL = classifier; PASS = passive marker; and NEG = negator.

2 A reviewer questions that the definiteness of the objects in (3a,b) is not consistent: one is indefinite, and the other is definite. The examples are from (Phan, 2013). However, a bare noun in Vietnamese can be definite or indefinite (Phan & Chierchia, 2022). Thus, we can replace the objects in (3a,b) with a numeral to make them both indefinite. The semantics remain the same.

(i) a. Ông lại viết một lá thư.
he again write one CL letter
'He wrote a letter again.'

b. Ông viết lại một lá thư.
he write again one CL letter
'He wrote a letter again.'

3 For resultative verbs such as "catch" or "open", the restitutive reading does not need to involve the same agent (or it does not need an agent at all). But, for verbs like "write", it is difficult to imagine a restitutive reading such that different agents are involved in the "written-state" of a same letter. This is why the restitutive reading of (3b) yields a "revisiting" or "rewriting" meaning. We thank an anonymous reviewer for bringing this issue to our attention.

4 Stechow (1996) does not provide translations for these two readings. The English translations for the two readings of this German sentence are taken from Stechow's (1996, pp. 94–95) discussion of this sentence.

5 Please note that the example sentence in (6) follows a dependent-clause word order (S-O-V), rather than the standard independent-clause word order (S-V-O) in German.

6 Phan (2024) also argues that the adverb *có lẽ*, "possibly", in Vietnamese adjoins to ModP_{Epi}, which is within the CP layer.

7 Following (Cao, 2003; Trinh, 2005), we assume that the sentence-final particle *rồi* is a perfect aspect marker.

8 When the perfect marker *rồi* is present, a stative sentence becomes dynamic, as shown in (i).

(i) a. Nam thích Hoa.
Nam like Hoa
'Nam likes Hoa.'

b. Nam thích Hoa rồi.
Nam like Hoa PERF
'Nam comes to like Hoa.'

9 We will not delve into questions about how CompletiveP and ResultativeP would be represented in our framework. One possibility, though, is to assume, with Phan (2013), that CompletiveP and ResultativeP occur in the phrase structure, and that the postverbal *lại* adjoins to them rather than to VP when they occur (assuming that VP is a complement of CompletiveP or ResultativeP). We leave the relevant questions aside.

10 The movement of the vP that *nữa* modifies, without *nữa* itself, could pose a problem in the Government-Binding (GB) theory, because, in the GB theory, only a maximal projection can move. However, we assume that, in the theory of Bare Phrase Structure (Chomsky, 1995), which does not formally distinguish maximal and non-maximal projections, this issue does not pose a serious problem.

References

Beck, S. (2005). There and back again: A semantic analysis. *Journal of Semantics*, *22*(1), 3–51. [CrossRef]

Beck, S. (2006). Focus on Again. *Linguistics & Philosophy*, *29*, 277–314. [CrossRef]

Beck, S., & Johnson, K. (2004). Double Objects Again. *Linguistic Inquiry*, *35*(1), 97–124. [CrossRef]

Beck, S., & Snyder, W. (2001). The resultative parameter and restitutive 'again'. In C. Féry, & W. Stemefeld (Eds.), *Auditur vox sapientiae: A festschrift for arnim von stechow* (pp. 48–69). Akademie Verlag.

Bruening, B., & Tran, T. (2015). The nature of the passive, with an analysis of Vietnamese. *Lingua*, *165*, 133–172. [CrossRef]

Cao, X. H. (2003). *Tiếng việt. mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa*. Nhà xuất bản Giáo dục.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Chomsky, N. (1995). Bare phrase structure. In G. Webelhuth (Ed.), *Government and binding theory and the minimalist program (generative syntax 1)* (pp. 383–439). Blackwell.

Csirmaz, A., & Slade, B. (2020). Anatomy of Hungarian aspectual particles. In V. Hegedűs, & I. Vogel (Eds.), *Approaches to hungarian: Volume 16: Papers from the 2017 Budapest conference* (pp. 26–45). John Benjamins.

Dowty, D. (1979). *Word meaning and montague grammar*. Reidel.

Fabricius-Hansen, C. (1983). Wieder ein wieder? Zur Semannk von wiede. In R. Bauerle, C. Schwarze, & A. von Stcchow (Eds.), *Meaning, use and interpretation of language* (pp. 97–120). Walter de Gruyter.

Giorgi, A. (2010). *About the speaker: Towards a syntax of indexicality*. Oxford University Press.

Ippolito, M. (2007). On the meaning of some focus-sensitive particles. *Natural Language Semantics*, *15*, 1–34. [CrossRef]

Kratzer, A. (1994). *The event argument and the semantics of voice* [Master's thesis, University of Massachusetts].

Lin, T.-H. J., & Liu, C.-M. (2009). 'Again' and 'again': A grammatical analysis of *you* and *zai* in Mandarin Chinese. *Linguistics*, *47*(5), 1183–1210. [CrossRef]

May, R. (1985). *Logical form*. MIT Press.

Nguyễn, Đ.-H. (1997). *Vietnamese*. John Benjamins Publishing Company.

Phan, T. (2013). *Syntax of Vietnamese aspect* [Ph.D. thesis, University of Sheffield].

Phan, T. (2024). Vietnamese modal system of dynamic possibility: A diachronic perspective. *Taiwan Journal of Linguistics*, *22*(1), 167–198.

Phan, T., & Chierchia, G. (2022). Identifying (in)definiteness in Vietnamese noun phrase. *Journal of the Southeast Asian Linguistics Society*, *15*(2), 27–49.

Simpson, A. (2001). Focus, presupposition and light predicate raising in East and Southeast Asia. *Journal of East Asian Linguistics*, *10*, 89–128. [CrossRef]

Simpson, A., & Wu, Z. (2002). IP-raising, Tone Sandhi and the Creation of S-final Particles: Evidence for Cyclic Spell-Out. *Journal of East Asian Linguistics*, *11*, 67–99. [CrossRef]

Stechow, A. v. (1995). Lexical decomposition in syntax. In U. Egli, P. E. Pause, C. Schwarze, A. von Stechow, & G. Wienold (Eds.), *The lexicon in the organization of language* (pp. 81–118). John Benjamins.

Stechow, A. v. (1996). The Different Readings of *Wieder* 'Again': A Structural Account. *Journal of Semantics*, *13*(2), 87–138. [CrossRef]

Thompson, L. C. (1987). *A Vietnamese reference grammar*. University of Hawaii Press.

Tovena, L., & Donazzan, M. (2008). On ways of repeating. *Recherches Linguistiques de Vincennes*, *37*, 85–112. [CrossRef]

Trần, K. P. (2023). LẠI 'come/again' in Vietnamese from syntactic, semantic and pragmatic perspectives. In M. J. Alves, L. Q. Đông, T. C. Lan, T. T. H. Hạnh, & D. X. Quang (Eds.), *Researching and applying linguistics and Vietnamese language studies* (pp. 251–269). Geolinguistic Society of Japan.

Trinh, T. (2005). *Aspect of clause structure in Vietnamese*. Humboldt University.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

Locative Inversion in Vietnamese

Trần Phan [1] and Chao-Ting Tim Chou [2,*]

[1] Institute of Linguistics, National Tsing Hua University, Hsinchu 300044, Taiwan
[2] Department of Foreign Languages & Literatures, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
* Correspondence: ctchou@nycu.edu.tw

Abstract: This paper has four primary objectives. First, we examine two Vietnamese constructions—the locative and existential constructions—in which a locative phrase appears clause-initially, preceding a verb-DP sequence. We demonstrate that, in the locative construction, the locative phrase is an argument of the locative verb, whereas, in the existential construction, it functions as a locative adjunct. Second, we show that this locative adjunct in the existential construction can undergo optional topicalization to the left periphery. More crucially, we argue that the fronting of the locative argument in the locative construction is neither Ā-topicalization to the left periphery nor A-movement targeting Spec,TP, supporting the featural approach to the A/Ā distinction. Third, by comparing the expletive *nó* with the fronting locative argument in the locative construction, we contend that both compete for the same specifier position of a hybrid A/Ā projection in the high TP domain. Fourth, to detail the movement of the locative argument in the locative construction, we posit a composite probe [D, δ] on the head of the A/ĀP to selectively move the locative argument to Spec,A/ĀP over the structurally higher theme DP.

Keywords: locative inversion; A/ĀP; composite probe; Vietnamese

1. Introduction

Locative phrases in Vietnamese are sometimes found sentence-initially preceding a verb-DP sequence, as exemplified by *trên bàn* 'on the table' in (1) and (2). The minimal difference between (1) and (2) lies in the choice of the main predicates: (1) contains a predicate of posture *nằm* 'lie', while (2) features an existential predicate *có* (lit. 'have; exist'). For expository convenience, we refer to the former as locative constructions (LC) and the latter as existential constructions (EC).[1]

(1) **Trên** **bàn** nằm một chồng sách (LC)
on table lie a pile book
'On the table lies a pile of books.'

(2) **Trên** **bàn** có một chồng sách (EC)
on table exist a pile book
'On the table there exists a pile of books."

In this paper, we argue that the derivations of the sentence-initial locative phrases in (1) and (2) do not fall under a uniform analysis. In particular, we propose that the locative phrase in LC is a locative argument that raises to a hybrid A/Ā functional projection in the high TP-domain extensively studied in Bošković (2024a), while that in EC is a locative adjunct.

The structure of this paper proceeds as follows. Section 2 examines the fundamental properties of LC and EC, focusing on the verb types and their syntactic relations with the locative phrases. We will see that the locative phrase in LC is an argument of the locative verb, while that in EC is a locative adjunct. Section 3 addresses the syntactic position(s) of the locative phrases in LC and EC. We show that the locative adjunct in EC may undergo optional topicalization to the left periphery. More importantly, we argue the fronting of the locative argument in LC is neither Ā-topicalization to the left periphery nor an A-movement targeting Spec,TP, under the traditional positional approach to A/Ā-distinction (Chomsky, 1981, p. 47; 1995, 2007, 2008; Mahajan, 1990; Déprez, 1989; Miyagawa, 2010, 2017). The goal of Section 4 is two-fold. First, based on the findings in Section 3, we show that the locative argument in LC competes with the expletive *nó* for the same specifier position of a hybrid A/Ā-projection in the high TP-domain (Bošković, 2024a). Second, we detail the derivational mechanics of the movement to this A/Ā-projection based on the composite probing mechanism in Coon and Bale (2014) and Van Urk (2015). Section 5 explores the theoretical implications of our analysis, including the Merge over Move principle (Chomsky, 1995) and Ā-opacity (Rezac, 2003). Section 6 concludes this paper.

2. Verb Classes and the EC/LC Distinction

An LC is typically composed of posture verbs (such as *nằm* 'lie', *ngồi* 'sit', and *đứng* 'stand'), verbs of placement (such as *đặt* 'place', *để* 'put, place', and *treo* 'hang') and dwell verbs (such as *ở* and *sống* 'dwell') as exemplified by (3).

(3) **Trên** **bàn** nằm/ để một chồng sách.
on table lie put a pile book
'On the table lies / is put a pile of books.'

These verbs must be used in their stative sense in LC, not their change-of-state sense, exhibiting "informational lightness" in the sense of Birner (1992). This property explains why LC is sometimes found accompanied by manner adverbs such as *lăn lóc* 'in a disordered/scattered manner' and *lủng lẳng* 'in a dangling manner' as in (4a) and (4b), respectively. These manner adverbs are incompatible with the dynamic/change-of-state interpretation of these verbs.[2]

(4) a. **Trên** **mặt** **bàn** nằm lăn lóc mấy vỏ chai bia 33.[3]
on surface table lie disorderly several shell bottle beer 33
'On the table lie several empty "33" beer bottles in a disorderly manner.'

b. **Dọc theo** **lan can** **đò** treo lủng lẳng một số áo phao.[4]
along railing boat hang danglingly some life.jacket
'Along the railing of the boat hang several life jackets in a danglingly manner.'

Collectively referred to as locative verbs here, the verbs forming LC are intrinsically tied to a location, necessitating subcategorization for a locative phrase. The subcategorization requirement of the locative verbs can be jointly evidenced by the obligatory occurrence of the locative phrase in (5) and (6). First, (5) shows that an LC is grammatical only with the overt realization of the locative phrase. Next, although (6) indicates that the locative phrase can stay in situ within the VP hosting the locative verb when the theme DP moves to Spec,TP, its presence within VP is mandatory.

(5) The clause-initial locative phrase is obligatory in LC:
*(**Trên** **bàn**) nằm một chồng sách.
on table lie a pile book

(6) The locative phrase is a core argument of the locative verb:

Một	chồng	sách	nằm	*(**trên**	**bàn**.)
a	pile	book	lie	on	table

'A pile of books lies *(on the table).'

On the other hand, an EC typically goes with verbs of existence, which denote general existence and duration, such as *có* 'exist' and *còn* 'remain' as in (7).[5] In principle, (7) allows two interpretations: it can be interpreted as a generic statement asserting the presence of water in general or as a contextualized statement asserting the presence of water in the specific speech context. Importantly, an EC remains grammatical without the overt presence of a clause-initial locative phrase, in contrast to (5). In addition, an EC allows both the theme DP and the locative phrase to appear post-verbally, as in (8). See the contrast between (8a) and its corresponding LC in (9).

(7) The locative phrase is optional in EC:

Có/	còn	nước.
exist	remain	water

'There is/ remains water.'

(8) EC allows a post-verbal locative phrase:

a.

Có	một	chồng	sách	**trên**	**bàn.**
exist	a	pile	book	on	table

'There is a pile of books on the table.'

b.

Xuất hiện	băng	tuyết	**tại**	**huyện**	**Bình Liêu,**	**Quảng Ninh.**[6]
appear	ice	snow	at	district	B.L.	Q.N.

'There appeared frost snow in Binh Lieu District of Quang Ninh Province.'

(9)

*Nằm	một	chồng	sách	**trên**	**bàn**.
lie	a	pile	book	on	table

Intended: 'There lied a pile of books on the table.'

In view of the contrast between LC and EC with respect to the requirement of an overt locative phrase, we assume that the locative phrase is a core argument of the locative verb in LC (cf. Bresnan, 1994), while that in EC is a locative adjunct. Thus, we argue that the locative argument in LC is not base-generated in the clause-initial position as in (1). Rather, it is base-generated within VP as in (10). The derivation starting from the base structure in (10) may be followed by moving either the locative argument or the theme DP, resulting in (1) and (6), respectively.[7,8]

(10) The base structure of LC:

$[_{CP}$ ..... $[_{vP}$ v $[_{VP}$ DP V LP]]]

An alternative hypothesis assumes that the locative argument is base-generated at a position structurally higher than the theme DP in LC. This alternative poses a theoretical problem with respect to Baker's (1997) Uniformity of Theta Assignment Hypothesis (UTAH): the locative phrase bears the oblique theta-role of Location, which should be lower than the Theme argument according to the UTAH's hierarchy in (11) (cf. Wu, 2008).

(11) Agent > Theme > Goal > Obliques (manner, location, time)

We offer two more arguments for our proposal in (10) regarding the subcategorization requirement of the locative verbs forming LC. The first argument is related to the selectional restrictions of locative verbs in LC. If the locative phrase is an argument of the locative verb in LC, there should be a c-selection restriction on the kind of locative information that is

permitted in LC. This prediction is borne out, as only locative phrases denoting a specified location via the use of a localizer are felicitous, while more general locations without a localizer are odd in LC, as in (12b). The contrast in (12) suggests that the locative verb in LC imposes a c-selection requirement on its locative argument. We assume that the c-selection requirement in (12) follows from the argument status of the locative phrase in LC.

(12) The localizer is necessary in LC:

a. **Trong công viên** ngồi một cụ già.
in park sit one old.person
'In the park sits an old person.'

b. ??/***Công viên** ngồi một cụ già.
park sit one old.person
Intended: 'At the park sits an old person.'

By contrast, the locative information in EC can but does not need to be realized by a locative phrase, as shown in (13). Consequently, EC can express either a specified location as in (13a) or a general location in (13b) (which can be contextually determined).

(13) The localizer is optional in EC:

a. **Trong công viên** có / xuất hiện một cụ già.
in park exist appear one old.person
'In the park there is/ appears an old person.'

b. **Công viên** có / xuất hiện một cụ già.
park exist appear one old.person
'At the park there is/ appears an old person.'

The second argument concerns the insertion of temporal adverbials or tense/aspectual markers in EC/LC. Specifically, note that although the locative phrase is optional in EC in (7), EC is most natural with some sort of overt temporal/locative anchoring.[9] In the absence of a locative phrase, the anchoring of EC can be achieved via temporal adverbials or tense/aspectual markers, as in (14) and (15), respectively.

(14) Anchoring of EC via temporal adverb(ial)s:

a. **Sáng nay** xuất hiện một bệnh nhân.
morning this appear one patient
'There appeared a patient this morning.'

b. **Hôm qua** rơi nhiềutuyết.
yesterday fall a.lot snow
'Yesterday fell a lot of snow.'

(15) Anchoring of EC via tense/aspectual markers:

a. **Vừa** đến một bệnh nhân.
just arrive one patient
'A patient just arrived.'

b. **Đang** tan nhiều tuyết.
PROG melt a.lot snow
'A lot of snow is melting.'

However, temporal adverb(ial)s and tense/aspectual markers cannot salvage an LC without a locative phrase, as evidenced by (16). Again, the ungrammaticality of (16) highlights the necessity of the overt presence of a locative phrase in LC.

(16) a. ***Sáng** **nay** nằm một bệnh nhân.
morning this lie one patient
Intended: 'This morning lied a pile of books.'
b. ***Vừa**nằm một chồng sách.
just lie one pile book
Intended: 'Just lied a pile of books.'

In summary, the optional presence of a locative phrase in existential constructions (EC), as shown in examples (7), (14), and (15), supports its status as an adjunct. In contrast, the predicates forming an LC require a locative argument, as demonstrated by the grammatical occurrence of the locative phrase within the VP in (6), the c-selection requirement in (12), and the mandatory inclusion of the locative phrase in (16). This distinction highlights the subcategorization differences between EC and LC with regard to the locative phrases.

3. The Raising of the Locative Argument in LC

Given the conclusion established in the previous section, note that adding the locative phrase at the post-verbal position, as in (17), does not save (16).

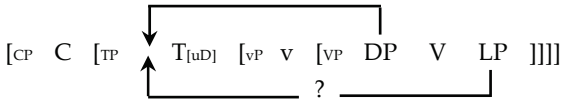
(17) a. *Sáng nay nằm một chồng **sách trên bàn**.
morning this lie one pile book on table
Intended: 'This morning lied a pile of books on the table.'
b. Vừa nằm một chồng sách **trên bàn.**
just lie one pile book on Table
Intended: 'Just lied a pile of books on the table.'

The ungrammaticality of (17) indicates that the locative argument of the locative verb has to raise out of the vP domain to construct a grammatical LC. Now, the question comes down to the landing site of the movement of the locative argument in LC. Two avenues can be envisaged to identify the nature of the position(s) targeted by the fronting locative argument in LC, as shown in (18): it raises either to the canonical subject position Spec,TP, or to a syntactic position other than Spec,TP.

(18) Possible analyses of the movement of the locative phrase in LC:
a. It targets Spec,TP,
b. It raises to a position other than Spec,TP.

The possibility in (18a) encounters two analytical challenges. The first challenge concerns the minimality of the movement. In particular, suppose the raising to Spec,TP is triggered by the satisfaction of a D-feature on T in Vietnamese as in (19); the raising of the locative argument to Spec,TP across the theme DP runs afoul of the relativized minimality constraint (Rizzi, 1990). This is due to the fact that both the locative argument and the theme DP can grammatically occupy this position according to the hypothesis in (18a); however, the theme DP is situated closer to T than the locative argument. This raises the question of how the locative argument can move to Spec,TP, surpassing the theme DP, which typically occupies that position.

(19) The raising to Spec,TP trigged by a D-feature:



Second, if we assume that vP is a phase (Legate, 2003), the locative phrase may move to the edge of the vP phase to become a closer goal to T than the theme DP is, as in (20). However, given that this type of raising to the phase edge is typically considered as an instance of Ā-movement, it is unclear how a subsequent raising to Spec,TP (as a canonical instance of A-movement triggered by D/φ-features) in (20) circumvents the constraint on improper movement which prohibits an Ā-A sequence of derivation (May, 1979; Chomsky, 1981; Abels, 2007; Neeleman & Van De Koot, 2010; Williams, 2011).

(20) [$_{CP}$ C [$_{TP}$ T$_{[uD]}$ [$_{vP}$ v [$_{VP}$ DP V LP]]]]
A Ā

Due to the minimality issue in (19) and the ban on improper movement in (20), we do not assume that the locative argument in LC raises to Spec,TP. In Section 4, we will see an additional empirical problem of the hypothesis in (18a): the raising of the theme DP and that of the locative argument in LC exhibit different semantic/pragmatic properties in terms of specificity.[10]

Given the challenges facing (18a), we pursue the possibility in (18b) to derive the fronting of the locative argument in LC. Our discussion of (18b) focuses on the relative order of the fronting locative argument and two elements: the complementizer in the left periphery and the expletive *nó* in the high TP-domain (Greco et al., 2017, 2018). We discuss the relative order with the complementizers *là* and *rằng* in this section. Note that the locative adjunct in EC may either precede or follow the complementizers *là* as in (21).

(21) a. Rõ ràng **là** **trên** **giường** có một bệnh nhân.
clearly C on bed exist a patient
'Clearly on the bed there is a patient.'

b. **Trên giường** rõ ràng **là** có một bệnh nhân.
on bed clearly C exist a patient
'On the bed there clearly is a patient.'

Phan (2024) observes that the complementizer *là* can occur in main clauses only if it is immediately preceded by CP-level sentential mood and modal markers such as the evidential phrases *rõ ràng* 'clearly' and *hình như* 'seemingly'. He proposes that *là* heads the Finiteness Phrase, marking the lowest bound of the Split-CP *à la* Rizzi (1997), and is licensed by these sentential mood/modal items. The grammaticality of the pair in (21) with respect to the relative order with *là* indicates that the locative adjunct in EC is adjoined TP-internally in (21a)[11] and may undergo optional topicalization to the left periphery above the FinP in (21b). Note that the locative adjunct is not inherently a topic. The non-topical nature of the locative adjunct in EC can be evidenced by its occurrence in out-of-the-blue contexts as a felicitous reply to a global 'what-happened' question as in (22a), unlike the topicalized variation in (22b) (see Rizzi, 2004; Belletti & Rizzi, 2017).

(22) Q: Chuyện gì vậy?
matter what SFP
'What happened?'

a. **Trên bàn** có/ xuất hiện một chồng sách.
on table exist appear a pile book
'On the table there is/ appears a pile of books.'

b. #**Trên bàn** **thì** có/ xuất hiện một chồng sách.
on table TOP exist appear a pile book

Similarly, notice that the position of locative adjunct in relation to *là* has a crucial effect on the grammaticality of the insertion of the topic marker, as shown by (23). The insertion of the topic marker *thì* is allowed only when the locative adjunct precedes the complementizer, as in (23a), suggesting that the topic marker is available only for topic expressions moving to the left periphery.

(23) The topic marker *thì* is available only for a topicalized phrase:

a. **Trên giường thì** có lẽ/ rõ ràng **là** có một bệnh nhân.
on bed TOP perhaps clearly C exist a patient
'Perhaps/ clearly on the bed there is a patient.'

b. *Có lẽ/ rõ ràng **là trên giường thì** có một bệnh nhân.
perhaps clearly C on bed TOP exist a patient

On the other hand, the pair in (24) shows that an LC sounds degraded if the locative argument precedes the head of FinP and its licensers but not when it follows them, in contrast to (21).[12] This suggests that the fronting of the locative argument in LC lands below CP and cannot keep moving to CP-level projections such as the Topic Phrase.

(24) a. Rõ ràng **là trên giường** nằm một bệnh nhân.
clearly C on bed lie a patient
'Clearly on the bed lies a patient.'

b. ??**Trên giường** rõ ràng **là** nằm một bệnh nhân.
on bed clearly C lie a patient
Intended: 'On the bed there clearly lies a patient.'

Our analysis of the locative phrases in LC and EC occupying different syntactic positions can be further supported by the insertion of the topic marker *thì* in Vietnamese. Note that the fronting locative argument in LC cannot be followed by the topic marker *thì* as in (25a), while the locative adjunct in EC is not subject to this constraint as in (25b).[13] The contrast between (25a) and (25b) clearly indicates that the fronting of the locative argument in LC is not an instance of Ā-topicalization to the left periphery.[14] Moreover, the ungrammaticality of (17) indicates the obligatory character of the fronting of the locative argument in LC, which is incompatible with the optionality of Ā-topicalization in general. This incompatibility makes an Ā-topicalization analysis of the fronting locative argument in LC difficult to maintain.

(25) a. *__Trên giường thì__ nằm một bệnh nhân. (LC)
on bed TOP lie a patient
Intended: 'On the bed is lying a patient.'

b. **Trên giường thì** có một bệnh nhân. (EC)
on bed TOP exist a patient
'On the bed there is a patient.'

Another argument against the Ā-topicalization analysis of the fronting locative argument in LC concerns hyperraising. Several languages have been demonstrated to allow for A-movement from the subject position out of a finite CP, which is referred to as hyperraising since Ura (1994). Hyperraising can be demonstrated by Zulu in (26), where the subject DP *Zinhle* of the embedded predicate *xova* "make" moves out of the finite CP complement across the matrix raising predicate *bonakala* "seem" (see also Ferreira, 2000, 2004, 2009; Rodrigues, 2004; Martins & Nunes, 2005, 2009, 2010; Nunes, 2008; Deal, 2017; Fong, 2019; Zyman, 2023).

(26) Hyper-raising in Zulu: (Halpert, 2019, ex.3b)
uZinhle$_1$ **u**-bonakala [$_{CP}$ ukuthi ___$_1$ **u**-zo-xova ujeqe].
AUG.1.Zinhle 1s-seem that 1s-FUT-make AUG.1steamed.bread
'It seems that Zinhle will make steamed bread.'

Lee and Yip (2024) argue that hyper-raising exists in Vietnamese in constructions involving (raising) attitude verbs such as *nghe nói* 'hear', as exemplified by (27b) (see their Section 3 for evidence for an A-movement analysis of (27b) and the finiteness of the CP complement).

(27) Hyperraising in Vietnamese:
a. Nghe nói [$_{CP}$ rằng/ là **cơn mưa này** sẽ không dừng.]
hear C CL rain this FUT not stop
'It is heard that the rain will not stop.'
b. **Cơn mưa này** nghe nói [$_{CP}$ rằng/ là sẽ không dừng.]
CL rain this hear C FUT not stop
'It is heard that the rain will not stop.' (adapted from Lee & Yip, 2024, p. 3)

What makes hyperraising relevant to our discussion is that LC is allowed to occur in the complement clause of *nghe nói*, as shown by (28a). More importantly, the locative argument is allowed to move out of the complement clause as in (28b).[15]

(28) Hyperraising of the locative argument:
a. Nghe nói [$_{CP}$ rằng/ là [**trên giường**]$_1$ nằm một bệnh nhân. t$_1$.]
hear C on bed lie one patient
'It is heard that on the bed lies a patient.'
b. [**Trên giường**]$_1$ nghe nói [$_{CP}$ rằng/ là nằm một bệnh nhân. t$_1$.]
on bed hear C lie one patient
'It is heard that on the bed lies a patient.'

Notice that the movement of the locative argument in (28b) cannot be analyzed as Ā-topicalization, as evidenced by the ungrammaticality of the insertion of *thì* (the marker of the Topic head in the left periphery) in (29).[16]

(29) *[**Trên giường**]$_1$ **thì** nghe nói [$_{CP}$ rằng/ là nằm một bệnh nhân. t$_1$.]
on bed TOP hear C lie one patient
Intended: 'It is heard that on the bed lies a patient.'

The hyperraising of the locative argument lends further support to our analysis that the fronting of the locative argument in LC cannot be analyzed as Ā-topicalization. If the fronting of the locative argument in LC were Ā-topicalization, the subsequent A-movement of hyperraising in (28b) would constitute a violation of the ban on improper movement, which prohibits an Ā-A sequence of derivation (May, 1979; Chomsky, 1981; Abels, 2007; Neeleman & Van De Koot, 2010; Williams, 2011).[17]

Given the resistance of topic marking by *thì* shown in (25a), one may assume that the movement of the locative argument in LC can be analyzed as an A-movement. Aside from functioning as the grammatical subject in LC, the raising of the locative argument does not show a Weak Crossover effect (WCO) as in (30), indicating a typical feature of A-movement (see Culicover & Levine, 2001, pp. 289–290 for the same line of reasoning regarding the A-property of English locative inversion).[18]

(30) [Trên giường của mỗi đứa **trẻ**$_2$]$_1$ đều để [sách của **nó**$_2$ t$_1$].
on bed POSS each CL child all place book POSS he
'On every child$_2$'s bed are placed his$_2$ books.'

Although the raising locative argument in LC resists overt topic marking by *thì* in (25a) and does not exhibit WCO[19], it aligns with an Ā-topicalized phrase with respect to topic island effects and the specificity requirement of indefinite nominals. Let us begin with the topic island effects. It is well-known that locative inversion in English parallels Ā-topicalization in inducing island effects that block *wh*-movement, as illustrated in (31a) and (31b/c), respectively. This parallelism has been cited to support the Ā-properties of English locative inversion (Bresnan, 1994, p. 87; den Dikken, 2006, p. 100; Rizzi & Shlonsky, 2006, p. 344).

(31) Topic island effects:
a. *When$_i$ did he say that **into the room** walked Jack t$_i$?
b. *When$_1$ did **to Lee** Robin give the pencil t$_1$?
c. *When$_1$ did **this book** everyone read t$_1$?

The same reasoning can be applied to demonstrate the Ā-properties of the fronted locative phrase in LCs. Since Vietnamese is a *wh*-in-situ language, Ā-focalization, as shown in (32a), is employed to construct the relevant argument. Notably, the Ā-topicalization of the locative adjunct in ECs blocks Ā-focalization, as illustrated in (32b).

(32) a. **Là [sách ngôn ngữ học]**$_1$ Tí nghĩ [CP rằng có [một chồng **t**$_1$] trên bàn].
FOC book linguistics Tí think C exist a pile on table
Literal: 'It is linguistics books that Tí thinks there is a pile of on the table.'
b. ***Là [sách ngôn ngữ học]**$_1$ Tí nghĩ [CP rằng **[trên bàn]**$_2$ có [một chồng **t**$_1$] **t**$_2$].
FOC book linguistics Tí think C on table exist a pile
Intended: 'It is linguistics books that Tí thinks on the table is a pile of.'

Crucially, the fronting of the locative argument in LCs similarly induces a topic island effect, as shown in (33). The parallelism between (32b) and (33) indicates that the fronting of the locative phrase in LCs and ECS, though targeting different structural positions, exhibits the same Ā-property in terms of the topic island effects.

(33) The fronting locative argument in LC creates a topic island:
***Là [sách ngôn ngữ học]**$_1$ Tí nghĩ [CP rằng **[trên bàn]**$_2$ nằm [một chồng **t**$_1$] **t**$_2$].
FOC book linguistics Tí think C on table lie a pile

Next, we turn to the argument concerning the specific reading for indefinite nominals. Note that nominal phrases that feature the indefinite article *một* 'a/an' followed by a classifier-noun sequence allow both a specific and a nonspecific/generic interpretation (see also Phan & Chierchia, 2022; Enç, 1991). For instance, the subject *một bệnh nhân* 'a patient' in (34) is allowed either a specific reading or a nonspecific/generic reading. The two possible interpretations of an indefinite NP with *một* can be further demonstrated by the two possible continuations in (35a) and (35b) (see Fodor & Sag, 1982, p. 355; Von Heusinger, 2002, p. 245).

(34) **Một bệnh nhân** nằm trên một cái giường.
a patient lie on A CL bed
a. 'A specific patient lies on a bed.' (specific)
b 'A/any patient lies on a bed.' (nonspecific/generic)

(35) **Một sinh viên** lớp Cú pháp gian lận trong giờ thi.
a student class syntax cheat in time test
'A student in Syntax cheated on the exam.'
a. Tên sinh viên đó là Tí. (specific)
name student that be Tí
'The student's name is Tí.'
b. Chúng tôi đang điều tra xem đó là ai. (nonspecific)
we PROG investigate see that be who
'We are trying to figure out who it was.'

Notice that the specificity ambiguity of indefinite nominals disappears under Ā-topicalization as in (36) because topics display definiteness/specificity effects (Enç, 1991; Erteschik-Shir, 1997, a.o.).[20] Similarly, Phan and Lander (2015, fn.3) suggest that topicality is a function of specificity (cf. Cresti, 1995; Portner, 2002) and that the topic particle *thì* is one of the specificity markers in Vietnamese. Thus, the topicalized indefinite nominal marked by *thì* in (36) receives only the specific reading.

(36) A topicalized indefinite nominal must receive a specific reading:
Context: Professor A is telling about his three favorite students…
Một sinh viên **thì** rất giỏi cú pháp.
a student TOP very good syntax
'As for one student, s/he is really good at syntax.' (specific/*nonspecific/*generic)

The specificity ambiguity of indefinite nominals is carried over to locative phrases containing indefinite NPs, as demonstrated by (37). When the locative phrase containing an indefinite NP stays within vP as in (37), it patterns with the indefinite NPs in (34)/(35) in displaying ambiguity with respect to specificity.

(37) Một bệnh nhân nằm **trên một cái giường.**
a patient lie on a CL bed
a. 'A patient lies on a specific bed.' (specific)
b 'A patient lies on a/any bed.' (nonspecific/generic)

Importantly, the ambiguity in (37) disappears when the locative phrase is fronted as in (38): the fronted LP is necessarily interpreted as being specific, relating to a specific bed in the speaker's mind. This specific spatial coordinate helps anchor the eventuality reported in the LC to a specific context.

(38) **Trên một cái giường** nằm một bệnh nhân.
on a CL bed lie a patient
a. 'On a specific bed lies a patient.' (specific)
b #'On a/any bed lies a patient.' (*nonspecific/*generic)

Now, we encounter a dilemma in the analysis of the trigger of the movement of the locative argument in LC. On the one hand, the fronting locative argument in LC patterns with a topicalized phrase in terms of the topic island effects and the specificity requirement of indefinite nominals. On the other hand, the fronting locative argument resists the topic marker *thì* and does not show WCO. How do we capture the fact that the movement of the locative argument, though not a case of Ā-topicalization to the left periphery, exhibits the specificity requirement seen in Ā-topicalization? What makes this problem trickier is the contrast between (34) and (38). The indefinite subject NP in the categorical sentence in (34) allows either a nonspecific/generic or a specific reading, while the fronted locative argument in (38) permits only a specific interpretation. This contrast indicates that the

movement of the indefinite subject NP in (34) and that of the locative argument in LC in (38) are driven by two distinct motivations and target different syntactic positions such that only the former retains the specificity ambiguity of an indefinite NP.

Summing up, the ungrammaticality of (25a), the contrast in (23)/(24), and the grammatical hyperraising of the locative argument in (28b) corroborate our analysis that the fronting locative argument in LC stays below the CP level (more precisely, the FinP headed by the complementizer *là*), whereas the locative adjunct in EC may undergo optional topicalization to the left periphery. Thus, the fronting locative argument does not target Spec,TP, and it cannot be analyzed as Ā-topicalization. Now the question is where exactly in the high TP-domain the fronting locative argument in LC moves to. In the next section, we propose that the locative argument in LCs moves to the specifier position of a hybrid A/Ā-projection in the high TP-domain extensively investigated by Bošković (2024a). Our proposal will be empirically corroborated by the similarities of the locative argument in LCs and the non-referential expletive *nó* (lit. 'it') in Vietnamese.

4. Proposal

4.1. Movement to a Hybrid A/Ā-Projection Between TP and CP

Bošković (2024a) claims that there are not just one but three distinct structural positions for subjects across languages. Quirky subjects such as Icelandic *mér* 'me' and *Þeim* 'them' in (39) are in Spec,TP. Agreeing/nominative subjects such as *Amy* in (40) merge in a higher Spec,AgrsP, and non-nominal subjects and locally Ā-moved subjects such as *who* in (41) move further to the Spec of an A/Ā projection above AgrsP.[21] The relative position of these three different projections hosting different types of subjects is shown in (42).

(39) **Mér**/ **Þeim** er kalt. (Bošković, 2024a, p. 26)
me$_{DAT}$ them$_{DAT}$ is cold

(40) I think **Amy** left. (Bošković, 2024a, p. 8)

(41) I wonder **who** left. (Bošković, 2024a, p. 8)

(42) [$_{A/\bar{A}P}$ *wh*-moved subject [$_{AgrsP}$ Mary [$_{TP}$ quirky subjects…
(adapted from Bošković, 2024a, p. 26)

Bošković (2024a) refers to the highest projection as the A/ĀP because it resides on the border of the traditional A and Ā field, with its Spec (where English focalized or *wh*-moved subjects are located) exhibiting mixed A/Ā-properties. Adopting this line of analysis of multiple landing sites of subject movement, we suggest that the locative argument in LC is moved to the specifier of the A/ĀP, a "subject-dedicated position" which Bošković (2024a, p. 23, fn.12) views as not having "the same featural content in all its instantiations." Note that Bošković (2024a) argues that Spec,A/ĀP is the dedicated position for non-nominal subjects, making it the ideal landing site for quirky PP subjects (with a null P) in Serbo–Croatian (see Bošković, 2024b and citations therein) and locative PPs in English locative inversion.

That the locative argument in Vietnamese LC is not canonically nominal, given that it can be headed by either a localizer (43) or a preposition (44), is well aligned with the characteristics of this A/Ā subject-dedicated position shown in Bošković (2024a, 2024b).[22]

(43) a. **Trên** **bàn** nằm một chồng sách. (=(1))
on table lie a pile book
'On the table lies a pile of books.'

b. [$_{LP}$ trên [$_{NP}$ bàn…

(44) a. **Ở** **trên** **bàn** nằm một chồng sách.
at on table lie a pile book
'On the table lies a pile of books.'
b. [PP ở [LP trên [NP bàn…

Therefore, in this paper, we propose that the fronting locative argument in LCs targets the specifier position of the hybrid A/Ā-projection proposed by Bošković (2024a, 2024b). In the next section, we discuss the similarities between the locative argument in LCs and the expletive *nó* in Vietnamese to corroborate this proposal.

4.2. The Similarities of the Fronting Locative Argument and the Expletive Nó

We propose that the occurrence of the non-referential expletive *nó* (lit. 'it') in Vietnamese, as discussed in Greco et al. (2017, 2018), sheds light on the question of the syntactic position occupied by the locative argument in LC. We find that LC is incompatible with the expletive *nó*, regardless of the word order of the locative argument and the expletive, as evidenced by (45).

(45) The complementary distribution of LC and the expletive *nó*:
a. ***Trên** **bàn** **nó** nằm một chồng sách.
on table EXPL lie a pile book
Intended: 'There lies a pile of books on the table.
b. ***Nó** **trên** **bàn** nằm một chồng sách.
EXPL on table lie a pile book

By contrast, Greco et al. (2017, 2018) find that the locative adjunct in EC precedes the expletive as in (46a). Reversing the word order of the locative adjunct and the expletive, as in (46b), results in an ungrammatical sentence.[23]

(46) The locative adjunct in EC precedes the expletive:
(adapted from Greco et al., 2018, p. 44)
a. **Trên** **bàn** (thì) **nó** sẽ không có cái bút nào.
on table TOP EXPL FUT NEG exist CL pen any
'On the table, there will be no pens.'
b. ***Nó** **trên** **bàn** (thì) sẽ không có cái bút nào.
EXPL on table TOP FUT NEG exist CL pen any
Intended: 'On the table, there will be no pens.'

The complementary distribution of LC and the expletive shown in (45) indicates that Bruening's (2010) null expletive analysis of English locative inversion cannot be extended to the derivation of Vietnamese LC as depicted in (47) (see also Postal, 1977, 2004; Maruta, 1985; Coopmans, 1989; Hoekstra & Mulder, 1990). The null expletive analysis suggests that Spec,TP is occupied by a null expletive, with the locative phrase undergoing Ā-topicalization. The pattern we see in LC refutes both of these two hypotheses. For one thing, the fronting locative argument in LC resists marking by *thì*, representing the Topic head in the left periphery (see (25a)). For another, the postulation of a null expletive in Vietnamese is implausible in light of the incompatibility of LC and the expletive *nó* (see (45)).

(47) A null expletive analysis of LC: (with the null expletive indicated by *pro*)
[TopP **Trên** **bàn** [TP *pro* nằm một chồng sách]].
on table EXPL lie a pile book
'There lies a pile of books on the table.'

The pattern in (45) is expected if the fronting locative argument and the expletive compete for the same syntactic position in the derivation of LC, so the obligatory fronting of the locative argument in the derivation of LC is blocked, resulting in the ungrammaticality of the pair in (45). A further piece of evidence supporting this line of analysis concerns Ā-topicalization. Specifically, Greco et al. (2018) observe that *nó* cannot undergo topicalization and precede the topic marker *thì*, as shown in (48). This ban on topicalization patterns with the locative argument in LC, as shown in (25a), repeated here in (49).

(48) *Nó* cannot undergo topicalization: (Greco et al., 2018, p. 41)

a. ***Nó** **thì** ngã thằng bé.
EXPL TOP fall CL little
'As for him, the boy fell.'

b. ***Nó** **thì** chết cá tao.
EXPL TOP die fish mine
'As for him/her, my fish died.'

(49) The locative argument in LC cannot undergo topicalization:
***Trên** **giường** **thì** nằm một bệnh nhân. (=(25a))
on bed TOP lie a patient
Intended: 'On the bed is lying a patient.'

The similarity between *nó* and the locative argument in LC goes beyond their syntactic distribution. Greco et al. (2017, 2018) find that the existential and thetic sentences featuring the expletive *nó* show a specificity effect. Concretely, the contrast in (50) shows that the presence of preverbal *nó* necessarily restricts an existential sentence to a contextual interpretation where the relevant eventuality is tied to some specific spatial/temporal coordinates.

(50) The specificity requirement of *nó*: (Greco et al., 2017, p. 78)

a. Không có ma.
NEG exist ghost
Generic: 'Ghosts do not exist.'
Contextual: 'There are no ghosts speaking of a certain place/time.'

b. **Nó** không có ma.
EXPL NEG exist ghost
#Generic: 'Ghosts do not exist.'
Contextual: 'There are no ghosts speaking of a certain place/time.'

To account for the ban on the generic reading of existential sentences in the presence of *nó*, Greco et al. (2018, p. 49) state that the referent of the postverbal DP is necessarily linked to a certain set of entities that was previously determined, and the addition of *nó* obligatorily "narrows the possible denotations to the values that can be assigned through a retrievable relation ranging over familiar entities, either present in the discourse or in the speaker's mind." As pointed out by a reviewer, this discourse property of *nó* resonates with Bolinger's (1973, 1977) view of ambient expletives or situation pronouns that have special referential content. That is, ambient expletives such as *nó* have a *deixis ad oculos*, "a reference to the immediate situation" (Bolinger, 1973, p. 263). Using Milosavljević and Milosavljević's (2022) wording, *nó* may be understood as having a discourse-topical referent, referring back to an epistemically specific Topic Situation.

Therefore, both the fronting locative argument in LC and the (ambient) expletive *nó* in existential/thetic sentences function to narrow down the domain of validity of the associated eventuality to a specific context. This is aligned with Birner's (1992, p. iii) conception of inversion (with locative inversion as an instance) as having the discourse function of introducing "relatively unfamiliar information to the prior context via the

clause-initial placement of information that is relatively familiar in the discourse" (see also Levin & Rappaport Hovav, 1995). In the case of LC, the more familiar information is encoded by the fronting locative phrase, while the less familiar information is conveyed by the postverbal DP. It is the latter that receives presentational focus (Bresnan, 1994 and citations therein) and is (re)introduced into the domain of discourse.[24] Using Enç's (1991, p. 24) language, specificity in LC involves a (weak) link to a previously established referent, with the denotation of the locative phrase "being a subset of or standing in some recoverable relation to a familiar object" that is already prominent in the discourse. In short, both the expletive *nó* and the locative argument in LC communicate a discourse-related meaning of specificity in that it denotes the speaker's ability to identify the specific eventuality reported.

Summing up, the complementary distribution of the locative argument in LC and *nó*, their incompatibility with topicalization, and their similarity in terms of the specificity requirement jointly suggest that the locative argument and *nó* target the same syntactic position.[25] Thus, the question comes down to the syntactic position occupied by the expletive *nó*. Greco et al. (2017, 2018) argue that syntactically, the expletive *nó* functions as a subject and occupies the highest position in the inflectional layer, although it still merges below CP. However, Greco et al. (2017, 2018) remain agnostic about the precise merge position of the expletive, mentioning three possibilities: *nó* could merge either in Kiss's (1996) RefP (a specialized projection hosting subjects endowed with a [+specific] feature), in a projection that also attracts referential subjects in categorical sentences (Cardinaletti, 2004) or in a dedicated SubjP for subjects *à la* Rizzi and Shlonsky (2006).

In this paper, we propose that these three possible positions of *nó* assumed by Greco et al. (2017, 2018) can be subsumed under a special subject position exhibiting mixed A/Ā-properties located between TP and CP investigated in Bošković (2024a).[26] Based on this assumption, we argue that the raising locative argument in LC and the merger of the expletive *nó* both target the specifier of this hybrid A/Ā-projection. We spell out the derivational details of the movement to this A/Ā-projection based on the composite probing mechanism in Coon and Bale (2014) and Van Urk (2015).

4.3. The Derivation of LCs

One important implication of adopting Bošković's (2024a) A/ĀP in the high TP-domain is that we need to abandon the traditional positional approach to A/Ā-distinction (Chomsky, 1981, p. 47; 1995, 2007, 2008; Déprez, 1989; Mahajan, 1990). According to the positional approach, A-positions include thematic positions and the subject position (i.e., spec,TP), while Ā-positions encompass positions in CP and the edge of a phase. However, the A/ĀP in the high TP domain does not fall into either of these two categories.

Therefore, to spell out the derivational details of the movement of the locative argument in LC, we adopt the featural approach to A/Ā-distinction in (51).

(51) The featural approach to A/Ā-distinction:
(Obata & Epstein, 2011; Van Urk, 2015; Fong, 2019)
A and Ā-properties derive from properties of the attracting feature: A-positions are crea ted by the valuation of D/φ-features, whereas Ā-positions by that of δ-features (e.g., *wh*, topic focus).

Considering the A-properties (such as case, φ-agreement, and anaphoric binding) of Ā-topicalization in Dinka, Van Urk posits that the distinction between A and Ā is not an intrinsic characteristic of a syntactic position. Instead, A and Ā properties emerge as derivative by-products of the specific features that trigger movement and/or establish the position. In particular, A-positions are generated through the valuation of D/φ-features, whereas Ā-positions result from the valuation of δ-features (for example, wh,

Focus, Topic, etc.). In this paper, D-feature is used in Vietnamese to be responsible for A-properties under the featural approach to A/Ā-distinction, in accordance with Preminger's (2019) No-Null-Agreement Generalization against postulating φ-agreement in languages that lack exponence of φ-features systematically throughout an entire paradigm (e.g., Vietnamese). One important theoretical implication of (51) is that an operation would exhibit a hybrid of A/Ā-properties if it is driven by both A- and Ā-features. Van Urk proposes that A/Ā operations can be realized through Coon and Bale's (2014) composite probing mechanism, which allows A-features and Ā-features to jointly probe from a single functional head, ultimately converging on a single goal that possesses both matching A-features and Ā-features.

As a consequence, operations guided by A/Ā-composite probing are expected to display a hybrid mixture of A/Ā-properties. This prediction aligns with the fronting of the locative argument in LC. Thus, we propose that Spec,A/ĀP hosts the locative argument in LC, and the head of A/ĀP is endowed with a composite A/Ā-probe [D, δ]. As for the δ-feature of this composite probe, we follow Bošković's (2024a) characterization of Spec,A/ĀP as not having "the same featural content in all its instantiations", and assume that the δ-feature can be instantiated by one of the following Ā-features: [*u*Topic], [*u*Focus] and [*u*wh]. In other words, the fronting locative argument in LC can deliver topicality, carry focus, or function as a *wh*-interrogative. As for topicality, recall that the fronting locative argument in LC exhibits a paradoxical pattern: it patterns with Ā-topicalization with respect to the specificity requirement as in (38), yet it resists the topic marker *thì* as shown in (25). We posit that this empirical dilemma can be reconciled by assuming that *thì* realizes the Topic head in the left periphery, and the specificity requirement of the fronting locative argument follows from the Topic feature on the head of the A/ĀP. Next, the compatibility of the locative argument with focus/*wh*-interrogative can be evidenced by (52) and (53): the locative argument may carry contrastive/identificational focus as in (52) or function as an interrogative *wh*-expression as in (53).

(52) Locative arguments in LC are able to encode contrastive/identificational focus:

Là	**TRÊN**	**BÀN**	để	sách	ngôn ngữ học	(chứ	không	phải	trên	ghế).
be	on	table	place	book	linguistics	CONJ	neg	be	on	chair

'ON THE TABLE are placed linguistic books (, not on the chair).'

(53) Locative arguments in LC can function as an interrogative *wh*-expression:

Trên	**cái**	**bàn**	**nào**	để	sách	ngôn ngữ học?
on	CL	table	which	place	book	linguistics

'On which table are linguistic books placed?'

The non-uniform pattern of the discourse function of the fronting locative argument is reminiscent of Van Urk's (2015, p. 107) hypothesis that Ā-probes may be flat or relativized (Rizzi, 1997; Abels, 2012).[27] Only a relativized Ā-probe is specified for what type of Ā-goal (topic, focus, or *wh*) is able to satisfy the probe. On the other hand, a flat Ā-probe is non-selective and free from such restrictions on the type of Ā-goal it requires. Van Urk demonstrates that the Ā-probe on the composite probe associated with C in Dinka is flat and non-selective, as the agreement correlation linked to topicalization also applies to relativization. Furthermore, topicalization and relativization exhibit complementary distribution in Dinka; that is, the flat and non-selective Ā-probe can be satisfied by either a topic or a relativized operator but not by both simultaneously. Following this line of analysis, we assume that the Ā-probe on the composite probe on the head of A/ĀP in Vietnamese is also a flat and non-selective probe.

With the hypothesis regarding composite probing on the head of the A/ĀP in place, we show the derivation of the fronting of the locative argument in LC. Note that the fundamental problem of the fronting of the locative argument within the generative framework is the featural basis that gives rise to the syntactic primacy of the locative argument to move over the DP to Spec,A/ĀP in (54).

(54) $A/\bar{A}_{[uD,\ u\delta]}$ ..... $[_{vP}\ v\ [_{VP}$ DP V LP]]

In order for the locative argument to be considered a valid target for the composite probe, it must possess both a D-feature and a δ-feature. In terms of the featural composition of locative phrases, Grimshaw (1991) and van Riemsdijk (1990, 1998) argue that locative PPs are more referential compared to other types of PPs, categorizing them as extended nominal categories that carry an incomplete set of φ-features.

Building on the work of Grimshaw (1991) and van Riemsdijk (1990, 1998), Doggett (2004, pp. 34–35) observes that locative PPs differ from other prepositional phrases in their capacity to move to Spec,TP. This distinction is illustrated by the ungrammatical movement of the PP object in double object unaccusative verbs, as shown in examples (55) and (56).[28]

(55) a. The play appealed to Mary.
b. *To Mary appealed the play.

(56) a. The grade mattered to John.
b. *To John mattered the grade.
(adapted from Pesetsky, 1995, p. 115)

Considering the difference between locative inversion and the examples in (55) and (56), Doggett (2004) argues that locative PPs must possess the ability to establish some form of agreement with T, which serves as the featural foundation for their movement to Spec,TP. In contrast, general prepositional phrases lack this capacity. Although Vietnamese lacks overt manifestation of φ-features/agreement, we follow the above-mentioned characterization of locative PPs to assume that the locative argument in LC bears a defective D-feature, *à la* Grimshaw (1991), van Riemsdijk (1990, 1998), and Doggett (2004) (see also Huang et al., 2009 on the analysis of localizers as a deficient sub-category of nouns in Mandarin Chinese). With the D-feature and an appropriate δ-feature compatible with the intended pragmatic function (i.e., topicality, focus, or *wh*-interrogative) of a locative construction, the locative argument is an eligible goal for the composite probe on the head of the A/ĀP as depicted in (57).

(57) $[_{A/\bar{A}P}\ A/\bar{A}_{[uD,\ u\delta]}$ … $[_{vP}\ v\ [_{VP}\ DP\ V\ LP_{[def\text{-}D,\ \delta]}]]]$

Finally, to move the locative argument to Spec,A/ĀP from within the vP phase (Legate, 2003), we adopt the weak version of Chomsky's (2001) Phase Impenetrability Condition in (58). In effect, one derivational consequence of (58) is that vP becomes a phase only when the next phase head up (say, C) enters the derivation, thereby delaying the opaqueness of the VP complement of v. Thus, the locative argument DP within VP remains visible to the composite probe on A/Ā and is able to move to Spec,A/ĀP as in (59) because C has not entered the derivation.

(58) Phase Impenetrability Condition: (Chomsky, 2001, p. 14)
The domain of H is not accessible to operations at ZP; only H and its edge are accessible to such operations (where H is a phase head, and ZP is the smallest phase dominating the HP phase).

(59) [A/ĀP LP[def-D, δ] A/Ā[~~uD, uδ~~] … [vP v [VP DP V t]]]

We want to end this section with two clarifications. The first clarification concerns whether the intervening DP induces a minimality problem for the raising of the locative argument to Spec,A/ĀP. It is important to note that neither the DP nor the locative argument fully satisfies the requirements of the composite probe. The theme DP is missing the δ-feature[29], while the locative argument is defective in its D-feature. Consequently, it is not immediately clear why the locative argument in (59) should be considered the optimal goal that aligns with the composite probe.

Coon and Bale (2014) offer a formal calculus of minimality for composite probing, specifying that the ranking of potential goals for a composite probe is determined by multiplying the ranks assigned by each sub-probe, as outlined in (60). Crucially, if either sub-probe assigns a rank of 0 to a potential goal, the overall rank becomes 0 due to the multiplicative nature of the calculation. This means that failure to match one sub-probe results in failure to meet the requirements of the composite probe as a whole. Consequently, the DP is excluded as a valid goal for the composite D+δ probe on the head of A/ĀP because it lacks the δ-feature.

(60) Conjoined Match of Rank n:
A syntactic object σ is a match of rank n for a fused probe ϱ composed of α and β, where i × j = n and where i is σ's rank with respect to α and j is σ's rank with respect to β.

Second, Bošković (2024a) views the locus of the EPP as not a fixed and unique position; instead, syntactic context determines the highest projection in an EPP domain, which is the position where the EPP is satisfied. Specifically, he states that the lower projection in the EPP domain, say Agrs seeking a nominal goal in (61a), is unable to trigger the movement of the *wh*-phrase because the *wh*-phrase is a QP, rather than a nominal to meet the probing requirement of Agrs, according to Cable (2010). As a consequence, the A/Ā head and Agrs in (61b) "undergo feature sharing in the sense of Frampton and Gutmann (2000), where two instances of an unvalued feature become one—when one of them gets valued they both do" (Bošković, 2024a, p. 51), so the A/Ā head, with its Ā-feature, is able to agree with the *wh*-expression and trigger the *wh*-movement.

(61) Bošković's (2024a, pp. 51–52) local computation of the contextuality of the EPP effect:
a. Agrs [$_{vP}$ [$_{QP}$ who] …]
b. A/Ā [$_{AgrsP}$ Agrs [$_{vP}$ [$_{QP}$ who] …]]

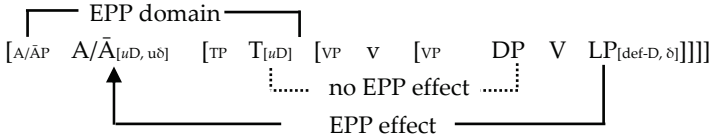
However, we do not see how this line of analysis can be extended to derive the contextuality of the EPP effect in the derivation of LC. In particular, when the T head enters the derivation of LC in (62a), the theme DP is clearly an eligible goal for the D-probe on T. Thus, the theme DP should undergo movement to merge with T, yielding a sentence similar to (62b). That is, Bošković's (2024a, pp. 51–52) deduction of the contextuality of the EPP effect is unable to derive the fronting of the locative argument in LC such as (62c): the A/Ā head and T in (62) do not "undergo feature sharing in the sense of Frampton and Gutmann (2000)" because the theme DP satisfies the D-probe on T in (62b). Even if we allow the A/Ā head to attract the locative argument based on its Ā-probe, the resulting sentence in (62d) is ungrammatical.

(62) a. **T[*u*D]** [VP v [VP **DP** V LP[def-D, δ]]]
b. [DP **Một** **chồng** **sách**] nằm [LP trên bàn.]
a pile book lie on table
'A pile of books lies on the table.'
c. [LP **Trên** **bàn**] nằm [DP **một** **chồng** **sách**].
on table lie a pile book
'On the bed lies a pile of books.'
d. *[LP **Trên** **bàn**] [DP **một** **chồng** **sách**] nằm.
on table a pile book lie

Therefore, although we adopt Bošković's (2024a) contextual approach to the EPP effect in the derivation of LC, we do not follow his deduction outlined in (61), at least not for the derivation of

LC. To derive the absence of the EPP effect at the TP level in the derivation of LC as in (63), we posit that the alternation between (62b) and (62c) boils down to the numeration choice for the derivation built based on a locative verb such as *nằm* 'lie'. If the A/Ā head is not selected in the numeration, TP would be the highest projection in the EPP domain. As a result, the theme DP undergoes movement to merge with T, yielding (62b). The locative argument will raise to derive LC, such as (62c) if the A/Ā head is selected to enter the derivation.

(63)



5. Theoretical Implications

In this section, we discuss two theoretical implications of our proposal of deriving Vietnamese LC based on Bošković's (2024a) A/ĀP and the composite probing mechanism in Van Urk (2015) and Coon and Bale (2014): the Merge over Move preference and the Ā-opacity of the locative adjunct to the composite probe on A/Ā.

5.1. Merge over Move

Chomsky (1995) argues for the Merge over Move preference in (64), where satisfying a requirement by Merge (external merge) is preferred over satisfying it by Move (internal merge) because the former is "simpler".

(64) Merge over Move (MoM):
If, at some point in the derivation, Merge (for lexical insertion) and Move can both apply, then Merge always applies first.

For instance, Chomsky (1995) explains the contrast between (65b) and (65c) in terms of MoM: both merging the expletive *there* and moving *a man* can satisfy the EPP on the embedded T, but the latter option is ruled out because Merge preempts Move.

(65) a. [TP A man1 seems [TP t1 to be t1 outside]].
b. [TP **There1** seems [TP **t1** to be a man outside]].
c. *[TP There seems [TP **a man1** to be **t1** outside]].

However, this preference has been questioned because it is not easy to maintain if Move is just a sub-class of Merge (e.g., Castillo et al., 2009; Epstein & Seely, 1999, 2006; Bošković, 2002, 2007). Specifically, Chomsky (2000, 2001, et seq.) recasts Merge and Move as external Merge and internal Merge, respectively, treating them as two sub-classes of Merge, an elementary and generalized syntactic operation. The consequence of this conceptual unification is that the contrast between (65b) and (65c) can no longer be explained in terms of MoM.[30] In fact, the considerations discussed in Chomsky (2019) would lead to preferring Move (internal merge) to Merge (external merge). That is, the opposite of (64) should be the preferred derivational choice when both Move and Merge are eligible operations to fulfill a syntactic requirement. Indeed, based on the evidence from reflexes of successive-cyclic movement in Wolof and Duala, Georgi (2014) argues that MoM is not a universal principle, and Move can precede Merge.

The derivation of LC offers additional evidence for the view that the application of movement could preempt the merger of an expletive. Specifically, at the derivational stage of LC in (66), the composite probe on the head of A/ĀP seeks an appropriate goal carrying matching A- and Ā-features. Assuming that the expletive *nó* is endowed with both a D-feature and a δ-feature (because of its specificity requirement), either moving the locative argument or merging the expletive *nó* may satisfy the composite probe on the head of A/ĀP.

(66) A/Ā[uD, uδ] … [vP v [VP DP V LP[def-D, δ]]]
a. Move: Moving the locative argument
b. Merge: Lexical insertion of the expletive *nó*

If MoM is an effective principle in Vietnamese, (66b) would be predicted to be the next step in the derivation. However, this prediction is not borne out, as evidenced by the ungrammaticality of (67). Therefore, we agree with Georgi's (2014) conclusion that MoM is not a universal principle, and Move may preempt Merge, as demonstrated by the choice of (66a) over (66b) in the derivation of LC.

(67) ***Nó** nằm một chồng sách trên bàn.
EXPL lie a pile book on table
Intended: 'There lies a pile of books on the table.'

5.2. Ā-Opacity

Interestingly, merging the expletive at Spec,A/ĀP is a grammatical step for the derivation of EC, as demonstrated in (68), in sharp contrast to (67).

(68) **Nó** có/ xuất hiện một chồng sách trên bàn.
EXPL exist appear a pile book on table
'There is/ appears a pile of books on the table.'

Given that the minimal difference between (67) and (68) lies in the argument/adjunct status of the locative phrase *trên bàn* 'on table', the contrast between LC and EC with respect to the grammaticality of inserting the expletive at Spec,A/ĀP suggests the ineligibility of the locative adjunct to serve as a goal for the composite probe on the head of A/ĀP. Concretely, assuming that the locative adjunct is right-adjoined to VP in the derivation of EC (69a) is not a possible option at this stage of the derivation of EC.

(69) $A/\bar{A}_{[uD,\ u\delta]}$ … [$_{VP}$ v [$_{VP}$ DP V] $\underline{LP_{[def\text{-}D,\ \delta]}}$]]
a. Move: Moving the locative adjunct
b. Merge: Lexical insertion of the expletive *nó*

In addition, recall that the fronting locative adjunct in EC has to precede the expletive, as shown by the contrast in (37), repeated below as (70). Therefore, it is not the case that the locative adjunct is immobile in the presence of the expletive. It just cannot move to Spec,A/ĀP.

(70) The locative adjunct in EC precedes the expletive:
(adapted from Greco et al., 2018, p. 44)
a. **Trên bàn** (thì) **nó** sẽ không có cái bút nào.
on table TOP EXPL FUT NEG exist CL pen any
'On the table, there will be no pens.'
b. ***Nó trên bàn** (thì) sẽ không có cái bút nào.
EXPL on table TOP FUT NEG exist CL pen any
Intended: 'On the table, there will be no pens.'

We want to point out that the opaqueness of the locative adjunct to the composite probe on the head of the A/ĀP is similar to the phenomenon of Ā-opacity discussed in Rezac (2003). Ā-opacity can be instantiated by the Icelandic sentence in (71), where the Ā-moved object quantifier *engu grjóti* 'no rock' appears opaque (i.e., invisible) as a (closer) goal for the φ-probe on T (see also Jayaseelan, 2001; Obata & Epstein, 2011). In short, the issue of Ā-opacity pertains to why Ā-interveners do not block A-agreement relations.

(71) Ā-opacity in Icelandic: (Svenonius, 2000; cited in Rezac, 2003, p. 178)
Strákarnir$_2$ höfðu [**engu grjóti**]$_1$ [$_{vP}$ t$_2$ [$_{VP}$ hent t$_1$ í bílana]].
the.boys had no rock throw in the.cars
'The boys had thrown no rocks at the cars.'

Importantly, this Ā-opacity effect is by no means a language-particular problem encountered only in specific constructions because it is also evident in other general derivations in other languages. For example, Obata and Epstein (2011, p. 130) note that the Ā-opacity effect also manifests itself in the derivation of *wh*-object movement similar to (72). In their own words, "when T (*does*) φ-agrees

with the subject *John*, the intervening DP *what cars* bearing "plural" can be mysteriously skipped. As in Icelandic data, elements at Ā-positions do not function as interveners of T-agreement".

(72) What cars does John buy?
[$_{CP}$ what cars does [$_{TP}$ John T [$_{vP}$ **<what cars>** [$_{vP}$ John [$_{VP}$ buy <what cars>]]]]]
singular

Similarly, the derivation of internal object topicalization in Chinese, such as (73), also exemplifies the Ā-opacity effect (Shyu, 1995; Ting, 1995; Paul, 2005; Kuo, 2009; Chen, 2023). In particular, why does the closer object DP not cause an interference problem for the D-probe on T to find and attract the lower subject DP, as shown in (74)?

(73) Internal topicalization:

Lisi	**zhe-ben**	**shu**$_1$	zaojiu	kan-guo	**t**$_1$	le
Lisi	this-CL	book	already	read-EXP		SFP

'As for this book, Lisi has already read it.'

(74) $T_{[uD]}$ $\mathbf{DP_{obj}}$ [$_{v*P}$ DP_{subj} ..... t_{obj}]
intervention?

Now, notice the parallelism between the Ā-opacity effect in (68)–(71) and the opacity of the locative adjunct to the composite probe on A/Ā in EC in (66)/(67). An A-probe (whether it is contained in a composite probe) skips an appropriate goal when the goal is at an Ā-position. Although we do not have a principled account for the Ā-opacity effect, a possible line of analysis is to attribute the opacity to the ban on improper movement, excluding the derivational sequence of the Ā-A movement (Chomsky, 1973; May, 1979; Chomsky, 1981; Abels, 2007; Neeleman & Van De Koot, 2010; Williams, 2011). Pursuing this line of analysis goes beyond the scope of this paper, so we leave it to future studies.[31]

6. Conclusions

The hybrid A/Ā properties of English locative inversion present a notable analytical challenge within the generative framework. For instance, Bresnan (1994) argues that English locative inversion cannot be adequately derived within Chomsky's generative framework, which posits that all human languages utilize universal syntactic operations on underlying structures (see also Postal, 2004). Two major analytical challenges posed by English locative inversion are (i) why the fronting locative phrase does not behave as a canonical subject (e.g., lack of agreement; see Bruening, 2010 for further discussion), and (ii) how the movement of the locative phrase, if analyzed as an instance of A-movement targeting Spec,TP, is able to move across the structurally higher theme DP, which canonically does so. To address these puzzles, some researchers have proposed a null expletive analysis, suggesting that Spec,TP is occupied by a null expletive, with the locative phrase undergoing Ā-topicalization (e.g., Postal, 1977, 2004; Maruta, 1985; Coopmans, 1989; Hoekstra & Mulder, 1990; Bruening, 2010).

In this paper, we demonstrate that the null expletive analysis cannot be extended to account for the derivation of LC in Vietnamese. Specifically, the fronting locative argument in LC resists marking by *thì*, representing the Topic head in the left periphery (see (25a)). Moreover, the postulation of a null expletive in Vietnamese is implausible in light of the incompatibility of LC and the expletive *nó* (see (45)). After establishing that the null expletive analysis is unsuitable for LC in Vietnamese (see Diercks, 2017 for cross-linguistic considerations), we explore parallels between the fronting locative argument and the expletive *nó* with regard to their shared specificity requirement (see Section 4.2). These findings lead us to a dilemma: the fronting locative argument in LC is not a topic in the left periphery, yet it patterns with the expletive *nó* regarding the specificity requirement, a property that is aligned with Ā-topicalization. We suggest a solution based on three assumptions. First, we adopt Bošković's (2024a) proposal of three distinct subject positions between TP and CP across languages and argue that both the fronting locative argument and *nó* target the specifier position of the A/ĀP, the uppermost of these subject positions. Second, to clarify the derivational process of locative inversion in LC, we adopt the featural characterization of the A/Ā-distinction (Obata & Epstein, 2011; Van Urk, 2015; Fong, 2019) and assume a composite probe [D, δ] on the head of the A/ĀP. Finally, we show that Coon and Bale's (2014)

formal calculus of minimality enables the composite probe to target and move the locative argument to Spec,A/ĀP over the structurally higher theme DP, achieving the desired derivation.

Author Contributions: Conceptualization, C.-T.T.C.; Validation, T.P.; Formal analysis, C.-T.T.C.; Investigation, C.-T.T.C. and T.P.; Resources, T.P.; Data curation, T.P.; Writing—original draft, T.P. and C.-T.T.C.; Writing—review & editing, T.P. and C.-T.T.C.; Supervision, C.-T.T.C.; Project administration, C.-T.T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We are very grateful to the anonymous reviewers for their constructive comments, which greatly improved our work. An earlier version of this paper has been presented at the 2024 Workshop on Vietnamese Linguistics (ISVL-4). We would like to thank the audience for their questions and suggestions. All errors and shortcomings are the sole responsibility of the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 The abbreviations used in this paper include: 1: class 1 subject agreement in Zulu; ANT: anterior marker; AUG: augment vowel; ACC: accusative case; C: complementizer; CONJ: conjunction; CL: classifier; EXPL: expletive; FUT: future tense marker; NEG: negation marker; PL: plural marker; POSS: possessive marker; PROG; progressive marker; TOP: topic marker.

2 This is aligned with Bresnan's (1994) claim that locative inversion constructions featuring posture verbs generally express a simple position, i.e., a physical disposition of a stationary figure (the referent of the post-verbal DP) with respect to a ground (as denoted by the locative phrase), or a maintain position which requires some sort of deliberate effort by an animate being to maintain the posture (see also Levin & Rappaport Hovav, 1995). In contrast, the assume position sense, which "describes an animate being coming to be in a particular position under his or her own control" (Levin & Rappaport Hovav, 1995, p. 127), is absent from locative constructions.

3 https://vietmessenger.com/books/?title=buong%20theo%20giot%20buon&page=11 (accessed on 3 November 2024).

4 https://tienphong.vn/ao-phao-lam-canh-post581205.tpo (accessed on 3 November 2024).

5 Other potential candidates for EC are (eventive) verbs of (dis-)appearance (cf. Nguyễn, 1997) such as *xuất hiện* 'appear', *xảy ra* 'happen', *mất* 'lose, disappear', other unaccusative verbs such as *rơi* 'fall' or *tan* 'melt' (also in their stative sense), and non-voluntary verbs of appearance, sound, smell such as *lấp lánh* 'sparkle'. As noted by Levin and Rappaport Hovav (1995), such verbs must also be "informationally light" in the context, typically not communicating anything that goes beyond a simple existence or appearance on the scene of the following DP (Landau, 2009).

6 https://www.vietnamplus.vn/xuat-hien-bang-tuyet-tai-huyen-binh-lieu-quang-ninh-post368147.vnp (accessed on 3 November 2024).

7 The grammatical occurrence of the locative phrase in the post-verbal position in (6) indicates that the fronting locative phrase in LC originates from within vP rather than being base-generated at the surface position (see Paul et al., 2020, pp. 23–24 on why Chinese locative construction does not involve the movement of the locative phrase).

8 The correct V-DP word order is derived after V-to-v raising. In addition, we assume that the postverbal DP in LCs is base-generated within VP as in (10). If the DP were base-generated at Spec,vP, the locative verb would have to move out of vP to derive the correct word order. However, data in (i–iii) show that the locative verb follows aspectual/temporal markers and negation outside the vP, indicating that the locative verb stays within vP.

(i)	Trên	bàn	**đang**	để	một	cuốn	sách	[aspect > V]	
	on	table	PROG	place	a	CL	book		
	'On the table is placing a book.'								
(ii)	Trong	phòng	**từng**	nằm	một	bệnh nhân	COVID	[temporal/aspectual > V]	
	in	room	used.to	lie	a	patient	COVID		
	'In the room used to lie a COVID patient.'								
(iii)	Trên	tường	**không**	treo	một	bức	tranh	nào	[negation >V]
	on	wall	not	hang	a	CL	painting	any	
	'On the wall does not hang any painting.'								

9 We assume, following Dư (2005), that EC without locative phrases such as (7) still have to be semantically tied to a locative/temporal frame (see also Greco et al., 2018). The apparent absence of such overt locative/temporal information is permitted only if it is retrievable from context.

10 An alternative to circumvent the locality issue under (18a) is to assume that the theme DP and the locative argument are equidistant from Spec,TP because they are within the same minimal domain, i.e., within the projection of the same head, e.g., within a small-clause/PredP (e.g., Hoekstra & Mulder, 1990; den Dikken, 2006; Broekhuis, 2008; Hartmann, 2008). However, Doggett (2004) argues that the concept of equidistance yields incorrect predictions across various syntactic constructions, including passivization in applicatives, movement to multiple specifiers, and Agree operations involving multiple specifiers. Similarly, Carstens and Diercks (2013), drawing on the morphosyntactic behavior of the agreeing wh-phrase *rieena* ('how') in the Bantu language Lubukusu, assert that equidistance fails to provide an empirically adequate resolution to the locality problem. Adjudicating the validity of equidistance and minimal domains as strategies to address apparent locality violations lies beyond the scope of this paper. Consequently, we do not evaluate whether the equidistance-based approach provides an adequate account of the locality issue in the derivation of LCs. Nevertheless, even if equidistance and minimal domains are ultimately validated as conditions on locality, the question of what motivates the movement of the locative phrase rather than the theme DP, as discussed later in the text, persists.

11 The locative adjunct in EC may also be adjoined vP-internally, as shown in (8).

12 (24b) is marked with "??" instead of "*" to indicate that it sounds degraded (and not necessarily bad) to some native speakers we consulted.

13 Similarly, there can be no prosodic break between the locative phrase and the succeeding verb-DP sequence, which form a single thetic judgment (Kuroda, 1972; see also Ladusaw, 1994 for the distinction between thetic sentences and categorical ones). That is, an LC conveys "the recognition of an event in one single act" (Paul et al., 2020, p. 25) and does not constitute a typical bipartite topic-comment configuration.

14 One may maintain that a *thì*-marked topic expression is base-generated in Vietnamese left-periphery. We do not dispute the fact that topics in Vietnamese may be base-generated in their position. This is especially true for cases of topics that have an "aboutness" relation with the comment (i.e., Aboutness topics). However, we believe that Vietnamese *thì*-marked topics may also be formed by movement. Note that the fronting of locative adjuncts in ECs is not clause-bound as in (i). Importantly, the presence of island effects shown in (ii-v) (see Trinh, 2009) indicates that the fronting of the *thì*-marked locative adjuncts bears characteristics of Ā-movement. If all *thì*-marked topics in Vietnamese were uniformly base-generated in their surface position, it is unclear how we can account for the island sensitivity of the fronting *thì*-marked locative adjunct.

(i) [Trên bàn]$_1$ thì tôi tin [là có một chồng sách t$_1$].
on table TOP I believe C exist a pile book
'On the table, I believe that there's a pile of books.'

(ii) *[Trên bàn]$_1$ thì tôi tin [chuyện có một chồng sách t$_1$]. (Complex NP island)
on table TOP I believe story exist a pile book
Intended: 'On the table, I believe the story that there's a pile of books.'

(iii) *[Trên bàn]$_1$ thì tôi tin [có một chồng sách t$_1$] là tốt. (Subject island)
on table TOP I believe exist a pile book be good
Intended: 'On the table, I believe that there's a pile of books is good.'

(iv) *[Trên bàn]$_1$ thì tôi vui [vì có một chồng sách t$_1$]. (Adjunct island)
on table TOP I happy because exist a pile book
Intended: 'On the table, I am happy because there's a pile of books.'

(v) *[Trên bàn]$_1$ thì tôi tiếc [là có một chồng sách t$_1$]. (Factive island)
on table TOP I regret C exist a pile book
Intended: 'On the table, I regret that there's a pile of books.'

15 Lee and Yip (2024) also use *cảm giác* 'feel' as another raising attitude verb to illustrate hyperraising in Vietnamese. However, we do not use this verb when discussing hyper-raising of the locative argument in (28) since it is hard to imagine how one could "feel" that "on the bed lies a patient".

16 EC is also allowed in the complement of the raising attitudinal verb *nghe nói* 'hear', as exemplified by (i). The locative adjunct is able to move out of the complement clause as in (ii). Since the locative adjunct can be followed by *thì*, the marker of the Topic head in the left periphery, we do not treat the movement of the locative adjunct in (ii) as an instance of hyperraising.

(i) Nghe nói [CP rằng/là [trên giường]$_1$ có một bệnh nhân].
Hear C on bed exist one patient
'It is heard that on the bed is a patient.'

(ii) [Trên giường]$_1$ (thì) nghe nói [CP rằng/là t$_1$ có một bệnh nhân].
on bed TOP hear C have one patient
'It is heard that on the bed there is a patient.'

17 The hyperraising of the locative argument also suggests that the fronting of the locative argument in LC cannot be analyzed as moving to a criterial position in the sense of Rizzi (2017). This is because the movement to a criterial position is subject to the criteria freezing effects, which would wrongly predict that the fronting locative argument in LC cannot undergo hyper-raising out of the CP complement in (28b).

18 That the intended coreference in (i) is not acceptable indicates that Ā-topicalization in Vietnamese (as overtly marked by the topic marker *thì*) displays WCO on a par with English topicalization in (ii).

(i) *[Trên giường của mỗi đứa trẻ$_2$]$_1$ thì đều có [sách của nó$_2$] t$_1$.
on bed POSS each CL child TOP all exist book POSS he
Intended: 'On every child$_2$'s bed there are his$_2$ books.'

(ii) *[Into every dog$_2$'s cage]$_1$ its$_2$ owner peered t$_1$. (Culicover & Levine, 2001, p. 289, (14a))

19 In addition to the lack of WCO and the resistance of marking by *thì*, the findings enumerated in (i) lend further credence to the A-properties of the fronting locative phrase in an LC.

(i) a. The locative phrase in an LC is obligatory.
b. The locative phrase in an LC must be below the complementizer.
c. The locative phrase in an LC can undergo hyperraising.

20 See Von Heusinger (2002) for a distinction between definiteness and specificity.

21 AgrsP and TP are split from the traditional IP under early minimalist assumptions.

22 We remain agnostic about the necessity of postulating an AgrsP in Vietnamese, a language without morphological markings of φ-agreement.

23 Given the syntactic position of the expletive within the high-TP domain assumed by Greco et al. (2017, 2018), the contrast between (46a) and (46b) strengthens our analysis of the fronting of the locative adjunct in EC as a case of topicalization to the left periphery.

[24] To quote Bresnan (1994, p. 90), "in presentational focus, a scene is set and a referent is introduced on the scene to become the new focus of attention. In the core cases, a scene is naturally expressed as a location, and the referent as something of which location is predicated-hence, a theme." Presentational focus can be employed to either introduce a new referent or reintroduce a previously mentioned referent into the domain of discourse (Birner, 1992; Bresnan, 1994).

[25] A reviewer asks whether *nó* patterns with the locative argument in LCs with respect to its position relative to complementizers and the possibility of hyperraising. Data in (i–ii) below confirm that *nó* patterns with the locative argument in LCs in terms of its position relative to the complementizer *là.* However, *nó* resists hyper-raising, as shown in (iii). Following the reviewer's suggestion, hyper-raising of an expletive might be ruled out for independent reasons. After all, the locative phrase in LCs is an argument of the locative predicate, while the expletive *nó* is non-argumental. The impossibility of hyperraising for *nó* might be related to its non-argumental nature. We will leave the ungrammaticality of (iii) for future research.

(i) Rõ ràng là nó ngã thằng bé.
clearly C EXPL fall CL little
'Clearly that the boy fell.'

(ii) ??Nó rõ ràng là ngã thằng bé.
EXPL clearly C fall CL little
Intended: 'Clearly that the boy fell.'

(iii) *Nó nghe nói là ngã thằng bé.
EXPL hear C fall CL little

[26] The three possible positions for *nó* mentioned in Greco et al.'s work all appear hybrid in their original characterization. Rizzi and Shlonsky's (2006, 2007) SubjP is treated as a specialized criterial projection, which serves the discourse function of encoding the "subject of predication" of a clause. The characterization of SubjP as "the A counterpart of Ā criterial positions in the left periphery" (Shlonsky & Rizzi, 2018) alludes to its hybrid nature. Rizzi & Shlonsky's conception of SubjP goes back to Cardinaletti's (2004) idea of a SubjP endowed with a subject-of-predication feature; the subject in Spec,SubjP is distinct from the grammatical subject, which is lower in the clausal spine. Similar to SubjP, Kiss's (1996) RefP is also assumed to merge between IP and CP, hosting specific subjects (in contrast, non-specific subjects stay in Spec,TP). Lund and Charnavel (2020) go as far as arguing that RefP is part of the lower end of CP, marking a topic position. Unlike TP, SubjP/RefP is not a purely A position. It is this gray area between the traditional CP and TP in which Bošković's (2024a, 2024b) hybrid A/ĀP is assumed to be located. An additional empirical argument against treating *nó* as occupying Spec,TP concerns its word order in relation to the high (negative) polarity marker *đâu*. If *nó* merges in Spec,TP, we should expect it to occur linearly to the right of the high (negative) polarity marker *đâu*, given that *đâu* merges immediately above TP in Vietnamese (Phan & Tsai, 2025). This is contrary to fact, as shown in (i).

(i) {Nó} đâu {*nó} có ma.
EXPL POL EXPL exist ghost
'It's untrue that there are ghosts.'

[27] The distinction between flat and relativized probes is originally inspired by patterns observed in omnivorous agreement in languages such as Nez Perce (Deal, 2014) and Kaqchikel (Preminger, 2014). These patterns suggest that various φ-features (such as person, number, and gender) can either combine into a flat, non-selective probe that is satisfied by any φ-feature bundle regardless of specific values, or they can be relativized to target specific features (e.g., [Participant] or [Plural]). Furthermore, the presence of a δ-feature associated with the fronting of the locative argument in LCs can be further evidenced by the infelicity of an LC in an out-of-blue context in (i). In particular, in (i), unless the table was mentioned in the previous discourse, B's response with a locative phrase fronted sounds odd. As such, the fronting of the locative phrase must involve the presence of a δ-feature.

(i) Context: B gasped upon stepping into the hotel room they just checked into. Standing just behind B, A asked
A: Có chuyệngì SFP
exist matter what SFP
'What happened?'
B: #Trên bàn nằm một con mèo đen.
on table lie a CL cat black
Intended: 'On the table lies a black cat.'

[28] We refer to the landing site of locative inversion in English as Spec,TP as in the original work we cite here. Under Bošković's (2024a) analysis of locative inversion, the locative phrase targets Spec,A/ĀP. We gloss over this difference here for expository convenience.

[29] The post-verbal theme DP cannot bear contrastive focus (indicated by a focus accent) nor be questioned in LCs, as demonstrated in (i–ii). This infelicity may be attributed to the communicative purpose of LCs. Specifically, the referent of the theme DP is intended to be simply introduced into the discourse through its locational relation to a more discourse-familiar location. As such, the theme DP in LCs is not meant to be contrasted with alternatives or subjected to questioning.

(i) *Trên bàn nằm một chồng sáchF (, không phải một chồng tạp chíF).
on table lie a pile book not correct a pile magazine
Intended: 'On the table lies a pile of book (, not a pile of magazines).'

(ii) *Trên bàn nằm hai cuốn sách nào?
on table lie two CL book which
Intended: 'On the table lie which two books?'

[30] The contrast between (62b) and (62c) turns into supporting evidence for the checking/valuation-based approach to A-movement, argued by Epstein and Seely (2006) and Bošković (2002, 2007). Since the infinitival T is not able to check/value the unvalued Case feature of a DP, the movement of *a man* to embedded Spec,TP, as in (62c) simply cannot apply.

[31] Obata and Epstein (2011) advance a feature-splitting analysis of Ā-opacity effects. In short, they argue that a DP does not carry its φ-features and valued Case feature when it undergoes Ā-movement because these features are not the goal of the Ā-probe triggering the Ā-movement. Thus, an Ā-moved DP is not visible to an A-probe because it does not carry an A-feature. Although well-motivated by theoretical considerations under Chomsky's (2007, 2008) feature-inheritance system, we maintain that this approach cannot be extended to capture Ā-opacity in (67) since locative phrases in Vietnamese are assumed to contain a (defective) D-feature in this paper and it is well-known that internal topicalization in Chinese exhibits A-properties (see Shyu, 1995; Ting, 1995; Paul, 2005; Kuo, 2009; Chen, 2023).

References

Abels, K. (2007). Towards a restrictive theory of (remnant) movement! *Linguistic Variation Yearbook*, *7*, 53–120. [CrossRef]

Abels, K. (2012). The Italian left periphery: A view from locality. *Linguistic Inquiry*, *43*(2), 229–254. [CrossRef]

Baker, M. (1997). Thematic roles and syntactic structure. In L. Haegeman (Ed.), *Elements of grammar* (pp. 73–137). Kluwer.

Belletti, A., & Rizzi, L. (2017). On the syntax and pragmatics of some clause-peripheral positions. In J. Blochowiak, C. Grisot, S. Durrleman, & C. Laenzlinger (Eds.), *Formal models in the study of language* (pp. 33–48). Springer.

Birner, B. (1992). *The discourse function of inversion in English* [Doctoral dissertation, Northwestern University].

Bolinger, D. (1973). Ambient it is meaningful too. *Journal of Linguistics*, *9*(2), 261–270. [CrossRef]

Bolinger, D. (1977). *Meaning and form*. Longman.

Bošković, Z. (2002). A–movement and the EPP. *Syntax*, *5*, 167–218. [CrossRef]

Bošković, Ž. (2007). On the locality and motivation of move and agree: An even more minimal theory. *Linguistic Inquiry*, *38*(4), 589–644. [CrossRef]

Bošković, Ž. (2024a). On wh and subject positions, the EPP, and contextuality of syntax. *The Linguistic Review*, *41*(1), 7–58. [CrossRef]

Bošković, Ž. (2024b). On (partially) quirky subjects, numeral subjects, and subject-oriented anaphor binding: Nominal and non-nominal subjects and their structural positions. In *Manuscript*. University of Connecticut.

Bresnan, J. (1994). Locative inversion and the architecture of universal grammar. *Language*, *70*(1), 72. [CrossRef]

Broekhuis, H. (2008). *Derivations and evaluations*. Mouton de Gruyter.

Bruening, B. (2010). Language-particular syntactic rules and constraints: English locative inversion and do-support. *Language*, *86*(1), 43–84. [CrossRef]

Cable, S. (2010). *The grammar of Q*. Oxford University Press.

Cardinaletti, A. (2004). Toward a cartography of subject positions. In L. Rizzi (Ed.), *The structure of CP and IP: The cartography of syntactic structures* (Vol. 2, pp. 115–165). Oxford University Press.

Carstens, V., & Diercks, M. (2013). Agreeing how? Implications for theories of agreement and locality. *Linguistic Inquiry*, *44*(2), 179–237. [CrossRef]

Castillo, J. C., Drury, J. E., & Grohmann, K. (2009). Merge over move and the extended projection principle: MOM and the EPP revisited. *Iberia: An International Journal of Theoretical Linguistics*, *1*, 53–114.

Chen, F. (2023). *Obscured universality in Mandarin* [Doctoral dissertation, Massachusetts Institute of Technology].

Chomsky, N. (1973). Conditions on transformations. In S. Anderson, & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 232–286). Holt, Rinehart and Winston.

Chomsky, N. (1981). *Lectures on government and binding*. Foris.

Chomsky, N. (1995). *The minimalism program*. MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays in minimalist syntax in honor of Howard Lasnik* (pp. 89–155). The MIT Press.

Chomsky, N. (2001). Derivation by phase. In K. Michael (Ed.), *Ken hale: A life in language* (pp. 1–52). MIT Press.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland, & H.-M. Gärtner (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics* (pp. 1–29). Mouton de Gruyter.

Chomsky, N. (2008). On phases. In R. Freidin, C. P. Otero, & M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud* (pp. 133–166). MIT Press.

Chomsky, N. (2019). *The UCLA lectures* [Unpublished manuscript]. Universit of Arizona.

Coon, J., & Bale, A. (2014). The interaction of person and number in Mi'gmaq. *Nordlyd*, *41*(1), 85–101. [CrossRef] [PubMed]

Coopmans, P. (1989). Where stylistic and syntactic processes meet: Locative inversion in English. *Language*, *65*(4), 728. [CrossRef]

Cresti, D. (1995). *Indefinite topics* [Doctoral dissertation, Massachusetts Institute of Technology].

Culicover, P., & Levine, R. (2001). Stylistic inversion in English: A reconsideration. *Natural Language and Linguistic Theory*, *19*, 283–310. [CrossRef]

Deal, A. R. (2014). *Properties of probes: Evidence from Nez Perce complementizer agreement*. NELS 45. MIT.

Deal, A. R. (2017). Covert hyperraising to object. In A. Lamont, & K. Tetzlof (Eds.), *Proceedings of the 47th annual meeting of the North East Linguistic Society (NELS 47)* (Vol. 3, pp. 257–270). University of Massachusetts.

den Dikken, M. (2006). *Relators and linkers*. MIT Press.

Déprez, V. M. (1989). *On the typology of syntactic positions and the nature of chains* [Doctoral dissertation, Massachusetts Institute of Technology].

Diercks, M. (2017). Locative inversion. In M. Everaert, & H. C. Riemsdijk (Eds.), *The Wiley Blackwell companion to syntax* (2nd ed., pp. 1–30). Wiley.

Doggett, T. B. (2004). *All things being unequal: Locality in movement* [Doctoral dissertation, Massachusetts Institute of Technology].

Dư, N. N. (2005). Bàn thêm về câu tồn tại trong tiếng Việt [More on existential sentences in Vietnamese]. *Ho Chi Minh: Tạp chí Khoa học Trường Đại học Sư phạm TP Hồ Chí Minh*, *5*, 55–65.

Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry*, *22*, 1–26.

Epstein, S. D., & Seely, T. D. (1999). *SPEC-ifying the GF "subject"; eliminating A-chains and the EPP within a derivational model* [Unpublished manuscript]. University of Michigan and East Michigan University.

Epstein, S. D., & Seely, T. D. (2006). *Derivations in minimalism*. Cambridge University Press.

Erteschik-Shir, N. (1997). *The Dynamics of focus structure*. Cambridge University Press.

Ferreira, M. (2000). *Argumentos nulos em português brasileiro* [Master's thesis, Department of Linguistics, University Estadual Campinas].

Ferreira, M. (2004). Hyperraising and null subjects in Brazilian Portuguese. In *MIT working papers in linguistics, Vol. 47: Collected papers on romance syntax* (pp. 57–85). MITWPL.

Ferreira, M. (2009). Null subjects and finite control in Brazilian Portuguese. In J. Amsterdam (Ed.), *Minimalist essays on Brazilian Portuguese syntax* (pp. 17–50). Benjamins.

Fodor, J. D., & Sag, I. A. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, *5*(3), 355–398. [CrossRef]

Fong, S. (2019). Proper movement through Spec-CP: An argument from hyperraising in Mongolian. *Glossa: A Journal of General Linguistics*, *4*(1), 30. [CrossRef]

Frampton, J., & Gutmann, S. (2000). *Agreement is feature sharing* [Unpublished manuscript]. Northeastern University.

Georgi, D. (2014). *Opaque interactions of Merge and Agree. On the nature and order of elementary operations* [Doctoral dissertation, Leipzig University].

Greco, C., Haegeman, L., & Phan, T. (2017). Expletives and speaker-related meaning. In M. Sheehan, & L. R. Bailey (Eds.), *Order and structure in syntax II: Subjecthood and argument structure* (pp. 69–93). Language Science Press.

Greco, C., Phan, T., & Haegeman, L. (2018). On nó as an optional expletive in Vietnamese. In F. Cognola, & J. Casalicchio (Eds.), *Null subjects in generative grammar: A synchronic and diachronic perspective* (pp. 31–51). Oxford University Press.

Grimshaw, J. (1991). *Extended projections* [Unpublished manuscript]. Brandeis University.

Halpert, C. (2019). Raising, unphased. *Natural Language & Linguistic Theory*, *37*, 123–165.

Hartmann, J. M. (2008). *Expletives in Existentials. English there and German da*. LOT.

Hoekstra, T., & Mulder, R. (1990). Unergatives as copular verbs; locational and existential predication. *The Linguistic Review*, *7*(1), 1–80. [CrossRef]

Huang, C. T. J., Li, Y. H. A., & Li, Y. (2009). *The syntax of Chinese*. Cambridge University Press.

Jayaseelan, K. A. (2001). IP-internal topic and focus phrases. *Studia Linguistica*, *55*(1), 39–75. [CrossRef]

Kiss, K. É. (1996). Two subject positions in English. *The Linguistic Review*, *13*(2), 119–142. [CrossRef]

Kuo, P.-J. (2009). *IP internal movement and topicalization* [Ph.D. dissertation, University of Connecticut].

Kuroda, S.-Y. (1972). The categorical and the thetic judgments. *Foundations of Language*, *9*, 153–185.

Ladusaw, W. (1994). Thetic and categorical, stage and individual, weak and strong. In M. Harvey, & L. Santelmann (Eds.), *Proceedings of SALT IV* (pp. 220–229). Cornell University, Department of Modern Languages & Linguistics.

Landau, I. (2009). *The locative syntax of experiencers*. MIT Press. [CrossRef]

Lee, T. T., & Yip, K. (2024). Hyperraising, evidentiality, and phase deactivation. *Natural Language & Linguistic Theory*, *42*(4), 1527–1578. [CrossRef]

Legate, J. A. (2003). Some interface properties of the phase. *Linguistic Inquiry*, *34*(3), 506–515. [CrossRef]

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. MIT Press.

Lund, G., & Charnavel, I. (2020). The syntax of concessive clauses: Evidence from exempt anaphora. *University of Pennsylvania Working Papers in Linguistics*, *26*(1), 159–168.

Mahajan, A. (1990). *The A/Ā-distinction and movement theory* [Doctoral dissertation, Massachusetts Institute of Technology].

Martins, A. M., & Nunes, J. (2005). Raising issues in Brazilian and European Portuguese. *Journal of Portuguese Linguistics*, *4*(2), 53–77. [CrossRef]

Martins, A. M., & Nunes, J. (2009). Syntactic change as chain reaction: The emergence of hyper-raising in Brazilian Portuguese. In P. Crisma, & G. Longobardi (Eds.), *Historical syntax and linguistic theory* (pp. 144–157). Oxford University Press.

Martins, A. M., & Nunes, J. (2010). Apparent hyper-raising in Brazilian Portuguese: Agreement with topics across a finite CP. In E. P. Panagiotidis (Ed.), *The complementizer phase: Subjects and operators* (pp. 143–63). Oxford University Press.

Maruta, T. (1985). Is stylistic inversion stylistic? *Cahiers Linguistiques d'Ottawa*, *14*, 111–120.

May, R. (1979). Must COMP-to-COMP movement be stipulated? *Linguistic Inquiry*, *10*, 719–725.

Milosavljević, A., & Milosavljević, S. (2022). The intensifying accusative clitic ga 'it' in Serbian: From syntax to pragmatics. *Proceedings of ConSOLE XXX*, 63–88.

Miyagawa, S. (2010). *Why agree? Why move?* MIT Press.

Miyagawa, S. (2017). *Agreement beyond Phi*. MIT Press.

Neeleman, A., & Van De Koot, H. (2010). A local encoding of syntactic dependencies and its consequences for the theory of movement. *Syntax*, *13*(4), 331–372. [CrossRef]

Nguyễn, K. T. (1997). *Nghiên cứu ngữ pháp tiếng Việt*. Giáo dục.

Nunes, J. (2008). Inherent case as a licensing condition for A-movement: The case of hyper-raising constructions in Brazilian Portuguese. *Journal of Portuguese Linguistics*, *7*(2), 83–108. [CrossRef]

Obata, M., & Epstein, S. D. (2011). Feature-splitting internal merge: Improper movement, intervention, and the A/A'-distinction. *Syntax*, *14*(2), 122–147. [CrossRef]

Paul, W. (2005). Low IP area left periphery in Mandarin Chinese. *Recherches Linguistiques de Vincennes*, *33*, 111–134. [CrossRef]

Paul, W., Lu, Y., & Lee, T. H. (2020). Existential and locative constructions in Mandarin Chinese. *The Linguistic Review*, *37*(2), 231–267. [CrossRef]

Pesetsky, D. 1995. *Zero syntax*. MIT Press.

Phan, T. (2024). *Where two ends meet: Non-at-issue meanings on the syntactic treetops of Vietnamese* [Doctoral dissertation, National Tsing Hua University].

Phan, T., & Chierchia, G. (2022). Identifying (in)definiteness in the Vietnamese noun phrase. *Journal of the Southeast Asian Linguistics Society*, *15*(2), 27–49.

Phan, T., & Lander, E. (2015). Vietnamese and the NP/DP parameter. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *60*(3), 391–415. [CrossRef]

Phan, T., & Tsai, W. T. D. (2025). Traversing the "where" landscape: Vietnamese *đâu* and the clausal architecture of Vietnamese. *Journal of East Asian Linguistics*. [CrossRef]

Portner, P. (2002). Topicality and (non-)specificity in Mandarin. *Journal of Semantics*, *19*(3), 275–287. [CrossRef]

Postal, P. (1977). About a nonargument for raising. *Linguistic Inquiry*, *8*, 141–154.

Postal, P. (2004). A paradox in English syntax. In *Skeptical linguistic essays* (pp. 15–82). Oxford University Press.

Preminger, O. (2014). *Agreement and its failures*. MIT Press. [CrossRef]

Preminger, O. (2019). What the PCC tells us about "abstract" agreement, head movement, and locality. *Glossa: A Journal of General Linguistics*, *4*(1), 13. [CrossRef]

Rezac, M. (2003). The fine structure of cyclic agree. *Syntax*, *6*(2), 156–182. [CrossRef]

Rizzi, L. (1990). *Relativized minimality*. MIT Press.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar. Kluwer international handbooks of linguistics* (pp. 281–337). Springer.

Rizzi, L. (2004). Locality and left periphery. In A. Belletti (Ed.), *Structures and beyond* (vol. 3, pp. 223–251). Oxford University Press.

Rizzi, L. (2017). Types of criterial freezing. *Rivista di Grammatica*, *39*, 1–43.

Rizzi, L., & Shlonsky, U. (2006). Satisfying the subject criterion by a non subject: English Locative Inversion and Heavy NP Shift. In M. Frascarelli (Ed.), *Phases of interpretation* (pp. 341–361). Mouton de Gruyter.

Rizzi, L., & Shlonsky, U. (2007). Strategies of subject extraction. In H. M. Gärtner, & U. Sauerland (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics* (pp. 115–160). Mouton de Gruyter.

Rodrigues, C. A. N. (2004). *Impoverished morphology and a-movement out of case domains* [Ph.D. dissertation, University of Marland].

Shlonsky, U., & Rizzi, L. (2018). Criterial freezing in small clauses and the cartography of copular constructions. In J. Hartmann, M. Jäger, A. Kehl, A. Konietzko, & S. Winkler (Eds.), *Freezing: Theoretical approaches and empirical domains* (pp. 29–65). Mouton de Gruyter.

Shyu, S.-I. (1995). *The syntax of focus and topic* [Ph.D. dissertation, University of Southern California].

Svenonius, P. (2000). Quantifier movement in Icelandic. In P. Svenonius (Ed.), *The derivation of VO and OV* (pp. 255–292). John Benjamins.

Ting, J. (1995). Deriving the secondary topic in Mandarin Chinese. In T.-F. Cheng, Y. Li, & H. Zhang (Eds.), *Proceedings of the 7th North America conference on Chinese linguistics (NACCL-7), volume 1: Pragmatics, syntax and semantics*. University of Southern California: GSIL.

Trinh, T. (2009). A constraint on copy deletion. *Theoretical Linguistics*, *35*(2/3), 183–227. [CrossRef]

Ura, H. (1994). Varieties of raising and the feature-based theory of movement. In *MIT occasional papers in linguistics 7*. MIT.

van Riemsdijk, H. (1990). Functional prepositions. In H. Pinkster, & I. Genee (Eds.), *Unity in diversity, papers presented to Simon C. Dik on his 50th birthday* (pp. 229–241). De Gruyter Mouton.

van Riemsdijk, H. (1998). Categorial feature magnetism: The endocentricity and distribution of projections. *Journal of Comparative Germanic Linguistics*, *2*, 1–48. [CrossRef]

Van Urk, C. (2015). *A uniform syntax for phrasal movement: A case study of Dinka Bor* [Doctoral dissertation, Massachusetts Institute of Technology].

Von Heusinger, K. (2002). Specificity and definiteness in sentence and discourse structure. *Journal of Semantics*, *19*(3), 245–274. [CrossRef]

Williams, E. (2011). *Regimes of derivation in syntax and morphology*. Taylor & Francis. [CrossRef]

Wu, H.-H. I. (2008). *Generalized inversion and the theory of agree* [Doctoral dissertation, Massachusetts Institute of Technology].

Zyman, E. (2023). Raising out of finite clauses (Hyperraising). *Annual Review of Linguistics*, *9*(1), 29–48. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

Whys and Wherefores: The Aetiology of the Left Periphery (With Reference to Vietnamese)

Nigel Duffield

Department of English Literature and Language, Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan; duffield@konan-u.ac.jp

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 in the cartographic literature. Whilst largely consistent with the Italian pattern, our findings diverge in certain respects, especially in suggesting a modification of previous treatments of the *'*why*-to' constraint observed in English and a number of other varieties.

Keywords: Vietnamese syntax; 'left periphery'; complementation; embedded topicalization; '*Wh*-questions'; '*Why*'-questions; scope and constituency; cartographic approaches

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! . . .

To the child, enamoured of prints and maps
The universe has the size of his vast appetite.
How large the world seems by the light of a lamp!
How small it is, now, in memory's sight!

Charles Baudelaire, Le Voyage [first stanza] (*Les Fleurs du Mal*, 1861)

1. 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 focusing 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).

Let us 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)[1] periphery? Indeed, *Why* is CP considered 'peripheral' at all—part of another territory in the satirical map of *The Derivation* in Figure 1, posted more than a decade ago?



Figure 1. Cartography of the Derivation (© Zac Smith: used with permission.).

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, 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 do not 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, 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 in Figure 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).)[2]

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):[3] 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 Note 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.[4]

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 ($v\mathrm{P}^{\mathrm{max}}$, cf. "Complete Functional Complex" (Chomsky, 1985)). 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 and Pierce (1993), Guilfoyle and Noonan (1992), though cf. Chung and 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 as integral—*within*, not *without*. Compare the location of 'The United States of Tense and Aspect' in Figure 1 above.[5]

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, but at least 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 suggest that borders need 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 focused 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 and 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 and 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.[6]

(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

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

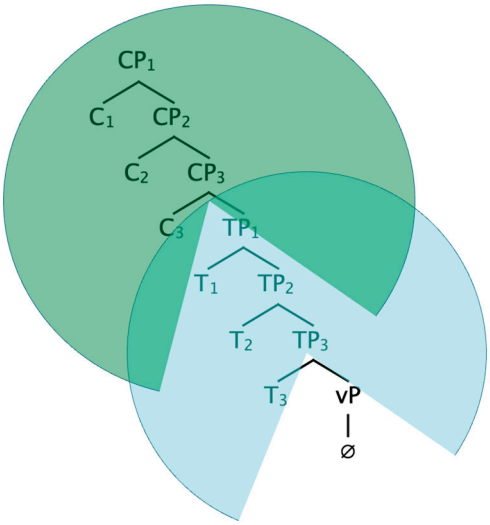
(2) $\text{Mood}_{\text{speech act}}$ > $\text{Mood}_{\text{evaluative}}$ > $\text{Mood}_{\text{evidential}}$ > $\text{Mod}_{\text{epistemic}}$ > T (Past) > T (Future) > $\text{Mood}_{\text{(ir)realis}}$ > Mod_{root}/$\text{Aspect}_{\text{habitual}}$/T (Anterior) > $\text{Aspect}_{\text{perfect}}$ > $\text{Aspect}_{\text{progressive}}$/ $\text{Aspect}_{\text{completive}}$ > 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*?[7] The problem is that cartographers—with the notable exception of Rizzi (2017) to which I will turn directly—are insufficiently concerned with what might be termed as 'mid-level' *Why* questions, even once the map is drawn.[8]

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,

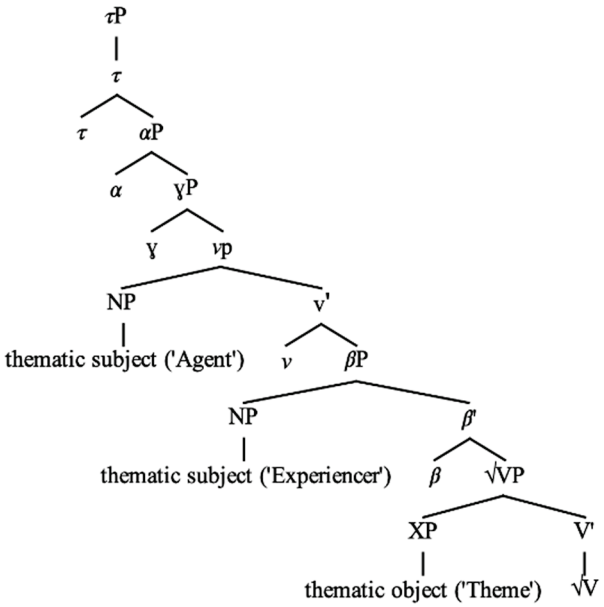
TP_2, TP_3, }, rather than in inverse, or intercalated, orders {TP_1 TP_2, TP_3. . .CP_1, CP_2, CP_3, }, {CP_1 TP_2, CP_3. . .TP_1, CP_2, TP_3 }; *why* thematic subjects always raise out of the predicate phrase, even in VSO languages (McCloskey, 1996, 2017); why T is supervenient on ASP; conversely, why T and Neg vary in their relative order cross linguistically (Ouhalla, 1991); *why* almost all lexical categories appear below almost all functional categories (an exception being "Inner Aspect" projections, see above; see also (Phan, 2013, 2024, for discussion of Inner Aspect in Vietnamese.).[9,10]

(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.[11] 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:[12]

> "Why is it that we typically find certain properties of ordering and cooccurrence restrictions, rather than others?. . .[. . .] Two broadly defined candidates come to mind:[13]

i. 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);

ii. 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 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* 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, 1993)).[14]

The alternative, functionalist proposal I have been promoting recently is cast in terms of four principles, two of which—*Supervenience* and *'1-Arg'*—offer a partial solution to the mid-level questions introduced above; see Duffield (2022). In this paper, I shall focus attention instead the core data question: "*Wherefore is* why *different?".* And before that, we have some mapping to do.[15]

2. Vietnamese Clause Structure

As may be familiar, Vietnamese is distinguished from other, better known, MSEA languages in having a rich inventory of pre-subject constituents, whose relative distributions provide evidence of an articulated left periphery.[16] While this super-structure is similar in many ways to that proposed by Rizzi and 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),*[17] *(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 note 1)}.

iii. Speaker/subject-oriented adverbial phrases (e.g., *quả thật*, 'indeed').

iv. Weak indefinites interpreted as universally quantified XPs (which in all probability have been moved to a pre-subject position, given the distribution of the same constituents in the general case);

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 that the overt phrasal constituents in (ii) occupy the 'specifier' position of their corresponding licensing heads (XP-**thì**, XP-**mà**, in the case of topicalization and relativization, respectively.)[18]

2.1. 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 $Q^{?}$ poss. go arrive party CAN NEG^{Q}
'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.'

2.1.1. 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 đàp: **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?'[19]

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

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^{Q}
'John asked if I wanted to go on a date with him.'

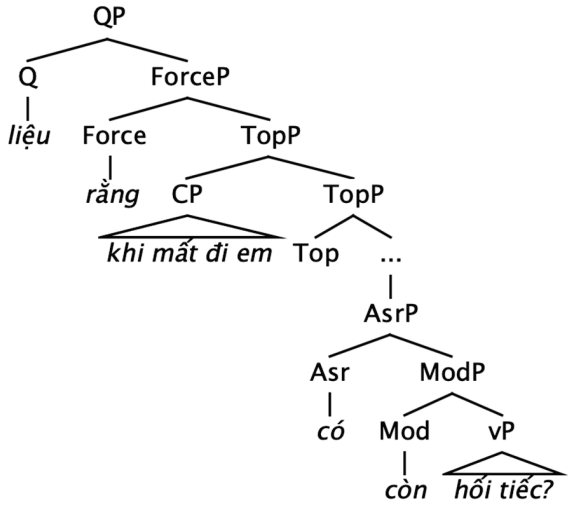
b. *John hỏi **liệu rằng** [tôi có muốn hẹn hò với anh ấy không]
John ask I Q want go.out with PRN.DEM NEG^{Q}
(as a)

c. John đã hỏi mẹ **rằng liệu** [mẹ có thể đón cậu bé không]
John ANT ask mother mother mother can child 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."'

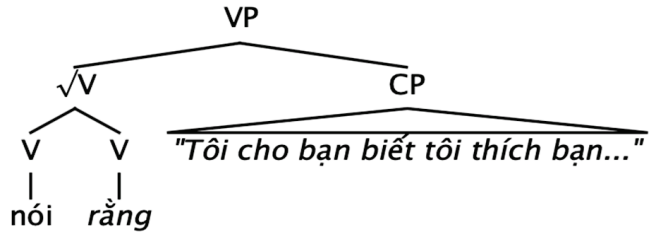
Given its verbal origins, it is reasonable to analyse contemporary examples where *nói rằng* introduces direct speech, and is immediately followed by a colon or comma—examples such as those in (10)—in terms of a similar matrix analysis.

(10) a. Murakami Haruki **nói** **rằng**: "Tôi "Tôi bạn biết tôi thích bạn . . ..."[20]
Murakami Haruku say ??: "I let friend know I like friend. . .
'Haruki Murakami said: "I'm telling you that I like you. . ."'

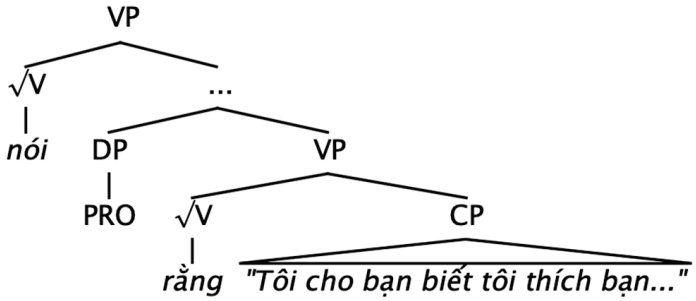
b. Tôi thấy lỗi **nói** **rằng**: "Một số quy tắc không được áp dụng nhất"
I see error say ??: 1 no. rule NEG CAN apply correct most
'I see an error saying "A number of rules are not most applicable"'[21]

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):

(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 KJV)"

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 factive predicates *tin* ('believe'), *nhận ra* ('realize'), *bực tức* ('resent'), and *tiếc* ('regret'), discussed in the section on embedded topicalization below.[22]

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.

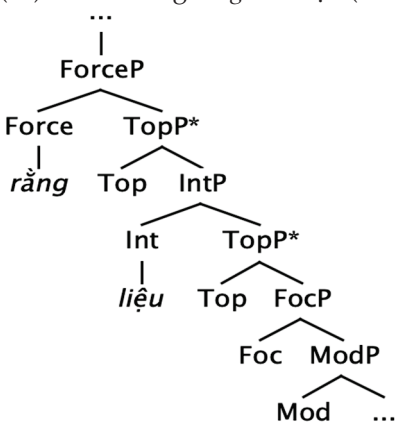
(13) a. Nó không nói gì, [**tuy rằng** [nó biết rất rõ]].
PRN NEG say what, [although PRN know very clear
'He didn't say anything, even though he clearly knows (the answer).'

b. [**Tuy rằng** [nhớ tên]] [nhưng lại không nhớ ra gương mặt]
although recall 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.

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)[23]



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 in (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 and Duffield (2022).[24]

(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 there COP John.
'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.'

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, since *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.[25]
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 một 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: *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, namely, {*rằng*. . .*là*}. (Example (19c) shows that {*rằng..là*} collocations contrast directly with {**mà..l!*)} sequences: *là* cannot occur within relative clauses, nor may it substitute for *mà* as a relativizing head.)

(19) a. Phải nói [(***là**) [rằng [(**là**) thế hệ trẻ của chúng.ta rất tài năng.]]][26]
must say ?? COMP ?? generation young POSS 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. . .]]][27]
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.[28] 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, the 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 C/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 C/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.)

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

d. Cô ấy không biết [rằng/***liệu** [tôi đang vui vẻ tại bữa tiệc]]
PRN.DEM NEG know C/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
PRN NEG know ?? I PROG happy at party
hay không.[29]
or NEG
'She doesn't know whether I am having fun at the party or not.'
(= it is unclear to her whether 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.

(23) a. ?Có **nên** *rằng/là đàn ông **là** trụ cột gia đình?[30]
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 death COP end final
'Is Death really the end?'

c. **Chắc** *rằng/là anh ấy bị tắc đường nên đến mượn.
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 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?[31]
man ASR MOD COP pillar family
'Should men be the family breadwinners?'

b. Cái chết ASR **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 mượn.
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 *rare* ('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 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 Italian *di* only appears with the Control use (of *sembrare* (25f)).[32]

(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.'

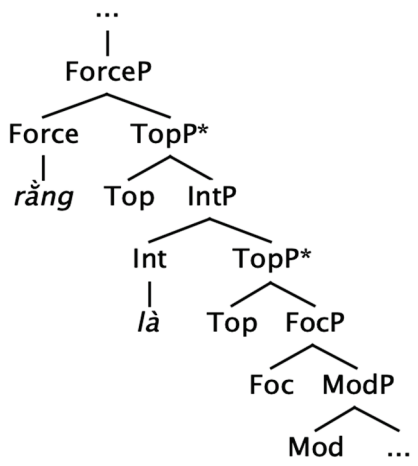
d. *ma (loro) sembrano [**che** use.SUBJNC.3.PL la violenza per altre cose].
but seem-3.PL that DET violence for other things
'. . .but they seem that (they) use violence for other things.'

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:

(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*: as Rizzi & Bocci point out, di 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.

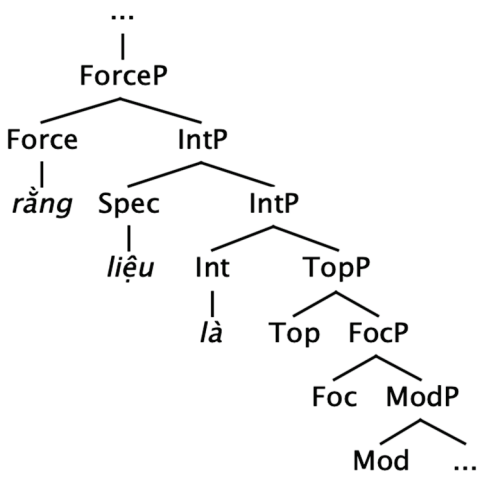
However, it should be clear that this yields a cartographic clash: the tree in (14) that derives {*rằng*. . .*liệu*} sequences is the same as the one that describes {*rằng*. . .*là*} order in (26). This leads to the prediction that *liệu* and *là* should be in complementary distribution, whenever they 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.'[33]

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 would be 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.[34]

(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 functional heads:

> '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, also in a lower position, lower than the *wh*-element in indirect questions. . . (Rizzi & Bocci, 2017, fn 32).

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

2.1.2. 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 (1992), Clark (1992), and Tran (2016, 2024). Nguyen (2021) provides an excellent summary; she also contributes some useful new data, particularly in respect of clause-initial *thì*.[35] 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.'

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 CONJ 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 PRN 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 and 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 and Thompson (1973, p. 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 and 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).[36]

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.[37] Developing his (2010) work, Miyagawa (2017) distinguishes two language types (cf. Li and Thompson's (1976): "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.[38]

Compare first the examples in (35) and (36) above with those in (40) and (41), from Miyagawa (2017, pp. 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).'

b. Class D:

*Hanako	wa	[**sono hon wa**	zibun ga	yonda	*koto*]	o	kookai shita.
Hanako	TOP	that book TOP	self NOM	read	C	ACC	regretted

'Hanako regretted that, as for that book, she herself read (it).'[39]

Having established that Japanese patterns with English in the default case ('Aboutness Topics'), Miyagawa then draws on work by Bianchi and Frascarelli (2010), demonstrating how the two languages diverge with respect to Contrastive Topics (Contrastive Focus) (see Kuno, 1976; Büring, 2003):[40] this is where the main empirical contribution lies. Miyagawa claims that Japanese diverges from English in allowing CTs in all environments—including Class {C, D} contexts, where ATs are blocked. Now compare the sentences in (41) with those in (42) below:[41]

(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**	zibun ga	yonda	*koto*]	o	kookai shita.
Hanako	TOP	that book CONTR.TOP	self NOM	read	C	ACC	regretted

'Hanako regretted that as for that book, she 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 immediately raises the question of which class Vietnamese belongs to. Intuitively, since Vietnamese is normally classified as a "topic-prominent" language,[42] 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 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,
Mary deny COMP PL CLS book DEM2 TOP PRN FUT read,
không phải những cuốn này.
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,
NEG fact ASR story PL CLS book DEM2 TOP John FUT read,
không phải những cuốn này.
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
Mary resent COMP. PL CLF book DEM2 TOP John read
trong kỳ nghỉ.
in vacation
*'Mary resents that those books, John read while on vacation.'

CT: Mary bực tức rằng **những cuốn sách đó** **thì** John đọc,
Mary resent COMP PL CLF book DEM2 TOP John read,
không phải những cuốn này.
NEG correct PL CLS DEM1
*'Mary resents that those books, John read, but not these.'

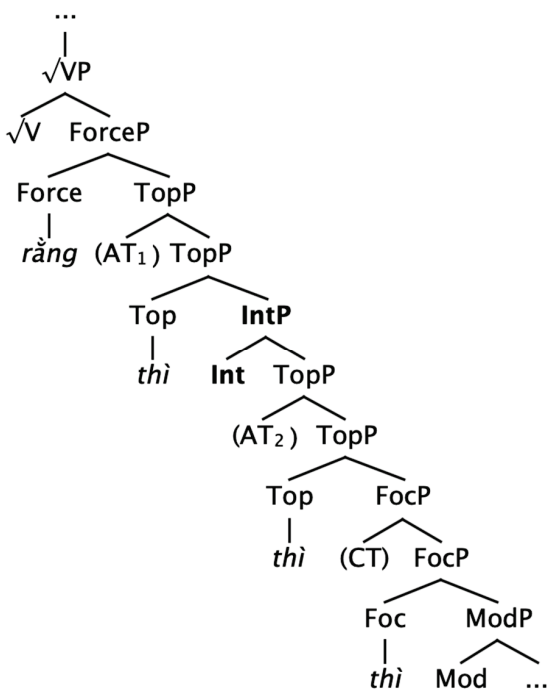
b. Class D (2)

AT: ?Tôi tiếc rằng **những cuốn sách đó** **thì** John đọc
I regret COMP PL CLS book DEM2 TOP John read,
mà không tham.khảo ý.kiến tôi.
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,
I regret COMP PL CLF book DEM2 TOP John read,
không phải những cuốn này.
NEG correct PL CLS DEM1
?*'I regret that those books, John read, but not these.'

A crucial point to bear in mind is that these are judgments of only two native speakers: follow-up investigation is clearly necessary.[43] Nevertheless, assuming the patterns to be valid, the Rizzi and 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:

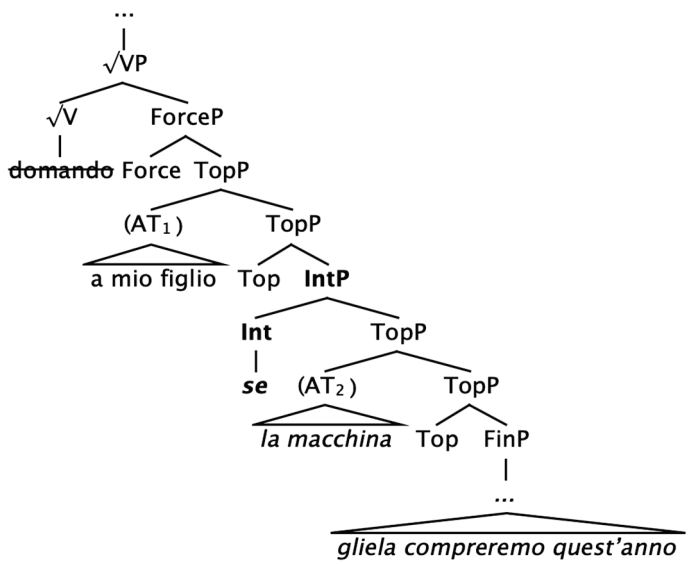
(46)



This diagram not only fits the data well, it also allows us to go further, to examine whether Vietnamese makes use of the higher or lower position for Aboutness Topics.[44] 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: AT_1 or AT_2, in (46)?

(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.



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

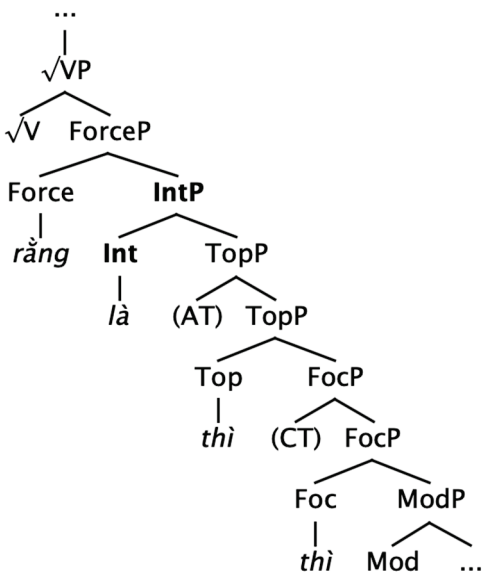
b. *Mary nói **rằng** [$_{AT1}$ những cuốn sách đó] **thì** **là** cô ấy sẽ đọc...
Mary say C PL CLF book DEM2 TOP INT PRN FUT read...
(as a)

c. Mary nói **rằng** **là** [$_{AT2}$ những cuốn sách đó] **thì** cô ấy sẽ đọc...
Mary say C INT PL CLF book DEM2 TOP PRN FUT read...
(as a)

These examples clearly 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:[45,46]

(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:[47]

(50) a. "không phải anh **thì.là** anh khác…"[48]
NEG right PRN CONJ-CONJ you different…
'If not you, then someone else…'

b. "ừ, **thì.là** em có nỗi buồn thật đẹp…" [song lyric][49]
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.'

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

The reason that 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.[50] 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 analysed as a separate clause:[51]

(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; Phan & Duffield, 2022). Both restrictions follow if sentential topics cannot be full clauses, unlike in English.

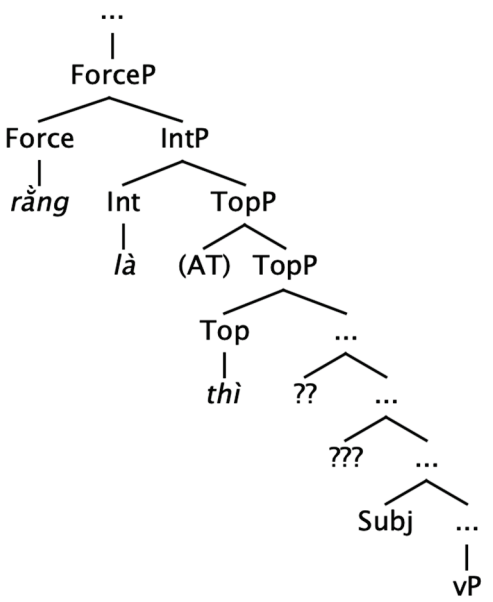
I postpone to future work discussion of the clause following *là*.

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

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)



2.2. Left Peripheral Adverbials

Even before we come to weak (universal) indefinites 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 and Duffield (2022). 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.'

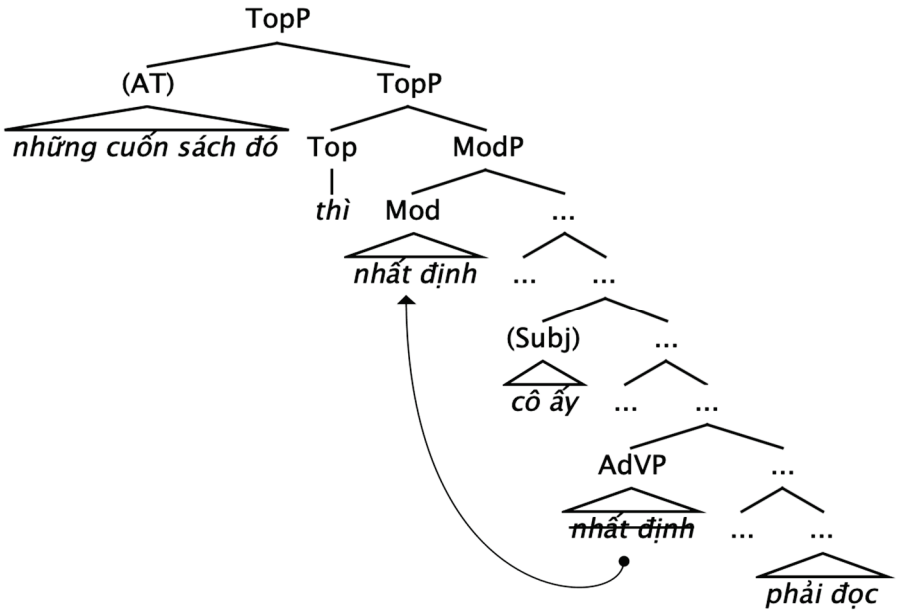
b. *__Quả thật__ [cô.ấy **thì** [(pro) đã nói với anh ấy]]
indeed PRN TOP ANT talk with PRN.DEM
intended (a)

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

From these data, it is plausible to conclude that adverbials are moved from a post-subject, pre-verbal slot to a pre-subject position, to the right of the Topic, as diagrammed in (58), consistent with the Rizzi/Bocci hierarchy:[52]

(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 Rizzi & Bocci 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 (59c,d) below (see Phan & Duffield, 2022):

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

b. Giahái 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.

2.3. "Quantifier-Raising"

Let us now consider fronted weak indefinites interpreted as universal quantifiers ("$QP_{\forall}$").[53] As discussed in Duffield (2007, 2014), Vietnamese disposes of a subset of weak indefinite 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 post-verbally. In the examples in (60), the displaced indefinite 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. [$\surd$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].[$\surd$S-$ADJN_{QP}$-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 exercise 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].] [$\surd$$O_{QP}$-S-cung-V order]
word WH PRN.DEM OP recall
'He remembers every word.'

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

c. **Bao giờ** [anh ấy [cũng đến muộn]]. [$\surd$$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.[54] This state-of-affairs contrasts with regular, fully specified, universal quantifiers (VN. *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**]. [$\surd$S-cung-V-O_{QP} (mọi)]
PRN DEM also recall 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**]. [$\surd$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), these patterns are accounted for in terms of 'scope evasion': $QP_{\forall}$-movement is 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 weak indefinite, 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.[55]

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):

(63) [Force [Int [Top [**Foc** [[Mod [Qemb [Fin [IP …]]]]]]]]]
↑ **$QP_{\forall}$**

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-$O_{QP\forall}$-V is topicalized; equally, (ii), the *S* in $O_{QP\forall}$-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 topicalization, the raised QP intervenes between the Topic and Subject position, as predicted by (63); the unacceptability of (64b) also implies that the weak indefinite cannot move beyond the topic position, again as predicted.[56] What is more, the acceptability of TOP_{SUBJ}-$O_{QP\forall}$-V in (64c) offers further support for $QP_{\forall}$-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 WH 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 recall
'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 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. [$_{AT}$ 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: on 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 hypothesized **Foc** landing-site. Whilst all 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 casting doubt on the universality of the Rizzi/Bocci hierarchy.

2.4. Where(fore) Why?

2.4.1. 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). First, 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*P*s* (thematic phrases containing a subject (67).[58] The examples in (68) show that other *wh*-expressions are 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.

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 purely grammatical. 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 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' construal of *why*, in what is otherwise an ambiguous string (74b). No similar constraint applies to other *wh*-adjuncts: if anything, the downstairs reading is favoured 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: see Culicover (1993), Douglas (2017); Sobin (1987, 2002); Salzmann et al. (2013); also Duffield (2018):

(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.
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 and 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 and 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 Shlonsky and 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.[61]

(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. 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, 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, 2001 and the revision in Shlonsky & 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 . . .]]]]]]]] '

2.4.2. Vietnamese *Why*

The previous discussion raises two kinds of questions: *where* is *why/perché*, initially? and *why* is 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.

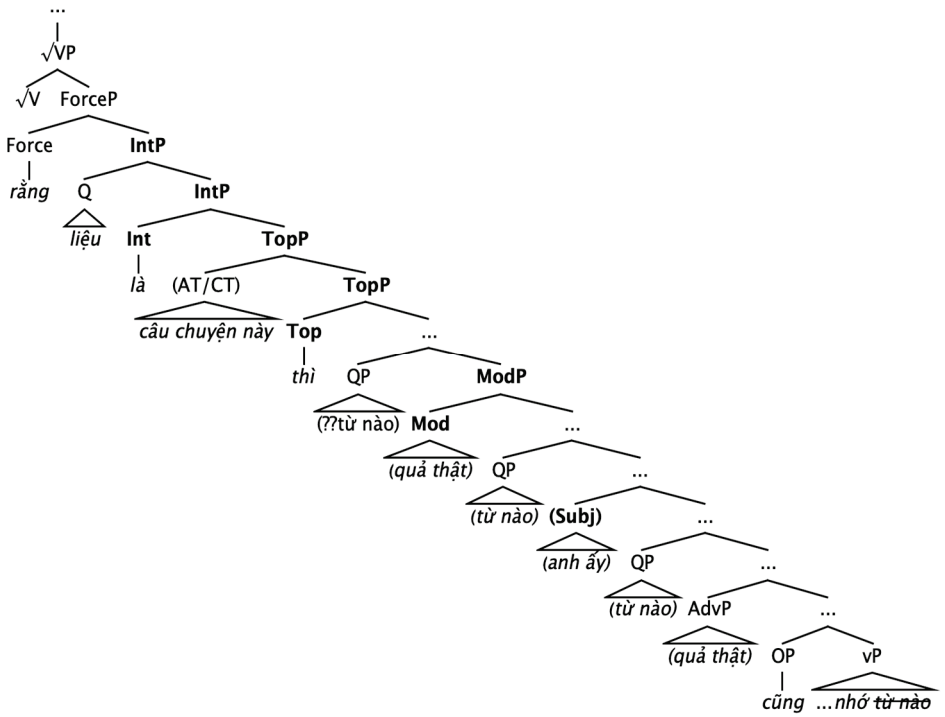
Before considering our final data sets, it is useful to summarize our interim progress, climbing down from the top of the left periphery. First, we have distinguished two head positions in the highest regions of CP: **Force**, typically lexicalized by the declarative complementizer *rằng*, and **Int**, 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 (2009, 2017). 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 note 57). 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 these weak indefinites 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…?[62]
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 ANT many times ask PRN C why
'I have asked you why, so many times.'

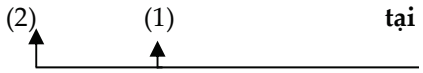
b. Có nhiều bạn hỏi [*rằng* [**tại sao** thầy lại quay được các Video TiếngHàn]][63]
ASR many friends ask C why teacher upload CAN CLF video lge. Korean
'I have many friends ask why my teacher can make Korean videos…'

c. [$_{DP}$ câu hỏi [*rằng* [**tại sao** bỗng nhiên thế giới lại có thang máy]]][64]
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? 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 modelled by Shlonsky and Soare (2011).

As noted earlier, cartographic treatments of *why/perché* analyse 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**} 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 (Shlonsky & Soare, 2011: [35])



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 behaviour.

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?[65]

Intuitively, the final answer is an interpretive one: unlike other *wh*-expressions, which probe gaps in 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.[66]

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 NEG know C INT why PRN 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 NEG know C why INT. PRN recall every word
(as a)

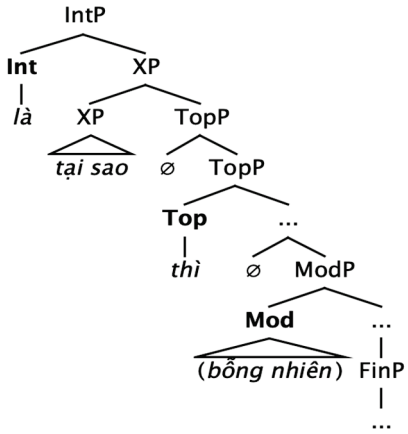
c. *Cô ấy không biết **(rằng)** **là** [anh ấy **tại sao** nhớ mọi từ]
PRN.DEM NEG know C INT PRN.DEM why recall every word
(as a)[67]

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

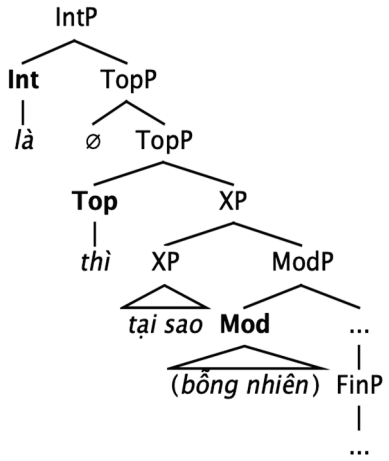
It is worth noting immediately that the unacceptability of (84b) is not due to any simple collocational restriction on {*tại sao. . .là*} sequences: we will consider grammatical instances of this directly.[68]

Moving down the tree, the next question concerns the position of *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.



b.



We will again consider these projections in turn. The alternations in (86)[69] reveal that both affirmative hypotheses are borne out, with respect of topic constituents: *tại sao* may appear either to the left (86a) or—marginally, to the right (86b)—of the topic:[70]

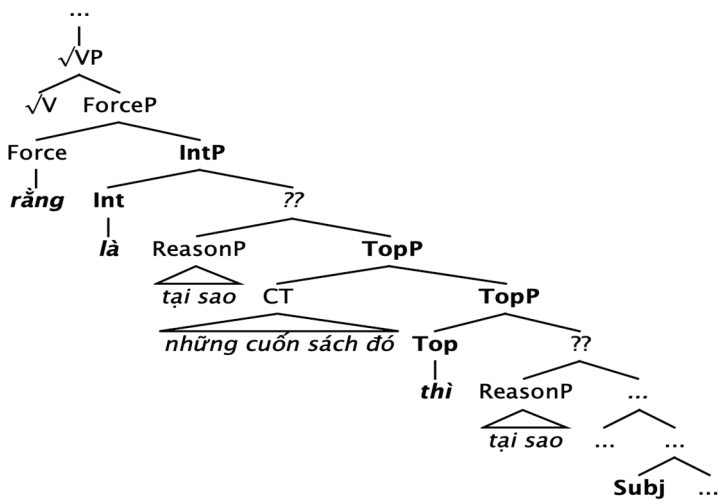
(86) Cô ấy không biết. . .

a. . . .**tại sao** [*những* *cuốn sách đó*] **thì** cô ấy có thể đọc,
why PL CLF book DEM_2 TOP PRN possible read,
không phải những cuốn này.
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,
PL CLF book DEM_2 TOP why PRN.DEM possible read,
hông phải những cuốn này.
EG 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]

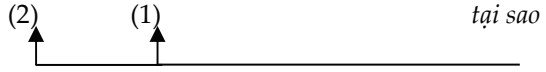
(87)



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 *nhất định* 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 analysed 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 (Shlonsky & Soare, 2011: [35])



The formal prediction is that 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): 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. Shlonsky & Soare, 2011: [35])

Although I remain 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, the **why-to* constraint could be derived 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 obviates the need to assume raising to **Int** (whether overt or covert movement): **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 the 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.[72]

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.'

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: first, that (94b) and (94d) are less than fully acceptable; second, 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 just 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.

2.5. 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 weak indefinites $QP_{\forall}$ (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 to say, 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 >

As for 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).[73]

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

3. Conclusions (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). Mapping is only a means, not an end in itsIlf.

It is nonetheless necessary work, since one can Inly 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 a famous medieval map of uncharted territory,[74] 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 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 a'verbs and weak indefinite (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 (as suggested in Clark, 1992), or whether they simply urge a more flexible and nuanced view of functional categories;
- There also remain unanswered questions concerning the position of (non-topicalized) subjects. One specific issue is whether canonical subjects occupy a position within 'IP' or in the intermediate zone of the low left periphery; alternatively, whether the region conventionally labelled 'IP' may be more highly articulated, to include a lower FocP (*cf*. Belletti, 2004);
- More narrowly, a conceptual puzzle remains about the "*Why-to" constraint. Splicing of non-finite clauses (Shlonsky & Soare, 2011), Truncation (Hooper & Thompson, 1973), Exfoliation (Pesetsky, 2019/2021) 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 standpoint, 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 [final stanza] (*Les Fleurs du Mal*, 1861)

Funding: This research was partially funded by a KAKEN(C) award 21K00538 "Progress in Vietnamese Linguistics", by the Japan Society for the Promotion of Science (JSPS), as well as by sabbatical leave funding from Konan University. I am grateful to both bodies for their support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: 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. I am also extremely grateful to Guglielmo Cinque for taking the time to correct errors in the original draft, and for his invaluable suggestions on further improvements. Finally, I should like to thank the reviewers for agreeing to review an article of this length, and for their extremely constructive and helpful remarks and suggestions.

Conflicts of Interest: The author declares no conflict of interest in the design or implementation of this study.

Abbreviations

The following abbreviations are used in this manuscript:

ANT	Anterior marker
ASR	Assertion Marker
C(OMP)	Complementizer
CLF	Noun Classifier
CONJ	Conjunction
COP	Copular/Linker
DEM	Demonstrative Marker
INF	Infinitive
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

Notes

1 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 et al. (2017), Roberts (2019), and others, 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—very probably erroneous—view, the left periphery is a universal domain, appearances to the contrary notwithstanding.

2 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,

2000); *cf.* Hyams (1986), Deprez and 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.

3 Lasnik and Uriagareka (2022, p. 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.

4 Emonds (1969, p. 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).

5 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.

6 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.

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

8 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's (2019) "^" feature), Rizzi's (2014) *Subject Criterion:* see below.

9 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.

10 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, 2006; Endo et al., 2016).

11 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.

12 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 Bach, 1974, p. 154).

13 Rizzi (2017) implies that Cinque and Rizzi (2015) is the original source of these speculations ("As pointed out in Cinque and Rizzi (2010 [sic]. . ."): 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.

14 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).

15 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, 1992)?

16 A reviewer points out that Vietnamese is probably not unique in this regard: they point to evidence of complementizer variation, and embedded topicalization in other MSEA varieties such as Thai, Lao and Burmese, also to clause-initial interrogative particles in main clauses in Khmer. It seems plausible, therefore, that further research will uncover left periphery structures in other varieties as intricate as those documented here. Time and space constraints preclude consideration of *mà* in this article.

17 Time and space constraints preclude consideration of *mà* in this article.

18 In both cases, I will adopt a base-generation (with co-indexation) analysis here, though nothing hangs on this: here, I am simply concerned to establish surface positions within the functional sequence.

19 *Elle* magazine (VN edition, 17 February 2024).

20 The full citation 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.") https://www.threads.net/@merrygoround.divongquanh/post/DAibDFoSw8B/murakami-haruki-n%C3%B3i-r%E1%BA%B1ng-t%C3%B4i-cho-b%E1%BA%A1n-bi%E1%BA%BFt-t%C3%B4i-th%C3%ADch-b%E1%BA%A1n-kh%C3%B4ng-h%E1%BA%B3n-l%C3%A0-mu%E1%BB%91n-%E1%BB%9F-c%E1%BA%A1nh, accessed on 30 September 2024.

21 https://support.microsoft.com/vi-vn/office/t%C3%B4i-th%E1%BA%A5y-l%E1%BB%97i-n%C3%B3i-r%E1%BA%B1ng-m%E1%BB%99t-s%E1%BB%91-quy-t%E1%BA%AFc-kh%C3%B4ng-%C4%91%C6%B0%E1%BB%A3c-%C3%A1p-d%E1%BB%A5ng-nh%E1%BA%A5t-a0b0f9d6-bb8b-4214-a079-64ceaa9435f7, accessed on 30 September 2024.

22 See the examples in (35)–(36) below, discussed in relation to embedded topicalization (ET).

23 We shall see directly that this is untenable: in contrast to what appears to be in Italian, Vietnamese does not allow the recursive higher topic position between **Force** and **Int**.

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

25 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 more typically associated with mental states and embedded propositions.

26 Example (19a), from Duffield (2013).

27 "*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').

28 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 never occurs sentence-initially.

29 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.

30 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?*

31 (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/, all accessed on 30 September 2024.

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

33 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.

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

35 I am grateful to Trang Phan for drawing my attention to Nguyen's (2021) dissertation. 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.

36 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, p. 25)". In this way, the restriction is explained by selection, as Rizzi suggests (see earlier discussion).

37 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 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 is the discussed example); and Class IV, where both features remain in C (e.g., Dinka). Only the first two types are relevant to the present discussion.

38 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.

39 It is noteworthy that Miyagawa 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. It is also unclear why the subject is changed in examples (41a) and (41b) (Miyagawa's examples [44] and [45]).

40 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.

41 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).

42 Li and Thompson (1976), Thompson (1978), Givón (1984), Hawkins (1986).

43 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 illustrates the fact that *mà*, like *thì*, can lexicalize several different positions.

44 At least on first inspection: when we come to *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.

45 There is indirect evidence that the opposite may be true of Japanese: i.e., 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 should you talk to"—Vietnamese speakers replicated NS judgments, whereas Japanese learners strongly preferred the grammatically unacceptable option). See Duffield and Matsuo (2019).

46 One question left unaddressed in this investigation is whether there is a difference between 'moved' vs. 'non-moved' topics, where the former are directly linked to lower argument positions. A reviewer points to work by Liao and Kao (2023), which compares topicalization structures in Italian, English and Chinese, and which finds just such a difference, 'moved' topics being unique in their clause. Whilst further tests are required for a conclusive answer about the Vietnamese, the data collected so far indicate that non-moved topics are no less restricted than their moved counterparts.

47 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).

48 The whole citation is interesting and relevant. "không phải anh thì là anh khác, tình là rác cớ sao phải buồn. . ..| #shorts"

49 https://www.youtube.com/watch?v=-A9PW_X7kJs (accessed 5 October 2024). At least, this example distinguishes *là* from *rằng*: the latter element would not be possible in this context.

50 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.'

51 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, re-accessed on 6 April 2025)

52 This tree only indicates possible 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. With respect to the subject position, a reviewer suggests examining the distribution of "Scope Bearing" subjects, such as 'nobody', 'few people' 'at most two people': at least in Chinese, these elements cannot undergo topicalization, in contrast to regular thematic subjects; see Ko (2005). Whilst closer examination of the Vietnamese equivalents of such items is obviously worthy of investigation, it is unclear how this would change the picture here, given the contrast between the marginal (57a) and the acceptable example (57b), which crucially does not involve subject topicalization. That said, such an inquiry might well shed some light on the architecture of the lowest levels of the left periphery. Unfortunately, time and space constraints prevent elaboration here.

53 The scare quotes around the section title indicates that we are not dealing here with true universal quantifiers here, but rather with underspecified, 'multifunctional' items that vary in their interpretation according to their scopal relationship to particular semantic operators: these elements might be better, if less felicitously, translated as (free choice) 'any', rather than 'every'. I am grateful to Nikolas Gisborne for drawing my attention to this point.

54 Alternatively, as negative polarity items (NPI), whenever they appear in the scope of negation.

55 The situation is further complicated by the fact that QPs themselves can be topicalized in Class II (Japanese-type) languages; see Miyagawa (2017).

56 A reviewer suggests a semantic explanation for the marginality of (64b): "…where a larger whole is introduced as an Aboutness Topic (AT), followed by a reference to a smaller part of that whole. If reversed, such an order may violate the expected part-whole discourse relation, leading to an unnatural interpretation". If this explanation were correct, then we might expect such orders would be acceptable if the AT were unrelated to the moved QP, that is, if the AT were the subject, as in (i) below. To the best of my knowledge, however, that sentence is, if anything, even less acceptable than (64b) above (cf. 64c):

(i) "Câu chuyện này? ??/***Từ** **nào** anh ấy **thì** cũng nhớ."
CLS story DEM1 word which PRN.DEM2 TOP ALSO remember
*'This story? Every word (of it), he's the one that remembers.'

57 Were it not for the facts in (65), we could account for (64d), by supposing that even non-topicalized subjects raise to a higher SUBJ position, outside of **FinP**. This would be consistent with ideas presented in Rizzi (2017). 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.

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 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.

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 Shlonsky and Soare (2011) refer to discussion in Cattell (1978) and Ko (2005, pp. 898–899.n31), citing David Pesetsky, pers. comm.).

62 https://vietjack.com/khoa-hoc-tu-nhien-6-ct/tai-sao-nguoi-ta-khong-trong-nam-tren-dat-ma-phai.jsp, accessed on 5 October 2024.

63 https://www.instagram.com/ha_manh_trung/p/C9lzXkDyD7o/?img_index=1, accessed on 5 October 2024.

64 https://tapchithangmay.vn/tai-sao-thang-may-ra-doi/, accessed on 5 October 2024. Note that this sentence provides additional evidence of the availability of {C-tại sao-ADVP-subject} order.

65 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.

66 Of course, languages do not always make complete 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.

67 Example (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.

68 Cf. Doubly-Filled-Comp-Filter (DFCF) that rules out {**wh**. . .*that*} in Standard English; *cf.* Bayer and Brandner (2008).

69 Compare the CT sentences in (44–45) above.

70 In these examples, Contrastive Topics are illustrated, but Aboutness Topics exhibit the same alternation.

71 As pointed out by a reviewer, Chinese *weishenme* ('why') seems to behave differently from Vietnamese tai sao, at least on the evidence presented in Yang (2021): Yang argues that Shlonsky and Soare's (2011) ReasonP can be dispensed with, with Int alone accounting for the scopal properties of *why*; the Vietnamese facts suggest the opposite conclusion.

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

73 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.

74 'Hic sunt dracones' in fact appears on no medieval map, though it is found on the Hunt-Lenox Globe (1510). Source: https://www.theatlantic.com/technology/archive/2013/12/no-old-maps-actually-say-here-be-dragons/282267/, accessed on 30 September 2024.

References

Abels, K. (2012). The Italian left periphery: A view from locality. *Linguistic Inquiry*, *43*, 229–254. [CrossRef]

Aoun, J., Hornstein, N., Lightfoot, D., & Weinberg, A. (1987). Two types of locality. *Linguistic Inquiry*, *18*, 537–578.

Bach, E. (1974). *Syntactic theory*. Holt, Rinehart & Winston.

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*. Cascadilla Press.

Belletti, A. (2004). Aspects of the low IP area. In L. Rizzi (Ed.), *The structure of CP and IP: The cartography of syntactic structures* (pp. 16–51). Oxford University Press.

Bianchi, V., & Frascarelli, M. (2010). Is topic a root phenomenon? *IBERIA: An International Journal of Theoretical Linguistics*, *2*, 43–88.

Bresnan, J. (1970). On complementizers: Towards a syntactic 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.

Cao, X. H. (1992). Some preliminaries to the syntactic analysis of the vietnamese sentence. *Mon-Khmer Studies*, *20*, 137–151.

Cattell, R. (1978). On the source of interrogative adverbs. *Language*, *54*, 61–77. [CrossRef]

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson, & P. Kiparsky (Eds.), *A festschrift for morris halle* (pp. 232–286). Holt, Rinehart & Winston.

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

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures* (Vol. 9). Foris.

Chomsky, N. (1985). *Knowledge of language: Its nature, origin, and use*. 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* (pp. 1–52). MIT Press.

Chomsky, N. (1995). *The minimalist program*. MIT Press.

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). MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*, 1–22. [CrossRef]

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). MIT Press.

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 University Press.

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

Cinque, G., & Rizzi, L. (2015). The cartography of syntactic structures. In B. Heine, & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 115–165). Oxford University Press.

Citko, B. (2014). *Phase theory: An introduction*. 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 University Press.
Deprez, V., & Pierce, A. (1993). Negation and functional projections in early grammar. *Linguistic Inquiry*, *24*(1), 25–67.
de Villiers, J. (1991). Why questions. In *University of Massachusetts occasional papers in linguistics (special editions): Papers in the acquisition of WH; Proceedings of the UMASS round table* (Vol. 17, Article 8). GLSA: University of Massachusetts at Amherst. Graduate Linguistics Student Association.
Douglas, J. (2017). Unifying the *that*-trace and anti-*that*-trace effects. *Glossa: A Journal of General Linguistics*, *2*(1), 1–28.
Duffield, N. (2007). Aspects of Vietnamese clause structure: Separating tense from assertion. *Linguistics*, *45*(4), 765–814. [CrossRef]
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). de Gruyter Mouton.
Duffield, N. (2014). Minimalism and semantic syntax: Interpreting multifunctionality in Vietnamese. In *Proceedings of international conference on the linguistics of Vietnam in the context of renovation and integration* (pp. 1155–1175). Social Sciences Publisher.
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). Science and Technics Publishing House.
Duffield, N. (2022). Whales, and other mammals: A naturalist's take on syntactic variation (with particular reference to Vietnamese). In T. Phan, J. Phan, & M. Alves (Eds.), *Vietnamese linguistics: State of the field* (pp. 127–167; JSEALS Special Publication Number 9). University of Hawai'i Press.
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 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, pp. 7–11). EICE.
Emonds, J. (1969). *Root and structure-preserving transformations* [Ph.D. dissertation, MIT].
Emonds, J. (1980). Word order in generative grammar. *Journal of Linguistic Research*, *1*, 33–54.
Emonds, J. (2004). Unspecified categories as the key to root constructions. In D. Adger, C. De Cat, & G. Tsoulas (Eds.), *Peripheries. Studies in natural language and linguistic theory* (Vol. 59, pp. 75–120). Springer.
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. 60–71). Cascadilla Proceedings Project.
Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics*, *1*, 293–320. [CrossRef]
Givón, T. (1984). *Syntax: A functional-typological introduction* (Vol. 1). John Benjamins.
Guilfoyle, E., & Noonan, M. (1992). Functional categories and language acquisition. *Canadian Journal of Linguistics*, *37*, 241–272. [CrossRef]
Haegeman, L. (2006). Conditionals, factives and the left periphery. *Lingua*, *116*(10), 1651–1669. [CrossRef]
Haegeman, L. (2010). The internal syntax of adverbial clauses. *Lingua*, *120*(3), 628–648. [CrossRef]
Hawkins, J. A. (1986). *A comparative typology of English and German: Unifying the Contrasts*. University of Texas Press.
Hawkins, J. A. (2001). Why are categories adjacent. *Journal of Linguistics*, *37*, 1–34. [CrossRef]
Hooper, J., & Thompson, S. (1973). On the applicability of root transformations. *Linguistic Inquiry*, *4*, 465–497.
Hyams, N. (1986). *Language acquisition and the theory of parameters*. Reidel.
Kayne, R. S. (1994). *The antisymmetry of syntax*. 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. [CrossRef]
Kuno, S. (1976). Subject, theme, and the speaker's empathy: A re-examination of relativization phenomena. In C. N. Li (Ed.), *Subject and topic* (pp. 417–444). Academic Press.
Lasnik, H., & Uriagareka, J. (2022). *Structure: Concepts, consequences, interactions*. 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). Academic Press.
Liao, R., & Kao, G. H.-T. (2023). Extraction asymmetries in topic structures: A comparative analysis. *Journal of East Asian Linguistics*, *32*, 63–90. [CrossRef]
McCawley, J. D. (1970). English as a VSO language. *Language*, *46*, 286–299. [CrossRef]
McCloskey, J. (1996). Subjects and subject positions in Irish. In R. D. Borsley, & I. G. Roberts (Eds.), *The syntax of the celtic languages: A comparative perspective* (pp. 241–283). Cambridge University Press.
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). 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). Springer International Publishing.
Nguyen, M. C. (2021). *The particle thì in Vietnamese* [Master's thesis, Liberty University].
Ouhalla, J. (1991). *Functional categories and parametric Variation*. Croom Helm.
Pesetsky, D. (2021). *Exfoliation: Towards a derivational theory of clause size* [Unpublished ms., MIT]. (Original work published 2019). Available online: https://ling.auf.net/lingbuzz/004440 (accessed on 12 October 2024).
Phan, T. (2013). *Syntax of Vietnamese aspect* [Ph.D. dissertation, University of Sheffield].
Phan, T. (2024). *The syntax of Vietnamese tense, aspect and negation*. Routledge.
Phan, T., & Duffield, N. (2022). A Roadmap to Vietnamese phrase structure. In C. Shei, & S. Li (Eds.), *The Routledge handbook of Asian linguistics*. Taylor & Francis.
Rizzi, L. (1990). *Relativized minimality*. MIT Press.
Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar*. Kluwer.
Rizzi, L. (2001). On the position "Int(errogative)" in the left periphery of the clause. In *Current studies in Italian syntax: Essay offered to Lorenzo Renzi*. North Holland Linguistic Series: Linguistic Variations, Volume: 59. Brill Publishers.
Rizzi, L. (2004). Locality and left periphery. In A. Belletti (Ed.), *Structures and beyond. The cartography of syntactic structure* (Vol. 3, pp. 223–251). Oxford University Press.
Rizzi, L. (2014). Some consequences of criterial freezing. In P. Svenonius (Ed.), *Functional structure from top to toe* (pp. 19–45). Oxford University Press.
Rizzi, L. (2017). Locality and the functional sequence in the left periphery. In E. Aboh, E. Haeberli, G. Puskás, & M. Schönenberger (Eds.), *Elements of comparative syntax: Theory and description* (pp. 319–348). De Gruyter Mouton. [CrossRef]
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* (2nd 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 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 Northeast Linguistic Society* (Vol. 2, pp. 149–162). 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. [CrossRef]
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. [CrossRef]
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*. Kluwer.
Starke, M. (2001). *Move dissolves into merge: A theory of locality* [Ph.D. dissertation, University of Geneva].
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). Cascadilla Press.
Thompson, L. C. (1987). *A Vietnamese reference grammar*. University of Hawaii Press. (Original work published 1965).
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 University Press.
Tran, T. (2024). Syntax-information structure interface in Vietnamese. In E. McCready, H. Nomoto, & Y. Takubo (Eds.), *Discourse particles in Asian languages*. Routledge.
Travis, L. (1984). *Parameters and effects of word order variation* [Ph.D. dissertation, MIT].
Yang, B. C.-Y. (2021). On the position of *ReasonP*. *International Journal on Chinese Linguistics*, *8*(2), 291–313. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

Reflexivization and Mình-Exceptional Local Binding by a Monomorphemic Anaphor?

Andrew Simpson * and Linh Pham

Department of Linguistics, University of Southern California, Los Angeles, CA 90089, USA; llpham@usc.edu
* Correspondence: andrew.simpson@usc.edu

Abstract: This paper considers the distribution of the anaphor *mình* in Vietnamese and whether *mình* can be locally bound in the absence of the reflexivizing element *tự*. The patterning of Vietnamese *mình* poses a potential challenge to theories of binding based on claims that anaphors that are monomorphemic reject local binding and are unable to reflexivize a predicate. The paper reports on an experiment designed to probe judgements of *mình* in local reflexive interpretations and concludes that, for many speakers from different regions of Vietnam, *mình* in object position may be interpreted reflexively with the subject of the same clause, without the need for *tự*. On the basis of patterns involving ellipsis and quantificational subjects, it is further shown that this is a genuine binding relation and not simple co-reference. Such conclusions are noted to have significant consequences for certain approaches to binding.

Keywords: binding; monomorphemic anaphors; co-reference; reflexivizers

1. Introduction

The inspiration and starting point for the current paper is an important, extensive study of anaphora in Vietnamese by Đoàn (2022): *Anaphoric dependencies in Vietnamese–a syntactic approach*. In this work, Đoàn describes how the anaphor *mình* is frequently paired with a pre-verbal reflexivizing morpheme, *tự*, and presents the claim in (1):

(1) '. . . the anaphor *mình* cannot be locally bound unless accompanied by the element *tự*.' (Đoàn, 2022, p. 114).

An illustration of this restriction is shown in (2) below. In (2a), *tự* is not present, and a reflexive interpretation between the subject and object of the embedded clause is not available—though a long-distance anaphoric relation is possible between *mình* and the subject of the higher clause.[1] In (2b), by way of contrast, *tự* occurs in the embedded clause, and this is shown to enable (and also force) a reflexive interpretation between *mình* and the subject of the embedded clause.

(2) a. Nam_i nghĩ Lan_k ngưỡng mộ $\text{mình}_{i/*k}$.
Nam think Lan admire body.
'Nam thinks Lan admires him/*herself.'

b. Nam_i nghĩ Lan_k tự ngưỡng mộ $\text{mình}_{*i/k}$.
Nam think Lan tu admire body.
'Nam thinks Lan admires *him/herself.' (Đoàn, 2022, p. 3)

A consideration of the broader patterning of *mình* might seem to suggest, however, that the characterization in (1) is actually too strict. Examples can regularly be found where *mình* is used to refer back to a local subject without the accompaniment of *tự*, as in (3):[2]

(3a) Mịch rất hối hận giận mình sao đã quá thật thà
Mich very regret mad REFL why PST too honest
đến nỗi thú thật với Long là đi ăn trộm.
that confess with Long is go stealing.
'Mịch deeply regretted angering herself for being so honest that she confessed to Long about stealing.'
Source: Giông tố (1936). Vũ Trọng Phụng. Tuyển tập Vũ Trọng Phụng, tập 1. Nhà xuất bản Văn học, Hà Nội, 1996.

(3b) Cô cố giấu mình giữa đám đông đảo
She try hide REFL middle crowd
'She tried to hide herself in the noisy crowd.'
Source: Thời xa vắng. Lê Lựu.

(3c) Còn lão Hứng thất kinh sinh ra hành tội con mèo vô tội
As old Hung terrified so frustrate cat innocent
để trấn an mình.
to calm REFL.
'As for old Hứng, terrified, he took his frustration out on an innocent cat to calm himself down.'
Source: Côi cút giữa cảnh đời (1988). Ma Văn Kháng.

(3d) Biết bao kẻ đang hăm hở chứng tỏ mình rên lên
Many people PROG eager prove REFL groan
như vậy.
same.
'Many people, eager to prove themselves, are groaning the same way!'
Source: Cơ hội của Chúa (3/1989–21/2/1997). Nguyễn Việt Hà.

(3e) Biết bao người anh tuấn còn chôn mình trong
Many people handsome still bury REFL in
lau sậy để đợi lúc gió mây.
reeds to wait when it's cloudy.
'So many handsome men still buried themselves in the reeds to wait for the cloudy weather.'
Source: Phan Khôi. Tác phẩm đăng báo 1928. Lại Nguyên Ân sưu tầm, biên soạn. Nhà xuất bản Đà Nẵng–2001. (lainguyenan.free.fr)

Furthermore, the judgements of sentences with *tự*-less local binding with *mình* in Đoàn (2022) are not necessarily shared by all speakers. The author of the present paper who is a native speaker of Vietnamese finds the following examples from Đoàn (2022) to be acceptable with a local binding relation between *mình* and the subject of the sentence, although such a dependency is starred as ungrammatical in Đoàn (2022):

(4) Nam_k chỉ trích mình_k
Nam criticize body
'Nam criticizes himself.' (* in Đoàn, 2022, p. 71)

(5) Lan_k trừng phạt mình_k
Lan punish body
'Lan punished herself.' (* in Đoàn, 2022, p. 4)

(6) Người đàn ông_k đã khen mình_k.
CL man PST praise body.
'The man praised himself.' (* in Đoàn, 2022, p. 116)

In a subsequent paper focused on long-distance binding in Vietnamese, Đoàn et al. (2023, p. 1) acknowledge the following:

> 'Like many languages, Vietnamese is not entirely homogeneous. The data in this article reflect the language spoken in the Center/Middle of the country . . . The reader should be aware that speakers from the Northern region or the Southern region might diverge in some of their judgments. So far, a comprehensive study of the variation within Vietnamese is not available.'

This raises the possibility that regional variation in the use of *mình* might perhaps underlie differences in acceptability judgments with examples that appear to violate (1), an issue examined in the current paper.

The question of whether there are speakers who permit local binding with *mình* in the absence of *tự* also has a broader significance, extending beyond Vietnamese to the cross-linguistic typology of binding relations and general theories of anaphora and reflexivity. It has frequently been suggested (Cole et al., 1990; Reinhart & Reuland, 1993; Reuland, 2001) that local binding between a subject and an object (or other co-argument) requires an anaphor that is morpho-syntactically complex, and anaphors that are monomorphemic 'simplex expressions/SEs' have been described as rejecting local binding relations, as exemplified in Dutch (7).

(7) Winnie bewonderde zich*(zelf)
Winnie admired self self
'Winnie admired herself.' (Volkova & Reuland, 2014, p. 594)

Such typological observations have led on to the development of a highly prominent and influential approach to binding in Reinhart and Reuland (1993) and Reuland (2001, 2011, 2017), which sets out to explain why monomorphemic anaphors are uniformly unable to reflexivize a predicate. In a recent extension of this body of work, Reuland et al. (2020), it is noted that Mandarin Chinese might appear to constitute an exception to the crosslinguistic patterning of SEs, as both the complex anaphor *ta ziji* and the (apparently) simplex anaphor *ziji* allow for local binding:

(8) Zhangsan$_k$ changchang piping ziji$_k$/ta ziji$_k$.
Zhangsan often criticize 3 self
'Zhangsan criticizes himself.'[3]

Reuland et al. (2020, p. 800) write as follows: 'If *zi-ji* were truly simplex. . .its behavior would not fit the overall pattern observed, and it would pose a challenge to the approach in Reuland (2011)'. Reuland et al. (2020) then go on to argue that Mandarin *ziji* is actually not monomorphemic and is instead a morphologically complex anaphor, consisting of two morphemes *zi-* and *-ji*, and that this explains its 'exceptional' behavior. However, the Vietnamese *mình* cannot be analyzed as bi-morphemic, unlike the Mandarin *ziji*. If *mình* can be used to reflexive-mark a verb, it would therefore pose a clear empirical challenge to theories of reflexivity built on the assumption that local binding is not possible between a monomorphemic anaphor and a co-argument.

With this as background, the primary goal of the current experimental investigation of *mình* is to achieve a better understanding of the cross-speaker patterning of *mình* in configurations of local reflexivity and to probe what regional variation may potentially occur in the use of *mình* as a reflexive object. We conclude that, for many speakers from different regions of Vietnam, *mình* in object position may indeed be interpreted reflexively with the subject of the same clause without the need for *tự*. On the basis of patterns involving ellipsis and quantificational subjects, it is further shown that this is a genuine

binding relation and not simple co-reference. Such conclusions are then noted to have significant consequences for approaches to binding such as that of Reuland (2011).

The structure of the paper is as follows. Section 2, the core of the paper, describes an experiment carried out with 95 speakers of Vietnamese from different parts of Vietnam, designed to examine the acceptability of local binding with *mình* in combination with different types of predicate, and discusses the results of the experiment, which present a different picture of binding relations in Vietnamese than the patterning described in Đoàn (2022). As the results of the experiment also call into question assumptions held about SEs in Reuland (2001) and other works, Section 3 considers whether and how it might be possible to reinterpret the current experimental data in ways that might avoid the conclusion that *mình* is extensively used as a local anaphor. Section 3 argues against such alternative analyses and for the conclusion that the patterns tested in the experiment display the properties of genuine syntactic binding. Section 4 summarizes the findings of the paper for the study of anaphora in Vietnamese and also for binding theory in general.

2. An Experimental Investigation of the Local/Non-Local Interpretation of Mình

The focus of the current study is to determine whether speakers from different parts of Vietnam accept co-reference between the subject of a verb and *mình* in object-of-verb position (or as the object of a preposition) without the occurrence of *tự*, resulting in a reflexive interpretation, as schematized in (9):

(9) subject_k verb $\textit{mình}_k$

In designing an experiment to examine the general acceptability of structures such as (9), we factored in considerations of the potential influence of predicate choice on the naturalness of reflexive interpretations. The verbs in Đoàn's unacceptable examples (where *mình* resists local co-reference in the configuration in (9)) all encode *heavily other-directed actions*—the set of verbs listed in (10) and illustrated with (11). The actions represented by such predicates are all perceived as naturally being carried out on other individuals, and are difficult to imagine as being directed toward the initiator of the action, i.e., being reflexive in orientation.

(10) *động viên*–encourage
khen–praise
ngưỡng mộ–admire
la mắng–scold
xúc phạm–offend

(11)	Mai_i	nghĩ	Nam_j	sẽ	động viên	$\text{mình}_i/{}_{*j}$.
	Mai	think	Nam	FUT	encourage	self

'Mai_i thought Nam_j would encourage her_i/*himself_j.' (Đoàn, 2022)

With heavily other-directed verbs, hearers may not consider interpretations in which the subject acts on herself, i.e., reflexive interpretations. This might lead hearers to regularly ignore possible construals of *mình* as referring to local subjects in favor of other available interpretations of *mình*. As is well-known (Bui, 2019; Đoàn, 2022; Phan & Chou, 2023), a significant complication in many sentences incorporating *mình* is that *mình* can frequently be interpreted as a first-person or second-person pronoun, referencing either the speaker or the hearer. Hence, in many instances, a [Subject Verb *mình*] configuration will allow for *mình* to be interpreted as either 'me' or 'you', as alternatives to any sentence-internal co-reference with the subject. It may be speculated that with heavily other-directed verbs, hearers might often discard potentially available reflexive interpretations of *mình* due to the easier avail-

ability of construing *mình* as representing the speaker or the hearer, an interpretation more readily in line with the typical use of heavily other-directed verbs. Such a confounding possibility suggests that the patterning of *mình* with *động viên* 'encourage', *ngưỡng mộ* 'admire', etc., should not immediately be taken to be representative of the general distribution of *mình*, and a broader range of predicates should be examined in any investigation of local binding in Vietnamese. Additionally, Đoàn's example sentences demonstrating that *mình* cannot be interpreted as co-referential with a local subject are presented in isolation from any embedding context, which may arguably increase the difficulties for hearers to access reflexive interpretations with heavily other-directed verbs. Considerations of this kind lead to two questions, which have guided the current experimental study of *mình*. First, how do heavily other-directed verbs pattern (in the interpretations accepted for *mình*) if a supporting context is added rather than sentences being presented in isolation, without any context? Second, what patterning is observed with *mình* if one considers more neutral verbs rather than just heavily other-directed verbs?

In connection with the above, prior to creating stimuli for the experimental study, the authors hypothesized, on the basis of initial, casual, native-speaker intuitions, that there may be differences among verbs with regard to the availability of local binding with *mình* relating to the typical use of verbs as heavily other-directed vs. not other-directed, as described below.

Type I. Verbs which are typically/heavily other-directed. These may be expected to either favor or perhaps require the presence of *tự* for reflexive interpretations in the absence of a supporting context. Henceforth, we refer to this group of verbs as the 'admire' class. Examples of this posited group are the following:

- *ngưỡng mộ*–admire
- *ghét*–hate
- *chửi mắng*–scold
- *bảo vệ*–protect

Type II. Verbs which are not heavily other-directed/are more neutral. Such predicates might permit reflexive interpretations for certain speakers without the necessary/preferred presence of *tự*. We refer to this group as the 'blame' class. Examples posited in this group (as a result of preliminary intuitions gathered from other speakers prior to the experiment) are the following:

- *trách*–blame
- *ngụy trang*–disguise
- *bồi bổ*–nourish
- *thể hiện*–show off

The stimuli created for the experiment made use of predicates from both posited types, in order to be able to report on differences in local binding that might potentially occur as a function of typical verb orientation. The parameters of the experiment are described in Section 2.1.

2.1. Experiment Design

2.1.1. Participants

One hundred and twelve Vietnamese native speakers were recruited online across three regions in exchange for a $2.05 gift card to the online store Tiki. Eighteen subjects having accuracy rates below 67% on the unambiguous conditions were excluded, leaving 94 subjects in the main data analysis. The study was reviewed and approved by the Institutional Review Board of the University of Southern California, which is fully accredited by the Association for the Accreditation of Human Research Protection Programs (AAHRPP).

Prior to the main experiment, the participants were asked to complete a questionnaire about their language background. The numbers of participants speaking different regional varieties of Vietnamese are summarized in Table 1.

Table 1. Number of participants from different regions.

Regional Variety of Vietnamese	Number of Participants
Northern	27
Middle	28
Southern	40

2.1.2. Procedure and Materials

The participants in the experiment were asked to read Vietnamese sentences in which two or more noun phrases were underlined in order to signal that these noun phrases should be construed as referring to the same individual/set of individuals. The participants were instructed to rate how natural the sentences were with the intended co-reference relation, using a 1–7 scale of naturalness, 1 being completely unnatural/unacceptable and 7 being completely natural/acceptable. Examples of items used in the experiment are presented in Table 2. The experiment included sentences containing examples of verbs of Types I and II, as described in Section 2. Eight of the test items made use of regularly other-directed verbs (Type I), with two of these items involving *mình* following a preposition. Twelve test items used verbs encoding actions that are not so regularly other-directed—Type II 'neutral' verbs—with two of these examples involving *mình* following a preposition, and two examples having quantificational subject antecedents for *mình*. All of the test items containing Type I and Type II verbs occurred without the use of *tự*.

All the test items occurred in linguistic contexts that would allow for a reflexive interpretation to be available and potentially natural. The set of test items involving Type I and Type II predicates specifically aimed to establish whether *mình* could be interpreted reflexively (referring to its immediately preceding subject) in the absence of *tự*, only assisted by a supportive, plausible context. In addition to the target sentences involving *mình* with Type I and Type II verbs, the experiment also included sentences using the pronoun *nó* 'it' in Vietnamese as fillers, with 10 acceptable and 10 unacceptable conditions (i.e., 10 instances in which participants should find co-reference between *nó* and a second, underlined noun phrase acceptable, and 10 examples where such co-reference should not be acceptable). These fillers were used as catch trials to determine whether participants were concentrating on the study or not.[4]

Table 2. Test examples containing Type I, II, III verbs.

Condition	Example							
Typically other-directed verbs (Type I)	**Other-directed verb (6 items)**							
	Hằng	rất	tự tin	vào	nhan sắc	của	bản thân.	
	Hang	very	confident	in	appearance	of	own.	
	Cô ta	thậm chí còn		ngưỡng mộ		**mình.**		
	She	even		admire		**REFL.**		
	'Hang is very confident in her own appearance. She even admires herself.'							
	Other-directed verb followed by a preposition (2 items)							
	Tiên	nói rằng	trong	dịch covid		bà chủ nhà		đã
	Tien	said that	during	covid pandemic		landlord		PST
	hỗ trợ	cho **mình**	rất nhiều.					
	support	for **REFL**	a lot.					
	'Tien said that during the COVID pandemic, the landlady supported herself a lot.'							

Table 2. *Cont*.

Condition	Example							
Neutral verbs (not heavily other-directed) (Type II)	**Neutral verb (8 items)**							
	Khi	cần	đưa ra	quyết định	quan trọng,	Hoàng		luôn
	When	need	give	decision	important,	Hoàng		always
	tin tưởng	**mình**	thay vì	đi xin	lời khuyên	từ	người khác.	
	trust	**REFL**	instead of	go ask	advice	from	other people.	
	'When making important decisions, Hoàng always trusts himself instead of seeking advice from others.'							
	Neutral verb followed by a preposition (2 items)							
	Trong	cuộc bầu cử		mới nhất	, ông	Huy	đã	bình chọn
	In	election		latest	, HON	Huy	PST	vote
	cho	**mình**.						
	for	**REFL**.						
	'In the latest election, Mr. Huy voted for **himself**.'							
	Neutral verb with quantifiers (2 items)							
	Tất cả	các	giáo viên	trong	trường	đều	đang	cố gắng
	All	CL	teacher	in	school	all	PROG	try
	hoàn thiện		**mình**.					
	complete		**REFL**.					
	'All the teachers in school are trying to complete **themselves**.'							

2.1.3. Data Analysis and Results

Figure 1 presents the results of the participants' judgements on the acceptability of reflexive interpretations of *mình*—where *mình* in object-of-verb position is construed as co-referential with the subject of its own clause. As seen in Figure 1, there was a very high level of acceptability/naturalness when *mình* occurred as the object of Type II/neutral verbs and was interpreted as referring to its local subject, i.e., a reflexive interpretation. Participants from all three regions of Vietnam judged such examples to be very natural and acceptable, with average scores exceeding 6/7 for all regional participants. This provides clear, strong support for the conclusion that reflexive interpretations of *mình* are both widely available and felt to be natural without the support of *tự*, in contrast to Đoàn's (2022) findings and generalization that such interpretations are only possible when licensed by the overt presence of *tự*. Considering the participants' judgements of examples involving Type I typically other-directed verbs, these also showed a high level of acceptability, exceeding 6/7 for speakers from southern Vietnam and close to 6/7 for speakers from central and northern Vietnam. These results confirm that reflexive interpretations of *mình* with a local subject are generally possible in the absence of *tự* even with typically other-oriented verbs once such predicates are embedded in plausible supporting contexts rather than being presented out of the blue as isolated sentences.

Figure 2 shows the acceptability values for the test items in which the subject in [Subject Verb *mình*] examples was a quantifier phrase/QP. Four different quantifiers occurred in subjects in these test items: *tất cả* 'all', *mỗi* 'every', *ít nhất* 'at least', and *hơn mười* 'more than ten'. As seen in Figure 2, a very high level of acceptability was recorded by speakers from all three regions for sentences in which *mình* occurred in a reflexive interpretation with a quantificational subject in the absence of *tự*, averaging over 6/7 for speakers from all regions.[5]

The experiment also included sentences in which *mình* occurred following a preposition in sequences of [Subject Verb [P *mình*]], rather than as the direct object of the verb. The expectation was that reflexive interpretations of *mình* with the subject in such examples might perhaps be found to have a higher level of acceptability than in configurations where *mình* was the direct object of the verb. The reason for such an expectation was the suggestion in Đoàn (2022) that the embedding of *mình* in PPs permits reflexive interpretations with

a local subject which are not available when *mình* occurs as the direct object of the verb (in the absence of *tự*)—see the discussion in Section 3.2 below. Interestingly, the results of the experiment did not align with this expectation and present a clear difference in acceptability when *mình* occurred in PPs with Type I and Type II verbs. As shown in Figure 3, with Type II verbs, acceptability ratings with *mình* following prepositions were as high as those in examples where *mình* occurred as the direct object of the verb, over 6/7. However, with Type I verbs, there was a significantly lower level of acceptability, ranging from 4–5/7.

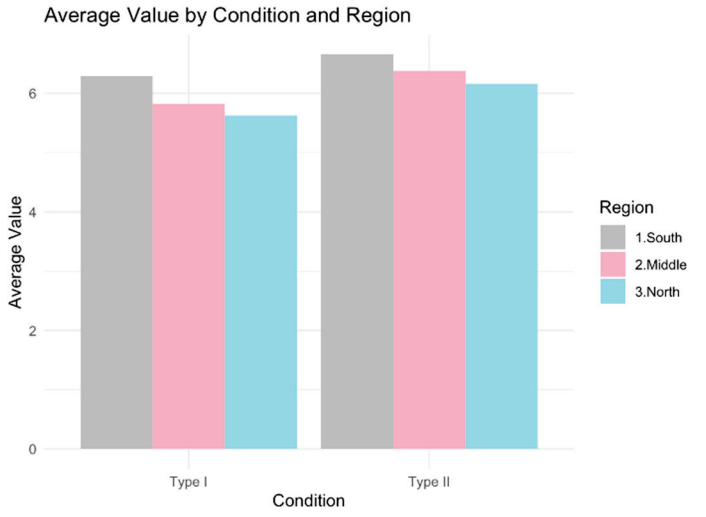


Figure 1. Acceptability of [Subject Verb *mình*] sequences with reflexive interpretations.

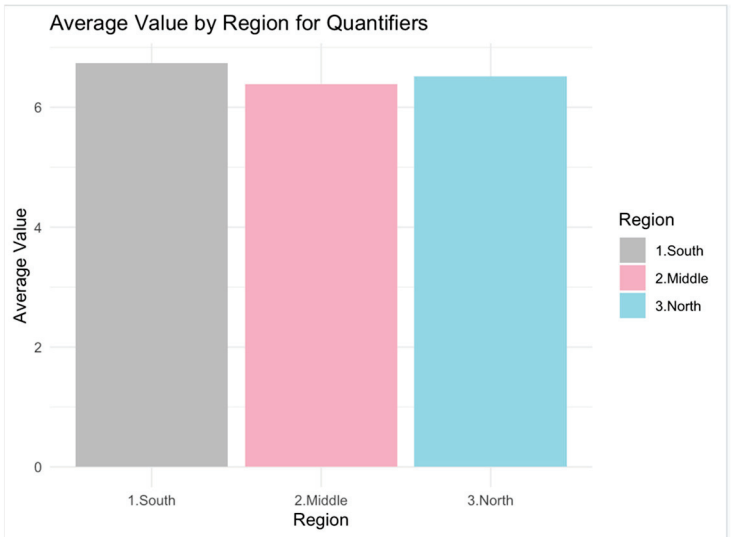


Figure 2. Acceptability of [QP Verb *mình*] sequences with reflexive interpretations.

This reduced acceptability rate relates to specific combinations of verb + preposition. The four pairs of verb + preposition used in examples in the experiment were *biện hộ cho* 'advocate for', *bình chọn cho* 'vote for', *quan tâm đến* 'care for', and *hỗ trợ cho* 'support'. As shown in Figure 4, acceptability ratings with the former two V + P combinations were as high as in examples where *mình* occurred as the direct object of the verb: over 6/7 for speakers from all regions. However, ratings for examples with *quan tâm đến* 'care for' and *hỗ trợ cho* 'support' were much lower: 4–5.5/7. This lower acceptability with certain [Subject Verb [P *mình*]] combinations casts doubt on there being a fully general, structural licensing of reflexive interpretations of *mình* when *mình* is 'protected' by a PP shell (see Section 3.2 below).[6]

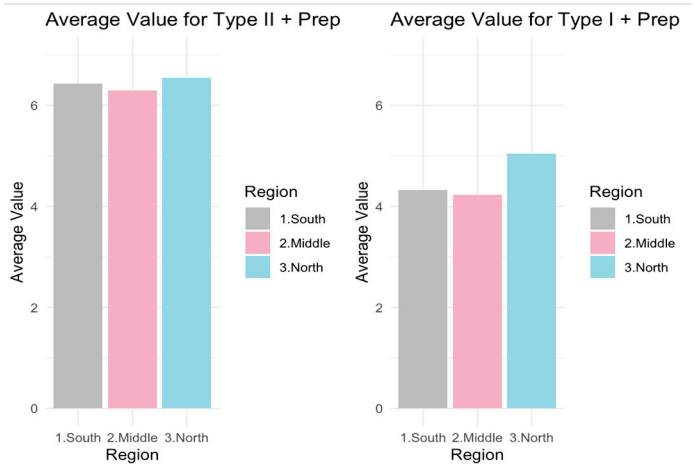


Figure 3. Acceptability of [Subject Verb [P *mình*]] sequences with reflexive interpretations.

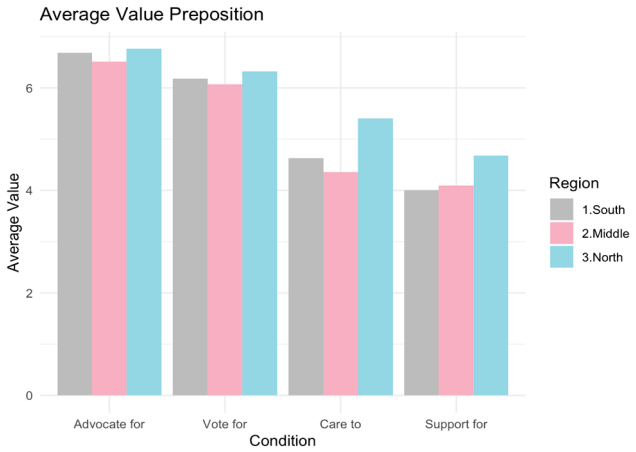


Figure 4. Acceptability of [Subject Verb [P *mình*]] sequences with reflexive interpretations for each [V + P] combination.

While the variable licensing of reflexive interpretations with *mình* in PPs may warrant further useful investigation, in general, the experiment presents strong evidence for the availability of reflexive interpretations with *mình* in (at least) object-of-verb configurations without any necessary presence of *tự* to license such interpretations, when relevant examples are presented in supportive contexts, both with 'neutral' Type II verbs and with typically other-directed Type I verbs, for speakers from different regions of Vietnam.[7] Section 3 considers these findings further, in the light of general theories of binding and reflexivity.

3. *Mình*—A Monomorphemic Anaphor Allowing Local Binding?

The results of the experiment reported in Section 2 present a clear challenge to claims that *monomorphemic* anaphors cannot be used in local binding relations between a subject and an object, to reflexivize the verb (Reuland, 2011), i.e., the configuration in (12) below. [Subject$_k$ Verb *mình*$_k$] patterns indeed appear to be robustly accepted by many speakers of Vietnamese.

(12) *Subject$_k$ Verb Object$_k$ Where Object = monomorphemic anaphor

A question that arises now is whether it might be possible to *reinterpret* the results of the experiment in ways that might preserve (12) as a cross-linguistic generalization in order to maintain theories of reflexivity built on (12) as a valid, cross-linguistic description. Here we will consider three potential counter-proposals that might be made in response to the

conclusion that *mình* permits a local binding relation with the subject of its clause. In each case, it will be argued that such potential reinterpretations of the experimental data are not well supported.

3.1. Inherent Lexical Reflexives

In English, verbs such as *wash, shower*, and *shave* have been classified as inherently reflexive predicates (Reinhart & Reuland, 1993; Reuland, 2001), because such verbs can occur without an object and still be construed as self-directed actions:

(13) John washed/shaved/showered.

However, an overt anaphor can also be added to such verbs, as seen in (14):

(14) John washed/shaved/showered himself.

One conceivable approach to the alternation in (13/14) might be to suggest that the anaphor object in (14) actually does not function to reflexivize the predicate, as *wash/shave/shower* are already lexically listed as having a potentially reflexive meaning.[8] Such a hypothetical approach to *wash/shave/shower*-type verbs in English would raise the question whether a similar analysis and categorization might be made of Vietnamese verbs which allow for local binding with *mình*. If *mình* were to be assumed not to have a reflexivizing function in these cases, similar to *himself* in (14), because the verbs that *mình* is combined with are lexically specified as reflexive predicates, it might be possible to avoid the conclusion that Vietnamese allows local binding with a monomorphemic anaphor.

However, such an approach appears implausible as a potential 'explanation' for the apparent well-formedness of sequences corresponding to (13) in Vietnamese, for the simple reason that many of the verbs which allow for (13) in Vietnamese do not exhibit properties typically found with lexically reflexive verbs.

Vietnamese has two clear types of lexical reflexives, as described in Đoàn (2022), and these show different patterns to the use of *mình* described so far and examined in the experiment.

The first set of lexical reflexives is made up of pairings of *tự* and a verbal root, as exemplified in (15).

(15) Type I lexical reflexives: *tự* + verb combinations

tự sát	*tự vệ*	*tự chủ*
SELF kill	SELF guard	SELF control
'commit suicide'	'protect onself'	'control oneself'
tự nhủ	*tự phê*	*tự cấp*
SELF say	SELF critique	SELF provide
'wonder'	'criticize oneself'	'be self-sufficient'

Significantly, lexicalized combinations of this type do not permit the realization of an object—they are lexically detransitivized by the presence of *tự*, hence *mình* cannot be added:

(16) Nam tự sát/ tự phê/ tự vệ (*/??mình).
Nam self-kill/ self-criticize/ self-protect self
'Nam killed himself/criticized himself/protected himself.'

This property distinguishes such compound reflexives from regular verbs such as *trách* 'blame', *bồi bổ* 'nourish', *thể hiện* 'show off', which, for many speakers, allow for local binding and interpretations of reflexivity with *mình* in object-of-verb position, either with or without *tự*:

(17) Nam (tự) trách mình.
Nam tự blame self
'Nam blamed himself.'

The set of lexical reflexives of the *tự sát* type is, furthermore, very restricted and has a small membership, whereas verbs which allow local binding with *mình* in object position are not obviously restricted in number. The latter consequently do not show the typical characteristics of lexical reflexives across languages, which are often found to be restricted to a small number of verbs that describe self-directed actions.

A second type of lexical reflexive described in Đoàn (2022:109 fn8) consists in simple combinations of a verb and *mình*, as illustrated in (18):

(18) Type II: verb + *mình* combinations

giật mình	*cúi mình*	*thẹn mình*
jerk body	bend body	shy body
'be startled'	'bow'	'be ashamed'

These lexically reflexive predicates are again different from the patterns with *trách* 'blame', etc., as they do not permit objects which are not *mình*, and hence they must be reflexive—they are lexically listed as reflexive-only predicates:

(19) *Nam thẹn/ giật/ cúi Nga.
Nam shame/ shake/ bend Nga
Intended: 'Nam shamed/shook/bent Nga.'

This patterning is quite unlike *trách* 'blame' and other verbs investigated in the current study, which readily permit objects that do not result in reflexive interpretations:

(20) a Nam trách Nga
Nam blame Nga
'Nam blames Nga.'

b Nam ghét Nga.
Nam hate Nga
'Nam hates Nga.'

We conclude that verbs such as *trách* 'blame', *trừng phạt* 'punish', and *ghét* 'hate', which, for many speakers, allow for local binding with *mình*, are not inherently reflexive, as they display properties that clearly distinguish them from lexicalized reflexives. A reviewer of this paper also helpfully notes that subject-experiencer verbs have thus far never been found to allow lexical reflexivization (Reinhart & Siloni, 2005), providing a further reason to reject a lexical reflexive analysis of verbs such as *ghét* 'hate' when they occur in local binding configurations.

3.2. Mình Is Covertly Complex and Not Monomorphemic

A second potential way to avoid the conclusion that *mình* is a monomorphemic anaphor allowing local binding might be to suggest that *mình* is part of a more complex nominal expression and hence not monomorphemic, despite appearances. It has been proposed that additional structural complexity allows monomorphemic anaphors to be legitimately used in reflexive configurations (the idea of 'protection' to avoid a constraint referred to in Reuland (2011, 2017) as the 'Inability to Distinguish Indistinguishables'). In Vietnamese, Đoàn (2022) suggests that the embedding of *mình* within a PP provides a structural way for *mình* to be able to refer reflexively to a local subject in the absence of *tự*: 'when *mình* is in a locative PP, it can be bound without the element *tự*' (Đoàn, 2022, p. 71). This is illustrated in (21):

(21) Nam_i đặt quyển sách phía sau mình_i.
Nam put CL book behind self.
'Nam put the book behind him(self).' (Đoàn, 2022, p. 71)

It might therefore, perhaps, be speculated that, for speakers who regularly accept local binding with *mình* (without *tự*), *mình* is actually part of a more complex nominal structure enabling its apparent reflexive use. Given the commonly assumed origin of *mình* from the noun 'body', a prime candidate for such an analysis could be the possibility that *mình* occurs as the head of a DP which also projects a *pro* possessor, as schematized in (22):

(22) $[_{DP}$ pro $[_{NP}$ mình]]

The interpretation of (23) below, for speakers who accept local binding, might therefore be licensed by such a complex protective structure and be loosely translated as 'Nam criticized his self/body.'

(23) Nam_k trách $[_{DP}$ pro_k $[_{NP}$ mình]].
Nam blame PRN body
'Nam blames himself.'

However, if *mình* were to project a larger, covert structure with a pro possessor, it might naturally be expected that this null element could be overtly lexicalized, but this is not possible, as shown in (22a/b):

(24)	a.	*Nam	trách	$[_{DP}$ nó	$[_{NP}$ mình]].		
		Nam	blame	3	body		
	b.	*Nam	trách	$[_{DP}$ $[_{NP}$ mình]	(của)	nó]	
		Nam	blame	body	of	3	

There is consequently no evidence that might support the assumption of a covert, complex structure with *mình* in such cases, so we continue to assume that *mình* is indeed a monomorphemic anaphor (as also maintained in (Bui, 2019) and (Đoàn, 2022)).

3.3. The Interpretation of Mình with Local Subjects Is Co-Reference, Not Binding

A final way to deflect the current conclusion that *mình* permits local binding, for many speakers, in the absence of *tự*, might be to suggest that the relation of *mình* to its local subject in such cases is actually 'co-reference' rather than 'binding'. In previous work on binding relations, it has been noted that the interpretation of objects as referring to the same individual/entity as a local subject (or some other sentence-internal c-commanding element) should sometimes be attributed to contextually licensed 'co-reference' rather than syntactically mediated binding, as in (112), where Principle C appears to be violated:

(25) Nobody likes John. Even John doesn't like John.

In order to rule out the possibility that two sentence-internal elements are interpretationally related to each other via simple discourse-dependent co-reference rather than binding, two kinds of test have been proposed to identify true instances of binding, involving the use of quantificational antecedents and the potential availability of sloppy interpretations in ellipsis(-type) structures.

If a quantifier phrase may legitimately occur as the antecedent for a pronoun/anaphor, as illustrated in (26), this is taken to signal that the relation between pronoun/anaphor and antecedent is one of genuine binding (Reinhart, 1983):

(26) a. Every $soldier_k$ abandoned his_k position.
b. $Nobody_k$ gave away his_k secret.
c. At least three $students_k$ will submit $their_k$ assignment late.

Where sloppy interpretations of pronouns/anaphors may arise in the interpretation of ellipsis patterns, this is similarly assumed to require and indicate a true variable-binding relation, as, for example, in (27) (Sag, 1976; Williams, 1977 among others):

(27) John likes his teacher. Bill does too.

In assessing the results of the present experimental investigation of *mình*, it is important to check and confirm that the construal of *mình* with a local subject is due to binding rather than discourse-mediated co-reference. In order to do this, we consequently need to examine whether *mình* accepts quantificational antecedents in local configurations, and whether interpretations of sloppy identity are locally possible with *mình*.

The first question is therefore whether *mình* in object-of-verb position can be interpreted as referring to a quantificational subject (without the presence of *tự*). The answer to this question is a clear 'yes'. It is both possible and natural for *mình* to refer to a local QP subject, as illustrated in (28) and (29), which were part of the experimental data and well accepted by participants in the experiment, as already reported in Section 2.

(28) Cuộc thi sắc đẹp diễn ra rất căng thẳng.
NZL contest beauty occur very competitive
'The beauty contest is very competitive.'
Mỗi ứng viên đều cố gắng thể hiện mình
Every candidate all try show off REFL
trước ban giám khảo
in front of judge
'Every candidate is trying to show themselves off in front of the judges.'

(29) Chủ nhật tuần rồi đội tuyển quốc gia đã thua trận bóng.
Sunday last week team national PST lose game
'Last Sunday, the national team lost the game.'
Mỗi cầu thủ đều trách mình vì đã bỏ lỡ
Every player all blame self because PST miss
cơ hội vào vòng bán kết.
opportunity enter semi-final
'Every player blamed himself because they missed the opportunity to be in the semi-final game.'

The second issue is whether *mình* (without *tự*) permits interpretations of sloppy reference in ellipsis-type constructions. Here we considered the interpretation of [Subject cũng vậy] 'subject also thus' structures, in which content from a preceding clause/sentence is associated with a different subject in a following *cũng vậy* clause, in a way similar to VP-ellipsis constructions in English. The critical question is whether an instance of *mình* in the antecedent clause can be interpreted as referring to the subject of the *cũng vậy* clause (via variable binding)—a sloppy interpretation.[9] Such construals are indeed judged to be very natural and straightforward, as shown in (30–31).

(30) Năm học này, toàn bộ học sinh đã rớt môn.
Year study all student PST fail exam
'This year, the students all failed their exams.'
Thầy Nam trách mình, và cô Nga cũng vậy.
Teacher Nam blame self, and teacher Nga also thus
'Teacher Nam blames himself and Teacher Nga does too.'
Sloppy interpretation: Teacher Nga blames herself.'

(31)	Trong	buổi bầu cử	trưởng thôn, Nam	đã	bầu chọn	cho	mình
	In	election	village chief, Nam	PST	vote	for	self
	và	Tâm	cũng vậy.				
	and	Tam	also thus				

'In the elections for village headman, Nam voted for himself and Tam also did.'
Sloppy interpretation: Tam voted for himself.

A consideration of quantificational antecedents and the availability of sloppy identity therefore both unequivocally support the conclusion that the construal of *mình* with a local subject is/can be genuine binding and cannot be dismissed as 'simple co-reference'. The element *mình* in Vietnamese consequently represents a genuine counter-example to assumptions underpinning certain prominent approaches to binding—*mình* is an anaphor which, for many speakers, permits local binding with the subject of its clause, despite its monomorphemic nature.

4. Summary and General Conclusions

The primary goal of the present investigation has been to examine and clarify the status of certain basic binding patterns in Vietnamese involving the anaphor *mình*. While the important study of anaphoric relations made by Đoàn (2022) reports that *mình* may only refer to the subject of its containing clause if paired with the element *tự* (32), the current paper has endeavored to probe whether such an intuition is fully shared by other speakers, or whether *mình* may perhaps allow for local binding in the configuration in (33), without the support of *tự*:

(32) subject$_k$ *(*tự*) verb *mình*$_k$ (Đoàn, 2022)
(33) subject$_k$ verb *mình*$_k$

We also aimed to examine whether there might be any consistent regional differences in speakers' acceptance of *mình* in local binding relations without *tự*. Our experiment consequently involved 94 participants from north, central, and south Vietnam and tested speakers' reactions to *mình* presented in structures corresponding to (32) with a variety of predicate types. We would like to stress that the current study involved a different approach to Đoàn (2022), testing the acceptability/naturalness of examples in supportive contexts rather than as isolated sentences. Consequently, the present study is not directly comparable to the judgements recorded by Đoàn (2022), though it can be assessed relative to the generalizations about possible/impossible anaphoric structures in Vietnamese which Đoàn draws from her findings. The results of the experiment show that binding between *mình* and its local subject (without *tự*) appears, in fact, to be widely accepted by speakers from all areas of Vietnam when presented in a natural, supporting context. Participants in the experiment from all regions also regularly accepted the variable construal of *mình* with a local quantificational phrase subject, confirming that the relation between *mình* and its local subject is binding rather than simple accidental co-reference. Such a conclusion is further supported by the occurrence of sloppy identity readings involving *mình* in *cũng vậy* constructions.

Taken together, the results of the experiment provide strong support for the position that *mình* can indeed engage in a binding relation with a local subject in the absence of *tự* for many speakers of Vietnamese. Such a conclusion will, we believe, be useful for future studies of anaphoric relations in Vietnamese and the use of binding patterns to investigate other aspects of Vietnamese syntactic structures. As noted earlier in the Introduction, it also has potentially broader importance for the cross-linguistic typology of binding by means of anaphors that are simplex expressions/SEs. In a series of works by Reinhart and Reuland (1993), Reuland (2001), and Reuland et al. (2020), the assumption is made that

SEs/monomorphemic anaphors are either heavily restricted to a set of self-directed actions (*washing oneself*, *shaving oneself*, etc.) or require the support of some other element (such as *tự*) in order to reflexivize a predicate. This does not seem to be the case in Vietnamese, where there are no obvious restrictions on verbs that can be interpreted reflexively with the use of *mình*, including verbs that are typically other-directed. Vietnamese therefore presents an interesting empirical challenge to theories of binding that assume that local binding via SEs should not be found in any language—such patterns were robustly attested in the current experimental investigation of *mình*. How such 'exceptionality' in Vietnamese may be accommodated with the patterning of SEs reported for other languages, and whether Vietnamese is truly exceptional among languages with monomorphemic anaphors, we must leave for future work.

Author Contributions: Conceptualization, A.S. and L.P.; methodology, A.S. and L.P.; software, L.P.; validation, A.S. and L.P.; formal analysis, L.P.; investigation, A.S. and L.P.; resources, A.S. and L.P.; writing—original draft preparation, A.S.; writing—review and editing, A.S. and L.P.; visualization, L.P.; supervision, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study has been approved by the Institutional Review Board of the University of Southern California (approval code UP-23-00492, approved on 22 June 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the second author.

Acknowledgments: We extend our sincere gratitude to Lecturer Thu Bui from FPT University and Dr. Van Le from Van Hien University for their assistance in data collection. Additionally, we appreciate the valuable feedback provided by participants at the 4th Annual International Symposium on Vietnamese Linguistics.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

1 Long-distance construals of *mình* are suggested by Phan and Chou (2023) to be licensed via logophoricity. The investigation of such long-distance interpretations is beyond the scope of the current paper. Interested readers may consult Phan and Chou (2023) for useful discussion of these patterns.

2 The paper makes use of the following abbreviations: CL = classifier, FUT = future, HON = honorific, NEG = negation, NZL = nominalizer, PRN = pronoun, PST = past, REFL = reflexive.

3 Example from Xiu-zhi Wu, p.c.

4 With regard to the use of short supportive contexts in the experiment, a reviewer of this paper comments that embedding target sentences within an explicit context may introduce variables that are hard to control for and might unnaturally elevate acceptance ratings. Different perspectives on this issue may be held by different researchers. Our own position here is the following. When example sentences are presented to informants with no supplied context (i.e., out of the blue), it is likely that informants will in many instances supply their own contexts to test out the acceptability of the sentences, and not all informants will necessarily try out the same set of contexts. We therefore believe that offering some kind of potentially facilitative context actually provides a higher degree of control than presenting sentences out of the blue and relying on speakers themselves to uniformly check for contexts which may make a patterning acceptable and natural. Having said this, we accept that it is quite possible that a higher rate of acceptability might result when an appropriate context is supplied than when example sentences are judged fully in isolation, although we do not expect such a higher rate to be artificial and misrepresentative of speakers' mental grammars. In other instances of anaphoric licensing, facilitative context does not seem to lead speakers to accept ill-formed patterns. For example, in (i), a context highly supporting co-reference between 'Bill' and 'himself' does not license such a relation:
(i) 'When Bill_{k} was ill, John_{m} kindly looked after $\text{himself}_{*k/m}$ every day with much care.'

[5] A reviewer asks about the quantificational subjects used in the experiment and whether there might be any effect of distributivity distinguishing their patterning. As noted in the text, four different quantifiers were used with subjects in the experiment: *tất cả* 'all', *mỗi* 'every', *ít nhất* 'at least', and *hơn mười* 'more than ten'. The acceptability rates for all of these quantifiers were (very) high, and there were no obvious effects of (non-)distributivity.

[6] We are currently not sure what might cause the difference in acceptability among the four types of V + P combinations. The variation found here suggests the need for a focused study with a more extensive range of V + P pairs to identify the factors potentially responsible for such patterns.

[7] A reviewer of this paper notes that while the use of classifications such as 'other-directed' and 'self-directed' are common in much work on verb semantics, there are no independent criteria for evaluating which category a verb might belong to. This is certainly correct. However, the 'other-directed' vs. 'neutral orientation' categorization which partially framed the current experiment turned out not to be significant in the results that were actually obtained—there was a high degree of acceptance for examples instantiating both 'types' of verb. The same reviewer adds that differences in thematic roles (agent-theme vs. subject-experiencer) might be useful to test the patterning of subclasses of verbs in Vietnamese with regards to reflexivity. We believe that it would be welcome for follow-up experiments to focus specifically on the testing of agent-theme vs. subject-experiencer verbs. In the current experiment, which included verbs of both types, no indications of differences in patterning were observed. Nevertheless, it would be worthwhile for future work to investigate this possibility further.

[8] It should be noted that Reinhart and Reuland (1993) and Reuland (2001) actually do not assume such an analysis. Instead, they suggest that elements such as 'himself' in (13) are combined with 'wash/shower/shave' as regular transitive verbs (not inherently reflexive) and that 'himself' does reflexive-mark these verbs when it appears with them.

[9] The sloppy interpretation of anaphors in contexts of ellipsis has been documented in many works since Sag (1976); Williams (1977). In (i) below, Ke and Pires (2022) note that only a sloppy interpretation is possible when an anaphor occurs in the antecedent clause of a VP-ellipsis structure:
(i) 'John$_j$ defended himself$_j$ and Bill$_k$ did [$_{\text{elided VP}}$ defend himself$_{j/*k}$] too.' (Ke & Pires, 2022, p. 278)

References

Bui, T. (2019). *Binding and coreference in Vietnamese* [Doctoral dissertation, University of Massachusetts Amherst].

Cole, P., Hermon, G., & Sung, L.-M. (1990). Principles and parameters of long-distance reflexives. *Linguistic Inquiry*, *21*, 1–22.

Đoàn, T. Q. N. (2022). *Anaphoric dependencies in Vietnamese—A syntactic approach*. LOT Dissertations in Linguistics.

Đoàn, T. Q. N., Reuland, E., & Everart, M. (2023). The blocking effect in Vietnamese. *Journal of East Asian Linguistics*, *33*, 153–189. [CrossRef]

Ke, A. H., & Pires, A. (2022). Local vs. long-distance bound implicit arguments of inalienable relational nouns in Chinese. *Journal of Linguistics*, *58*, 269–305. [CrossRef]

Phan, T., & Chou, C.-T. (2023, May 12–14). *Vietnamese bare reflexive and the blocking effect*. 13th Workshop on Theoretical East Asian Linguistics (TEAL 13), Taipei, Taiwan.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. Croon Helm.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, *24*(4), 657–720.

Reinhart, T., & Siloni, T. (2005). The lexicon-syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry*, *36*, 389–436.

Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry*, *32*(3), 439–492. [CrossRef]

Reuland, E. (2011). *Anaphora and language design*. MIT Press.

Reuland, E. (2017). Why is reflexivity so special? Understanding the world or reflexives. *Studia linguistica*, *71*, 12–59.

Reuland, E., Wong, S. C. H., & Everart, M. (2020). How the complexity of Mandarin zi-ji simplifies the grammar. *Linguistic Inquiry*, *51*(4), 799–814. [CrossRef]

Sag, I. (1976). *Deletion and logical form* [Ph.D. dissertation, MIT].

Volkova, A., & Reuland, E. (2014). Reflexivity without reflexives? *The Linguistic Review*, *31*, 587–633. [CrossRef]

Williams, E. (1977). Discourse and Logical Form. *Linguistic Inquiry*, *8*(1), 101–139.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

Article

The Syntax of Null Possessors with Kinship Terms and Body Part Nouns in Vietnamese

Andrew Simpson * and Linh Pham

Department of Linguistics, University of Southern California, Los Angeles, CA 90089, USA; llpham@usc.edu
* Correspondence: andrew.simpson@usc.edu

Abstract: Bare nouns representing kinship terms (KNs) and body parts (BPNs) can be assumed to project a null possessor argument, which allows for the interpretation of such nouns relative to other linguistically present NPs. In Vietnamese, the distribution of KNs and BPNs is subject to different locality conditions and leads to the analysis of null possessors with KNs as covert anaphors, while null possessors with BPNs are null pronominals (pro). This contrasts with Mandarin Chinese, where it has been suggested that null possessors of KNs and BPNs are two different types of null anaphors. The observed distributional differences and analyses of bare KNs and BPNs in Vietnamese vs. Chinese raise questions of parametric variation with regard to null elements with parallel interpretive properties and also whether linking mechanisms may occur with other bare nouns without the projection of null possessors that are subject to binding theoretic locality restrictions.

Keywords: kinship terms; body part nouns; null possessors

1. Introduction

Kinship term nouns (KNs) and body part nouns (BPNs) in Vietnamese do not need an overt possessor, unlike English, and may occur as simple bare nouns, as illustrated in (1) and (2):[1]

(1) Lan nói chuyện với chồng.
Lan talk with husband
'Lan talks to her husband'.

(2) Lan đang xoa vào tay.
Lan ASP rub at arm
'Lan is rubbing her arm'.

However, the distribution of bare noun KNs and BPNs is not free, and the relation of KNs/BPNs to their understood possessors can be shown to be constrained in various ways. For example, KNs need to be c-commanded by their antecedent possessor, as seen in the restrictions on interpretation shown in (3) and (4)—the NPs 'Nam' and 'Ngọc' may not be interpreted as the antecedent possessors for *con trai* 'son' in (3) and (4) even though this might be a pragmatically natural interpretation given the discourse situation.

(3) Giáo viên của Nam trách mắng con trai.
teacher of Nam criticized son
OK: 'Nam's teacher criticized his son (the teacher's son)'.
NOT: 'Nam's teacher criticized Nam's son'.

(4) Lan, trong thời gian mà Ngọc nằm bệnh viện,
Lan, at time REL Ngọc, be hospitalized
đã đưa con trai đi du lịch Thái Lan.
PST take son go travel Thailand.
OK: 'Lan, at the time Ngọc was hospitalized, took Lan's son on a trip to Thailand'.
NOT: 'Lan, at the time Ngọc was hospitalized, took Ngọc's son on a trip to Thailand'.

In what follows, we will use the simple term 'antecedent' to refer to the NP that is construed as controlling the reference of a null possessor hypothesized to be present with bare noun KNs and BPNs. The paper's investigation of KNs and BPNs in Vietnamese has the following goals. First, we aim to descriptively map out the distribution of KNs and BPNs relative to NPs that serve as potential antecedent possessors across three different construction types—topic constructions, passive constructions, and biclausal structures—attempting to see how uniformly KNs and BPNs pattern in these different environments. Second, we will describe and analyze a range of syntactic and pragmatic factors that govern and affect antecedent KN/BPN relations in different ways in Vietnamese. Third, the study will compare KN/BPN patterns in Vietnamese with equivalent patterns in Chinese and the analyses of null possessors in Chinese that have been offered for such elements. Synthesizing all of this information will then lead to an analysis of the syntactic status of null possessors in Vietnamese and a discussion of cross-linguistic variation with bare noun KNs and BPNs, in particular the different status of null possessors with KNs and BPNs in Mandarin Chinese. Finally, the paper will reflect on issues of parametric variation with regard to null elements that have common interpretive properties and whether (and how) linking mechanisms may occur with other types of nouns without the necessary projection of null possessors subject to binding theoretic locality restrictions.

2. The Distribution and Interpretation of Bare Noun KNs and BPNs

In order to generate descriptions of the distribution of KNs and BPNs that can be used to analyze the syntactic status of their null possessors, we will consider the patterning of KNs and BPNs relative to their antecedents in three different constructions: (a) topic constructions where the topic is the intended antecedent of a null possessor in a KN/BPN; (b) two-clause constructions where the subject of a matrix clause is the intended antecedent for a lower clause subject or object KN/BPN; and (c) (long) passive constructions, where an NP preceding the passive morpheme bị is the intended antecedent of a KN/BPN in the object-of-main-verb position. For each construction type, a considerable variety of examples were tested, with typically representative examples being presented here in the paper.

2.1. Bare Noun KNs and BPNs in Topic Constructions

With topic constructions, two major patterns were examined, varying the position of the KN/BPN in subject and object positions, with the intended antecedent for the KN/BPN always being placed in the sentence's initial topic position.

The common pattern found when a KN occurs in matrix subject position and its intended antecedent is the sentential topic is that the KN can, without difficulty, be construed

as relating to the topic NP. For example, in (5) below, the bare KN *bố* 'father' can be easily understood to be the father of the topic *Nam*. We will make the assumption that this relation involves a null possessor element projected with the KN, which we will represent as 'e':

(5) Namk, [e_k bố] khá là già rồi.
Nam father quite is old ASP
'Nam's father is quite old'.

The same patterning occurs with BPNs—a BPN in the subject position is naturally understood to be referentially linked to the topic NP:

(6) Nga_k, [e_k tóc] rất là dài.
Nga hair very be long.
'Nga's hair is very long'.

However, a regular difference in acceptability occurs when a KN or BPN is placed in an object position relating to an antecedent topic. As illustrated in (7) and (8), such patterns are significantly degraded.

(7) ?? Nam_k, tôi rất thích [e_k bố].
Nam I very like father
Intended: 'I like Nam's father a lot'.

(8) ?? Nga_k, tôi rất thích [e_k đôi mắt].
Nga I very like eyes
Intended: 'I like Nga's eyes very much'.

The generalization resulting from a consideration of many such examples is noted in (9):

(9) *Generalization I (Vietnamese)*
In topic constructions, KNs and BPNs relating to the topic as antecedents are acceptable as subjects but significantly degraded as objects.

If this patterning in Vietnamese is now compared with Mandarin Chinese (henceforth simply referred to as 'Chinese'), it is found that the same generalization applies with KN/BPN-antecedent relations in Chinese topic constructions. If a KN or BPN occurs in the subject position, it can naturally relate to the preceding topic as its antecedent, as seen in (10) and (11) (Huang, 1984; Zhang, 2009; Ke & Pires, 2022; Ke, 2023):

(10) $Huangrong_m$, [e_m meimei] zhang de hen piaoliang.
Huangrong sister grow DE very pretty
'Huangrong's sister is beautiful' (Ke, 2023, p. 135).

(11) $Huangrong_m$, [e_m yanjing] zhang de hen piaoliang.
Huangrong eye grow DE very pretty
'Huangrong's eyes are beautiful' (Ke, 2023, p. 135).

However, if a KN or BPN is placed in an object position, it cannot relate to the topic as its antecedent:

(12) *Huangrong$_m$, Guojing hen xihuan [e$_m$ meimei].
Huangrong Guojing very like sister
Intended: 'Guojing likes Huangrong's sister very much' (Ke, 2023, p. 135).

(13) *Huangrong$_m$, Guojing hen xihuan [e$_m$ yanjing].
Huangrong Guojing very like eye
Intended: 'Guojing likes Huangrong's eyes very much' (Ke, 2023, p. 135).

We will postpone analysis of these contrasts in both Vietnamese and Chinese until we have documented patterns that occur in the two other constructions to be considered.

2.2. Bare Noun KNs and BPNs in Two-Clause Constructions

The second syntactic environment to be considered here is two-clause constructions consisting of a matrix clause whose main verb embeds a finite complement clause.[2] When a KN/BPN is placed in an embedded clause object position and its intended antecedent is the matrix clause subject, there are differences in acceptability between KNs and BPNs. Sentences with KNs in this configuration are degraded, as seen in (14), whereas BPNs are fine in the same position, as shown in (15) and (16):

(14) ??Có một bạn học$_k$ (trong lớp) nghĩ Nam không thích
be one classmate (in class) think Nam NEG like
[e$_k$ bố].
father
NOT OK: 'One student$_k$ thinks Nam dislikes their$_k$ father'.

(15) Nam$_k$ nói rằng Nga đã làm gãy [e$_k$ mũi].
Nam say that Nga PST make break nose.
'Nam$_k$ said that Nga broke his$_k$ nose'[3].

(16) Nam$_k$ nói rằng Nga thường chạm vào [e$_k$ chân/mông].
Nam say that Nga often touch into leg/butt
'Nam$_k$ says that Nga often touches his$_k$ leg/butt'.

Such patterns in Vietnamese contrast with what is found in Chinese, where KNs are reported to be acceptable in lower clause object positions, relating to a matrix antecedent (17)/(18), but BPNs are not acceptable (19).

(17) Yuehan$_j$ zhidao Tangmu$_k$ hen aihu [e$_{j/k}$ fuqin].
John know Tangmu very care.for father
'Yuehan$_j$ knows that Tangmu$_k$ takes good care of his$_{j/k}$ father'
(Ke & Pires, 2022, p. 282).

(18) Zhangsan$_j$ yishi-dao Lisi$_k$ kuai nongshang [e$_{j/k}$ fuqin] le.
Zhangsan realize Lisi soon hurt father SFP
'Zhangsan$_j$ realized that Lisi is about to hurt his$_{j/k}$ father' (Ke & Pires, 2022, p. 282).

(19) Zhangsan$_j$ yishi-dao Lisi kuai nongshang [e$_{*j/k}$ shou] le.
Zhangsan realize Lisi soon hurt hand SFP
'Zhangsan$_j$ realized that Lisi$_k$ is about to hurt his$_{*j/k}$ hand' (Ke & Pires, 2022, p. 282).

The relevant generalization for 2-clause structures in Vietnamese is given in (20).

(20) *Generalization II (Vietnamese)*
In 2-clause structures, BPNs in embedded object position may relate to the matrix subject as their antecedent, but KNs are significantly degraded in the same configuration.

This patterning in Vietnamese is, somewhat surprisingly, the opposite of Chinese, where BPNs are not acceptable in embedded object position (relating to the matrix subject as antecedent), but KNs are judged to be fine in this position.

2.3. Bare Noun KNs and BPNs in Passive Constructions

The third structure which will be examined relative to antecedent NP-KN/BPN linking is the *bị* (long) passive construction, in which a passive-like meaning arises from the use of the morpheme *bị* in the schema represented in (21) and illustrated in (22), where NP_1 is typically a patient undergoing the action of the verb, and NP_2 is the agent of the action (Simpson & Ho, 2013):

(21) NP_1 *bị* NP_2 V

(22) Nam bị Nga đánh.
Nam PASS Nga hit
'Nam was hit by Nga' (Simpson & Ho, 2013, p. 160).

In various instances, it may be possible for a third NP to occur in such strucures, following the main verb. This allows for the testing of KNs and BPNs in this position, as schematized in (23).

(23) NP_1 bị NP_2 V [$_{NP3}$ e KN/BPN]

When a BPN occurs in the post-verbal object position in (23), NP_1 may serve as its antecedent, as shown in the examples (24–25):

(24) Nga_k bị Nam giật [e_k tóc].
Nga PASS Nam pull hair
'Nga_k had her$_k$ hair pulled by Nam' (Simpson & Ho, 2013, p. 162).

(25) Nam_k bị Nga tát vào [e_k mặt].
Nam PASS Nga slap on face
'Nam's face was slapped by Nga'.

However, it is not possible for a KN to be placed in the NP3 position linked to NP1 as its intended antecedent, as seen in (26–27):

(26) ??Nga_k bị một người găngxtơ giết [e_k ba].
Nga PASS one CL gangster kill father
Intended: 'Nga suffered her father being killed by a gangster' (Simpson & Ho, 2013, p. 163).

(27) ??Nga$_k$ bị Ông thầy giáo phạt [e$_k$ con trai].
Nga PASS Ong teacher criticise son
Intended: 'Nga's son was criticized by teacher Ong' (Simpson & Ho, 2013, p. 163).

The generalization relating to passive structures is noted in (28).

(28) *Generalization III (Vietnamese)*
In passive structures, BPNs in object-of-verb position may relate to the pre-bị subject NP as their antecedent, but KNs are severely degraded in the same configuration.

Here again, there is a notable contrast with KNs and BPNs in the similarly structured Chinese long passive construction using the morpheme *bei* (29). Both BPNs and KNs are acceptable in the NP3 position in Chinese (30–31):

(29) NP$_1$ bei NP$_2$ V [$_{NP3}$ e KN/BPN]

(30) Zhangsan$_k$ bei tufei da-si-le [e$_k$ baba]
Zhangsan PASS bandit hit-kill-ASP father
'Zhangsan suffered his father being killed by bandits' (Huang, 1999).

(31) Lao Zhang$_k$ bei Lisi da-diao-le [e$_k$ yachi].
Lao Zhang PASS Lisi hit-lose-ASP teeth
'Zhang had his teeth knocked out by Lisi' (adapted from Shi, 1997).

2.4. Mid-Point Summary

The statements in (32) summarize the core observations which been made thus far, and the patterns reported in Sections 2.1–2.3 are represented in Table 1).

(32) a. The distribution of KNs and BPNs in Vietnamese is not free or consistently parallel when a KN or BPN is linked to an intended antecedent across another NP in topic structures, 2-clause sentences, and passive constructions.
b. The linking of KNs and BPNs to antecedent NPs in Vietnamese shows some similarities with equivalent relations in Chinese but also a range of notable differences.
c. In Vietnamese, KNs are more restricted than BPNs in their distribution relative to their antecedents in complex structures. This difference is seen in both 2-clause structures and passives.

Table 1. The distribution of KNs and BPNs to antecedents in Vietnamese and Chinese.

Construction Type:	Noun Type		Vietnamese	Chinese
Topic	KNs, BPNs	as:	✓Subjects	✓Subjects
	KNs, BPNs	as:	??Objects	*Objects
2-clause	KNs	as:	??Objects	✓Objects
	BPNs	as:	✓Objects	*Objects
Passive	KNs	as:	??Objects	✓Objects
	BPNs	as:	✓Objects	✓Objects

The generalizations in (9), (20), and (28) and Table 1 record the typical acceptability of KNs and BPNs in the three syntactic environments considered. Before incorporating this information into an analysis of KNs and BPNs in Vietnamese, we would like to point out and describe certain factors that may further affect perceptions of acceptability and which need to be controlled for in assessing the general patterning of KNs and BPNs. These additional, potentially complicating factors are noted in Sections 2.5–2.8 below.

2.5. Blocking Effects

The interpretation of KNs and BPNs may frequently be restricted in topic constructions when there is a structurally closer potential antecedent for the null possessor in configurations of the form represented in (33).

(33) $NP1_j \ldots NP2_k \ldots [e_{??j/k}$ KN/BPN].

This intervention effect typically occurs when both NP1 and NP2 are human in reference.

If NP2 is inanimate, the linking of NP1 to a null possessor with BPNs and KNs may sometimes improve, particularly in topic constructions. BPNs may be significantly improved in such circumstances, as shown in (34) and (35):

(34) Nga_k, máy nhận diện đã không nhận ra [e_k gương mặt].
Nga the machine PST NEG recognize face.
'The machine didn't recognize Nga's face'.

(35) Nam_k, loại thuốc này giúp chữa lành [e_k cánh tay].
Nam, medicine this help heal arm
'This medicine healed Nam's arm'.

The acceptability of KNs in topic constructions may also be improved when NP2 is inanimate, though KNs still may not be felt to be fully natural:

(36) (?)Nam_k, USC vừa đuổi học [e_k con trai].
Nam USC just expel son.
'USC just expelled Nam's son'.

2.6. Effects of Verb Type

Some effects on the acceptability of BPNs in topic constructions may relate to verb type encoding other-directed as opposed to neutral or self-directed actions. If an unambiguously other-directed verb (e.g., 'kiss (on the lips)') is used in topic constructions, this will effectively eliminate its subject (NP2) as a potential antecedent for a BPN in object position.[4] The construal of a BPN with a topic may subsequently improve, although not necessarily to the point of being fully natural:

(37) (?)Nga_k, tôi đã hôn lên [e_k má].
Nga I PST kiss on cheek.
'I just kissed Nga on her cheek'.

2.7. Effects of Context/Contextual Support

The long-distance linking of KNs over intervening potential antecedents may be ameliorated with rich contextual information. Without special context, KNs are restricted to taking local antecedents:

(38) Nam_j biết rằng Nga_k yêu $[e_{*j/k}$ cha] rất nhiều.
Nam know that Nga love father very much
'Nam_j knows that Nga_k loves *his_j/her_k father a lot'.

However, if (38) is judged within an explicit (preceding) context, such as 'While Nam was not at home, Nga (the neighbor) helped Nam taking care of her father', a long-distance construal of the KN becomes acceptable, and *cha* 'father' may be construed as relating to the matrix subject (as 'Nam's father'). With such rich context situations, we believe that KNs may be interpreted in a name-like way, as referring to individuals already established in the discourse situation and providing a reference value for the KN. If this is correct, and the context essentially supplies the identity of a bare noun KN rather than an NP present in the sentence, such rich context examples do not provide clear information on how KNs may depend on other sentence-internal elements for their interpretation.

2.8. Effects of Focus and Contrast

There are quite significant effects of contrast and focus that affect the acceptability and naturalness of KNs and BPNs in topic constructions. Most typically, KNs and BPNs that occur in the object-of-verb position (39) cannot be linked to a topic, as illustrated in (40):

(39) ??$Topic_k$ Subject V [e_k KN/BPN]

(40) ?(?)Nga_k, tôi thích [e_k đôi mắt].
Nga I like eyes
Intended: 'I like Nga's eyes'.

However, if both contrast and focus are added, such patterns may frequently become fully acceptable. Two components seem to jointly create the improvement effect, involving (a) the addition of the contrastive topic marker thì to the topic and (b) the addition of a focus operator such as *chỉ* 'only' or *cả* 'also' scoping over the object. The amelioration effect of adding contrast on the topic and focus on the operator is shown in (41), which improves the BPN in (40) to a high level of acceptability:

(41) Nga_k thì, tôi chỉ thích [e_k đôi mắt], không thích [e_k tóc].
Nga CT I only like eyes NEG like hair
'I only like Nga's eyes, not her hair'.

KNs in object positions in topic constructions are also significantly improved when contrast occurs on the topic and the object is focused with *chỉ* 'only', as seen in (38):

(42) Nam_k thì tôi chỉ thích [e_k bố] chứ/mà không thích [e_k mẹ].
Nam CT I only like father but/but NEG like mother
'I only like Nam's father. I don't like his mother'.

The contrast+focus effect additionally occurs with the pairing of contrastive topic marker *thì* and the focus operator *cả* 'also':

(43) Nga_k thì, tôi thích cả [e_k mái tóc].
Nga CT I like also hair
'I also like Nga's hair'.

Similar patterns have been observed in Chinese in Zhang (2009), Ke and Pires (2022), and Huang and Yang (2024), raising the question of how such effects should inform the analysis of KNs and BPNs. In Ke and Pires (2022) and Ke (2023), focus effects are acknowledged but then set aside and not considered as part of the core patterns to be modeled, whereas in Huang and Yang (2024), the role of focus in improving examples of the type schematized in (39) figures as a key concern, motivating a particular syntactic approach. We will discuss how the focus effects found in Vietnamese may plausibly be addressed later on in Section 3.

All of the factors identified in Sections 2.5–2.8 may serve to improve otherwise degraded structures with KNs and BPNs to different degrees and need to be controlled for in assessing the general patterning of KNs and BPNs. As mentioned earlier in Section 2.4, the generalizations about Vietnamese presented in (9), (20), and (28) and Table 1 encode what we take to be the typical acceptability of KNs and BPNs in the three syntactic environments investigated, taking the factors described in Sections 2.5–2.8 into account. Now, we will turn to consider what kind of syntactic analysis of KNs and BPNs in Vietnamese is supported by the data patterns considered thus far.

3. The Analysis of KNs and BPNs in Vietnamese

We make the assumption that locality restrictions and constraints on the interpretation of KNs and BPNs relate to the syntactic properties of null possessors projected with KNs and BPNs in the absence of overt possessor NPs. Such null possessors might be instantiations of one (or more) of the null NP types which have been posited to occur elsewhere in different languages, listed in (44):

(44) Commonly hypothesized null NP types
- (i) null pronominals;
- (ii) null anaphors;
- (iii) null operator trace;
- (iv) ellipsis of NPs.

Two of the types noted in (44) may, however, be unlikely candidates for the identity of null possessors with KNs/BPNs in Vietnamese—(iii) and (iv). With regard to (iii), the analysis of NP gaps arising from the movement of null topic operators proposed in Huang (1984) was critically supported by the island sensitivity of null objects in Chinese and Brazilian Portuguese. However, in Vietnamese, KNs and BPNs can occur within islands such as relative clauses and show no sensitivity to the CED/Subjacency (45/46), suggesting that a movement analysis of null possessors in Vietnamese is not appropriate. This additionally rules out an analysis of possessor raising, in which the possessor would be moved from within the DP headed by the possessum (Deal, 2013).

(45) Nam_k kiện cô gái mà đã làm gãy [e_k mũi].
Nam sue the girl who PST make break nose.
'Nam_k knows the girl who broke his_k nose'.

(46) Nga_k biết người đàn ông mà đã chỉ trích [e_k cha].
Nga know the man who PST criticized father.
'Nga_k knows the man who criticized her_k father'.

As for possibility (iv), it might be hypothesized that null possessors of KN/BPNs result from ellipsis applying to a syntactically projected possessor, akin to the process of argument ellipsis, which is argued to produce null subject and objects in a range of

languages (Takahashi, 2008a; H.-T. Cheng, 2013). However, a typical property of ellipsis is that it applies in structures where a constituent such as a VP, TP, or DP would otherwise redundantly be (overtly) repeated in the second of two broadly parallel structures, as in (47–50). The content of elided constituents is consequently old/given and regularly recoverable from a preceding clause with a similar/related structure. The patterning of null possessors with KNs and BPNs does not share these common properties of ellipsis, decreasing the plausibility of an ellipsis analysis of such elements.[5]

(47) John asked me to open the vault, and so I will [$_{VP}$ ~~open the vault~~]!

(48) Bill invited someone to dinner, but I don't know who [$_{TP}$ ~~Bill invited to dinner~~].

(49) Mary reviewed John's paper, and I reviewed [Mary's ~~paper~~].

(50) a. Taroo-wa sannin-no sensei-o sonkeishiteiru
Taroo-TOP 3-gen teacher-ACC respects
'Taroo respects three teachers'.

b. Hanako-mo ~~sannin-no~~ ~~sensei-o~~ sonkeishiteiru
Hanako-also 3-gen teacher-ACC respects
'Hanako also respects three teachers' (adjusted from Takahashi, 2008b).

Considering the possibility that Vietnamese null possessors may instead have the syntactic status of null pronominals or null anaphors, it is useful to reflect on the analysis that has been proposed for such elements in Chinese by Huang (1984, 1989) and Ke and Pires (2022). Huang (1984, 1989) suggests that null possessors in Mandarin Chinese are instances of pro and subject to the Generalized Control Rule/GCR, which essentially requires that a pro must be coreferential with the closest c-commanding antecedent. Such assumptions are able to capture the observation that KNs and BPNs with null possessors are acceptable in topic constructions in Mandarin in the subject position (10, 11) but not possible in object positions (12, 13). In the former case, the pro null possessor is directly c-commanded by its antecedent in the topic position, satisfying the GCR, whereas when a pro null possessor occurs in a KN/BPN in object position, it cannot take the topic as antecedent due to the GCR, because there is a closer c-commanding NP—the subject. Such an approach has more recently been critiqued in Ke (2023, pp. 134–135) on the grounds that it may not satisfactorily account for the patterning of KNs and BPNs in embedded clause subject positions, where KNs are unacceptable (predicted by the GCR) but BPNs are acceptable, which is unexpected for a GCR account (see Ke, 2023, pp. 134–135 for details). Ke and Pires (2022) offer a different characterization of null possessors with KNs and BPNs in Mandarin as null anaphors of two distinct types. Comparing restrictions on the distribution and interpretation of KNs and BPNs relative to their antecedent (possessor) NPs, it is argued that the null possessors taken to be present with KNs have all of the regular properties of the overt long-distance anaphor *ziji* in Mandarin (subject-oriented, logophoric, permitting long-distance binding), while the null possessors of BPNs share a parallel patterning with the local anaphor *ta-ziji* (not subject-oriented, not logophoric, requiring local binding). Ke and Pires (2022) consequently propose that the null possessor projected with KNs is a covert equivalent to the long-distance anaphor *ziji* and a null counterpart to the local anaphor *ta-ziji* with BPNs.

If we now consider Vietnamese null possessors in the light of models proposed for Mandarin, an adoption of Huang's *pro* + GCR account would not seem to be viable for Vietnamese. While such an approach would make correct predictions for patterns found in topic structures in Vietnamese (which are essentially parallel to those found in Mandarin),

the acceptable occurrence of BPN objects in embedded clauses relating to subjects in higher clauses is unexpected and would violate the GCR. In the configuration in (47), a pro occurring as a null possessor of the BPN should only be able to be controlled by the closer subject NP_2 according to the GCR, yet instances of (51) where the BPN takes NP_1 as its antecedent are fine in Vietnamese, as already illustrated in examples (15) and (16).

(51) NP_1 V [NP_2 V [$_{NP3}$ $e_{1/2}$ BPN]]

The analysis of Mandarin null possessors developed by Ke and Pires (2022) would also not seem to be appropriate for Vietnamese. Although such an approach provides an accurate account of the patterning of null possessors found in Mandarin with KNs and BPNs, as we have shown in Sections 2.1–2.3 and summarized in Table 1, the distribution of null possessors in Vietnamese is significantly different from Mandarin in both two-clause and passive structures. In Vietnamese, the patterns found with KNs and BPNs consequently require a different analysis from Mandarin. For example, BPNs in Vietnamese do not have the distribution of local anaphors, unlike their equivalents in Mandarin, and KNs in Vietnamese show greater locality restrictions than Mandarin KNs, analyzed by Ke & Pires as long-distance anaphors.

Various of the basic diagnostics used by Ke & Pires to analyze Mandarin can, however, usefully be considered in approaching Vietnamese KNs and BPNs and how to capture their distinctive properties and differences from Mandarin. Here, we will describe sloppy identity interpretations, subject orientation, and c-command in antecedent/null possessor relations and subsequently combine this information with the locality phenomena noted in Table 1 to finally arrive at an analysis of Vietnamese null possessors with KNs and BPNs in Section 3.4.

3.1. Sloppy Identity

Whether sloppy and/or strict interpretations are possible or required in patterns of ellipsis is sometimes suggested to be determined by the syntactic status of relevant elements as either pronouns or anaphors. As noted by Ke and Pires (2022, p. 278), 'It has long been known that elided bound reflexives have only a sloppy reading, whereas elided bound pronouns have both a strict and a sloppy reading' (Sag, 1976; Williams, 1977). This contrast is illustrated in the pair of sentences (52) and (53). The anaphor in (52) only permits a sloppy reading when VP ellipsis occurs, but the pronoun in (53) allows either a strict or a sloppy interpretation:

(52) John$_j$ defended himself$_j$, and Bill$_k$ did [$_{elided\ VP}$ ~~defend himself~~$_{*j/k}$] too.
Only sloppy: 'Bill defended Bill'.
No strict reading: 'Bill defended John'.

(53) John$_j$ likes his$_j$ car and Bill$_k$ does [$_{elided\ VP}$ ~~like his~~$_{j/k}$ ~~car~~] too.
Sloppy OK: 'Bill likes Bill's/his won car'.
Strict OK: 'Bill likes John's car'.

Ke & Pires show that null possessors in Mandarin permit only sloppy interpretations in ellipsis constructions with both KNs and BPNs, supporting their analysis as null anaphors in both cases.

Turning to Vietnamese, the *cũng vậy* 'also thus' copying construction can potentially be used to probe the status of null possessors with KNs and BPNs (see Simpson & Ngo, 2022 for strict/sloppy interpretations of elided/copied objects in cũng vậy constructions). With KNs, it is found that only sloppy interpretations of a null possessor are available, as illustrated in (54):

(54) Trong chuyến dã ngoại, Ngọc rất nhớ mẹ, Lan cũng vậy.
during camping, Ngọc very miss mom, Lan also thus
'During the camping, Ngọc misses her mom, and Lan does too'.
Only sloppy: 'Lan misses Lan's mom'.
No strict reading: "Lan misses Ngọc's mom'.

This patterning would be consistent with the null possessors of Vietnamese KNs also being an anaphor of some type. With BPNs, however, it is found that either a strict or a sloppy interpretation is possible in *cũng vậy* constructions, as shown in (55):

(55) Ông tổng thống đã bảo vệ đầu. Người vệ sĩ cũng vậy.
president PST protected head bodyguard also thus.
'The president protected his head, and the bodyguard did too'.
Sloppy OK: 'The bodyguard protected his own head'.
Strict OK: 'The bodyguard protected the president's head'.

This different patterning with BPNs (different from KNs in Vietnamese and also different from BPNs in Mandarin) would potentially align with null possessors in Vietnamese being null pronominals rather than null anaphors.

3.2. Subject Orientation

Ke and Pires (2022) claim that subject orientation patterns in Mandarin support their analysis of the null possessor in Chinese KNs as being a null equivalent of the overt anaphor *ziji* 'self'. They note that overt *ziji* is subject-oriented and then demonstrate that KNs also appear to be subject-oriented. In (56), the KN *erzi* 'son' can only be construed as relating to the subject *Zhangsan*, not the indirect object *Lisi*:

(56) Zhangsan$_j$ songgei-le Lisi$_k$ yi-zhang [e$_{j/*k}$ erzi] de zhaopian.
Zhangsan give-ASP Lisi 1-CL son DE photo
'Zhangsan$_j$ gave Lisi$_k$ a photo of his$_{j/*k}$ son' (Ke & Pires, 2022).

This contrasts with the patterning of BPNs, which Ke & Pires suggest have null possessors that are local anaphors equivalent to *ta-ziji*, which is not subject-oriented. In the same configuration as (57), a BPN can be (easily) construed as relating to the indirect object; hence, the null possessor of a BPN is not subject-oriented, as expected if it indeed is a null equivalent of the *ta-ziji*:

(57) Zhangsan$_j$ gei Lisi$_k$ hua-le yi-zhang lian$_{?j/k}$ de sumiao.
Zhangsan for Lisi draw-ASP 1-CL face DE photo
'Zhangsan$_j$ drew Lisi$_k$ a sketch of his$_{j/*k}$ face' (Ke & Pires, 2022).

Considering Vietnamese, the overt long-distance anaphor *mình* is subject-oriented, like Chinese *ziji* (Ivan & Bui, 2019; Chou and Phan, to appear):

(58) Ginny$_j$ nói với Luna$_k$ về mình$_{j/*k}$.
Ginny talk with Luna about self
'Ginny$_j$ talked with Luna$_k$ about self$_{j/*k}$' (Ivan & Bui, 2019, p. 52).

This subject orientation is maintained when *mình* occurs as the overt possessor of a KN, as seen in (55):

(59) Ngọc$_j$ đưa cho Lan$_k$ bức tranh của [ba mình$_{j/*k}$].
Ngọc give for Lan picture of father self.
'Ngoc$_j$ gives Lan$_k$ the picture of her$_{j/*k}$ father'.

However, when null possessors are (by hypothesis) projected with KNs and BPNs, neither KNs nor BPNs show a subject orientation in their interpretation:

(60) Ngọc$_j$ đưa cho Lan$_k$ bức tranh của [e$_{j/k}$ ba].
Ngọc give for Lan picture of father.
'Ngọc$_j$ gives Lan$_k$ the picture of her$_{j/k}$ father'.

(61) Ngọc$_j$ vẽ cho Lan$_k$ một bức phát hoạ [e$_{j/k}$ khuôn mặt].
Ngọc drew for Lan 1 CL sketch face.
'Ngoc$_j$ drew a sketch of her$_k$ face for Lan$_k$'.

This patterning suggests that (a) neither of the null possessors with KNs and BPNs in Vietnamese are simple null equivalents to the overt long-distance anaphor *mình*, and (b) KNs in Vietnamese again show differences in their interpretation from KNs in Mandarin and so should be analyzed in a different way.

3.3. C-Command Revisited

Ke and Pires (2022) observe that both KNs and BPNs must be c-commanded by their antecedents in Mandarin. This is taken to further support an analysis of the null possessors of both KNs and BPNs as null anaphors rather than null pronouns, as pronouns do not require c-commanding antecedents. At the outset of the current paper, it was noted that KNs in Vietnamese similarly require c-commanding antecedents, as illustrated in examples (3) and (4), repeated here:

(3) Giáo viên của Nam trách mắng con trai.
teacher of Nam criticized son
OK: 'Nam's teacher criticized his son (the teacher's son)'.
NOT: 'Nam's teacher criticized Nam's son'.

(4) Lan, trong thời gian mà Ngọc nằm bệnh viện, đã đưa con trai đi
Lan, at time REL Ngọc be hospitalized, PST take son go
du lịch Thái Lan.
travel Thailand.
OK: 'Lan, at the time Ngọc was hospitalized, took Lan's son on a trip to Thailand'.
NOT: 'Lan, at the time Ngọc was hospitalized, took Ngọc's son on a trip to Thailand'.

Such patterns would potentially support the analysis of null possessors with KNs in Vietnamese as being null anaphors of some type rather than null pronouns. With regard to BPNs, the patterning is different, however, and BPNs may relate to antecedents that do not c-command them, as shown in (62) and (63):

(62) [Khi Linh$_j$ đang đọc sách], thì Hoàng$_k$ chải [e$_{j/k}$ tóc].
when Linh PROG read book, CONJ Hoang comb hair.
'While Linh$_j$ was reading a book, Hoang$_k$ comb her$_{j/k}$ hair'.

(63) [Con dao của Nam_k] rơi vào [e_k chân].
knife of Nam fell onto foot
'Nam_k's knife fell onto his_k foot'.

This contrastive patterning with BPNs argues against any analysis of their null possessors as null anaphors, which should require c-commanding antecedents, and instead presents a distribution that is more like that of pronominal elements.

3.4. Putting All the Pieces Together: The Analysis of Vietnamese KNs and BPNs

We are now in a position to reach an analysis of the null possessors assumed to be projected with KNs and BPNs in Vietnamese, combining insights into their distribution described in Section 2 with other information just detailed above in Sections 3.1–3.3. Our proposals are made in (64) and (65), and this is followed by a Table that provides a summary of the properties of KNs and BPNs from the discussion in Sections 2.1–2.4 and 3.1–3.3.

KNs have been noted to have the following properties. First, they need to be c-commanded by their antecedents, which is a patterning typical of anaphors, not pronominals. Second, they only give rise to sloppy interpretations in ellipsis-type constructions, which is again a property characteristic of anaphors rather than pronouns. KNs are additionally restricted in their distribution and must occur in a very local relation with their antecedents, not being separated from their antecedent by any intervening, c-commanding NP. Specifically, as noted in Section 2, KNs are degraded as objects in topic constructions: [??Topic_m Subject_k V [e_m KN]]. They are degraded as the lower objects of 2-clause constructions relating to the higher clause subject: [??Subject_m V [Subject V [$_{Ob}$ e_m KN]]]. They are also degraded as the objects of passive constructions: [??Subject_m bị NP_k V [[$_{Ob}$ e_m KN]]]. KNs consequently have the profile of anaphors but not that of long-distance anaphors such as Vietnamese *mình* or Mandarin *ziji*. Bare noun KNs are additionally not subject-oriented, unlike the overt long-distance anaphor *mình*. From all of the above, we now conclude that the identity of null possessors is that of a local anaphor, approximate to the Vietnamese local anaphor *chính mình*.

(64) *Proposal 1: KNs in Vietnamese*
The null possessors projected by KNs are covert local anaphors which must be bound by the structurally closest, c-commanding antecedent NP.

As for BPNs, the null possessors assumed to occur with BPNs are quite unrestricted in their properties, in contrast to the patterning of KNs. BPNs in Vietnamese do not need to be c-commanded by their antecedents, they do not exhibit subject orientation, and they permit either sloppy or strict interpretations in ellipsis-like *cũng vậy* constructions, a property which aligns them with pronouns rather than anaphors. In terms of locality and the relation of BPNs to their antecedents, BPNs are fine as the lower objects of two-clause constructions relating to the matrix subject as an antecedent (unlike KNs): [Subject_m V [Subject V [$_{Ob}$ e_m BPN]]]. BPNs are also fine as objects in passive constructions relating to a pre-*bị* subject (again, unlike KNs): [Subject_m bị NP_k V [[$_{Ob}$ e_m BPN]]]. The sole obvious locality restriction with BPNs is that they are sometimes degraded as objects in topic constructions: [??Topic_m Subject_k V [e_m BPN]]. However, this restriction disappears when the intervening subject is inanimate (30/31).

While something additional needs to be said about those instances of topic constructions where BPNs are degraded in object positions, and we will turn to this shortly, the general patterning of BPN null possessors otherwise demonstrates a consistent matching with the common properties of pronouns—no c-command requirement relative to antecedent NPs, no strictly local relation to an antecedent required, and the possibility of

strict as well as sloppy readings in ellipsis-type constructions. We, therefore, propose that the null possessors of BPNs in Vietnamese are null pronominals/instances of pro.

(65) *Proposal2: BPNs in Vietnamese*
The null possessors projected by BPNs are covert pronominal elements – instances of pro.

Table 2 provides a summary of the various properties of KNs and BPNs observed in Vietnamese.

Table 2. Summary of Properties of KNs and BPNs in Vietnamese.

Bare Noun Type	**KNs**	**BPNs**
Must be c-commanded by antecedent	Yes	No
Only sloppy interpretations in ellipsis-type constructions	Yes	No
Subject-oriented	No	No
OK as the object in an embedded complement clause	No	Yes
OK as the object in passive structures	No	Yes
Status as the object in topic constructions	??	?

3.5. Two Outstanding Issues

In Section 4, we will reflect on how to interpret the variation found in Vietnamese vs. Chinese in the patterning and analysis of their KNs and BPNs. First, having laid out our claims on the identity of null possessors with KNs and BPNs in Vietnamese, we return to two related remaining issues that are in need of some discussion, posed as questions in (66) and (67).

(66) If the null possessors of BPNs are indeed pronominal, why are they sometimes degraded in topic constructions?

(67) How does focus and contrast improve degraded topic constructions with BPN/KNs?

Concerning (62), we have reported that speakers often find BPNs unnatural in object position in topic constructions when the antecedent for the BPN is the topic, as schematized in (68) and illustrated with example (8) repeated here:

(68) ?Topic$_k$, Subject V [e$_k$ BPN]

(8) ?Nga$_k$, tôi rất thích [e$_k$ đôi mắt].
Nga I very like eyes
Intended: 'I like Nga's eyes very much'.

If the null possessor of BPNs is an occurrence of pro, the question is why the configuration in (67) should be felt to be (somewhat) unnatural. For Huang (1984, 1989), examples parallel to (8) in Mandarin would be accounted for by means of the GCR—the pro null possessor in the object would need to be controlled by the nearest c-commanding NP, which is the subject and not the topic. However, we are not in a position to invoke the GCR to rule out examples such as (8) in Vietnamese because there are other configurations in Vietnamese that are perfectly well-formed but violate the GCR—BPNs may occur as objects in embedded clauses in two clause structures relating to the matrix clause subject, as already seen in examples (15) and (16) repeated below. If the GCR were to constrain the null possessors of BPNs as instances of pro, the BPN objects in (15) and (16) should not be able to take the matrix clause subject as their antecedents, because there is a closer

c-commanding NP—the subject of the embedded clause. As examples like (15) and (16) are fully well-formed in Vietnamese, we have to conclude that the GCR does not constrain the distribution of BPNs relative to their antecedents and, hence, cannot be responsible for the degraded nature of topic construction examples like (8).

(15) Nam_k nói rằng Nga đã làm gãy [e_k mũi].
Nam say that Nga PST make break nose.
'Nam_k said that Nga broke his_k nose'[6].

(16) Nam_k nói rằng Nga thường chạm vào [e_k chân/mông].
Nam say that Nga often touch into leg/butt
'Nam_k says that Nga often touches his_k leg/butt'.

In considering alternative explanations for the slight deviance of (68), we make the observation that even examples where a BPN occurs in subject position in topic constructions (schematized in (69)) are perceived to sound more natural when interpreted in a context that involves contrast. For example, (6) will be most natural in a context in which the speaker is trying to distinguish 'Nga' from other people.

(69) Topic_k, [e_k BPN] V...

(6) Nga_k, [e_k tóc] rất là dài.
Nga hair very be long.
'Nga's hair is very long'.

The configuration in (69) should be well-formed according to the GCR as a putative constraint on pro null possessors, as (69) is a structure in which the BPN is immediately c-commanded by its antecedent, with no other intervening NPs. The fact that contrast is nevertheless required for instances of structure (69) as well as (68) to sound natural suggests to us that topic constructions of such a type are conditioned by a general pragmatic well-formedness requirement that contrast and focus be present in the construction in order for it to be used in a felicitous and natural way. Such a property may be a distinctive feature of Vietnamese topic constructions involving KNs and BPNs that are construed in relation to the topic (without movement) or perhaps are more widespread among other languages. We note that in English, overt pronominal possessors of BPNs sometimes sound unnatural in topic constructions in the absence of a clearly contrastive context (70) and improve in their naturalness when contrast and focus are added, as in (71), and some speakers of Mandarin have reported similar feelings to us about Chinese equivalents to (69).

(70) ?Tom, Bill looked at his knee.

(71) Tom, Bill looked at his knee. Sue, he stared at her *head.* They had all three been injured in the crash.

The 'explanation' for the degraded status of (69/9) that we would tentatively like to advance is that they are felt to sound unnatural simply due to a perceived lack of contrast and focus, which are properties regularly licensing the use of such non-movement, hanging topic constructions. As noted earlier, when contrast and focus are explicitly added into (68/9), as in (41), repeated here as (72), such structures become well-formed and are felt to be acceptable.[7]

(72) Nga_k thì, tôi chỉ thích [e_k đôi mắt], không thích [e_k tóc].
Nga CT I only like eyes NEG like hair
'I only like Nga's eyes, not her hair'.

This now leads us to our second, related question in (63): 'How does focus/contrast improve degraded examples of BPN/KNs in such constructions?'

An interesting syntactic analysis of this issue, as it manifests itself in Chinese, is developed by Huang and Yang (2024). This work builds on the proposal in Huang (1984, 1989) that null possessors of KNs/BPNs are instances of pro and must be bound by the closest c-commanding antecedent due to the GCR. As mentioned above, structures such as (68), repeated below, where a KN or BPN occurs in object position in a topic construction, will violate the GCR if the null possessor pro attempts to relate to the topic as its antecedent because the subject is a closer antecedent. Examples such as (73) are felt to be unacceptable.

(68) $Topic_m$ $Subject_k$ V [$pro_{k/*m}$ KN/BPN] . . .

(73) *Zhangsan, wo kanjian le [pro hou naoshao]
Zhangsan I saw PERF back of head
Intended: 'As for Zhangsan, I saw the back of his head' (Huang & Yang, 2024, p. 15).

However, when focus is added to examples of this type, as in (74), they are judged to be well-formed:

(74) Zhangsan, wo zhi kanjian le [pro hou naoshao]
Zhangsan, I only saw PERF back of head
OK: 'As for Zhangsan, I only saw the back of his head' (Huang & Yang, 2024, p. 15).

Huang and Yang's novel approach to this patterning is the suggestion that focused objects in examples like (74) undergo QR-like *covert focus movement* to a position higher than the subject, making the topic into the closest antecedent for the null possessor pro, as schematized in (75):

(75) $Topic_m$ [$pro_{k/*m}$ KN/BPN]$_p$ $Subject_k$ V t_p …

Huang and Yang's covert focus movement proposal offers a principled and genuinely interesting analysis of focus effects with KNs and BPNs in object positions in Mandarin and might be thought to be potentially appropriate for Vietnamese, too. However, we will not adopt such an analysis for Vietnamese due to the following concerns. First, Huang and Yang's modeling critically assumes the GCR as a restriction on pro in Chinese (LF focus movement serves to overcome the GCR-violating configuration in (68)), but in Vietnamese, the GCR seems to be violated regularly with BPNs in Vietnamese two-clause and passive constructions—see the examples (15) and (16) repeated above, and the passive examples (24) and (25) below. In all these instances, a null possessor pro is able to take as its antecedent an NP that is not the closest c-commanding NP:

(24) Nga_k bị Nam giật [e_k tóc].
Nga PASS Nam pull hair
'Nga_k had her_k hair pulled by Nam' (Simpson & Ho, 2013, p. 162).

(25) Nam$_k$ bị Nga tát vào [e$_k$ mặt].
Nam PASS Nga slap on face
'Nam's face was slapped by Nga'.

We, therefore, conclude above that the GCR does not govern the distribution of BPNs relative to their antecedents. It consequently cannot be the cause of degraded topic construction examples such as (8), and we are not able to invoke LF focus movement of objects over subjects as an explanation for the amelioration effects of focus and contrast in these constructions.

Instead, we would like to briefly sketch a potential *information structure* approach to this patterning, which is currently a speculation but one that we feel is deserving of consideration. It has frequently been remarked that null elements such as pro may often be naturally linked to continuing *discourse topics*—the reference of pro corresponds to the ongoing discourse topic. When topic constructions of the type in (64) [Topic, Subject V [$_{OBJ}$ e BPN]] are presented to listeners without any embedding context/prior discourse, it is plausible that listeners will tend to interpret the initial NP as a new discourse referent rather than as a continuing discourse topic, which may reduce its natural availability as an antecedent for a pro in the continuation of the sentence. The addition of the contrastive topic marker *thì* to the topic will signal that the topic is to be construed as a member of an existing contrast set assumed in the context, not a fully new discourse referent, facilitating its linking to a null possessor element. Additionally, if focus is placed on a following BPN, this will serve to highlight an intended *pairing* of the BPN with the topic, as it does in other contrastive topic constructions, for example: 'As for FISH, I only eat *salmon*'. The combination of contrast on a topic (with *thì*) and focus on a BPN can, therefore, be suggested to better support interpretations in which null possessors of BPNs are construed as *connected* to (coreferential with) topic NPs, resulting in the observed improvement 'effect' (and we remind readers that many speakers find that special context and focus/contrast is also necessary to make BPNs and KNs in the *subject* position sound natural as well—configuration (65))—hence such information structure licensing requirements would apply to KN/BPN topic constructions in general, not simply the object cases corresponding to (64)). While such a pragmatic approach to the effects of focus/contrast in these constructions certainly needs to be fleshed out further, we are hopeful that it ultimately will allow for a plausible (non-syntactic) understanding of the focus alternations noted here for Vietnamese (and see also (Ke, 2023) for a useful related discussion of discourse structure constraints on the construal of BPNs in topic constructions in Chinese).

4. Conclusions and Directions for Future Work

The primary goal of the current study is to develop a description and understanding of the occurrence of bare nouns encoding kinship terms and body parts in Vietnamese, investigating how such elements relate to antecedent possessors present in a variety of linguistic structures. In the course of the investigation, we highlighted how a variety of pragmatic factors may affect the natural use and interpretation of bare noun KNs and BPNs beyond other properties relating to the fundamental identity of null possessors projected with KNs and BPNs, which we take to be syntactic. Attempting to control for discourse factors as well as possible, we reached a number of generalizations concerning the basic patterning of KNs and BPNs in topic, two-clause, and passive structures. Collectively, this information reveals that KNs and BPNs have significantly different properties in Vietnamese and that KNs and BPNs in Vietnamese are also different from KNs and BPNs in Mandarin Chinese. We analyzed these differences as resulting from distinctions in

the null possessors taken to be present with bare noun KNs and BPNs, with KNs in Vietnamese projecting a null possessor that is a local anaphor and BPNs occurring with a null pronominal/pro possessor. This empirically-driven analysis of Vietnamese contrasts with conclusions reached about Mandarin in Ke and Pires (2022) and Ke (2023), where the null possessors of KNs are argued to be covert long-distance anaphors, and those with BPNs are shown to pattern like local anaphors. These characterizations of null possessors in Vietnamese and Mandarin are shown in Table 3.

Table 3. The syntactic identity of null possessors in Vietnamese and Mandarin.

	Vietnamese	**Mandarin**
KNs	local anaphor	long-distance anaphor
BPNs	pro	local anaphor

One important result of the present study is, consequently, the conclusion that null possessors projected by bare noun KNs and BPNs do not have the same syntactic identity and distribution across languages. The current investigation of Vietnamese, therefore, contributes to the initiation of descriptions of cross-linguistic variation among KN/BPN null possessors—the development of a typology of bare KNs/BPNs, which may hopefully grow with studies of other languages.

One might naturally also ask if the conclusions about Vietnamese KNs/BPNs relative to Mandarin are surprising or even concerning. From a certain perspective, Vietnamese and Mandarin are interestingly observed to show *opposite* patternings with null possessors, as noted in (76).

(76) Vietnamese KNs are more restricted in their distribution than BPNs.
Chinese BPNs are more restricted in their distribution than KNs.

We suggest that there is actually no theoretical reason why the null possessors of KNs vs. BPNs should be expected to pattern uniformly within a single language or in an identical way across languages with regard to locality (resulting from syntactic identity)—in other words, no reason why KNs should necessarily be consistently more or less restricted in their locality than BPNs. Possessors and possession relations may be structurally realized in different ways across languages, as found, for example, in the variation that occurs in 'possessor raising' patterns both across and within individual languages (Deal, 2013; Tyler, 2019). We submit that the syntactic identity of null possessors with kin terms and body part nouns may well vary across languages, just as the identity of other types of null elements in argument positions may be non-uniform and result from different instances of pro, ellipsis, operations of movement, etc. In brief, there is no fundamental property of KNs and BPNs that should require their possessors to exhibit fully parallel morpho-syntactic properties either within or across languages, and variation, when attested, should not be viewed as strikingly unexpected.

Though KNs and BPNs may consequently develop null possessor syntax that is independent and distinct, we believe that a closer parallelism might be anticipated to occur between the patterning of BPNs and a third 'type' of noun which involves a regular possession relation—Part-Whole Nouns/PWNs, such as those illustrated in (77) (PWNs underlined):

(77) 'a/this car and its <u>steering wheel</u>'
'a/this book and its <u>cover</u>',
'a/this house and its <u>roof</u>'

BPNs are, in essence, simply the animate equivalent of a Part/Whole relation, with an animate/human rather than an inanimate antecedent possessor—in both instances, there is a quasi-inalienable relation of a part to a whole:

(78) 'the man—his arm'
'John—his hair'
'Nga—her eyes'

Given the semantic similarity between BPNs and PWNs, it might, therefore, be expected that BPNs and PWNs would exhibit a similar/parallel distribution and be assumed to project null possessors with shared properties. This, indeed, is found to be the case in Vietnamese. PWNs in Vietnamese pattern very much like BPNs and are very free in their distribution. They do not need to be c-commanded by their antecedents, like BPNs:

(79) Khi mấy tên cướp biển lên được tàu, chúng đã cướp cái mỏ neo.
When pirates board ship PL PST steal CL anchor
'When the pirates boarded the ship, they stole its anchor'.

(80) Sau khi tên lửa cất cánh, khoang nhiên liệu rơi xuống Trái Đất.
after rocket took off, fuel pod fell to Earth.
'After the rocket took off, its fuel-pod fell to earth'.

Additionally, PWNs can occur separated from their antecedents by other, closer NPs, as BPNs can:

(81) Quyển sách này, tôi rất thích trang bìa.
CL book DEM I very like cover
'This book, I like its cover a lot'.

(82) Chiếc xe này, tôi rất thích cái bánh xe sau.
CL car DEM, I very like CL steering wheel
'This car, I like its steering wheel a lot'.

In this respect, it may be suggested that Vietnamese internally follows an expected patterning. Intriguingly, Mandarin Chinese is different in this regard. Ke (2023) reports that PWNs in Mandarin are very free in their distribution, as in Vietnamese. However, this contrasts with the distribution of BPNs in Mandarin, which is much more restricted and analyzed as resulting from the presence of null possessors, which are local anaphors. We believe it will be interesting for future studies on KNs and BPNs to further track the relation of BPNs to PWNs so as to establish whether Vietnamese indeed presents a more unmarked pattern, with a close distributional connection between BPNs and PWNs.

Another potentially fruitful topic for future studies focused on Vietnamese and certain other languages may be an investigation of the use of *classifiers* with KNs and BPNs. In the course of the present study, it was observed that, in certain instances, KNs and BPNs may only be acceptable/natural if a classifier is combined with them (without any numeral), resulting in a [CL + N] constituent—a 'bare classifier' structure (L. L.-S. Cheng & Sybesma, 2005; Simpson et al., 2011; Jenks, 2018). Hence, the use of a classifier (which is optional in other contexts) makes an unacceptable structure acceptable. This improvement effect is illustrated in (83–85) below, where the target sentences are felt to be degraded without the presence of a classifier on the KN or BPN:

(83) A: Nam is very upset right now.
B: Why? What happened?
A: Vài kẻ bắt cóc đã bắt giữ ?(đứa) con trai vào hôm qua.
Some kidnapper PST seize CL son yesterday.
'Yesterday some kidnappers seized Nam's son'.

(84) A: What's wrong with Roro? Why is he howling?
B: Lỗi của tôi. Tôi dẫm lên ?(cái) đuôi.
fault of I. I step on CL tail.
'Its my fault. I stepped on his tail'[8].

(85) Nga_k thì tôi thậm chí còn chưa thấy mặt ?(người) bạn trai.
Nga thì I even still not yet see face CL boyfriend.
'I still haven't even seen Nga's boyfriend'.

This naturally raises the question of why/how the use of classifiers sometimes improves degraded occurrences of BPNs/KNs. At this point, pending further careful investigation, which will have to be carried out in future work, we speculate that the use of classifiers with KNs and BPNs may enable a different mode of linking between nouns and 'antecedent' NPs, similar to the use of definite articles in place of overt possessors in other languages, which is not subject to obvious locality or c-command requirements. In languages such as English, KNs and BPNs most commonly occur with possessive adjectives such as 'his father/leg', but in certain contexts, such modifiers may also be replaced with the definite determiner 'the' resulting in [Det N] constituents of the type 'the father/the leg', as illustrated in (86–87). PWNs also regularly permit the use of definite determiners as an alternant to possessive adjectives (88):

(86) KNs. Context: A teacher speaking to a colleague about who is scheduled to collect a group of children from school after a special activity:
'As for this student, his/the father is going to pick him up'.

(87) BPNs. Context: A nurse is reporting information about patients to a doctor who is touring a hospital ward:
'As for this patient, his/the leg was injured by falling debris'.

(88) PWNs. Context: Someone considering purchasing a house comments on the positive and negative features of individual houses:
'As for this house, I like its/the roof and its/the windows'.

In Vietnamese, simple classifier + noun combinations ('bare classifier' constructions without numerals [CL N]) frequently have definite interpretations that equate with determiner + N pairs in English. We believe that this is an interesting parallel to explore further—whether the referential linking of possessed KNs, BPNs, and PWNs realized as [CL+N] pairs in Vietnamese converges with the use of [Det+N] forms with possessed KN/BPN/PWN nouns in English and makes use of a common mode of association with antecedent possessors which appears to be less constrained than the use of null possessors. As bare classifier constructions also occur in various regional varieties of Chinese (including Cantonese and many varieties of Wu Chinese—L. L.-S. Cheng & Sybesma, 2005; Li & Bisang, 2012; Jian, 2015) as well as north Indic languages (including Bangla, Assamese, Oriya—Dayal, 2012; Simpson & Biswas, 2016) and Hmong (Bisang, 1993), comparative work on the use of [CL+N] patterns as alternants to null possessor patterns with KNs and BPNs could very usefully be into extended into Sinitic and Indian languages. We will need to leave such a broad typological investigation for a future time.

Author Contributions: Conceptualization, A.S. and L.P.; methodology, A.S. and L.P.; validation, A.S. and L.P.; investigation, A.S. and L.P.; resources, A.S. and L.P.; writing—original draft preparation, A.S.; writing—review and editing, A.S. and L.P.; formating, L.P.; supervision, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The eight Vietnamese consultants contacted by the authors agreed to share their intuitions on data examples.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: We gratefully acknowledge the Vietnamese consultants for their valuable judgments and insightful suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

1 The Vietnamese patterns described in the paper represent the collective judgments of the second author of the paper and eight other native speakers consulted (in total, three speakers from the north of Vietnam, two from central Vietnam, and four from the south part of Vietnam).

2 Where the potential occurrence of overt subjects, modals, and complementizers is assumed to correlate with the finiteness of a clause.

3 It is also possible for the lower subject, Nga, to be construed as the antecedent of *mũi* 'nose'. We are interested in probing potential long-distance construals of KNs/BPNs and their antecedents, so we do not represent local construals between an object KN/BPN and subject antecedent in the same clause, which are generally always available.

4 There are no parallel effects with KNs because the use of other-directed verbs does not eliminate their subjects from being potential antecedents for a KN in object position.

5 Ellipsis may also sometimes be diagnosed via the possibility of extraction of an element from within an ellipsis site, indicating that a gap position contains a complex syntactic structure and is not just a null pro-form. Unfortunately, for various reasons, such a test cannot be reliably checked with null possessor gaps in Vietnamese.

6 See Note 3

7 If such intuitions are on the right track, the contrast felt between (68) and (69) still needs some understanding—why is it easier for structure (69) to be interpreted as involving contrast/focus without this explicitly being added in? We suspect that one relevant factor here may be a potentially confusing perspective shift that arises in (68), where two fully distinct NP referents occur in the topic and subject positions, creating uncertainty as to which referent the sentence is to be interpreted relative to—the topic or the subject. In (69), by way of contrast, the subject does not introduce a second individual into the discourse/sentence, just an NP relating to the topic, and consequently, there is less potential confusion relating to point-of-view and how the predication of the sentence is to be understood/evaluated.

8 Examples (83/84) were originally created to test whether null topics might pattern differently from overt topics in the topic configuration. For this reason, (83/84) have an additional embedding context—the first sentence was added to set up a new discourse topic in the following target sentence ('Nam' in (83), 'Roro' in (84)). The use of a null topic turned out not to improve the acceptability of topic structures with bare noun KNs and BPNs, but we noticed that adding in a classifier did seem to have an interesting effect of improvement.

References

Bisang, W. (1993). Classifiers, quantifiers and class nouns in Hmong. *Studies in Language*, *17*(1), 1–51. [CrossRef]

Cheng, H.-T. (2013). *Argument ellipsis, classifier phrases, and the DP parameter* [Doctoral dissertation, University of Connecticut].

Cheng, L. L.-S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. In G. Cinque, & R. S. Kayne (Eds.), *The Oxford handbook of comparative syntax* (pp. 259–292). Oxford University Press.

Chou, C.-T. T., & Phan, T. to appear. Vietnamese mình: Attitudes, empathy and the blocking effect. *Language and Linguistics*.

Dayal, V. (2012, March 17–18). *What can South Asian Languages tell us about classifier systems?* FASAL-2, Cambridge, MA, USA.

Deal, A. R. (2013). Possessor raising. *Linguistic Inquiry*, *44*, 391–432. [CrossRef]

Huang, C.-T. J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, *15*, 531–574.

Huang, C.-T. J. (1989). Pro-drop in Chinese: A generalized control theory. In O. Jaeggli, & K. Safir (Eds.), *The null subject parameter* (Vol. 15). Studies in Natural Language and Linguistic Theory. Springer.

Huang, C.-T. J. (1999). Chinese passives in comparative perspective. *Tsing Hua Journal of Chinese Studies*, *29*, 423–509.

Huang, C.-T. J., & Yang, B. (2024). Locality, focus and covert movement. *Journal of East Asian Linguistics*, *33*, 3–33. [CrossRef]

Ivan, R., & Bui, T. (2019). Vietnamese anaphora: Binding principles and the lack thereof. *Proceedings of Triple A*, *5*, 47–61.

Jenks, P. (2018). Articulated definiteness without articles. *Linguistic Inquiry*, *49*, 501–536. [CrossRef]

Jian, W. (2015). Bare classifier phrases in Sinitic languages: A typological perspective. In H. Chappell (Ed.), *Diversity in sinitic languages* (pp. 110–133). Oxford University Press.

Ke, A. H. (2023). Syntax and semantics of NPs in Chinese possessive topic constructions. *Journal of East Asian Linguistics*, *32*, 133–167. [CrossRef]

Ke, A. H., & Pires, A. (2022). Local versus long-distance bound implicit arguments of inalienable relational nouns in Chinese. *Journal of Linguistics*, *58*, 269–305. [CrossRef]

Li, X., & Bisang, W. (2012). Classifiers in Sinitic languages: From individuation to defininteness-marking. *Lingua*, *122*, 335–355. [CrossRef]

Sag, I. (1976). *Deletion and logical form* [Ph.D. dissertation, MIT].

Shi, D. (1997). Issues on Chinese passive. *Journal of Chinese Linguistics*, *25*(1), 41–69.

Simpson, A., & Biswas, P. (2016). Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis*, *40*(3–4), 167–198.

Simpson, A., & Ho, T. (2013). Vietnamese and the typology of passive constructions. In D. Hole, & E. Loebel (Eds.), *Linguistics of vietnamese* (pp. 155–185). Mouton de Gruyter.

Simpson, A., & Ngo, B. (2022). Null anaphora in Vietnamese: Pro and argument ellipsis. In C. Shei, & S. Li (Eds.), *The Routledge handbook of Asian linguistics* (pp. 186–203). Routledge.

Simpson, A., Soh, H. L., & Nomoto, H. (2011). Bare classifiers and definiteness: A cross-linguistic investigation. *Studies in Language*, *35*, 168–193. [CrossRef]

Takahashi, D. (2008a). Noun phrase ellipsis. In S. Miyagawa, & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 394–422). Oxford University Press.

Takahashi, D. (2008b). Quantificational null objects and argument ellipsis. *Linguistic Inquiry*, *39*, 307–326. [CrossRef]

Tyler, M. (2019). Two kinds of "possessor raising" in Choctaw. In D. K. E. Reisinger, & R. Y.-H. Lo (Eds.), *Proceedings of the workshop on the structure and constituency of languages of the Americas 23* (pp. 106–120). UBCWPL.

Williams, E. (1977). Discourse and logical form. *Linguistic Inquiry*, *8*(1), 101–139.

Zhang, N. N. (2009). The syntax of relational-nominal second constructions in Chinese. In *Yuyanxue Luncong* [Essays on Linguistics] (Vol. 39, pp. 257–301). Shangwu Yingshu Guan.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

MDPI

Article

Varieties of Polar Question Bias: Lessons from Vietnamese

Michael Yoshitaka Erlewine [1,2,*] and Anne Nguyen [1]

[1] Department of English, Linguistics, and Theatre Studies, National University of Singapore, 7 Arts Link, Singapore 11571, Singapore
[2] Helsinki Collegium for Advanced Studies, University of Helsinki, Fabianinkatu 24 A, 00014 Helsinki, Finland
* Correspondence: mitcho@mitcho.com

Abstract: This paper describes the use conditions of different polar question constructions in Vietnamese and their consequences for the description and analysis of polar question bias. We argue that the behavior of questions with the final particle *à* highlights the utility and relevance of the theoretical notion of *projected bias* for describing polar question bias, distinct from and in addition to *original bias* based on the speaker's prior beliefs and *contextual bias* based on evidence available in the context. We also argue that some but not all bias requirements may be described as due to pragmatic competition between different question forms.

Keywords: Vietnamese; polar questions; question bias; projected bias

1. Introduction

Languages frequently have more than one morphosyntactic strategy for polar question formation. In particular, many polar question constructions are limited in their use to situations where the propositions p and $\neg p$ do not have equal standing, hence reflecting some form of *bias* between these two alternatives. Question bias is not simply a matter of privileging one possibility in $\{p, \neg p\}$ over the other; for instance, particular constructions may make reference to the speaker's prior beliefs (Ladd, 1981, et seq) or to evidence available from the conversational context (Büring & Gunlogson, 2000, et seq). See Bill and Koev (2023) and Romero (2024) for two recent overviews, focused on the facts of English polar question forms and the history of their analysis.

In this short paper, we contribute to this literature by describing the bias profiles of several polar question constructions in Vietnamese. We first introduce the morphosyntax of polar questions in Vietnamese in Section 2 and then present the bias profiles of several such constructions in Section 3. We then discuss two consequences of our study that have broader relevance for the description and analysis of question bias:. First, in Section 4, we discuss the use conditions of *à* particle questions, which highlights the importance of Bill and Koev's (2023) notion of *projected bias*, which reflects the speaker's current expectation for how the question will be resolved. Second, in Sections 5 and 6, we argue that some but not all of the bias requirements of Vietnamese polar question constructions may be derived from pragmatic competition. We then conclude in Section 7.

2. Polar Question Constructions in Vietnamese

We begin by briefly introducing the range of Vietnamese polar question constructions that we will investigate here. Note that we here take a broad, functionally motivated definition of polar questions as those constructions which conventionally elicit an addressee

commitment towards the *prejacent* proposition p or to its negation $\neg p$, without reference to syntactic clause typing distinctions.

Broadly speaking, polar question constructions in Vietnamese can be placed in two categories, motivated both by their syntax and pragmatics. What we will call *standard polar questions* involve a negative marker in a sentence-final position, in contrast to the preverbal position of standard negation; see (1). The negative marker (*không* or *chưa*) must agree with the aspect of the clause, which itself cannot include negation. An affirmative auxiliary *có* frequently precedes the verb phrase in *không* questions.[1]

(1) **Standard polar questions:** (based on Trinh, 2005, p. 30)

a. Nó **có** đọc sách **không**?
he/she AFF read book NEG
'Does he/she read books?'

b. Nó **đã** đọc sách **chưa**?
he/she PFV read book NEG.PFV
'Has he/she read books (yet)?'

The other class of polar question constructions all involve the addition of an invariant and aspectually unrestricted sentence-final expression. These sentence-final expressions can be one of a number of monosyllabic sentence-final particles (2), which we describe further below, or the expression *(có) phải không* (3), which can also be used as an independent polar question meaning 'Is it right?'. We refer to the former as *particle polar questions* and, following prior work (e.g., Nguyen, 1997, p. 240; Trinh, 2005; Duffield, 2013), refer to the latter as *tag questions*. Unlike in standard polar questions, the main clause in particle polar questions and tag questions may include negation and has no restriction on its aspectual specification.

(2) **Particle polar questions:**

Nó đọc sách { **á** / **à** / **hả** / chăng / ư / sao / nhỉ / … }?
he/she read book Á À HẢ …
≈ 'Does he/she read books?'

(3) **Tag question:**

Nó đọc sách **phải không**?
he/she read book correct NEG
≈ 'He/she reads books, does not he/she?'

As indicated in (2) above, there are many polar question particles, some of which are associated with particular dialects.[2] In our discussion below, we will concentrate on the uses of *á*, *à*, and *hả*, bolded in (2), as well as the tag question with *phải không* as in (3). These are the forms for which the speakers that we consulted (mostly speakers of Northern varieties) and the second author (from Hanoi, and hence, also Northern) could most comfortably give clear judgments of felicity or infelicity in context, and the discussion of their use conditions will suffice to motivate our theoretical consequences. Particle polar questions and tag questions are all associated with some form of bias, unlike the standard polar question, which is in some sense neutral, as we describe in detail below.

Finally, we note that standard polar questions are the only one of these polar question forms that can be embedded.[3] See the contrast between examples (4a,b) below, as well as Phan (2024, pp. 169–170) for additional supporting data. We will return to this difference again in Section 5.

(4) **Only standard polar questions can be embedded:** (based on Trinh, 2005, p. 31)

a. Tôi muốn biết [nó **có** đọc sách **không**].
I want know he/she AFF read book NEG
'I want to know whether he/she reads books.'

b. * Tôi muốn biết [nó đọc sách { **á** / **à** / **hả** / … / **phải không** }].
I want know he/she read book Á À HẢ correct NEG

3. Varieties of Question Bias

We now discuss different notions of question bias and demonstrate their relevance for describing the use conditions of different polar question particles in Vietnamese. Prior work on question bias in English and other languages has established the need to distinguish between bias due to the speaker's initial epistemic state versus bias due to evidence from the conversational context. We follow Domanesci et al. (2017), Bill and Koev (2023), and Romero (2024) in referring to these as *original (speaker) bias* and *contextual (evidence) bias*, respectively. We introduce each in turn in Sections 3.1 and 3.2, adopting the definitions provided by B&K, which are designed to facilitate their comparison and joint discussion. We then provide an interim summary in Section 3.3.

Bill and Koev (2023) additionally introduced a third notion, which they called *projected bias*. We will introduce this notion and argue for its relevance in Section 4 below.

3.1. Original Bias

The first notion of bias we introduce, *original bias*, reflects the speaker's prior beliefs or knowledge regarding the status of the proposition p. The relevance of this type of bias has been established in works such as Ladd (1981).

(5) **Original bias** (definition from B&K, p. 4)
A question Q uttered in a context c conveys *original bias* for a possible answer p iff Q indicates that the prior epistemic state of the speaker in c supports p over $\neg p$.

We can systematically test for sensitivity to original bias by explicitly manipulating the presence or absence of reasons for the speaker's prior belief in minimally contrasting situations, as in S1–S3 below. For ease of presentation, we present these situations here in English and underline crucial points of manipulation.

(S1) **Situation 1: No bias**
You are sitting in a windowless room with no information about current weather conditions. You call a co-worker in a room that has a window to check on the weather.

(S2) **Situation 2: Original bias for *p***
You are sitting in a windowless room. A while ago, a friend told you on the phone that it was rainy outside. You call a co-worker in a room that has a window to check.

(S3) **Situation 3: Original bias for ¬*p***
You are sitting in a windowless room. A while ago, a friend told you on the phone that it was sunny outside. You call a co-worker in a room that has a window to check.

The effect of original bias is reflected in the felicity and infelicity of different question forms, as in (6–7). As we see below, the standard polar question must be neutral with respect to original bias, whereas the *à* and *hả* particle questions and the *phải không* tag question require positive original bias.[4] (We also introduce *á* particle questions below).

(6) **Standard question requires no bias:**
Trời (có) đang mưa **không**?
it AFF PROG rain NEG
≈ 'Is it raining?' S1 ✓ S2 # S3 #

(7) **Original bias for *p* licenses *à* and *hả* particle questions and tag questions:**
Trời đang mưa { **à** / **hả** / **phải** **không** }?
it PROG rain À HẢ correct NEG
≈ 'Is it raining?' S1 # S2 ✓ S3 #

3.2. Contextual Bias

A speaker might also be positively or negatively disposed towards a proposition based on shared evidence that is available in the present conversational context, as discussed in, e.g., Büring and Gunlogson (2000). We call this *contextual bias*:

(8) **Contextual bias** (definition from B&K, p. 4)
A question Q uttered in a context c conveys *contextual bias* for a possible answer p iff Q indicates that some piece of evidence that has just become available in c supports p over $\neg p$.

Situation 4 below is modeled after contexts discussed in Gunlogson (2001, pp. 95–96), parallel to S1 above in providing no original bias but with a manipulation providing contextual bias for p. S4 licenses the use of *à* and *hả* particle questions, but not the standard question or tag question, as indicated in (9) below.[5]

(S4) **Situation 4: Contextual bias for p**
You are sitting in a windowless room. You see your co-worker enter the office wearing a wet raincoat.

(9) **Felicity judgments for question forms in S4, with contextual bias for p:**

Trời đang mưa { **#không** / **✓à** / **✓hả** / **#phải không** }?
it PROG rain NEG À HẢ correct NEG
≈ 'Is it raining?'

This difference in felicity between the *à* and *hả* particle questions and the *phải không* tag question, which all pattern together with respect to original bias (7), establishes the relevance of contextual bias in addition to original bias.[6] Note that we will also show in Section 4 below that *à* and *hả* particle questions do not always pattern together.

For completeness, we demonstrate that none of the question forms considered here are licensed in a situation that provides *negative* contextual bias, in S5, just as with S3 above, which provides negative original bias. In such situations, speakers may form a polar question with a different prejacent instead. Note that a red, sweating face indicates a hot, sunny day, and therefore, no rain in the shared background of our speakers.

(S5) **Situation 5: Contextual bias for $\neg p$**
You are sitting in a windowless room. You see your co-worker enter the office with a red sweating face.

A conversational context might also support both original and contextual biases, in particular pointing to opposite conclusions. The last particle question we describe here, with final *á*, is specifically licensed in such situations, but not in situations providing only original or contextual bias alone:

(S6) **Situation 6: Original bias for $\neg p$ and contextual bias for p**
You are sitting in a windowless room. A while ago, a friend told you on the phone that it was sunny outside. You see your co-worker enter the office with a wet raincoat.

(10) ***Á* question requires both original bias for $\neg p$ and contextual bias for p:**

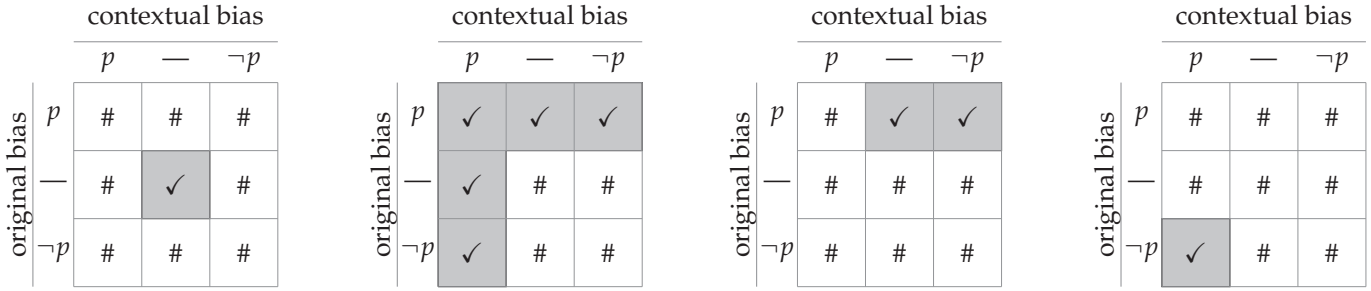
Trời đang mưa **á**?
it PROG rain Á
≈ 'Is it raining?' S1 # S2 # S3 # S4 # S5 # S6 ✓

3.3. Interim Summary

Having introduced the notions of original and contextual bias, as well as methods for their contextual manipulation, we now expand on the data above by considering nine types of situations, which systematically cross positive, negative, and no bias of these two types.[7] See Figure 1 below. We report the felicity of each polar question type discussed in this section in each of these nine situations in Figure 2 below. These figures give a visual representation of the *bias profile* associated with each polar question form, which we also summarize below each figure.

(You are sitting in a windowless room.)	contextual bias: p	contextual bias: —	contextual bias: $\neg p$
original bias	You see your co-worker enter the office wearing a wet raincoat.	You call a co-worker in a room that has a window to check.	You see your co-worker enter the office with a red sweating face.
p — A while ago, a friend told you on the phone that it was raining outside.		S2	
— — You have no information about current weather conditions.	S4	S1	S5
$\neg p$ — A while ago, a friend told you on the phone that it was sunny outside.	S6	S3	

Figure 1. Crossing original and contextual bias for p = 'it's raining'.



(**a**) Matrix standard: must be unbiased

(**b**) *hả*: original or contextual bias for p

(**c**) *phải không*: private bias for p

(**d**) *á*: surprise towards p

Figure 2. Felicity of each polar question construction.

One generalization that these figures help make apparent is that no polar question construction with prejacent p is felicitous in a situation where there is only negative bias, whether original or contextual; see the right, bottom, and bottom right cells. Practically, speakers use the prejacent's negation or yet another proposition to form polar questions in such situations. That is, if there is original or contextual bias *against* the proposition 'it's raining', a speaker may form a polar question based on a prejacent such as 'it's sunny'

instead, or with the negation 'it isn't raining' as the prejacent in the case of particle and tag questions. There is also a similar asymmetry in the use conditions of *á*, which specifically reflects contextual evidence for *p* when *p* is originally unexpected (bottom left cell)—which we may succinctly describe as expressing "surprise" towards *p*—but is infelicitous when original and contextual bias conflict in the opposite way (top right cell).

4. Projected Bias and *À* Questions

In addition to the notions of original and contextual bias, Bill and Koev (2023) introduced a third notion that they called *projected bias*, as defined in (11). In this section, we will discuss this third notion of question bias and argue for its relevance in the description of *à* particle questions.

(11) **Projected bias** (definition from B&K, p. 4)
A question *Q* uttered in a context *c* conveys *projected bias* for a possible answer *p* iff *Q* indicates that the speaker in *c* expects answer *p* rather than $\neg p$.

As B&K discussed, projected bias has a somewhat different status from both original and contextual bias above, both in its status within the literature and as a conceptual notion. First, the idea of projected bias has appeared as a component of particular theoretical analyses of certain bias requirements (as in Krifka, 2015; Malamud & Stephenson, 2015; and Jeong, 2018), rather than as a descriptive notion itself. Second and relatedly, the notion of projected bias is different in being "forward-looking" (that is, "forward-looking as it is about the answer that is expected to be chosen by the addressee," B&K p. 4), and in requiring reference to a mental model of the discourse, unlike both original and contextual bias, which may be described solely in terms of the speaker's own epistemic state without reference to the discourse itself.

Concretely, in cases where there is only original or contextual bias, or where both biases are aligned, we may expect projected bias to follow these biases. However, in situations where original and contextual bias conflict (that is, where one suggests *p* and another suggests $\neg p$), the direction of projected bias is in principle open. Such situations are exactly where we might see subtle contrasts in the acceptability of constructions that are sensitive to projected bias.

We argue that the behavior of Vietnamese *à* particle questions highlights the value of this notion of projected bias, as they are best described as felicitous when there is projected bias for *p*. Recall that we have already seen that *à* particle questions are licensed by positive original bias but not with no bias (see S1 and S2 in (7)), or else by positive contextual bias (as in S4; see (9)). What will be particularly informative is contrasts such as in (12–13) below. Both contexts provide original bias for $\neg p$ and contextual bias for *p*, but they differ in the speaker's understanding of the nature of this conflict and its resolution.

(12) **Surprised but willing to believe *p*:**
Minh is a classmate at school. He's discussed his family before but did not mention any siblings, so you think he's an only child. Your friend tells you that Minh went to the airport to pick up his (Minh's) sister. You reply:

Minh có chị { ✓**à** / ✓**á** / ✓**hả** }? Tớ cứ tưởng nó là con một.
Minh have sister À Á HẢ I thought he COP only.child
'Minh has a sister? I thought he's an only child.'

(13) **Surprised and unwilling to believe *p*:**
You and Minh have been friends since grade school, so you have known that he's an only child. But you have been out of town recently and had not heard that Minh's parents have adopted a girl. Your friend tells you that Minh went to the airport to pick up his (Minh's) sister. You reply:

Minh có chị { **#à** / **✓á** / **✓hả** }? Không thể nào!
Minh have sister À Á HẢ NEG possible PRT
'Minh has a sister? That's impossible!'

The *à* question is felicitous in (12) because, based on the friend's claim and their general reliability, the speaker is willing to believe that their friend must know something that they do not. The speaker reasons therefore that the addressee would resolve the question of $\{p, \neg p\}$ in the direction of p: that is, that there is projected bias for p, and hence, the *à* question form is licensed. In contrast, in (13), the speaker has a strong original belief that $\neg p$, and hence, does not expect p to be the answer of $\{p, \neg p\}$. There is then no projected bias for p and so the *à* question is infelicitous.

Empirically, our description here improves upon previous descriptions of the bias profiles of these polar question forms in Vietnamese. Trinh (2010) notes that *à* particle questions can be used to question a proposition that was presupposed by another speaker's utterance but is not common ground. Both contexts in (12–13) satisfy this description, but the *à* question is not felicitous in (13) (Trinh does not discuss other polar question particles). In addition, some works have described *à* and *hả* as equivalent and simply associated with different dialects (see, e.g., Tran, 2009, p. 18), but we see in these examples that their behaviors diverge. The *hả* question is felicitous in both of (12–13), as it simply tracks the existence of original or contextual bias for p under our description.

The notion of projected bias is also relevant in situations without conflicting sources of bias, but instead reflecting a sensitivity to the predictive power of such information. Consider example (14) below. Here, there is original bias for p and no contextual bias, but there is still perceived uncertainty as to how the addressee will reply. In other words, there is original bias for p but no projected bias for p, and the use of an *à* particle question here is accordingly degraded.

(14) **Double-checking context:** (based on Rudin, 2018, p. 37; Rudin, 2022, p. 346)
You and your friend made plans two days ago to get drinks tonight. You have not spoken about it since and worry that the friend may have forgotten.

Chúng ta vẫn đi tối nay { ??**à** / **✓hả** / **✓phải** **không**}
1pl still go tonight À HẢ correct NEG
'We're still on for tonight?'

We can summarize the bias profile of *à* questions as in Figure 3 below. Projected bias for p may be supported by original or contextual bias for p, which then leads the speaker to expect answer p as the resolution of Q. As the mere presence of original or contextual bias does not guarantee projected bias, we shade these cells in Figure 3 with diagonal lines. We can, however, be certain that, in cases with no source of bias for p or else only sources of bias for $\neg p$, there cannot be projected bias for p, and hence, *à* questions will always be infelicitous in these four bottom right cells.

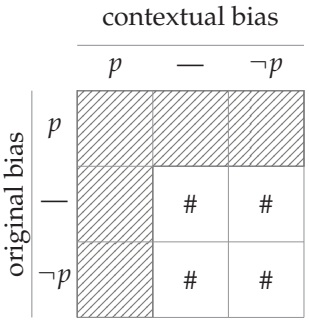


Figure 3. *à* requires projected bias for p, possible in the diagonal shaded cells.

The behavior of Vietnamese *à* questions thus attests to the possibility of a question's use conditions making conventionalized reference to B&K's notion of projected bias, in addition to specifications of original and contextual bias. Therefore, we advocate for the reification of this notion for the description of question bias.[8]

5. Bias in Embedded Questions

Next, we turn to the behavior of embedded questions, which leads to an important consequence for the theoretical analysis of question bias in Vietnamese. We have shown that particle questions with *á*, *à*, and *hả* as well as *phải không* tag questions all exhibit some form of bias, whereas matrix standard polar questions require an unbiased situation. Recall as well that only standard polar questions can be embedded (Duffield, 2013; Phan, 2024; Phan & Starke, 2022, among others), as illustrated in (4a,b) above. Interestingly, embedded standard polar questions are *not* limited to unbiased contexts.

We first illustrate the bias-insensitivity of embedded questions with respect to original bias. The example in (15) includes an embedded standard polar question with prejacent p = 'it's raining', and is felicitous in all three variant contexts below: (a) without original bias, (b) with original bias for p, and (c) with original bias for $\neg p$. The contexts here are based on situations S1–S3 above, but with both the speaker and addressee in a windowless room. (We thank an anonymous reviewer for suggesting this example.)

(15) **Embedded standard question is insensitive to original bias:**
You are sitting in a windowless room.

a. You have no information about current weather conditions. (no bias)

b. A while ago, a friend told you on the phone that it was rainy outside. (original bias for p)

c. A while ago, a friend told you on the phone that it was sunny outside. (original bias for $\neg p$)

You ask your officemate, in the same room as you:

Nếu tớ muốn biết [trời có đang mưa không] thì tớ phải hỏi ai?
if I want know it AFF PROG rain NEG then I must ask who
'If I want to know whether it's raining, who should I ask?' a ✓ b ✓ c ✓

Next, we illustrate this bias-insensitivity with respect to contextual bias, with example (16) below. This utterance embeds a standard polar question with prejacent p = 'Sam got the first prize' and is felicitous in all three contexts: (a) without contextual bias, (b) with contextual bias for p, or (c) with contextual bias for $\neg p$.

(16) **Embedded standard question is insensitive to contextual bias:**
You know Sam wanted to get the first prize in the contest.

a. You don't know how Sam did. (no bias)

b. You see Sam smiling afterwards, so you think he got it. (contextual bias for p)

c. You see Sam frowning afterwards, so you think he didn't get it.
(contextual bias for $\neg p$)

Nếu tớ muốn biết [Sâm có đạt giải nhất không] thì tớ phải hỏi ai?
if I want know Sam AFF get prize first NEG then I must ask who
'If I want to know whether Sam got the first prize, who should I ask?' a ✓ b ✓ c ✓

The striking difference between the bias sensitivity of matrix and embedded standard polar questions lends itself to an analysis in terms of pragmatic competition between different question forms. To facilitate this discussion, we present a summary of the bias profiles of all question forms discussed here, in Figure 4 below. Note that we illustrate the bias profiles for matrix standard questions (a) and embedded standard questions (f) separately.

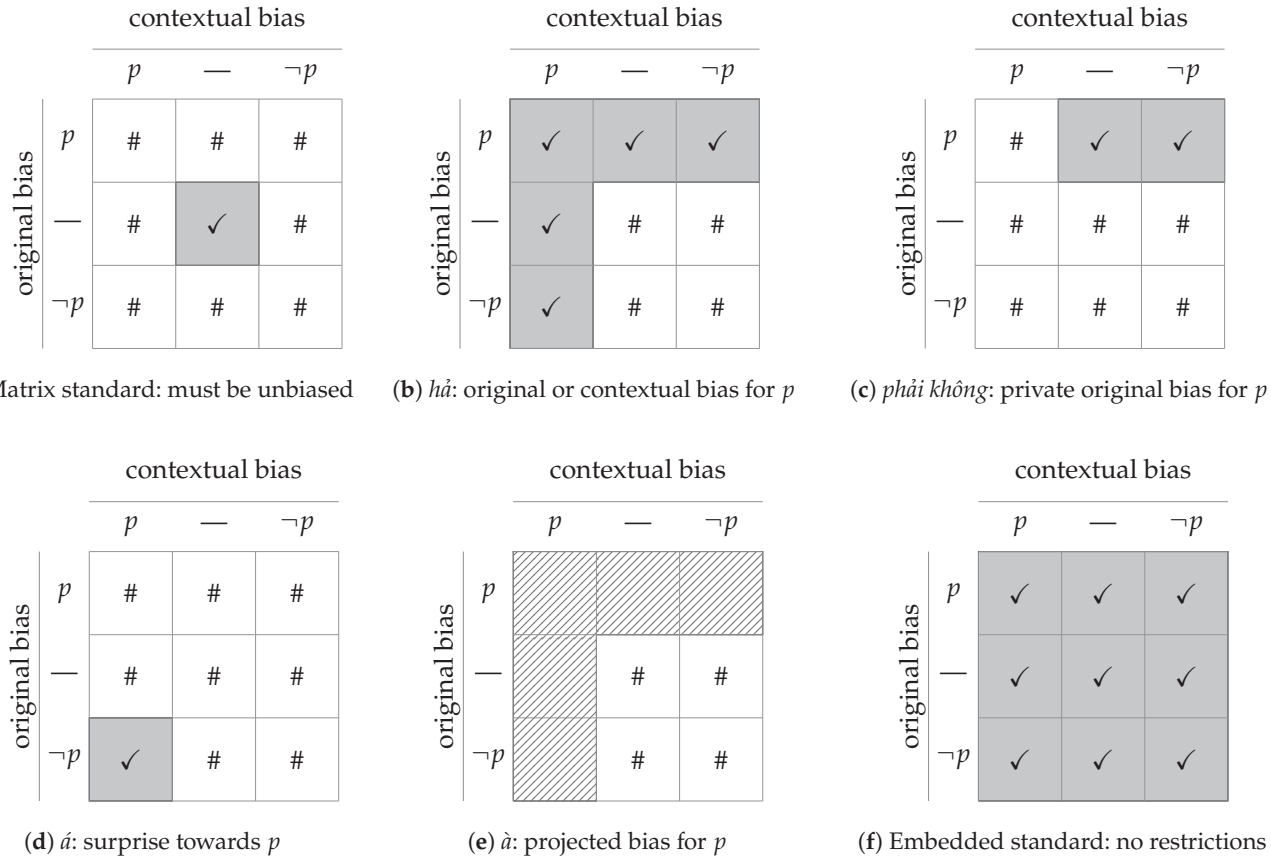


Figure 4. Felicity of each polar question construction ((**a**–**d**) repeated from Figure 2).

Suppose that the basic semantics of standard questions has no conventionalized bias sensitivity. Matrix standard polar questions must be neutral because in contexts with original or contextual bias, a corresponding matrix particle or tag question can be used instead.[9] However, for embedded questions, such pragmatic blocking does not occur, as the particle and tag questions are unavailable.[10] This derives the difference in bias profiles for matrix versus embedded standard questions, as in Figure 4a vs. Figure 4f. Such competition-based accounts have been proposed for explaining the bias profiles of English rising declaratives (Rudin, 2018, 2022) and high negation questions (Goodhue, 2022).

6. The Limits of Pragmatic Competition

Although we have proposed that pragmatic competition can explain the difference in bias-sensitivity between matrix and embedded standard polar questions, we also caution that this type of explanation is unlikely to provide further insight into the bias profiles of other polar question forms.

Concretely, consider Situation 6, repeated in (17) below, where there is both original bias for $\neg p$ and contextual bias for p = 'it's raining'. Such contexts where new, surprising evidence can lead to belief revision are the canonical contexts for *á* particle questions, as we reported above in (10). However, as we indicate here in (17), other question forms with the particles *à* or *hả*, as well as the *phải không* tag question, are all also acceptable here. In other words, it must not be the case that the availability of any of these particle or tag question forms blocks the use of another in the same context.

(17) **Surprise context supports various question forms:**
You are sitting in a windowless room. A while ago, a friend told you on the phone that it was sunny outside. You see your co-worker enter the office with a wet raincoat.

✓ Trời đang mưa { **á** / **à** / **hả** / **phải** **không** }?
it PROG rain Á À HẢ correct NEG
≈ 'Is it raining?'

This point is of particular note for *hả* questions, which roughly pattern with English rising declaratives in their use conditions. (We thank two anonymous reviewers for noting this similarity between *hả* and English rising declaratives.) Rudin (2018, 2022) has proposed that we can derive the bias profile of English rising declaratives pragmatically, by considering rising declarative utterances as being in competition with corresponding falling declaratives and standard polar questions. The overlapping contexts of use between *hả* questions and other particle polar questions makes it difficult, if not impossible, to import Rudin's account to the analysis of Vietnamese *hả* questions. Although this does not directly challenge Rudin's account for English itself, this comparative analysis demonstrates that it must be possible for a language to arrive at a polar question construction with the same or similar bias profile as English rising declaratives, without the proposed form of pragmatic competition.

We conclude then that the observed bias profiles of Vietnamese polar question constructions reflect a combination of conventionalized (i.e., learned, construction-specific) requirements on the use of individual particle and tag question constructions, as well as dynamically induced restrictions on the use of standard questions when unembedded.

7. Summary

Like many languages, Vietnamese has multiple morphosyntactic strategies for the expression of polar questions, with different pragmatic restrictions on their use. In this short paper, we have investigated the contexts of use for a number of such constructions, including standard (negation-final) questions, *phải không* tag questions, and *á*, *à*, *hả* particle questions, in Northern Vietnamese.

Previous work on biased polar questions in other languages have highlighted the importance of distinguishing between so-called *original bias*, based on the speaker's prior beliefs, and *contextual bias*, based on shared evidence available in the conversational context (see overviews in Bill & Koev, 2023; and Romero, 2024). We have shown that the contexts of use for a number of polar question constructions in Vietnamese can effectively be described in these terms. We additionally argue that the bias profile of *à* particle questions is best described by adopting Bill and Koev's (2023) notion of *projected bias*, reflecting the

speaker's current expectation for how the question will be resolved. As B&K discussed, this notion of projected bias has played a role in the conceptualization of question bias in some previous theoretical proposals, but has not been clearly reified as a necessary notion for the description of question bias.

Our study also contributes to the question of whether, and to what extent, the bias profiles of different question types can be derived in terms of pragmatic competition, as has been pursued in recent work on English questions (see, e.g., Goodhue, 2022; Rudin, 2018, 2022). We observe that embedded standard questions have no bias requirements, whereas matrix standard questions must be unbiased, and suggest that this difference can be captured via pragmatic competition, where competing, more specific forms are available for matrix questions but not in embedded contexts. However, the different bias profiles of individual polar question particles cannot be derived in the same way, thereby highlighting the need for multiple types of mechanisms and explanations for observed bias effects.

Author Contributions: Investigation, M.Y.E. and A.N.; formal analysis, M.Y.E. and A.N.; writing, M.Y.E. and A.N.; funding acquisition, M.Y.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National University of Singapore (grant number A-0007220-01-00) and by a Core Fellowship at the Helsinki Collegium for Advanced Studies to Erlewine.

Institutional Review Board Statement: Data collection followed protocol LA-16-225, approved by the Institutional Review Board of the National University of Singapore.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: For comments and discussion that improved this work, we thank the audience at the 9th TripleA workshop (2022), Judith Tonhauser, Ka-Fai Yip, and our anonymous reviewers. All remaining errors are our own.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used here:

AFF	affirmative
NEG	negative
PROG	progressive
PFV	perfective
PRT	particle
COP	copula
B&K	Bill and Koev (2023)

Notes

1 For discussion of the syntax of standard polar questions and their properties, such as the aspect-agreement effect and unavailability of negation, we refer readers to (Trinh, 2005, 2025; Duffield, 2013; Phan & Starke, 2022; Phan, 2024; Trinh et al., 2024; and Liao & Lin, 2024).

2 For instance, *à* is attributed to Northern and Central dialects (see Tran, 2009, pp. 42–43; Le, 2015, p. 28), whereas Nguyen (1997, p. 240) and Trinh et al. (2024, p. 77), respectively, describe *sao* and *hông* as associated with the Saigon dialect (Southern). See also Alves (2012, pp. 9–10), for some discussion of other forms in Central Vietnamese.

3 Embedded standard polar questions can also be introduced by a complementizer *liệu* (Phan & Starke, 2022, p. 193).

4 Note that the relevant contextual manipulations in S2 and S3 support the plausibility of the speaker having original bias for p or $\neg p$, but do not guarantee it. The speaker may or may not take a particular piece of prior information (here, an earlier report from

a friend) to affect their epistemic state with regard to p and $\neg p$. The judgments we report here are for where the manipulations in S2 and S3 *do* lead to such a relevant change.

For speakers who do *not* take situations S2 and S3 to support a relevant original bias in the speaker's mind, we expect both situations to pattern with S1 in the felicity of particular question constructions

5 An anonymous reviewer has made the interesting observation that the *phải không* tag question improves here if the context is changed to S4′ below. The second author shares this judgment.

(S4′) You now see your co-worker entering your shared office without a raincoat, but you notice clear wet shoe prints on the floor and wet spots on their pants.

We suspect that this variant, with no raincoat but wet shoes and spots, opens up the possibility that the addressee is not aware of the relevant evidence. In this case, it does not count as contextual evidence, which Büring and Gunlogson (2000) describe as what "has just become mutually available to the participants in the current discourse situation" (p. 7), clarifying its non-private nature. This information is then instead a source of private evidence for the speaker, and hence, licenses the use of *phải không*, which is associated with private bias.

6 See also Sudo (2013) for an earlier articulation of the idea that original and contextual bias must be distinguished, based on the behavior of different polar question types in English and Japanese. Note that Sudo (2013) refers to the two notions using the terms *epistemic* and *evidential* bias, which are also found in much of the literature. We choose to avoid these terms, because they are both "'epistemic' in some broad sense," as B&K note (p. 6).

In their study of different polar question types in Medumba (Grassfields; Cameroon), Keupdjio and Wiltschko (2018) motivated a distinction between bias that tracks information from *past* versus *present* situations. From our reading of the facts there, it appears that Keupdjio and Wiltschko's (2018) discussion of past- versus present-oriented bias is another statement of the original versus contextual bias distinction here. We thank Harold Torrence for bringing this work to our attention.

7 Domanesci et al. (2017) and Šimík (forthcoming) have employed a similar mode of presentation for the bias profiles of different polar question strategies, considering nine possible situations that cross three values each of original bias and contextual bias, for English and German and for Serbian and Czech, respectively.

8 Adopting the notion of projected bias may also allow for a reanalysis of cases in the literature that have been described as making conventional reference to the "strength" of biases, i.e., distinguishing between strong vs. weak bias of a particular type. See for example (Oshima, 2017; Keupdjio & Wiltschko, 2018; and Bill & Koev, 2023). Such apparent strength distinctions could perhaps instead be described in terms of the alignment between one form of bias and the overall projected bias. For instance, what these prior works describe as "strong" contextual bias for p may refer to the combination of contextual bias for p as well as projected bias for p, whereas "weak" contextual bias for p may describe situations with contextual bias for p but without projected bias for p, which would have to be due to perceived unreliability of the contextual evidence or the presence of conflicting information. We leave the full exploration of this possibility open for future work.

9 For situations with negative original or contextual bias, the speaker can instead use a particle or tag question built from the negation of p (as these question forms allow negation in their prejacent, unlike standard questions), or else another proposition q that is contextually equivalent or similar to $\neg p$. See the end of Section 3 for discussion of this point.

10 The unembeddability of particle and tag questions may in fact be related to their conventionalized bias. In her discussion of the articulated left periphery of interrogatives, Dayal (forthcoming) discussed English rising declaratives and similar biased polar questions in other languages and proposes that biased questions necessarily involve the outermost possible projection of questions, called the Speech Act Phrase (SAP; see also Speas & Tenny, 2003), which cannot be embedded. Phan (2024) also located a number of unembeddable sentence-final particles in Vietnamese in SAP. We leave the more detailed analysis of Vietnamese particle and tag questions and their unembeddability for future work.

References

Alves, M. J. (2012). Notes on grammatical vocabulary in Central Vietnamese. *Journal of the Southeast Asian Linguistics Society*, *5*, 1–11.

Bill, C., & Koev, T. (2023). *Question bias.* Available online: https://www.corybill.com/publications/BillKoev2023may.pdf (accessed on 14 September 2025).

Büring, D., & Gunlogson, C. (2000). *Aren't positive and negative polar questions the same?* [Unpublished manuscript]. Available online: https://urresearch.rochester.edu/fileDownloadForInstitutionalItem.action?itemId=1347&itemFileId=1635 (accessed on 15 September 2025).

Dayal, V. (Forthcoming). The interrogative left periphery: How a clause becomes a question. *Linguistic Inquiry*. [CrossRef]

Domanesci, F., Romero, M., & Braun, B. (2017). Bias in polar questions: Evidence from English and German production experiments. *Glossa*, *2*(26), 1–28.

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. 127–154). de Gruyter.

Goodhue, D. (2022). Isn't there more than one way to bias a polar question? *Natural Language Semantics*, *30*, 379–413. [CrossRef]

Gunlogson, C. (2001). *True to form: Rising and falling declaratives as questions in English* [Unpublished doctoral dissertation]. University of California Santa Cruz.

Jeong, S. (2018). Intonation and sentence type conventions: Two types of rising declaratives. *Journal of Semantics*, *35*, 305–356. [CrossRef]

Keupdjio, H., & Wiltschko, M. (2018). Polar questions in bamileke medumba. *Journal of West African Languages*, *45*(2), 17–40.

Krifka, M. (2015). Bias in commitment space semantics: Declarative questions, negated quetions, and question tags. *Proceedings of Semantics and Linguistic Theory*, *25*, 328–345. [CrossRef]

Ladd, D. R. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. *Proceedings of the Chicago Linguistic Society*, *17*, 164–171.

Le, G. H. (2015). *Vietnamese sentence-final particles* [Unpublished master's thesis]. University of Southern California.

Liao, Y.-L. I., & Lin, T.-H. J. (2024). The left-peripheral nature of the right-edge particle *không* in Vietnamese. *Taiwan Journal of Linguistics*, *22*(1), 115–133.

Malamud, S. A., & Stephenson, T. (2015). Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics*, *32*, 275–311. [CrossRef]

Nguyen, D. H. (1997). *Vietnamese: Tiếng Việt Không Son Phấn*. London Oriental and African Language Library.

Oshima, D. Y. (2017). Remarks on epistemically biased questions. *Proceedings of the Pacific Asia Conference on Language, Information and Computation*, *31*, 169–177.

Phan, T. (2024). *Where two ends meet: Non-at-issue meanings on the syntactic treetops of Vietnamese* [Unpublished doctoral dissertation]. National Tsing Hua University.

Phan, T., & Starke, M. (2022). Yes-no questions and Vietnamese clause structure. *Journal of the Southeast Asian Linguistics Society*, *15*(4), 192–211.

Romero, M. (2024). Biased polar questions. *Annual Review of Linguistics*, *10*, 279–302. [CrossRef]

Rudin, D. (2018). *Rising above commitment* [Unpublished doctoral dissertation]. University of California Santa Cruz.

Rudin, D. (2022). Intonational commitments. *Journal of Semantics*, *39*, 339–383. [CrossRef]

Speas, P., & Tenny, C. (2003). Configurational properties of point of view roles. In A. M. Di Sciullo (Ed.), *Asymmetry in grammar, volume 1: Syntax and semantics* (pp. 315–344). John Benjamins.

Sudo, Y. (2013). Biased polar questions in English and Japanese. In D. Gutzmann, & H.-M. Gärtner (Eds.), *Beyond expressives: Explorations in use-conditional meaning* (pp. 275–295). Brill.

Šimík, R. (Forthcoming) Polar question semantics and bias: Lessons from Slavic/Czech. In B. Gehrke, & R. Šimík (Eds.), *Topics in the semantics of Slavic languages.* Language Science Press.

Tran, T. (2009). Wh-*quantification in Vietnamese* [Unpublished doctoral dissertation]. University of Delaware.

Trinh, T. (2005). *Aspects of clause structure in Vietnamese* [Unpublished master's thesis]. Humboldt University.

Trinh, T. (2010, June 10–11). *Asking with assertions*. Program of the 20th SEALS Meeting, Zürich, Switzerland.

Trinh, T. (2025). Partition by exhaustification and polar questions in Vietnamese. *Languages*, *10*(9), 233. [CrossRef]

Trinh, T., Phan, T., & Vu, D. N. (2024). Negation and polar questions in Vietnamese: Present and past. *Taiwan Journal of Linguistics*, *22*(1), 67–88.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

languages

Article

Partition by Exhaustification and Polar Questions in Vietnamese

Tue Trinh

School of Humanities, University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia;
tuetrinh@alum.mit.edu

Abstract: This note presents a series of contrasts pertaining to Vietnamese polar questions: (i) The subject can be definite but not quantificational; (ii) the subject can be plain but not *only* -focused; (iii) the modal adverb *chắc chắn* ('certainly') can follow but not precede verum focus. I argue that a monoclausal analysis, advocated in several previous works, will have difficulties accounting for these contrasts and propose a bi-clausal analysis that explains them in a natural way. The explanation relies on the assumption of a general condition on questions, Partition by Exhaustification (PbE), in conjunction with some other independently motivated semantic and pragmatic constraints.

Keywords: polar questions; Vietnamese; syntax; semantics; pragmatics

1. Introduction

1.1. The Syntactic Profile of Vietnamese Polar Questions

A polar question in Vietnamese appears as an affirmative sentence followed by the negation. An affirmative sentence is a declarative with (matrix) verum focus. In English, verum focus is expressed by inserting the light verb *do* in the auxiliary position: *John came* becomes *John did come*, for example. Vietnamese has a similar strategy, using the light verb *có*, whose lexical meaning is 'have' (Phan, 2024; Trinh, 2005). As *có* is also used as a positive response particle, i.e., as 'yes,' I will gloss it as YES.[1]

(1) a. Nam đến
Nam came
'Nam came'
b. Nam có đến
Nam YES came
'Nam did come'

Negation in Vietnamese is expressed by the word *không*, which I will gloss as NO, since it is also used as the negative response particle, i.e., as 'no.' It also occurs in the auxiliary position; hence, in complementary distribution with YES.[2]

(2) Nam không đến
Nam NO came
'Nam didn't come'

Here is how the polar question meaning 'did Nam come?' is expressed in Vietnamese.

(3) Nam có đến không
Nam YES came NO
'Did Nam come?'

I will call the affirmative preceding *no* in a polar question the 'prejacent' of the question.

1.2. Some Puzzles About Vietnamese Polar Questions

The puzzles about Vietnamese polar questions that this note aims to resolve are of the following form: Some affirmative sentences that are independently acceptable become deviant as prejacents of polar questions. Let me now discuss the specific cases one by one.[3]

1.2.1. Definite vs. Quantificational Subjects

A natural expectation is that if an affirmative sentence ϕ is independently acceptable, the polar question ϕ *không*, expressing the meaning 'whether ϕ,' should be acceptable as well. This expectation is met in (4), where the subject of the affirmative is a definite noun phrase: Both the affirmative (4a) and the polar question (4b) are acceptable.

(4) a. bọn con trai có đến
the boys YES came
'the boys did come'
b. bọn con trai có đến không
the boys YES came NO
'did the boys come?'

However, such a pattern is disrupted when the subject is quantificational: Even though both (5a) and (6a) are acceptable affirmative sentences, the corresponding polar questions in (5b) and (6b), whose intended meanings are 'whether every boy came' and 'whether some boys came,' respectively, are both unacceptable.

(5) a. đứa con trai nào cũng có đến
every boy YES came
'every boy did come'
b. #đứa con trai nào cũng có đến không
every boy YES came NO
'did every boy come?'

(6) a. một đứa con trai có đến
some boy YES came
'some boys did come'
b. #một đứa con trai có đến không
some boy YES came NO
'did some boys come?'

1.2.2. Plain vs. Only-Focused Subjects

When the subject is a 'plain' noun phrase, e.g., *Nam*, things work as expected: The affirmative is acceptable in isolation and as the prejacent of a polar question. This is exemplified by (1b) and (3), reproduced below in (7a) and (7b), respectively.

(7) a. Nam có đến
Nam YES came
'Nam did come'
b. Nam có đến không
Nam YES came NO
'did Nam come?'

However, when the subject is focused by *chỉ* 'only,' the affirmative sentence is acceptable and means 'only Nam did come,' while the polar question, whose intended meaning is 'whether only Nam came,' is not.

(8) a. chỉ Nam có đến
only Nam YES came
'only Nam did come'

b. #chỉ Nam có đến không
only Nam YES came NO
'did only Nam come?'

1.2.3. Adverb Following vs. Preceding YES

The sentential adverb *chắc chắn* 'certainly' in affirmative sentences can follow the verum focus *có*, as in (9a), or precede it, as in (9b), with no semantic consequence.[4]

(9) a. Nam có chắc chắn đến
Nam YES certainly came
'Nam did certainly come'
b. Nam chắc chắn có đến
Nam certainly YES came
'Nam certainly did come'

The natural expectation, then, is that the two polar questions in (10) are both acceptable. The fact, however, is that only (10a) is acceptable.

(10) a. Nam có chắc chắn đến không
Nam YES certainly came NO
'is it certain that Nam came?'
b. #Nam chắc chắn có đến không
Nam certainly YES came NO
'is it certain that Nam came?'

1.3. The Explanation in Outline

My explanation of the puzzles above will relate the contrast in acceptability of sentences in Vietnamese to a contrast in availability of readings in English.

1.3.1. Some Facts About English Polar Questions

Let me be more specific. Consider the English polar questions in (11) and (12). Intuitively, (11) has to be read in such a way that its answer set is (11a), not (11b), and (12) has to be read in such a way that its answer set is (12a), not (12b).[5]

(11) did every boy come?
a. {every boy came, not every boy came} → available reading
b. #{every boy came, every boy didn't come} → unavailable reading

(12) did some boy come?
a. {some boys came, no boys came} → available reading
b. #{some boys came, some boys didn't come} → unavailable reading

In other words, a 'no' answer to (11) has to mean 'not every boy came' ($\neg\forall xPx$) and cannot mean 'every boy didn't come' ($\forall x\neg Px$). Similarly, a 'no' answer to (12) has to mean 'no boys came' ($\neg\exists xPx$) and cannot mean 'some boys didn't come' ($\exists x\neg Px$). Next, consider (13).

(13) did only John come?

a. {only John came, not only John came} → available reading
b. #{only John came, only John didn't come} → unavailable reading

The question in (13) has to be read in such a way that its 'no' answer means 'not only John came' ($\neg only(p)$), not 'only John didn't come' ($only(\neg p)$). Finally, consider (14).

(14) did John certainly win?

a. {John certainly won, John did not certainly win} → available reading
b. #{John certainly won, John certainly did not win} → unavailable reading

The question in (14) must be read in such a way that its 'no' answer means it is not certain that John won ($\neg\Box p$), not 'it is certain that John did not win,' i.e., 'it is certain that John lost' ($\Box\neg p$). Since this claim might not be as obvious as those regarding the missing readings of (11), (12), and (13), let me corroborate it with an example. Compare the two responses by A to B's negative answer in (15).

(15) A: Did John certainly win?
B: No.
A: (i) Why do you think he might have lost?
(ii) #Why do you think he must have lost?

The contrast between (i) and (ii) shows that B's answer is read as 'John might have lost,' i.e., 'John did not certainly win' ($\neg\Box p$), not as 'John must have lost,' i.e., 'John certainly did not win' ($\Box\neg p$).

1.3.2. Claims of the Account

My account of the Vietnamese facts in Section 1.2 consists of the two claims in (16).

(16) a. The readings in (11b), (12b), (13b), and (14b) are unavailable because they are necessarily infelicitous.
b. The sentences in (5b), (6b), (8b), and (10b) are unacceptable because they are necessarily associated with the readings in (11b), (12b), (13b), and (14b), respectively.

I will derive (16a) in Section 2 and derive (16b) in Section 3.

2. Partition by Exhaustification

2.1. Presenting PbE

Fox (2019, 2020) proposed the following felicity condition on questions:

(17) Partition by Exhaustification (PbE)
A question Q is felicitous in a context C only if the members of Q, once exhaustified with respect to Q, partition C

As is usually the case with theoretical statements, there is a simple and a sophisticated interpretation. For this discussion, it suffices to apply the simple intepretation of PbE, which I will now present. The reader is advised to consult the above-cited works for the unabridged version and empirical arguments in its favor.

The 'context' is the set of possible worlds in which all mutual assumptions of the discourse participants are true, i.e., worlds that are considered candidates for the actual world (Stalnaker, 1978, 1998). A question Q is a set of propositions, namely, those that count as its possible answers. A 'proposition p exhaustified with respect to a question Q,' written as $exh(Q)(p)$, is true if p is true and every other proposition in Q is false. Finally, a set $\mathcal{P}$ of propositions 'partitions' a set $\mathcal{W}$ of possible worlds only if $\bigvee \mathcal{P} = \mathcal{W}$, i.e., only if every world in $\mathcal{W}$ is such that at least one member of $\mathcal{P}$ is true in it. Let us consider an example. Suppose $Q = \textit{which girl came?} = \{\textit{Alice came } (a), \textit{Billie came } (b), \textit{Clara came } (c)\}$. What PbE predicts is that Q is felicitous in C only if every world w of C is such that either only Alice came in w, or only Bille came in w, or only Clara came in w.

(18) a. $Q = \{a, b, c\}$

b. $C = \underbrace{(a \wedge \neg b \wedge \neg c)}_{exh(Q)(a)} \vee \underbrace{(b \wedge \neg a \wedge \neg c)}_{exh(Q)(b)} \vee \underbrace{(c \wedge \neg a \wedge \neg b)}_{exh(Q)(c)}$

In other words, PbE predicts, correctly, that *which girl came?* requires a context in which exactly one girl came (Dayal, 1996).

2.2. Deriving the Facts in Section 1.3.1 from PbE

How does PbE help with the facts in Section 1.3.1? Let us look at the 'good' questions first: (11a), (12a), (13a), and (14a). All of these questions are sets containing a proposition and its logical negation. Now, if $Q = \{p, \neg p\}$, then $exh(Q)(p) = p \wedge \neg\neg p = p \wedge p = p$, and $exh(Q)(\neg p) = \neg p \wedge \neg p = \neg p$. This means, given PbE, that these questions require a context $C = exh(Q)(p) \vee exh(Q)(\neg p) = p \vee \neg p$. Since $p \vee \neg p$ is the set of all possible worlds, these questions presuppose nothing: They can be asked 'out of the blue,' i.e., without any preconceived ideas about how the actual world is.

So PbE predicts, correctly, that (11a), (12a), (13a), and (14a) are good questions. Does it predict the bad questions—(11b), (12b), (13b), and (14b)—to be bad? The answer is yes. Let us consider these one by one, starting with (11b).

(19) a. #Q = {every boy came $(\forall x Px)$, every boy didn't come $(\forall x \neg Px)$}

b. $C = \underbrace{\underbrace{exh(Q)(\forall x Px)}_{\forall x Px \ \wedge \ \neg\forall x \neg Px}}_{\forall x Px} \vee \underbrace{\underbrace{exh(Q)(\forall x \neg Px)}_{\forall x \neg Px \ \wedge \ \neg\forall x Px}}_{\forall x \neg Px}$

We see that given PbE, the question in (11b), which is reproduced in (19a), requires a 'homogeneous' context in which either every boy came $(\forall x Px)$ or no boys came $(\forall x \neg Px)$. The same holds for the question in (12b), reproduced in (20a).

(20) a. #Q = {some boys came $(\exists x Px)$, some boys didn't come $(\exists x \neg Px)$}

b. $C = \underbrace{\underbrace{exh(Q)(\exists x Px)}_{\exists x Px \ \wedge \ \neg\exists x \neg Px}}_{\forall x Px} \vee \underbrace{\underbrace{exh(Q)(\exists x \neg Px)}_{\exists x \neg Px \ \wedge \ \neg\exists x Px}}_{\forall x \neg Px}$

It is known that homogeneous contexts militate against quantifiers in favor of definites (Löbner, 1985; Heim, 1991; Magri, 2009, 2011). Thus, in the (realistic) context where it is presupposed that if the Argentinian team won the World Cup, then every Argentinian (on the team) won, and if they did not win, then no Argentinian (on the team) won. In this context, (21a) is acceptable, while (21b) and (21c) are infelicitous.

(21) Context: either every Argentinian won or none of them won

a. Did the Argentinians win?
b. #Did every Argentinian win?
c. #Did some Argentinians win?

Thus, the questions in (11b)/(19a) and (12b)/(20a) violate either PbE or the constraint against using quantifiers in homogeneous contexts. In other words, they are necessarily infelicitous. Next, consider (13b), reproduced in (22a).

(22) a. $\#Q = \{$only John came $(j \wedge no\ others)$, only John didn't come $(all\ others \wedge \neg j)\}$

b. $$C = \underbrace{\underbrace{exh(Q)(j \wedge no\ others)}}_{\underbrace{(j \wedge no\ others) \wedge \neg(all\ others \wedge \neg j)}_{j \wedge no\ others}} \vee \underbrace{\underbrace{exh(Q)(all\ others \wedge \neg j)}}_{\underbrace{(all\ others \wedge \neg j) \wedge \neg(j\ \wedge no\ others)}_{all\ others \wedge \neg j}}$$

Given PbE, the question in (13b)/(22a) requires a context where either John came and no one else did, or John did not come but everyone else did.[6] This is a context where *John came* is equivalent to *only John came*, i.e., where *only* is semantically vacuous. It is a fact that vacuous use of *only* is infelicitous (Alxatib, 2013, 2020): It is odd to ask the question in (23b), given the (realistic) context that the American presidential election has a unique winner.

(23) Context: there is a unique winner

a. Did Trump win?
b. #Did only Trump win?

Thus, the question in (13b)/(22a) is necessarily infelicitous: It violates either PbE or the constraint against vacuous use of *only*.

Finally, we come to the case of (14b), reproduced in (24a).

(24) a. $\#Q = \{$John certainly won $(\Box p)$, John certainly did not win $(\Box\neg p)\}$

b. $$C = \underbrace{\underbrace{exh(Q)(\Box p)}}_{\underbrace{\Box p \wedge \neg\Box\neg p}_{\Box p}} \vee \underbrace{\underbrace{exh(Q)(\Box\neg p)}}_{\underbrace{\Box\neg p \wedge \neg\Box p}_{\Box\neg p}}$$

Given PbE, the question in (14b)/(24a) requires an 'opinionated' context where it is either certain that John won $(\Box p)$ or certain that John did not win $(\Box\neg p)$. It has been noted that the use of modal adverbs such as *certainly* is infelicitous in such contexts (von Fintel & Gillies, 2010). Suppose, for example, that I am on the phone with my friend and I know he is looking out the window. My friend, then, is opinionated about the weather: Either he is certain that it is raining or he is certain that it is not raining. In this context, it would be acceptable for me to ask him the question in (25a), but asking (25b) is odd.

(25) Context: My friend is opinionated about the weather.

a. Is it raining where you are?
b. #Is it certainly raining where you are?

Thus, the question in (14b)/(24a) is necessarily infelicitous: It violates either PbE or the constraint against using *certainly* in opinionated contexts.

2.3. The Scope of 'Whether'

Invoking PbE and the other (semantic and/or pragmatic) constraints to rule out the unavailable readings in (11b), (12b), (13b), and (14b) presupposes that they are not ruled out by syntax. I assume, following many works, that English polar questions contain the operator *whether*, which is base-generated as sister to propositional constituents and moves to [Spec,C], leaving a trace, as illustrated in (26) (Bennett, 1977; Guerzoni, 2004; Higginbotham, 1993; Krifka, 2001; Larson, 1985). The trace of *whether* is of type $\langle st, st \rangle$, and the interpretation rule for $whether_n\ \phi$ is provided in (26a), where $\phi^{n \to f}$ is the result of replacing expressions bearing index n in ϕ with f, YES is the 'affirmation' function $[\lambda p.\ p]$, which maps a proposition to itself, and NO is the 'denial' function $[\lambda p.\ \neg p]$, which maps a proposition to its negation. The meaning of (26) is provided in (26b).

(26) [$_\psi$ whether$_n$ [$_\phi$ t_1 John talked to Mary]]

a. whether$_n$ ϕ = $\{\phi^{n \to f} \mid f$ is YES or NO$\}$

b. ψ = $\{[t_1$ John talked to Mary$]^{1 \to f} \mid f$ is YES or NO$\}$
= {YES John talked to Mary, NO John talked to Mary}
= {John talked to Mary, ¬John talked to Mary}

When the prejacent contains more than one propositional constituent, the question is ambiguous, as the trace of *whether* can be construed at different positions inside the prejacent. A case discussed in Guerzoni (2004) focuses on prejacents containing *even*, exemplified by (27).

(27) Did John even solve problem 2?

a. [whether$_1$ [t_1 [even [John solved problem 2]]]]
= {John even solved problem 2, it's not the case that John even solved prob 2}

b. [whether$_1$ [even [t_1 [John solved problem 2]]]]
= {John even solved problem 2, John didn't even solve problem 2}

Guerzoni noted two things about (27): (i) It is ambiguous between a reading in which solving problem 2 is hard and a reading in which solving problem 2 is easy; (ii) the 'easy' reading comes with a 'negative bias,' i.e., the implication that John did not solve the problem. Assume the LFs in (27a) and (27b) provide a natural explanation, as the reader is invited to consult Guerzoni (2004) and see for herself.[7]

For a question such as (11), reproduced in (28), two logical forms can be syntactically generated that correspond to two different base positions of *whether*.[8] The first LF, (28a), represents the felicitous reading and the second LF the infelicitous one, as indicated by '#'. The question, (28), is perceived as acceptable because it has one LF that induces an acceptable reading.

(28) did every boy come?

a. [whether$_1$ [t_1 [every boy$_2$ [t_2 came]]]] $= \{\forall xPx, \neg\forall xPx\}$

b. #[whether$_1$ [every boy$_2$ [t_1 [t_2 came]]]] $= \{\forall xPx, \forall x\neg Px\}$

The same holds for (12), (13), and (14), reproduced in (29), (30), and (31), respectively.

(29) did some boys come?

a. [whether$_1$ [t_1 [some boys$_2$ [t_2 came]]]] $= \{\exists xPx, \neg\exists xPx\}$
b. #[whether$_1$ [some boys$_2$ [t_1 [t_2 came]]]] $= \{\exists xPx, \exists x\neg Px\}$

(30) did only John come?

a. [whether$_1$ [t_1 [only [John$_2$ [t_2 came]]]]] $= \{only(p), \neg only(p)\}$
b. #[whether$_1$ [John$_2$ [only [t_1 [t_2 came]]]]] $= \{only(p), only(\neg p)\}$

(31) did John certainly win?

a. [whether$_1$ [t_1 [certainly John$_2$ [t_2 won]]]] $= \{\Box p, \neg\Box p\}$
b. #[whether$_1$ [John$_2$ [certainly [t_1 [t_2 won]]]]] $= \{\Box p, \Box\neg p\}$

2.4. Interim Summary

Recall that my account of the Vietnamese facts in Section 1.2 consists of the two claims in (16), reproduced in (32).

(32) a. The readings in (11b), (12b), (13b), and (14b) are unavailable because they are necessarily infelicitous.
b. The sentences in (5b), (6b), (8b), and (10b) are unacceptable because they are necessarily associated with the readings in (11b), (12b), (13b), and (14b), respectively.

I have just explained (32a): The readings are necessarily infelicitous because they are questions that either violate PbE or some other constraints, specifically those against the use of quantifiers in homogeneous contexts, against the vacuous use of *only*, and against the use of *certainly* in opinionated contexts. In the next section, I will propose a theory of Vietnamese polar questions that derive (32b). The theory will say, essentially, that the unacceptable questions in Vietnamese, unlike their English counterparts, have only the LF that induces the infelicitous reading.

3. An Analysis of Polar Questions in Vietnamese

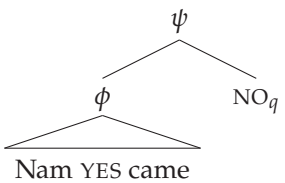
3.1. Previous Works

Recall the surface profile of Vietnamese polar questions: an affirmative sentence followed by the negation.

(33) Nam có đến không
Nam YES came NO
'did Nam come?'

Analyses that have been proposed include Duffield (2007); Phan (2024); Trinh (2005). These accounts differ in details about the affirmative morpheme YES but have in common the idea that underlying this profile is a monoclausal structure where the clause–final negation NO functions as a 'question particle' akin to *whether* in English. I will gloss NO in this use as NO$_q$, where the subscript q is a mnemonic for 'question.'

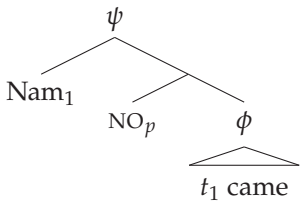
(34) a.



b. NO$_q$ $\phi = \{\phi, \neg\phi\}$
c. $\psi = \{$Nam came, ¬Nam came$\}$

The advantage of this analysis is that it makes Vietnamese polar questions syntactically similar to their English counterparts. There are, however, two problems with this approach. One is that it has to postulate two different NOs for Vietnamese: NO_q for questions, as exemplified in (34) above, and NO_p for propositional negations, as exemplified in (35) below.

(35) a.

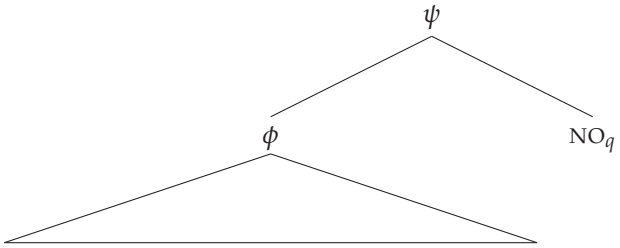


b. $\text{NO}_p\ \phi = \neg\phi$

c. $\psi = \neg$Nam came

The morphological identity between the two NOs would be essentially accidental. The second problem, which I think is more serious, is that this analysis would make it very difficult to state the selectional requirements of NO_q such that its complement, i.e., the prejacent of the question, can have a definite but not a quantificational subject, a plain but not an *only*-focused subject, and the modal adverb *certainly* following but not preceding the verum focus YES.

(36)



the boys/#every boy/#some boys YES came
Nam/#only Nam YES came
Nam (#certainly) YES (certainly) came

In other words, the monoclausal analysis offers no clear solution to the puzzles presented in Section 1.2. While I am not ruling out the possibility of revising this analysis so that it does account for these puzzles, I will explore a different approach, according to which Vietnamese polar questions have a bi-clausal structure.

3.2. Proposal: A Bi-Clausal Analysis

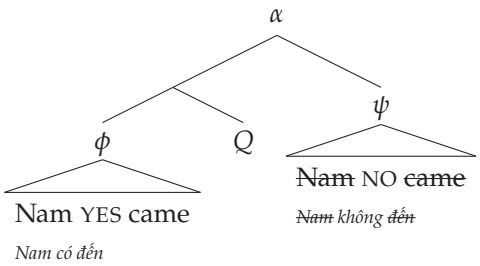
I propose that a polar question in Vietnamese is derived in the following steps.[9]

(37) Derivation of Vietnamese polar questions:

(i) Construct an affirmative sentence ϕ;
(ii) Merge ϕ with Q;
(iii) Merge $[\phi\ \ Q]$ with ψ, where ψ is derived from ϕ by replacing YES with NO;
(iv) Phonologically delete everything in ψ except NO.

The question in (33), then, has the structure in (38a), where strikethrough represents phonological deletion. The meanings of Q and (38a) are provided in (38b) and (38c), respectively.

(38) a.



b. $Q = \lambda p.\ \lambda q.\ \{p, q\}$

c. $\alpha = \{\text{Nam came}, \neg\text{Nam came}\}$

This analysis assigns Vietnamese polar questions a bi-clausal structure that is very different from that of their English counterparts. In English, there is a question operator, *whether*, which, in similar fashion to other question words in this language, moves from its base position, leaving a trace. Since the trace is silent, the English question is ambiguous if the trace can be construed at different scope sites, as illustrated by (27). In Vietnamese, there is no operator that moves and leaves a silent trace. The position of YES in the affirmative sentence indicates, unambiguously, the position of NO in the negated sentence. This means that the pronunciation of a Vietnamese polar question leaves no doubt as to what the positive and negative answers to that question are. This fact plays a crucial role in my resolution of the puzzles in Section 1.2, to which I now turn.

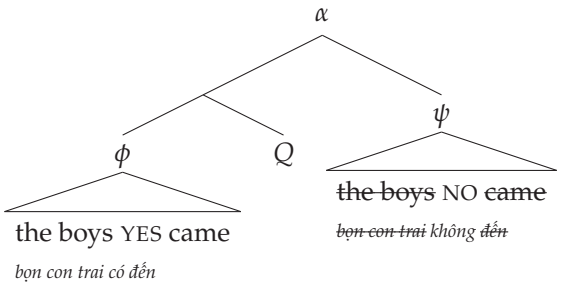
3.3. Resolving the Puzzles in Section 1.2

3.3.1. Explaining the Facts in Section 1.2.1: Definite vs. Quantificational Subjects

Polar questions with definite subjects are exemplified by (4b), reproduced in (39a). The structure of this question is (39b). Its meaning is (39c).

(39) a. bọn con trai có đến không
the boys YES came NO

b.



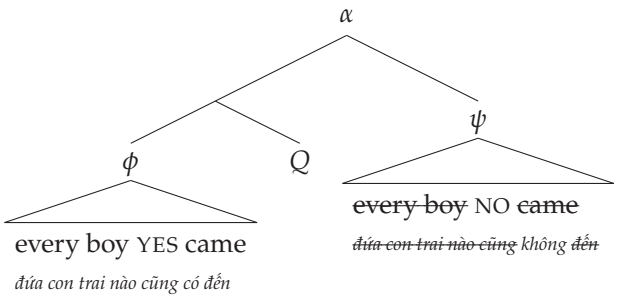
c. $\alpha = \{\phi, \psi\} = \{\text{the boys came, the boys didn't come}\}$

It is a fact of Vietnamese that ϕ = *bọn con trai có đến* means 'the boys came' and ψ = *bọn con trai không đến* means 'the boys did not come.' Thus, (39a) has the exact same meaning as the English question *did the boys come?* and is predicted to be acceptable, as observed.

Let us now deal with the observation that Vietnamese polar questions do not tolerate quantificational subjects. Consider first the case of universal quantifiers, exemplified by (5b), reproduced in (40a). The structure of this question is (40b).

(40) a. #đứa con trai nào cũng có đến không
every boy YES came NO

b.



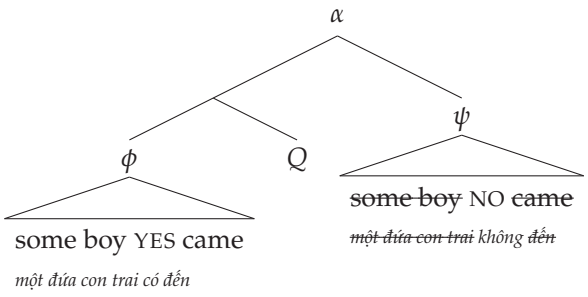
c. $\alpha = \{\phi, \psi\} = \{$every boy came $(\forall xPx)$, every boy didn't come $(\forall x \neg Px)\}$

It is a fact of Vietnamese that $\phi =$ *đứa con trai nào cũng có đến* means 'every boy came' $(\forall xPx)$ and $\psi =$ *đứa con trai nào cũng không đến* means 'every boy did not come' $(\forall x \neg Px)$. However, as shown in Section 2.2, $\{\forall xPx, \forall x \neg Px\}$ is necessarily an infelicitous question. Since (40a) unambiguously denotes this question, (40a) is predicted to be unacceptable, as observed.

Consider next the case of existential subjects, exemplified in (6b), reproduced in (41a). The structure of this question is (41b).

(41) a. #một đứa con trai có đến không
some boy YES came NO

b.



c. $\alpha = \{\phi, \psi\} = \{$some boys came $(\exists xPx)$, some boys didn't come $(\exists x \neg Px)\}$

It is a fact of Vietnamese that $\phi =$ *một đứa con trai có đến* means 'some boys came' $(\exists xPx)$ and $\psi =$ *một đứa con trai không đến* means 'some boys didn't come' $(\exists x \neg Px)$. However, as shown in Section 2.2, $\{\exists xPx, \exists x \neg Px\}$ is necessarily an infelicitous question. Since (41a) unambiguously denotes this question, (41a) is predicted to be unacceptable, as observed.

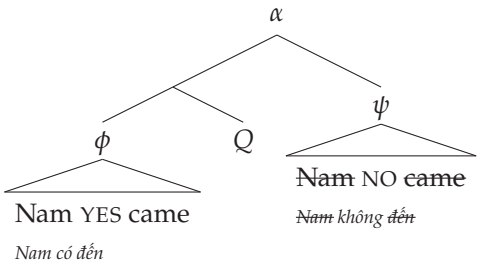
I have thus accounted for the contrast, discussed in Section 1.2.1, between polar questions with definite subjects and polar questions with quantificational subjects.

3.3.2. Explaining the Facts in Section 1.2.2: Plain vs. *Only*-Focused Subjects

Polar questions with plain subjects are exemplified by (33), reproduced in (42a). The structure of this question is (38a), reproduced in (42b), and its meaning is (38c), reproduced in (42c).

(42) a. Nam có đến không
Nam YES came NO

b.



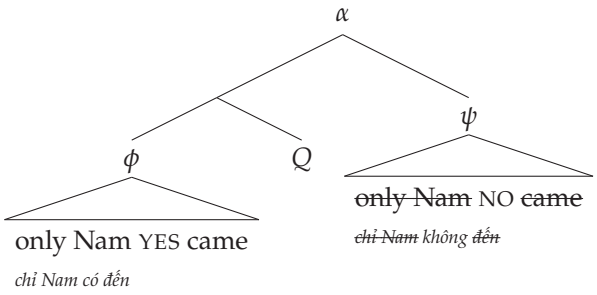
c. $\alpha = \{\phi, \psi\} = \{$Nam came, ¬Nam came$\}$

It is a fact of Vietnamese that ϕ = *Nam có đến* means 'Nam came' and ψ = *Nam không đến* means 'Nam did not come.' Thus, (42a) has the exact same meaning as the English question *did Nam come?* and is predicted to be acceptable, as observed.

Polar questions with *only*-focused subjects are exemplified by (8b), reproduced in (43a). The structure of this question is (43b). Its meaning is (43c).

(43) a. #chỉ Nam có đến không
only Nam YES came NO

b.



c. $\alpha = \{\phi, \psi\} = \{$only Nam came, only Nam didn't come$\}$

It is a fact of Vietnamese that ϕ = *chỉ Nam có đến* means 'only Nam came' $(only(p))$, i.e., 'Nam came but no one else did' and ψ = *chỉ Nam không đến* means 'only Nam didn't come' $(only(\neg p))$, i.e., 'Nam didn't come but everyone else did.' As shown in Section 2.2, $\{only(p), only(\neg p)\}$ is necessarily an infelicitous question. Since (43a) unambiguously denotes this question, (43a) is predicted to be unacceptable, as observed.

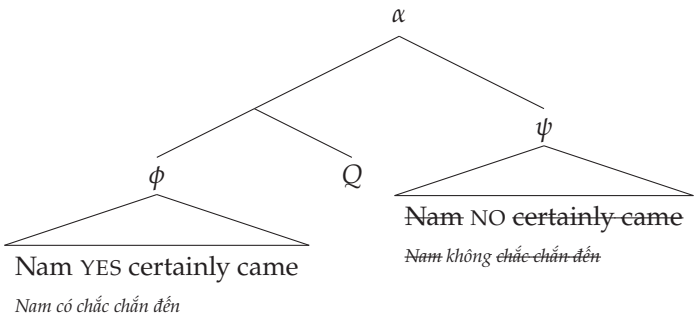
I have thus accounted for the contrast, discussed in Section 1.2.2, between polar questions with plain subjects and polar questions with *only*-focused subjects.

3.3.3. Explaining the Facts in Section 1.2.3: Adverb Following vs. Preceding YES

Polar questions in which the modal adverb *chắc chắn* 'certainly' follows YES are exemplified by (10a), reproduced in (44a). The structure of this question is (44b) and its meaning is (44c).

(44) a. Nam có chắc chắn đến không
Nam YES certainly came NO

b.



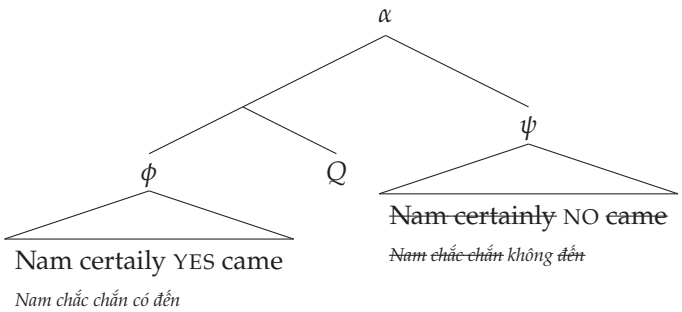
c. $\alpha = \{\phi, \psi\} = \{\Box$Nam came, $\neg\Box$Nam came$\}$

It is a fact of Vietnamese that ϕ = *Nam có chắc chắn đến* means 'it is certain that Nam came' $(\Box p)$ and ψ = *Nam không chắc chắn đến* means 'it is not certain that Nam came' $(\neg\Box p)$. The two answers to this question, then, are a proposition and its negation. The question is predicted to be without any problem, and it is.

Now let us consider polar questions where the adverb precedes *yes*. These are exemplified by (10b), reproduced in (45a). The structure of this question is (45b) and its meaning is (45c).

(45) a. #Nam chắc chắn có đến không
Nam certainly YES came NO

b.



c. $\alpha = \{\phi, \psi\} = \{\Box$Nam came, $\Box\neg$Nam came$\}$

It is a fact of Vietnamese that ϕ = *Nam chắc chắn có đến* means 'it is certain that Nam came' ($\Box p$) and ψ = *Nam chắc chắn không đến* means 'it is certain that Nam didn't come' ($\Box\neg p$). As shown in Section 2.2, $\{\Box p, \Box\neg p\}$ is necessarily an infelicitous question. Since (45a) unambiguously denotes this question, (45a) is predicted to be unacceptable, as observed.

I have thus accounted for the contrast, discussed in Section 1.2.3, between polar questions where the modal adverb *chắc chắn* follows YES and polar questions where this adverb precedes YES.

3.4. Wide-Scope Polarity and Discourse Particles

At this point, the reader may be asking how the *felicitous* readings of the English polar questions in (28)–(31), reproduced here with minimal modification in (46a)–(46d), are expressed in Vietnamese.

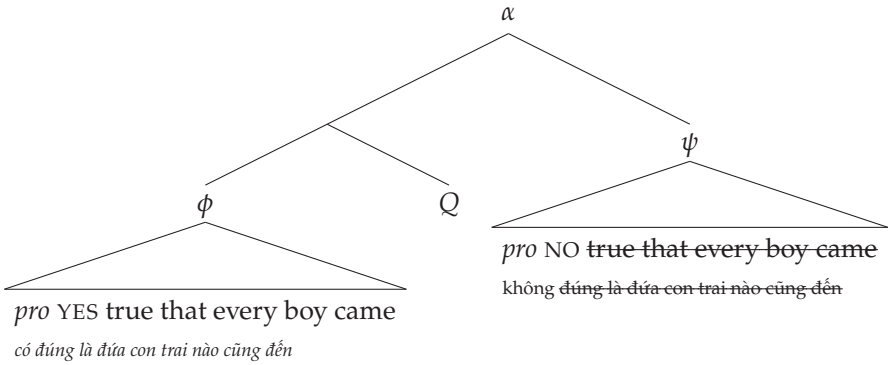
(46) a. did every boy come?
b. did some boys come?
c. did only Nam come?
d. did Nam certainly come?

There are several options. One is to place a declarative inside a larger YES ... NO structure to effect a meaning akin to *is it true that ϕ?*. Take (46a), for example. One can say (47) to ask this question in Vietnamese.

(47) có đúng là đứa con trai nào cũng đến không
YES true that every boy came NO
'is it the case that every boy came?'

I assume that (47) has the structure in (48), where *pro* is a silent expletive subject.

(48)



Another strategy is to resort to the discourse particle *à*, which is appended to a declarative ϕ to ask whether the addressee agrees that ϕ is true. The meaning that is effected is similar

to that of an English 'declarative' question, i.e., a yes/no question that has declarative word order and rising intonation (Gunlogson, 2003; Trinh & Crnic, 2011). In most situations, this meaning is, for all practical purposes, close enough to that of a canonical polar question.

(49) đứa con trai nào cũng đến à
every boy came A
'every boy came?' (with rising intonation)

4. Loose Ends and Conclusions

Before concluding, I will discuss two loose ends to which attention was drawn in the recent round of reviews of this note.

4.1. Other Focus Particles

One issue concerns focus particles other than 'only,' specifically 'also' and 'even.' To my ear, there is definitely a contrast between (8b)/(43a), reproduced in (50a), which is severely degraded, and (50b) and (50c), which are slightly degraded.

(50) a. #chỉ Nam có đến không
only Nam YES came NO
'{only Nam came, only Nam didn't come}'
b. ?cả Nam có đến không
also Nam YES came NO
'{also Nam came, also Nam didn't come}'
c. ?kể cả Nam có đến không
even Nam YES came no
'{even Nam came, even Nam didn't come}'

The theory proposed here predicts the contrast. The context for (50b) would consist of worlds where Nam came and someone else came ('yes') and ones where Nam did not come and someone else did not come ('no'). The context for (50c) would consist of worlds where Nam came even though him having come was unlikely ('yes') and ones where Nam did not come even though him not having come was unlikely ('no').[10] These contexts correspond to my intuition about what the answers mean and, as far as I can see, do not conflict with other grammatical aspects of the questions. However, I admit that (50b) and (50c) are not perfectly acceptable. At this time, I will have to leave this fact, and other facts, about 'also' and 'even,' as well as other focus particles, to future work.[11]

4.2. Other Quantifiers

Questions were also raised in the reviews about determiners such as 'more than half.' It turns out that (51) is unacceptable.

(51) #hơn một nửa số sinh viên có đến không
more than half student YES came NO
'{more than half of the students came, more than half of the students did not come}'

The context, given our account, would consist of worlds where more than half of the students came ('yes') and worlds where more than half of the students did not come, i.e., where less than half of the students came ('no'). Crucially, worlds where half of the students came are excluded. I tentatively propose that this is the reason for the oddness of (51). It seems that the expression *more than X* is generally odd if *X* is excluded as a possibility. Thus, if we know that Nam is a fastidious person who never loses any of his beautiful handcrafted wooden chopsticks, it would definitely be weirder to ask him (52b) than to ask him (52a).[12]

(52) Context: The addressee has an even number of chopsticks.

a. Do you have more than 10 chopsticks?

b. #Do you have more than 11 chopsticks?

The reader is invited to verify for herself that this argument applies generally to *more than X* and *less than X*.[13] A comprehensive survey of all subject quantifiers in polar questions in Vietnamese will have to await further research.

4.3. Conclusions

I propose a bi-clausal analysis of polar questions in Vietnamese to explain a set of data pertaining to such questions. The explanation relies crucially on Partition by Exhaustification (PbE), a felicity condition on questions in general. To the extent that my account is correct, an interesting picture emerges of the difference between English and Vietnamese with respect to which polar questions are 'feasible' in which language. The descriptive statement 'question *Q* can be asked in English but not Vietnamese' turns out to mean 'the syntax of *Q* in English allows a felicitous meaning, but the syntax of *Q* in Vietnamese does not.' In other words, the two languages differ with respect to the syntactic strategies they employ to formulate a question, but not with respect to the semantic/pragmatic constraints on the use of linguistic expressions in general.

Funding: This work is financially supported by the Slovenian Research Agency (ARIS) project no. J6-4615.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: I thank Itai Bassi, Nigel Duffield, Dan Goodhue, Manfred Krifka, Trang Phan and the audience at ISVL-5 for valuable input and discussion.

Conflicts of Interest: The author declares no conflicts of interest.

Notes

1 Auxiliary *có* and main verb *có* can co-occur: *Nam có có sách* (Nam YES have book) means 'Nam does have books.' Note that what I call 'polar question' in this note might be considered a subkind of polar question by other scholars who would also include so-called '*đã-chưa* questions' as polar. A *đã-chưa* question has the same profile as the polar questions I am describing here, with *đã* in place of *có* and *chưa* in place of *không*. I will not discuss *đã-chưa* questions in this paper.

2 I am describing the standard dialect (Hanoi) of Vietnamese, where *Nam có không đến* and *Nam không có đến* are unacceptable. In some southern dialects, *Nam không có đến* is acceptable (Duffield, 2007). I have nothing to say about these in this note.

3 The reader may wonder how the intended meaning of the unacceptable questions—(5b), (6b), (8b)—is expressed. I will answer this question in Section 3.4.

4 Which means that each can be used in the same discourse contexts as the other.

5 I take the meaning of a question to be the set containing its possible answers (Hamblin, 1958). I will discuss this view in more detail presently.

6 The reader will have noticed that the presupposition of *only* is here locally accommodated as part of the assertion. Do we still predict the question to be infelicitous if the presupposition is not locally accommodated? The answer is yes. In that case, the positive answer presupposes no others came and the negative answer presupposes all others came, which means the question ends up with a contradictory, hence unsatisfiable, presupposition. Thus, the account I am presenting here can be considered as the argument that, given PbE, even if the presuppositions of the answers can be locally accommodated, which they can be, the question is still predicted to be infelicitous.

7 A reviewer asks whether the questions in (27) satisfy PbE. The answer is yes: In each case, the positive and negative answers carve out two different regions in logical space, and the union of these regions is accommodated as the context. PbE, as it is formulated in this note, can always be 'satisfied' in isolation since context accommodation is available. The question is whether satisfaction of

PbE, i.e., accommodation of a certain context, leads to violation of some other principles. As far as I can see, that is not the case for the questions in (27) (assuming that the presuppositions of the answers can be locally accommodated (see note 6)).

[8] Note that the quantifier *every boy* itself moves, leaving a trace of type *e*.

[9] Note that an 'affirmative sentence' is one containing *matrix* verum focus, i.e., one where YES occupies the highest auxiliary position. Thus, (53a) is not an affirmative sentence, and (53b) is predicted, correctly, to be unacceptable as a polar question.

(53) a. Nam nghĩ Mai có đến
Nam think Mai YES came
'Nam thinks Mai did come'

b. #Nam nghĩ Mai có đến Q ~~Mai~~ không ~~đến~~
Nam think Mai YES come Q ~~Mai~~ NO ~~come~~
'{Nam thinks Mai did come, Nam thinks Mai did not come}'

More generally, polar questions cannot scope out of embedded positions (modulo parentheticals, which I will not discuss here). The fact that the derivation of polar questions starts from an affirmative sentence, not from a sentence containing an affirmative sentence, or a negative sentence, does not follow from other claims in my account and will have to be taken as a primitive for now. I hope to investigate this point further in future work.

[10] Again, I assume that it is possible to locally accommodate the presuppositions of 'also' and 'even' in polar questions (see note 6).

[11] One observation should perhaps be mentioned here: (50c) does not seem to have the 'negative bias' reading available for English (27). My hunch is that this is because there is no Vietnamese counterpart of (27a). Note that Guerzoni's (2004) explanation of the 'negative bias' reading of (27) depends crucially on both (27a) and (27b) being possible parses of (27).

[12] A question then arises with respect to the general acceptability of the English question *did more than half of the students come?*, which, hypothetically, partitions the context into worlds where more than half came and worlds where half or less than half came. Suppose the number of students is even, then there is no problem. However, what if the number is odd? Do we not then predict that the question should be unacceptable, since the possibility that half came is excluded? My very tentative answer is that at the level of grammatical analysis where judgments of acceptability are generated, the scale underlying measurement is dense, even for students, and hence, the possibility that half came is always existent (Fox & Hackl, 2006). Note that given density, (51) is predicted to be infelicitous independently of the parity of the number of students, as 'half' is always available as a possibility. Note, also, that this line of reasoning might present a conceptual problem: PbE relates questions to contexts, but density is context-independent. I thank a reviewer for drawing my attention to this important point and will leave an appropriate calibration of context and grammar regarding this issue to future work.

[13] Additionally, to the extent that 'many' and 'few' can be analyzed as 'more than half' and 'less than half,' I predict that replacing *hơn một nửa số* 'more than half' in (51) with *nhiều* 'many' or *ít* 'few' will not improve the sentence. This prediction is borne out of facts.

References

Alxatib, S. (2013). *'Only' and association with negative antonyms* [Ph. D. thesis, Massachusetts Institute of Technology].

Alxatib, S. (2020). *Focus, evaluativity, and antonymy: A study in the semantics of 'only' and its interaction with gradable antonyms*. Springer.

Bennett, M. (1977). A response to Karttunen. *Linguistics and Philosophy*, *1*, 279–300. [CrossRef]

Dayal, V. (1996). *Locality in wh-quantification: Questions and relative clauses in Hindi*. Kluwer Academic Press.

Duffield, N. (2007). Aspects of Vietnamese clausal structure: Separating tense from assertion. *Linguistics*, *45*, 765–814. [CrossRef]

Fox, D. (2019). Partition by exhaustification: Comments on Dayal 1996. *Proceedings of Sinn und Bedeutung 22*, 403–434. [CrossRef]

Fox, D. (2020). Partition by exhaustification: Towards a solution to Gentile and Scharwarz's puzzle. Manuscript, MIT.

Fox, D., & Hackl, M. (2006). The universal density of measurement. *Linguistics and Philosophy*, *29*, 537–586. [CrossRef]

Guerzoni, E. (2004). *Even*-NPIs in yes/no questions. *Natural Language Semantics*, *12*(4), 319–343. [CrossRef]

Gunlogson, C. (2003). *True to form: Rising and falling declaratives as questions in English*. Routledge.

Hamblin, C. L. (1958). Questions. *The Australasian Journal of Philosophy*, *36*, 159–168. [CrossRef]

Heim, I. (1991). Artikel und Definitheit. In A. von Stechow, & D. Wunderlich (Eds.), *Semantik: Ein internationales handbuch der zeitgenössischen Forschung* (pp. 487–535). De Gruyter.

Higginbotham, J. (1993). Interrogatives. In K. Hale, & S. J. Keyser (Eds.), *The view from building 20* (pp. 195–228). MIT Press.

Krifka, M. (2001). For a structured account of questions and answers. In C. Fery, & W. Sternefeld (Eds.), *Audiatur vox sapientiae. A festschrift for arnim von stechow* (pp. 287–319). Akademie Verlag.

Larson, R. (1985). On the syntax of disjunction scope. *Natural Language and Linguistic Theory*, *3*, 217–264. [CrossRef]

Löbner, S. (1985). Definites. *Journal of Semantics*, *4*, 279–326. [CrossRef]

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*(3), 245–297. [CrossRef]

Magri, G. (2011, November). Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics and Pragmatics*, *4*(6), 1–51. [CrossRef]

Phan, T. (2024). *The syntax of vietnamese tense, aspect, and negation*. Routledge.

Stalnaker, R. (1978). Assertion. *Syntax and Semantics*, *9*, 315–332.

Stalnaker, R. (1998). On the representation of context. *Journal of Logic, Language and Information*, *7*(1), 3–19. [CrossRef]

Trinh, T. (2005). Aspects of clause structure in Vietnamese [Master's thesis, Humboldt-Universität zu Berlin].

Trinh, T., & Crnic, L. (2011). The rise and fall of declaratives. *Proceedings of Sinn und Bedeutung 15*, 645–660.

von Fintel, K., & Gillies, A. S. (2010). Must ... stay ... strong! *Natural Language Semantics*, *18*, 351–383. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-6192-7