languages

MDPI

Article

From Chinese 隻 to Vietnamese *chiếc*: Diachronic Grammatical Integration of a Borrowed Classifier

Trang Phan [1,*] and Masaaki Shimizu [2]

[1] Department of Asian and North African Studies, Ca' Foscari University of Venice, 30123 Venice, Italy
[2] Graduate School of Humanities, The University of Osaka, Minoh 562-8678, Osaka, Japan; shmz.hmt@osaka-u.ac.jp
* Correspondence: thihuyentrang.phan@unive.it

Abstract

This article traces the written history of Vietnamese *chiếc*, a classifier of early Chinese origin that coexists with the later literary Sino-Vietnamese reading *chích*. This study examines 63 occurrences in nine texts, from early Hán Nôm Buddhist materials to Quốc ngữ prose from the mid-twentieth century, and codes them by construction, semantic domain, discourse function, and textual setting. The earliest clear attestation, *chiếc giày* "a shoe" in *Thiền tông Khóa hư ngữ lục*, occurs beside the Buddhist expression 隻履, *chích lí*, which provides a specific entry context mediated by contact. Subsequent texts do not reproduce a single Chinese distribution. Instead, *chiếc* enters Vietnamese quantified phrases, classifier noun phrases without numerals, phrases marked for plurality, proper name constructions, and elliptical reference. Its tokens remain sparse in most texts, but nineteenth-century travel prose supplies a dense vessel pattern, while *Sống mòn* extends the classifier to pointable artifacts, clothing, and teeth. This article therefore distinguishes grammatical integration from a change in which a lexical item becomes grammatical: a borrowed classifier is redistributed through constructions in the recipient language and through discourse shaped by particular genres. The findings support a modest account of constructional expansion shaped jointly by contact, Vietnamese nominal syntax, semantic persistence, and textual ecology. Tracing one classifier across different textual settings makes it possible to distinguish its Chinese source, the textual channel through which it entered Vietnamese, and the Vietnamese constructions that shaped its later use.

Keywords: Vietnamese; classifiers; *chiếc*; Sinitic contact; Nôm; early Chinese borrowing; grammatical integration; diachronic construction grammar

Check for updates

Academic Editors: Pui Yiu Szeto and Umberto Ansaldo

Received: 18 August 2026
Revised: 16 October 2026
Accepted: 2 November 2026
Published: 11 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Numeral classifiers are widespread in East and Mainland Southeast Asia, but their functions cannot be reduced to counting. They individuate referents, interact with noun phrase structure, and may contribute to discourse tracking, specificity, and definiteness (Aikhenvald, 2000; Bisang, 1999; Her et al., 2022). Vietnamese is especially instructive because classifiers occur not only between numerals and nouns, but also directly before nouns and in phrases whose lexical head is recoverable from context (Bisang & Quang, 2020; T. H. Nguyen, 2004; T. Phan & Chierchia, 2022; Simpson & Ngo, 2018; Trinh, 2011). Because classifiers enter several nominal configurations, a historical study must ask not simply when a form appears, but which nominal constructions license it.

This article examines Vietnamese *chiếc*. In present-day usage, *chiếc* commonly classifies inanimate count nouns, including vehicles, furniture, clothing, tools, and other bounded

objects. It also occurs when one member of a pair is singled out, as in *chiếc giày* "a shoe" and *chiếc đũa* "a chopstick."[1] Lexicographic traditions relate the form to the Sino-Vietnamese reading *chích* and Chinese 隻. Historical phonological research treats *chiếc* not as a later reshaping of *chích*, but as an Early Chinese loan from the first millennium CE that predates the regular Sino-Vietnamese reading system (Alves, 2018, 2022, 2025; Alves & Vũ, 2025). That connection is plausible, but etymology alone does not explain how the item became embedded in Vietnamese grammar.

The central question is thus how a classifier of early Chinese origin entered, persisted, and expanded within Vietnamese constructions. The earliest clear occurrence in the dataset belongs to a Buddhist Hán Nôm setting and corresponds to the expression 隻履, *chích lí* "single shoe." Later occurrences appear in a tale collection, catechetical prose, narrative verse, Catholic correspondence, travel writing, early fiction, and realist prose. These genres do not provide equivalent windows on the language. They differ in script, translation practice, topic, preservation, and the kinds of referents that their discourse repeatedly introduces. The textual record examined here runs from a Buddhist text provisionally dated to the thirteenth century to the 1944 novel *Sống mòn*; the first clear occurrence is in a Hán Nôm textual witness from the fourteenth century. The borrowing therefore probably predates the surviving written attestation. A plausible setting is sustained bilingual contact involving local Annamese Chinese communities and their shift toward early Vietnamese (J. D. Phan, 2013, 2025).

The dataset contains 63 occurrences across nine texts. Most individual texts yield few tokens, whereas *Sách sổ sang chép các việc* contains a marked concentration of *chiếc* with boats and ships: 29 of its 31 tokens receive a vessel code. *Sống mòn* supplies the second largest group, but its tokens occur with a wider range of pointable artifacts and body parts. The contrast suggests that token density must be read together with topic and construction. A maritime text repeatedly creates contexts for counting and distinguishing vessels; a realist novel repeatedly presents objects in scenes and tracks them across discourse.

The argument has three components. First, the Buddhist expression provides an entry context shaped by contact, not proof that Vietnamese copied a complete Chinese paradigm. Second, the form is redistributed by Vietnamese nominal syntax, including [Num/Q + CL + (N)], [CL + N] without a numeral, [plural marker + CL + (N)], [CL + modifier], and [CL + proper name]. Third, recurrent semantic domains support local stabilization. Paired objects, detachable parts, vessels, and later pointable artifacts share a construal as bounded units, but recurrence in these domains does not establish unrestricted classifier productivity.

The term textual ecology refers here to the conjunction of script, genre, translation practice, transmission, topic, and discourse task within a particular textual witness. It is not a claim that genre mechanically causes change. Rather, textual ecologies affect which referents are available, which constructions are repeated, and which uses survive in the written record. A Buddhist translation environment makes one type of contact relation visible; a travel notebook makes repeated vessel reference visible; a novel makes object reference within scenes visible.

This approach supports a distinction between grammaticalization and grammatical integration. The data do not document a Vietnamese lexical noun becoming a classifier. The form is already used as a classifier at its first clear attestation. What changes is its distribution across Vietnamese constructions and semantic domains. The article therefore uses grammaticalization only in a broad sense that includes change under contact and gives analytical priority to integration and constructional expansion. This formulation situates *chiếc* alongside other grammatical items of Chinese origin that developed further inside

Vietnamese, while avoiding the claim that their Vietnamese histories reproduce Chinese patterns (Alves, 2007, 2022; T. Phan et al., 2024).

Focusing on one classifier makes it possible to separate four questions that broad surveys often combine: etymological origin, textual transmission, Vietnamese constructional distribution, and later semantic extension. This is the main value of the narrow scope; the article does not claim that *chiếc* follows a unique or universal pathway.

Section 2 sets out the Vietnamese and Chinese nominal background. Section 3 defines the theoretical and corpus methods, including the treatment of frequency and negative evidence. Section 4 presents the dataset and quantitative profiles. Section 5 traces the attested trajectory. Section 6 develops the comparative implications for classifier borrowing, restructuring in Vietnamese, and textual ecology. Section 7 concludes.

2. Source Language Background and Vietnamese Nominal Syntax

2.1. Vietnamese Classifier Constructions

The canonical Vietnamese counting order is [Num + CL + N], as in *hai chiếc thuyền* "two boats." Yet Vietnamese also permits classifier plus noun sequences without an overt numeral, as in *chiếc thuyền* "a/the boat, depending on context," and classifier phrases in which the noun is omitted but recoverable, as in *chiếc ấy* "that one." Classifiers can further combine with plural markers, demonstratives, and postnominal modifiers. These possibilities are central to the present analysis because they give a borrowed classifier more than one structural entry point.

The interpretation of Vietnamese [CL + N] constructions remains debated. Some accounts relate them to definiteness or specificity, while others emphasize discourse anchoring and the interaction of position, context, and nominal semantics (Bisang & Quang, 2020; Đoàn et al., 2019; T. H. Nguyen, 2004; D. H. Nguyen, 1957; P. P. Nguyễn, 1995; T. C. Nguyễn, 1975; T. Phan & Chierchia, 2022; Simpson & Ngo, 2018; Trinh, 2011). The present article does not treat *chiếc* as an article. It uses the more limited term referential packaging for cases in which a classifier phrase singles out or tracks a contextually available unit. Đoàn et al. (2019) analyze Vietnamese bare [CL + N] as definite, in contrast to the corresponding Mandarin pattern, but their comparison also shows that bare nouns and [Num + CL + N] phrases may receive different interpretations across contexts.

This caution also applies to syntactic labels. Analyses that project NumP, ClP, and DP provide useful ways to distinguish numerals, classifiers, and higher nominal material, but the literature does not agree on a single structure for every classifier language. T. Phan and Chierchia (2026), for example, place plural markers and material with functions similar to those of articles above a classifier projection in Vietnamese, whereas other accounts differ in the status of bare nouns and classifier phrases. The historical coding in this article therefore records overt linear constructions before assigning any deeper structural analysis.

Vietnamese classifier choice is sensitive to lexical convention, animacy, shape, function, and discourse construal. The general inanimate classifier *cái* is more frequent and structurally more flexible than *chiếc*. Corpus research also shows that classifier distribution varies by genre and modality (Tran, 2021). The contrast matters because the history of *chiếc* should not be modeled on a presumed single pathway for all classifiers. It should not, however, be read as a major difference in borrowing date. Alves (2022) places both *cái* and *chiếc* in the broad first millennium layer of Early Chinese loans. Their later divergence is consistent with differences in the semantic and syntactic properties of the source items and with the Vietnamese constructions in which each became established. The attested development of *chiếc* remains more semantically selective and more firmly within nominal reference.

2.2. *Chinese* 隻/只 *and the Nominal Background*

The usual etymological comparison identifies Vietnamese *chiếc* with the Sino-Vietnamese reading chích and Chinese 隻. Historical and modern sources associate 隻 and its simplified counterpart 只 with individual units, one member of a pair, animals, birds, and various objects. Peyraube (1998) places classifier development within broader changes in Chinese nominal order, while Liu (2024) traces the modern semantic network of 只 to "bird" and "one out of a pair."[2] These studies clarify the semantic background in Chinese, but they do not establish that every Vietnamese use continues a Chinese construction.

Earlier scholarship distinguished an older borrowing layer from a category called "Vietnamized Sino-Vietnamese," which was assumed to result from changes after the regular Sino-Vietnamese layer (Wang, 1948). More recent historical phonological research does not retain that latter category as a separate stratum. Most items assigned to it show regular features of Early Chinese loans and were borrowed before the regular Sino-Vietnamese readings (Alves & Vũ, 2025; Alves, 2025). The present article therefore does not derive *chiếc* from *chích*.

Chiếc and *chích* are instead treated as a historical doublet from two borrowing strata. Alves (2022) lists *chiếc* among Early Chinese grammatical loans from the first millennium CE and compares it with Mường *chiếc*, Chinese 隻, Old Chinese **tek*, Middle Chinese **tsyek*, and regular Sino-Vietnamese *chích*. Alves (2025) shows more generally that diphthongs in Early Chinese loans can reflect Early Chinese vowel patterns rather than irregular changes after the Sino-Vietnamese period. The exact borrowing date and donor variety remain uncertain. The source item was already grammatical: by the late first millennium CE, Chinese classifier constructions were well established, and 隻/只 could individuate one member of a pair or another discrete unit (Peyraube, 1998). Vietnamese thus borrowed a classifier or unitizer and later extended its distribution within Vietnamese.

Chinese nominal syntax itself has competing analyses. C. N. Li and Thompson (1981) provide a functional description of numeral plus classifier phrases; Cheng and Sybesma (1999) distinguish the interpretations of bare nouns and phrases that contain classifiers; Y.-H. A. Li (1999) examines plurality in a classifier language. Under some formal analyses, NumP and ClP occupy distinct projections, and a DP layer is proposed above them. Other analyses question whether DP is uniformly present or assign different semantic roles to bare nouns, demonstratives, and classifier phrases. The relevant point here is limited: Chinese provides the source form and the relevant distributions in Chinese, but those distributions cannot be transferred wholesale to Vietnamese.

Some distributions overlap. Modern Mandarin 几只狗 *jǐ zhī gǒu* "several CL dog(s)" has the order [Q + CL + N], which is also available in Vietnamese quantified classifier phrases. Bare [CL + N] sequences also occur in Sinitic varieties such as Cantonese (Cheng & Sybesma, 1999). The comparison therefore does not support a categorical Chinese–Vietnamese contrast. The historical question is narrower: which later Vietnamese uses can be explained by general classifier syntax, and which require evidence from Vietnamese?

The contrast is clearest for Vietnamese phrases without numerals and elliptical phrases. The Buddhist combination *chiếc giày* has an immediate relation to 隻履. Expressions attested in Vietnamese texts from the seventeenth to the twentieth century, such as *các chiếc kia* "those other ones," *chiếc sau hết* "the last one," and *những chiếc răng* "teeth" depend on Vietnamese plural marking, postclassifier modification, and discourse recoverability. Their existence is compatible with contact history but cannot be derived from the early Chinese expression alone.

Historical order is also relevant, although not decisive. Chinese and Southeast Asian languages exhibit several surface arrangements of nouns, numerals, and classifiers. Simpson (2012) compares constituent orders involving classifiers across Southeast Asian lan-

guages and discusses analyses that distinguish numeral and classifier positions. Such comparison identifies structural possibilities but does not by itself demonstrate contact inheritance. Direct borrowing, areal convergence, restructuring in Vietnamese, and parallel innovation require independent historical evidence. This comparison also helps frame the [N + Num + CL] sequence in Example 6, which may reflect topic fronting or verse order rather than a new basic classifier pattern; the excerpt alone does not decide between these analyses.

2.3. Contact, Script, and Textual Transmission

Vietnamese written evidence survives in texts written in Nôm and Quốc ngữ. Many Hán Nôm texts present a Literary Sinitic line together with a vernacular Vietnamese line in Nôm; the examples below also give a modern Quốc ngữ transcription and an English gloss. In this bilingual presentation, the vernacular line may respond closely to a Sinitic model while presenting Vietnamese word order and grammatical structure. Such evidence is neither purely Chinese nor purely independent. It records a bilingual textual practice in which a form can be selected because of a source expression and simultaneously integrated into a Vietnamese construction.

The broader contact setting is also relevant. J. D. Phan (2013, 2025) argues that a resident Annamese Chinese community in northern Vietnam shifted over time toward early Vietnamese. This hypothesis provides a plausible social setting for early grammatical borrowing and bilingual transmission, but the present dataset cannot identify the speaker group or date responsible for *chiếc*.

The earliest clear token of *chiếc* occurs in this bilingual setting. The line about Bodhidharma carrying a shoe westward shows Vietnamese *chiếc giày* as the vernacular counterpart of 隻履, *chích lí*. Shimizu (2020) demonstrates more generally that Nôm phonetic components can preserve evidence relevant to the history of Chinese readings in Vietnamese. This article does not posit a phonological transition from *chích* to *chiếc*, because the two forms are treated as loans from different strata. The aligned expression instead provides a philological bridge with semantic and textual specificity. In the vernacular line, the Nôm character is read *chiếc*; in the Literary Sinitic line, the corresponding expression is read *chích lí*. Their shared source relation rests on the comparative phonological evidence summarized in Section 2.2, not on graphic identity alone.

Later witnesses provide different contact conditions. Catholic prose may reflect translation choices and early romanized conventions. Travel writing from the nineteenth century repeatedly names vessels, routes, and foreign places. Early fiction places a foreign ship name in a Vietnamese classifier phrase whose classifier selects a vessel referent. These textual settings may reinforce or extend the classifier without preserving an identifiable Chinese model. Contact helps explain the entry context and the textual settings in which the classifier appears, but it does not explain every later token.

Lexicographic sources support a persistent semantic association with singleness, paired objects, vessels, and bounded inanimate units. de Rhodes (1651) records *chiếc* with paired or matched objects; Taberd (1838) includes vessel and chopstick examples; Huình-Tịnh Paulus Của (1895–1896), Hoàng Phê (2011), and Nguyễn Quang Hồng (2014) continue relevant senses. Dictionaries cannot date productivity on their own. They are used here to clarify semantic persistence, not to replace textual attestations.

3. Framework and Method

3.1. Grammatical Integration and Grammaticalization

Grammaticalization research emphasizes gradual change, contextual reinterpretation, persistence, and layering (Diewald, 2002; Heine, 2002; Hopper, 1991; Hopper & Traugott,

2003). Those concepts are relevant to *chiếc*, but the classical cline from lexical to grammatical status is not attested. The form enters the available record in a classifier position. The more precise claim is that an already grammatical source item underwent further grammatical development after borrowing.

This pattern has precedents in research on grammatical vocabulary of Chinese origin in Vietnamese. Alves (2005, 2007, 2022) describes grammatical loans whose distributions were reshaped in Vietnamese, and T. Phan et al. (2024) show further development of negative markers of Chinese origin in primary Nôm texts. The present study adds a classifier case and an account at the token level of the constructions in which that development is visible.

Grammatical integration is defined here as the increasing licensing of a borrowed grammatical form across constructions of the recipient language. This definition separates origin from subsequent distribution. A form related to the source language may arrive with a narrow semantic orientation and one salient textual model, yet its later combinations are selected from the recipient grammar. Integration is demonstrated by constructional diversity, not by phonological erosion or obligatory status.

Constructional expansion refers to an increase in the attested range of pairings of form and meaning, nouns with which the classifier occurs, and discourse functions. It is not assumed to be monotonic, and it does not require the emergence of a wholly new schematic construction. The movement from *chiếc giày* to quantified vessel phrases, [CL + modifier] ellipses, body parts marked for plurality, and a phrase with [CL + proper name] is an expansion of distribution. Whether speakers at any one stage generalized beyond the attested range remains an open question.

Bisang (2015) warns that East and Mainland Southeast Asian languages often complicate distinctions between primary and secondary grammaticalization because multifunctionality, optionality, and pragmatic inference are prominent. The present analysis follows that caution. It does not rank tokens on a single scale of grammaticality. Instead, it asks which syntactic environments and semantic domains are attested and how they relate to the residue of singular boundedness associated with the source language.

3.2. Diachronic Construction Grammar

Diachronic Construction Grammar treats constructions as conventional pairings of form and meaning and asks how their internal properties and network relations change over time (Traugott & Trousdale, 2013). For this dataset, the relevant units are deliberately shallow: [Num/Q + *chiếc* + (N)], [*chiếc* + N], [plural marker + *chiếc* + (N)], [CL + modifier] and [CL + proper name]. These labels identify overt configurations without presupposing a single derivational analysis.

The constructional distinction is analytically useful because the patterns place different demands on the classifier. Quantified phrases test compatibility with counting. The [CL + N] pattern tests integration into nominal reference without an overt numeral. Elliptical phrases test whether a classifier can support a recoverable nominal without an overt noun. Phrases marked for plurality test compatibility with higher number material, while proper name phrases show how the classifier frames a named entity. For example, *chiếc sau hết* means "the last one," and *những chiếc răng* means "teeth."

Productivity is not inferred from token totals alone. Barðdal et al. (2026) define diachronic productivity in terms of extensibility and show why metrics based on types require large, comparable corpora. Their caution is decisive here. The dataset is too small and too uneven for robust productivity indices. Repetition can demonstrate local entrenchment, especially in a recurring vessel construction, but it cannot establish unrestricted extension to all inanimate nouns.

3.3. Borrowing, Replication, and Recipient Language Restructuring

Contact linguistics distinguishes the transfer of linguistic matter from the replication of structural patterns (Matras & Sakel, 2007; Thomason & Kaufman, 1988). Heine and Kuteva (2005) further show that contact may channel grammatical development without requiring direct copying of individual morphemes. In the earliest Vietnamese example, matter and pattern are both relevant: *chiếc* is historically related to 隻, and the textual expression 隻履 motivates the choice of a classifier with a single shoe. At that point, 隻 already had classifier or unitizer uses in Chinese; the Vietnamese history therefore concerns the further integration of borrowed grammatical material, not an independent shift from noun to classifier.

The later record requires a third term: restructuring in the recipient language. Once the form is available, Vietnamese constructions determine where it can appear. A source expression can explain the entry point, while Vietnamese [Num + CL + N], [CL + N], plural, demonstrative, and elliptical patterns explain subsequent distributions, although some of these, especially the order [Q + CL + N], also overlap with Chinese. This model avoids treating Vietnamese as a passive receptacle and avoids the opposite error of denying contact whenever later structures differ from Chinese.

Three questions are therefore kept separate. First, is the form historically related to Chinese 隻/只? Second, is a particular token linked to a textual model in the source language or in sinographic form? Third, does the Vietnamese token participate in a construction independently licensed in Vietnamese? The answer can differ by token. Example 1 supports all three relations; many later examples support the first and third but not the second.

Typological similarity is likewise distinguished from direct borrowing. Classifier languages across East and Southeast Asia share ordering possibilities and discourse functions, but similar surface patterns can arise through inheritance, areal convergence, parallel innovation, or internal restructuring. Comparative evidence is used below to delimit the Vietnamese claim, not to collapse these histories into one pathway.

3.4. Dataset, Textual Ecology, and Frequency

The dataset comprises nine texts or text collections spanning the thirteenth century to the middle of the twentieth century. The textual witnesses differ in chronology, script, genre, length, and transmission. Some are bilingual or sensitive to translation; others are vernacular narrative, correspondence, travel prose, or fiction. These differences are treated as constitutive properties of the evidence rather than noise that can be removed by normalization.

Taavitsainen et al. (2026) argue that genre and register shape the linguistic choices recoverable from historical writing. Whitt (2018) similarly emphasizes that diachronic corpora are structured by genre composition. Applied here, the lesson is that a token count is partly a count of discourse opportunities. A travel text with many ships creates more occasions for *chiếc* than a scripture whose topics rarely require bounded artifacts, even if both are extensive.

Token counts were normalized per 1000 syllables to provide a descriptive measure of density. The denominator permits comparison within the dataset, but it does not make the texts statistically interchangeable. A normalized value is affected by topic, quotation practices, the mix of Literary Sinitic and Vietnamese, text length, editorial segmentation, and the survival of relevant passages. No significance test is appropriate for the present sample.

Zero counts are interpreted conservatively. A zero means that no token was attested in the sampled text. It does not show that *chiếc* was absent from the language, unavailable to speakers, absent from every genre, or impossible in the period. Likewise, a low normalized rate does not entail grammatical marginality. Negative evidence becomes informative only

when the search domain and the competing opportunities are well characterized, which this heterogeneous dataset cannot fully provide.

Textual transmission introduces further uncertainty. Dates may refer to composition, copying, printing, or the edition used for analysis. A token in a later witness can preserve an older reading, and an absence can result from loss, revision, or editorial selection. The analysis consequently gives greatest weight to philologically specific alignments, recurrent constructional patterns, and transparent source locations.

Comparison is therefore made at the level of attested opportunities, not assumed availability across the language. Dense vessel passages and scriptures with zero tokens can coexist without implying an abrupt chronological shift or a grammatical threshold in Vietnamese. Such coexistence is expected in historically uneven textual evidence whose distribution is sensitive to topic. A defensible comparison with present-day usage would require a balanced synchronic corpus and the same coding at the token level. Unsystematic web or corpus searches would not be comparable with the historical dataset.

3.5. Coding and Category Definitions

Each occurrence was coded for text, period, script or language, genre, construction, semantic domain, discourse function, and interpretive confidence. The constructional table uses two mutually exclusive broad construction categories. [Num/Q + CL + (N)] includes an overt numeral or quantifier, even when the noun is omitted or follows an unusual order. [CL + N] without a numeral includes tokens in which *chiếc* directly precedes an overt noun without an overt numeral or quantifier. Together, these two columns exhaust the 63 tokens.

[Plural marker + CL + (N)], elliptical CL phrase, [CL + modifier], and [CL + proper name] are additional structural annotations used in the qualitative analysis. They may overlap with the two broad construction categories and are therefore not added as independent count columns without separate totals at the token level. For example, *chiếc sau hết* "the last one" has no numeral and is elliptical, while *những chiếc răng* "teeth" is marked for plurality and still contains an overt noun. This separation prevents double counting.

Semantic coding identifies vessel, artifact/tool, part/body part, and collective unit. A referential or discourse code is added when the phrase tracks, contrasts, or retrieves a contextually available entity, especially with ellipsis, a demonstrative, or a postclassifier modifier. These semantic and discourse codes are not mutually exclusive. A vessel can also be referential, and a tooth can be both a body part and an individually tracked member of a set.

A collective unit is operationally defined as an internally plural, organized grouping counted as one unit. The two qualifying tokens occur in *Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời* ("Catechism in Eight Days for Those Who Wish to Receive Baptism and Enter the Holy Faith of God"), including *chín chiếc lớn quân mình* "nine great units of his army." The code differs from ordinary singular individuation because the referent contains multiple members, and it differs from grammatical plurality because *chiếc* counts organizational units rather than directly pluralizing persons. The code may overlap with referential function but not with the vessel or artifact categories in the present table. Two further tokens, *6 chiếc kẻ cướp phalansa* and *chiếc ngựa*, are coded uncertain/other because their immediate contexts do not securely identify the intended referent.

4. Dataset and Quantitative Profile

As shown in Table 1, the 63 occurrences are distributed unevenly. Five of the first six rows contain four tokens or fewer, while *Sách sổ sang chép các việc* contains 31, including 29 tokens coded as vessels. Its normalized density, 0.49 per 1000 syllables, is also the highest. This concentration is interpretable only in relation to the text's repeated discussion of vessels,

routes, fleets, and maritime movement. The table does not support a smooth chronological frequency curve.

Table 1. Corpus of texts analyzed for *chiếc*, token counts, and normalized frequency.

No.	Text	Period	Script/Language	Genre	*chiếc*	Syllables	Per 1000
1	*Phật thuyết đại báo phụ mẫu ân trọng kinh* [Sutra on the Profound Gratitude Owed to Parents]	13th c.[3]	Literary Sinitic/Vietnamese in Nôm	Buddhist scripture	0	12,852	0.00
2	*Thiền tông Khóa hư ngữ lục* [Recorded Sayings of Zen Teachings]	14th c.	Literary Sinitic/Vietnamese in Nôm	Zen Buddhist prose	1	12,576	0.08
3	*Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú* [Newly Edited Tales of the Marvelous, with Added Explanations and Phonetic Annotations]	16th–17th c.	Literary Sinitic/Vietnamese in Nôm	Tales and glosses	4	44,678	0.09
4	*Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời* [Catechism in Eight Days for Those Who Wish to Receive Baptism and Enter the Holy Faith of God]	1651	Latin/Vietnamese in early Quốc ngữ	Catholic catechetical prose	2	56,812	0.04
5	*Thiên Nam ngữ lục* [Records of Southern Language]	late 17th c.	Vietnamese in Nôm	Narrative verse chronicle	3	28,205	0.11
6	34 Catholic letters in *Chữ Quốc ngữ thế kỉ XVIII* [34 Catholic Letters in Eighteenth Century Quốc ngữ Writing]	18th c.	Vietnamese in early Quốc ngữ	Catholic correspondence	1	52,820	0.02
7	*Sách sổ sang chép các việc*[4] [Book Recording Various Matters]	1822	Vietnamese in early Quốc ngữ	Travel and administrative prose	31	62,940	0.49
8	*Thầy Lazaro Phiền* [Father Lazaro Phiền]	1887	Vietnamese in modern Quốc ngữ	Early prose fiction	3	9596	0.31
9	*Sống mòn* [Worn Out Life]	1944	Vietnamese in modern Quốc ngữ	Realist fiction	18	83,825	0.21
Total					63	364,304	0.17

Note. Frequencies are descriptive densities within the sampled texts. A zero records lack of attestation in that text, not absence from the language.

Sống mòn contributes 18 occurrences, but its greater length yields a lower density of 0.21. *Thầy Lazaro Phiền* has only three occurrences but a density of 0.31 because the text is short. These contrasts illustrate why raw and normalized counts answer different questions. Raw counts show the amount of attested evidence; normalized rates show density within a witness. Neither directly estimates usage in a historical speech community.

Table 2 shows a shift in constructional balance rather than a simple replacement. Early Hán Nôm examples are predominantly [CL + N] without a numeral. Quantified phrases become prominent with collective units and vessels, especially in *Sách sổ sang chép các việc*. *Sống mòn* reverses the balance within its row: thirteen of eighteen tokens are [CL + N] phrases without a numeral, consistent with reference to salient objects within scenes.

Table 2. Constructional profile of *chiếc*.

Text	[Num/Q + CL + (N)]	[CL + N] Without Numeral	Total
Phật thuyết đại báo phụ mẫu ân trọng kinh	0	0	0
Thiền tông Khóa hư ngữ lục	0	1	1
Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú	0	4	4
Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời	2	0	2
Thiên Nam ngữ lục	3	0	3
34 Catholic letters in *Chữ Quốc ngữ thế kỉ XVIII*	1	0	1
Sách sổ sang chép các việc	28	3	31
Thầy Lazaro Phiền	2	1	3
Sống mòn	5	13	18
Total	41	22	63

Note. The two broad construction categories are mutually exclusive. Each row total equals their sum. Plural marking, ellipsis, modification, and proper name framing are overlapping annotations discussed qualitatively.

Table 3 makes the vessel concentration explicit. 36 tokens receive a vessel code, 29 of them in *Sách sổ sang chép các việc*. Uses with artifacts or tools and with parts or body parts are especially important in Sống mòn. 24 tokens receive the referential/discourse code, which overlaps strongly with vessels because the travel narrative repeatedly distinguishes members of an established set. The two collective unit tokens form a small domain shaped by translation that should not be generalized to human classification. Two tokens remain in uncertain/other: *6 chiếc kẻ cướp phalansa* and *chiếc ngựa*. The overlap among the semantic codes does not alter the dataset total of 63 or the article's principal conclusions.

Table 3. Semantic profile of *chiếc*.

Text	Vessel	Artifact/Tool	Part/Body Part	Referential/Discourse	Collective Unit	Uncertain/Other	Unique Tokens
Phật thuyết đại báo phụ mẫu ân trọng kinh	0	0	0	0	0	0	0
Thiền tông Khóa hư ngữ lục	0	1	0	0	0	0	1
Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú	0	2	1	3	0	0	4
Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời	0	0	0	0	2	0	2
Thiên Nam ngữ lục	3	0	0	1	0	0	3
34 Catholic letters in *Chữ Quốc ngữ thế kỉ XVIII*	1	0	0	0	0	0	1
Sách sổ sang chép các việc	29	1	0	13	0	1	31
Thầy Lazaro Phiền	3	0	0	1	0	0	3
Sống mòn	0	12	5	6	0	1	18
Total	36	16	6	24	2	2	63

Note. Semantic and discourse columns are not mutually exclusive, so their totals cannot be added to obtain the number of unique tokens. The final column gives the number of unique tokens in each text. Uncertain/other records tokens whose referent cannot be securely assigned from the immediate context.

The quantitative pattern therefore supports a layered account. The earliest philological bridge is rare but diagnostic. Vessel tokens become locally dense in texts that repeatedly count and track ships. Later fiction broadens the semantic range while making greater use of classifier phrases without numerals. The counts describe where constructions are visible, not a uniform rise in frequency across the language.

5. Diachronic Analysis

5.1. Buddhist Entry Point and 隻履

Phật thuyết đại báo phụ mẫu ân trọng kinh ("*Sutra on the Profound Gratitude Owed to Parents*") contains no occurrences in the sampled text. This zero provides a baseline only. It does not show that *chiếc* was unavailable in Vietnamese in the thirteenth century or in other Buddhist genres. The first clear attestation occurs in *Thiền tông Khóa hư ngữ lục* ("*Recorded Sayings of Zen Teachings*"), where a Literary Sinitic expression corresponds to a Vietnamese classifier phrase.

(1) Source: *Thiền tông Khóa hư ngữ lục*, 24b

Literary Sinitic	達磨	隻	履	西	歸			
Sino-Vietnamese	*đạt ma*	*chích*	*lí*	*tây*	*quy*			
Nôm	祖師	達磨	𢬣	捇	隻	跱	術	西
Quốc ngữ	*tổ sư*	*đạt ma*	*tay*	*xách*	*chiếc*	*giày*	*về*	*tây*
	patriarch	Bodhidharma	hand	carry	CL	shoe	return	west

"Patriarch Bodhidharma carried a shoe back to the West."

The bilingual correspondence is clear. Literary Sinitic 隻履 and Sino-Vietnamese chích lí identify a single shoe, while the Vietnamese line uses *chiếc giày* in the ordinary [CL + N] order. A shoe is one member of a pair, so the semantic relation between source expression and recipient phrase is transparent. The token is best treated as a Buddhist textual bridge rather than as an isolated lexical resemblance. The Nôm reading *chiếc* is supported by the vernacular transcription and by the correspondences of the earlier loan discussed in Section 2.2, rather than by the shared character alone.

The example does not demonstrate broad productivity in the fourteenth century. It establishes a narrower point: a classifier form related to Chinese could be selected in a Vietnamese construction where semantics in the source language and Vietnamese syntax converged. The subsequent history must therefore be traced from later Vietnamese evidence rather than projected backward from modern usage.

5.2. Early Hán Nôm Expansion: Paired Objects, Shadows, and Parts

Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú ("*Newly Edited Tales of the Marvelous, with Added Explanations and Phonetic Annotations*") contains four occurrences. They extend the form beyond the Bodhidharma formula but remain semantically constrained. The attested nouns include a chopstick, a shadow, and a wheel. These referents support three related construals: one item from a pair, a visually bounded counterpart, and a detachable part of a larger artifact.

(2) Source: *Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú*, TKML2, 49b

Literary Sinitic	隻	箸	山	
Sino-Vietnamese	*chích*	*trợ*	*sơn*	
Nôm	類	𡶀	隻	𥫗
Quốc ngữ	*loài*	*núi*	*chiếc*	*đũa*
	kind	mountain	CL	chopstick

"the place called mountain chiếc đũa"

(3) Source: *Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú*, TKML3, 3a

Literary Sinitic	隻	影	零丁
Sino-Vietnamese	*chích*	*ảnh*	*linh đinh*
Nôm	隻	霿	冷汀
Quốc ngữ	*chiếc*	*bóng*	*lênh đênh*
	CL	shadow	drifting

"a drifting shadow"

(4) Source: *Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú*, TKML3, 66b

Literary Sinitic	隻	輪	無	返	
Sino-Vietnamese	*chích*	*luân*	*vô*	*phản*	
Nôm	隻	餅車	拯	特	衛
Quốc ngữ	*chiếc*	*bánh xe*	*chẳng*	*được*	*về*
	CL	wheel	NEG	get	return

"not a wheel could return"

The chopstick expression preserves the association with paired objects expected from the source semantics and later dictionaries, although the phrase occurs within an expression similar to a place name rather than as an ordinary free noun phrase. The shadow examples are a mild extension because a shadow is not manufactured, yet it can be construed as a solitary visible outline. The wheel example adds a clear relation between a part and a whole. The relevant commonality is perceptual boundedness, not physical lightness: ships may be very heavy, and the dataset provides no independent evidence that low weight forms part of *chiếc*'s semantic contribution.

These tokens do not establish a general inanimate classifier. Their common denominator is a singled-out, bounded, or detachable unit. They are therefore best interpreted as early constructional and semantic bridging evidence. The classifier is available in [CL + N] constructions, but the host range remains narrow in the surviving sample.

5.3. Quantified Uses in the Seventeenth and Eighteenth Centuries

The two tokens in *Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời* ("Catechism in Eight Days for Those Who Wish to Receive Baptism and Enter the Holy Faith of God") occur in religious enumeration and are not ordinary classifications of individual persons. They count organized groups within a heavenly hierarchy. Their semantics, which are sensitive to translation, warrant the separate collective unit code introduced in Section 3.5.

(5) Source: *Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời*, p. 62

chín	*chiếc*	*lớn*	*quân*	*mình*
nine	CL.unit	great	army	own

"nine great units of his army"

In Example 5, *chiếc* individuates divisions or ranks within an army. The second token, *nhiều chiếc đức thánh thiên thần* "several units or companies of holy angels", likewise denotes companies or hosts of angels rather than separate human or animate individuals. These uses show that a classifier associated with units could be recruited for organized collectives. They do not support productive human classification.

Thiên Nam ngữ lục ("Records of Southern Language") supplies three vessel tokens. The classifier occurs in counting contexts involving ships and boats, sometimes with the head noun before the numeral plus classifier sequence. In Example 6, the surface order is

[N + Num + CL]. It may involve topic fronting or ordering conditioned by verse rather than a new basic pattern; a full analysis of this order is beyond the scope of this article.

(6) Source: *Thiên Nam ngữ lục*, line 6164

Nôm	艚	餘	寂	隻
Quốc ngữ	*tàu*	*dư*	*trăm*	*chiếc*
	ship	remain	hundred	CL

"there remained a hundred ships"

The vessel pattern continues in the Catholic correspondence from the eighteenth century published in *Chữ Quốc ngữ thế kỉ XVIII* ("Eighteenth Century Quốc ngữ Writing"). The token is structurally canonical and shows that *chiếc* was available with a numeral and a boat noun in early romanized prose.

(7) Source: *Chữ Quốc ngữ thế kỉ XVIII*, letter 31

có	*2*	*chiếc*	*thuyền*
have	two	CL	boat

"there were two boats"

Taken together, the evidence from the late seventeenth and eighteenth centuries is small but coherent. Vessels are bounded mobile artifacts that often occur in fleets or travel sequences, making them natural targets for exact counting. The evidence precedes the much denser cluster from the nineteenth century and therefore shows continuity across script and genre without implying a uniform rate of use.

5.4. Vessel Ecology in the Nineteenth Century and Referential Expansion

Sách sổ sang chép các việc ("Book Recording Various Matters") is the quantitative center of the dataset, with 31 occurrences. Of these, 29 receive a vessel code. The text repeatedly introduces boats, ships, warships, routes, and encounters at sea. Its density is therefore not an abstract sign of modernization. It reflects a maritime discourse in which vessels are counted, hired, sent, compared, followed, and retrieved from context.

The constructional range is broader than the lexical concentration. Exact numerals occur with vessel nouns, while classifier phrases without numerals distinguish already established ships. Elliptical expressions omit the lexical noun, and an artifact that is not a vessel appears in a comparison. The recurrence of the same semantic field thus supports several Vietnamese nominal constructions.

(8) Source: *Sách sổ sang chép các việc*, p. 6

mới	*thuê*	*một*	*chiếc*	*thuyền*
then	hire	one	CL	boat

"then hired a boat"

(9) Source: *Sách sổ sang chép các việc*, p. 216

liền	*sai*	*18*	*chiếc*	*tầu chiến*
immediately	send	eighteen	CL	warship

"immediately sent eighteen warships"

(10) Source: *Sách sổ sang chép các việc*, p. 53

tôi	*đi*	*chiếc*	*ấy*
I	go	CL	DEM

"I went in that one"

(11) Source: *Sách sổ sang chép các việc*, p. 59

chèo	*đến*	*chiếc*	*sau hết*	*mà*	*hỏi*
row	to	CL	last	and	ask

"rowed to the last one and asked"

(12) Source: *Sách sổ sang chép các việc*, p. 558

bằng	*một*	*chiếc*	*hòm gánh*
equal	one	CL	carrying chest

"about the size of a carrying chest"

Examples 10 and 11 show referential packaging beyond counting. In *chiếc ấy* and *chiếc sau hết*, the classifier identifies a member of an established vessel set while the noun is omitted. The demonstrative or modifier supplies the contrast required for retrieval. The result is not necessarily articlehood, but it is a clear nominal use headed by a classifier licensed by Vietnamese discourse syntax.

Example 12 confirms that the text is not mechanically repeating a ship formula. *Chiếc* combines with an artifact in a size comparison. Even so, the overwhelming vessel concentration shows local semantic stabilization rather than unrestricted generalization. The repeated vessel references help explain why this construction is especially frequent in the text.

Thầy Lazaro Phiền ("Father Lazaro Phiền") continues the vessel pattern and adds a construction with a proper name. The ship name is used metonymically for the vessel, a relation also found across languages. The specifically Vietnamese point is narrower: the name follows *chiếc* in a classifier phrase, and the classifier selects a vessel referent in this context.

(13) Source: *Thầy Lazaro Phiền*, part 1

chiếc	*jean-dupuis*	*định*	*mười giờ*	*mới*	*chạy*
CL	Jean-Dupuis	plan	ten o'clock	only.then	run

"the Jean-Dupuis was not scheduled to leave until ten o'clock"

The proper name token has a limited interpretation. It does not show a uniquely Vietnamese metonymic construction or that *chiếc* freely classifies foreign names. It shows that a ship name used for the vessel could occupy the nominal position in a construction already associated with ships. The name remains lexically foreign, while its placement after *chiếc* follows Vietnamese classifier syntax.

5.5. Broadening in Sống mòn During the Mid-Twentieth Century

Sống mòn ("Worn Out Life") contains 18 occurrences, with no vessel tokens. The classifier appears with furniture, containers, clothing, and teeth. Thirteen tokens are [CL + N] without an overt numeral, and several participate in reference within scenes. The semantic profile is broader than in the maritime texts, but it remains centered on discrete, pointable units.

(14) Source: *Sống mòn*, part 8

chỉ	*chiếc*	*ghế*	*còn*	*bỏ không*
point	CL	chair	still	left empty

"pointed to the empty chair"

(15) Source: *Sống mòn*, part 10

những	*chiếc*	*răng*	*thưa*	*chẳng còn*	*chiếc nào*	*dính*	*chiếc nào*
PL	CL	tooth	sparse	no longer	CL which	touch	CL which

"teeth so sparse that none touched another"

(16) Source: *Sống mòn*, part 13

trong	*chiếc*	*áo dài*	*[…] với*	*chiếc*	*quần lĩnh*	*đen*
in	CL	long dress	with	CL	silk trousers	black

"in a long dress with black silk trousers"

(17) Source: *Sống mòn*, part 8

Bà	*mời*	*Mô*	*ngồi*	*ghé*	*bên*	*chiếc*	*ngựa*
Madam	invite	Mô	sit	briefly	beside	CL	bench

"The madam invited Mô to sit briefly beside the bench."

The tooth example is constructionally important. Những *chiếc răng* places the classifier below a plural marker, and *chiếc nào* then retrieves individual members of the set. Teeth are body parts, not animate referents. Their individuation continues the earlier pattern linking a part to a whole seen with a wheel while also showing compatibility with a modern plural construction.

Clothing and furniture confirm broader artifact selection. In Example 17, the edition used here reads *ngựa* as a bench or another item similar to a stool. The immediate context, however, does not securely resolve object versus animal reference, so the token is coded as uncertain/other and is not used as central evidence for inanimate classification.[5] Across the securely coded *Sống mòn* tokens, the classifier presents referents as bounded objects available for perception, comparison, or local discourse tracking.

The modern broadening is therefore real but limited. The dataset does not show a productive human or animal classifier, nor does it show *chiếc* replacing the general classifier *cái*. Informal contemporary combinations with human nouns require a separate synchronic and sociopragmatic analysis. Table 4 summarizes the attested layers described above.

Table 4. Layered diachronic profile of *chiếc* in the dataset.

Layer	Witnesses	Representative Constructions	Interpretation
0. Sampled lack of attestation	*Phật thuyết đại báo phụ mẫu ân trọng kinh*	No token	No inference beyond this text.
I. Buddhist bridge	*Thiền tông Khóa hư ngữ lục*	*chiếc giày* beside 隻履	Entry mediated by contact in a Vietnamese [CL + N] construction.
II. Limited early expansion	*Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú*	*chiếc đũa*; *chiếc bóng*; *chiếc bánh xe*	Paired item, visible counterpart, and a unit defined through a relation between a part and a whole.
III. Collective and vessel uses	*Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời*; *Thiên Nam ngữ lục*; *Chữ Quốc ngữ thế kỉ XVIII*; *Sách sổ sang chép các việc*	Quantified units and vessels	Collective uses shaped by translation and vessel stabilization supported by genre.

Table 4. *Cont.*

Layer	Witnesses	Representative Constructions	Interpretation
IV. Named vessel	*Thầy Lazaro Phiền*	*chiếc Jean-Dupuis*	Classifier frames a foreign proper name denoting a ship.
V. Pointable countables	*Sống mòn*	*chiếc ghế*; *những chiếc răng*; *chiếc nào*	Broader artifact, body part, plural, and elliptical reference.

Note. The layers describe attested distributions and may overlap. They are not replacement stages or a complete periodization of spoken Vietnamese.

6. Discussion

6.1. Integration Conditioned by Contact

The clearest historical result is the asymmetry between entry and later distribution. At the entry point, Chinese evidence is concrete: 隻履 and *chích lí* align with *chiếc giày* in a Buddhist narrative. In later texts, the classifier remains historically related to the source form, but immediate Chinese templates are not recoverable. Later uses must therefore be explained mainly through Vietnamese constructions rather than direct source replication.

This pattern supports integration conditioned by contact rather than simple copying. Borrowed matter enters through a bilingual textual routine, while Vietnamese supplies the structural network in which the item persists. Quantified phrases, [CL + N] phrases without a numeral, phrases marked for plurality, ellipsis, and proper name framing are not a transferred Chinese inventory. They are Vietnamese resources that redistribute a form related to the source. The point is not that every one of these configurations is absent from Chinese: the quantified order [Q + CL + N] overlaps, and bare [CL + N] constructions occur in some Sinitic varieties. Rather, the Vietnamese record shows that the borrowed form was redistributed through the combinations and discourse functions licensed in Vietnamese.

The distinction also clarifies the term grammaticalization. The data do not show a lexical noun becoming a classifier inside Vietnamese. They show an item already used as a classifier acquiring wider constructional and discourse integration. Persistence remains visible in pairedness and bounded singularity, while expansion occurs through new hosts and contexts. The contrast with *cái* makes the item-specific nature of this development clear. Both classifiers belong to the same broad first millennium layer of Early Chinese loans, but they entered Vietnamese with different source semantics and later became established in different constructions (Alves, 2022). Their divergence therefore does not imply that *chiếc* is a later reshaping of *chích*. In this sense, the history of *chiếc* is one instance of a wider pattern in which grammatical material of Chinese origin undergoes additional development in the recipient language (Alves, 2005, 2007, 2022; T. Phan et al., 2024).

The case is useful because the same item can be followed across four levels: etymological origin, textual mediation, constructional licensing, and semantic extension. This separation is not unique to *chiếc*, but it offers a method that can be applied to other grammatical loans.

6.2. Recipient Language Restructuring in a Wider Perspective

The Vietnamese case illustrates a broader contact principle: borrowed grammatical material is not necessarily reproduced with its distribution in the source language. Recipient languages provide their own slots, restrictions on co-occurrence, and discourse functions. A classifier borrowed into a language that already licenses [CL + N] without a numeral can enter referential constructions unavailable, differently analyzed, or less central in the source language.

Classifier systems create several possible sites of integration. A form may first occur in exact counting, then in noun phrases without numerals, in expressions marked for

plurality, or as the recoverable center of an elliptical nominal. Which pathway is available depends on the recipient grammar. Similar classifier noun strings across languages may consequently reflect direct borrowing, areal convergence, typological similarity, or parallel use of a common nominal strategy. Formal similarity alone cannot decide among them.

Yan et al. (2026) show, in a different setting, that Chinese influence on Dai Lue varies across lexical and syntactic domains and across generations. Their evidence from apparent time is not a historical parallel to *chiếc*, but it reinforces the need to analyze contact within a local multilingual ecology. The comparison concerns method: source influence must be separated from recipient patterns and from the social conditions that make particular variants available.

This case does not establish a universal classifier pathway. It shows more narrowly how recipient constructions and recurring discourse needs can redistribute a borrowed form in an uneven written record.

6.3. Textual Ecology and the Interpretation of Frequency

The maritime cluster demonstrates why historical frequency must be ecologically interpreted. *Sách sổ sang chép các việc* contains nearly half of all tokens because it repeatedly discusses the referent class most strongly associated with *chiếc* in the sample from the nineteenth century. Its high density documents a stable vessel pattern within that text. It does not by itself show an abrupt increase across the language in the early 1800s.

Genre affects both semantic opportunities and constructional repetition. Travel prose requires exact enumeration, comparison, and anaphoric retrieval of vessels. Realist fiction requires deictic presentation and tracking of objects in scenes. Buddhist and catechetical texts create different referential demands. The historical record can therefore make one construction prominent not because it is newly grammatical, but because a text repeatedly needs it.

This point also constrains negative evidence. The zero in *Phật thuyết đại báo phụ mẫu ân trọng kinh* is compatible with several explanations: absence from the sampled witness, lack of relevant referents, textual replacement by another classifier, or historical absence. The dataset cannot choose among them. A responsible claim is limited to attestation and distribution, not speaker competence or categorical availability. A comparable endpoint for recent usage would require a balanced corpus and the same coding at the token level; a few search engine or social media queries would not provide a defensible frequency comparison.

6.4. Semantic Persistence and Local Stabilization

The semantic trajectory is neither random nor a simple progression from narrow to general. A single shoe, a chopstick, a wheel, a ship, a tooth, and a chair differ ontologically, but each can be construed as a bounded unit. Paired objects and parts provide especially transparent bridges because they foreground one member of a larger set or whole. Vessels add mobility and repeated set membership, while modern artifacts extend pointability in discourse.

Repetition within one domain may stabilize a local construction. The pairing [Num/Q + *chiếc* + vessel noun] recurs frequently in *Sách sổ sang chép các việc*, and the same discourse supports elliptical vessel phrases. Yet repetition within one domain does not prove that the classifier generalized freely to all inanimates. The absence of a broad inventory of noun types in the earlier texts and the continued semantic concentration argue for a modest account of extensibility.

The tokens coded as collective units show a different bridge. Their referents are internally plural groups, but the construction counts organized units. They should not be assimilated either to ordinary singular objects or to human classification. Their isolation in

a catechetical context shaped by translation further cautions against treating every semantic outlier as a stage in a single generalization chain.

6.5. Chinese and Southeast Asian Comparisons

Chinese contributes both a plausible historical form and a specific early expression. Liu's (2024) semantic study of 只 confirms the continuing importance of "one of a pair" and also documents playful Internet uses with human referents. The comparison is relevant because it shows that classifier extensions can be pragmatically marked and socially situated. It does not establish that contemporary Vietnamese combinations with human nouns developed directly from Chinese.

At the syntactic level, Chinese and Vietnamese differ in the distribution and interpretation of bare nouns, classifier phrases, plural material, and demonstratives (Cheng & Sybesma, 1999; Đoàn et al., 2019; Y.-H. A. Li, 1999; T. Phan & Chierchia, 2022; Simpson, 2012; Trinh, 2011). The literature remains divided on the exact DP structure, so the historical argument does not depend on one tree. The robust contrast is distributional: Vietnamese *chiếc* occurs in several recipient constructions whose later uses cannot be read off from 隻履. Đoàn et al. (2019) argue that bare [CL + N] is definite in Vietnamese while the corresponding Mandarin pattern is indefinite. The historical tokens examined here often occur in contexts of discourse retrieval, but the sample is not sufficient to identify an inherited Chinese indefinite reading as a stable remnant in Vietnamese.

Across Southeast Asia, classifier order and bare classifier constructions vary. Such patterns can be areal without being genealogically or historically identical. The Vietnamese evidence points to borrowing of the form and an early textual model, followed by restructuring in Vietnamese. Typological parallels in neighboring languages clarify the range of structural possibilities, but they do not supply missing historical links. Bare classifier noun patterns in Cantonese and other Sinitic varieties also caution against treating Vietnamese [CL + N] as typologically unique (Cheng & Sybesma, 1999). Such parallels are not evidence of direct Yue influence without independent historical support.

6.6. Limits and Evidential Scope

This study is limited by the size and composition of the dataset. Several periods are represented by one witness, and the early evidence is sparse. Text length, topic, copying, edition, and preservation all affect the apparent distribution. The article therefore reconstructs an attested written trajectory, not a complete history of spoken Vietnamese. The amount of premodern Nôm evidence is also permanently constrained by preservation and transmission. Possible historical losses cannot, however, be converted into a specific causal explanation for any individual zero count.

The corpus does not independently date the phonological borrowing of *chiếc*. Its classification as an Early Chinese loan rests on comparative historical phonology (Alves, 2018, 2022, 2025; Alves & Vũ, 2025). The Buddhist correspondence provides a later textual bridge, but it cannot identify the original borrowing event or donor variety. The article therefore separates a well-supported textual connection from the broader etymological reconstruction.

7. Conclusions

Vietnamese *chiếc* is best understood as an Early Chinese loan from the first millennium CE, distinct from the later regular Sino-Vietnamese reading *chích*. It became grammatically integrated through Vietnamese constructions. The earliest clear token in *Thiền tông Khóa hư ngữ lục* stands beside 隻履, *chích lí*, and provides a specific Buddhist contact bridge. That bridge explains an entry context, not the full later distribution.

Subsequent witnesses show layered expansion. *Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú* extends the form to paired and figurative units, as well as units defined through a relation between a part and a whole. *Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời* supplies two collective units shaped by translation. *Thiên Nam ngữ lục* and *Chữ Quốc ngữ thế kỉ XVIII* establish an early vessel pattern, which becomes locally dense and constructionally diverse in *Sách sổ sang chép các việc*. *Thầy Lazaro Phiền* frames a named foreign ship, and *Sống mòn* broadens the classifier to pointable artifacts, clothing, and teeth.

The strongest evidence concerns constructional distribution rather than frequency growth. A small historical corpus cannot establish productivity across the language, and normalized rates cannot neutralize genre, topic, or preservation. What the record does show is that *chiếc* participates in quantified phrases, [CL + N] without a numeral, phrases marked for plurality, elliptical reference, modification, and proper name framing. These uses demonstrate restructuring in Vietnamese.

The broader implication is that grammatical borrowing need not preserve distributions in the source language. A form may enter through one textual setting, while the recipient grammar shapes its later combinations. Recurrent semantic domains can stabilize local patterns, but local density does not prove unrestricted generalization. The history of *chiếc* therefore shows where evidence from the source language ends and restructuring in Vietnamese begins.

Author Contributions: Conceptualization, T.P. and M.S.; methodology, T.P. and M.S.; validation, T.P. and M.S.; formal analysis, T.P. and M.S.; investigation, T.P. and M.S.; resources, T.P. and M.S.; data curation, T.P. and M.S.; writing, original draft preparation, T.P.; writing—review and editing, T.P. and M.S. For academic clarity, T.P. was primarily responsible for Sections 2, 3, 5.3–5.5 and 6, while M.S. was primarily responsible for Sections 1, 4, 5.1, 5.2 and 7. All authors have read and agreed to the published version of the manuscript.

Funding: The second author (Masaaki Shimizu)'s work was partially supported by the National Institute for Japanese Language and Linguistics (NINJAL) under the Collaborative Research Project "Co-creation of Research Infrastructure Through the Integration of Diverse Lexical Resources," subproject "Development of a Japanese Function Expression Wordbank for Japanese Language Learners and a Theoretical Contrastive Study of Lexicalization and Nominalization." The first author (Trang Phan) received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

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 thank the anonymous reviewers and the guest editors for their constructive comments. The authors take full responsibility for the content of this publication.

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

Notes

[1] Contemporary social media examples in which *chiếc* occurs with human referents, such as *chiếc bác sĩ* "the doctor," *chiếc học sinh* "the student," *and chiếc em bé* "the baby," fall outside the historical dataset examined here. These uses may reflect pragmatic, evaluative, affective, playful, or other extensions that express stance, but the evidence presently available is too limited to determine their frequency, productivity, social distribution, or grammatical status. They are therefore best treated as suggestive synchronic observations rather than as evidence for an established extension of the classifier system. Further research based on systematically sampled social media data, with attention to discourse context, speaker stance, genre, and distribution across user groups, would be needed before stronger conclusions could be drawn.

[2] A potentially relevant comparison is reported for Mandarin 只 zhī. Liu (2024, sec. 5.2) discusses creative Internet language uses of the classifier with human referents and relates them to playful stylistic effects. The comparison is suggestive because it raises the possibility that classifier choice with human nouns may be exploited pragmatically in more than one Sinitic or Southeast Asian context. At present, however, it should not be taken as evidence of direct borrowing, shared grammaticalization, or equivalent grammatical status in Mandarin and Vietnamese. A fuller comparison would require independent synchronic corpora and closer analysis of the discourse conditions under which such uses with human referents occur in each language.

[3] Phật thuyết đại báo phụ mẫu ân trọng kinh (佛說大報父母恩重經 'Sūtra on the Profound Gratitude Owed to Parents') is among the earliest Vietnamese texts combining vernacular Nôm with Literary Sinitic. Its date remains disputed. Shimizu (2020) examines Sino-Vietnamese initials reflected in the phonetic components of Nôm characters from the fifteenth century, but this evidence does not by itself settle the date of the entire text. The text is therefore assigned to a stratum dated provisionally to the thirteenth century in this analysis.

[4] The present study examines pages 1–228 (ending partway through page 228) and 549–597 of *Sách sổ sang chép các việc*. Approximately half of the text remains outside the scope of this analysis; we hope to extend our investigation to the entire work in future research.

[5] The interpretation of *chiếc ngựa* in Example 17 requires caution. In the edition used here, *ngựa* has been read as a bench or another item similar to a stool, but the immediate context does not securely rule out the animal noun. The dataset therefore codes the token as uncertain/other and excludes it from the central evidence for inanimate classification. Vũ (2023) cites a different line from the nineteenth century, *chiếc ngựa trắng đi về góc bể* "the white horse goes towards the edge of the sea," to question a strict animate/inanimate division. That example may involve construal through the established domain of vehicles or means of conveyance, but one isolated token does not establish productive animate classification.

References

Primary Corpus Sources

1. *Phật thuyết đại báo phụ mẫu ân trọng kinh* 佛說大報父母恩重經 [Sutra of the Buddha's teaching on the profound gratitude owed to parents]. Call number PD.2350, Bibliothèque de la Société Asiatique. See also Hoàng Thị Ngọ. (2022). *Chữ Nôm và tiếng Việt qua bản giải âm Phật thuyết đại báo phụ mẫu ân trọng kinh*. Văn học.
2. *Thiền tông Khóa hư ngữ lục* 禪宗課虛語錄 [Recorded sayings of Zen teachings]. Call number A.268, Institute of Sino-Nom Studies, Hanoi. See also Trần Trọng Dương. (2009). *Thiền tông Khóa hư ngữ lục*. Văn học.
3. *Tân biên Truyền kỳ mạn lục tăng bổ giải âm tập chú* 新編傳奇漫錄增補解音集註 [Newly edited tales of the marvelous, with added explanations and phonetic annotations]. Call numbers HN.257 and HN.258, Institute of Literature. See Nguyễn Quang Hồng. (2001). *Truyền kỳ mạn lục giải âm*. Nhà xuất bản Khoa học Xã hội.
4. de Rhodes, A. (1651). *Phép giảng tám ngày cho kẻ muốn chịu phép rửa tội, mà vào đạo thánh Đức Chúa Trời* [Catechism in eight days for those who wish to receive baptism and enter the holy faith of God]. Sacra Congregatio de Propaganda Fide. See also Nguyễn Khắc Xuyên (Ed.). (1993). *Phép giảng tám ngày*. Tủ sách Đại Kết.
5. *Thiên Nam ngữ lục* 天南語錄 [Records of Southern language]. Call number AB.478/1–2, Institute of Sino-Nom Studies, Hanoi. See Nguyễn Thị Lâm. (2001). *Thiên Nam ngữ lục*. Văn học.
6. Đoàn Thiện Thuật (Ed.). (2008). *Chữ Quốc ngữ thế kỉ XVIII*. Giáo dục.
7. Philippe Bỉnh. (1968 [1822]). *Sách sổ sang chép các việc*. University of Đà Lạt.
8. Nguyễn Trọng Quản. (1887). *Thầy Lazaro Phiền*. J. Linage.
9. Nam Cao. (1956 [1944]). *Sống mòn*. Văn nghệ.

Lexicographic Sources

1. de Rhodes, A. (1651). Dictionarium annamiticum lusitanum et latinum: Ope Sacrae Congregationis de Propaganda Fide in lucem editum. Typis et sumptibus eiusdem Sacrae Congregationis.
2. Taberd, J.-L. (Ed.). (1838). Dictionarium Anamitico-Latinum. Ex typis J. C. Marshman.
3. Huình-Tịnh Paulus Của. (1895–1896). Đại Nam quấc âm tự vị (2 vols.). Imprimerie Rey, Curiol & Cie.
4. Hoàng Phê (Ed.). (2011). Từ điển tiếng Việt: 45.850 mục từ. Nhà xuất bản Đà Nẵng & Trung tâm Từ điển học.
5. Nguyễn Quang Hồng. (2014). Tự điển chữ Nôm dẫn giải (2 vols.). Nhà xuất bản Khoa học Xã hội.

Secondary Sources

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. Oxford University Press.

Alves, M. J. (2005). Sino-Vietnamese grammatical vocabulary and triggers for grammaticalization. In *The 6th Pan-Asiatic international symposium on linguistics* (pp. 315–332). Nhà xuất bản Khoa học Xã hội.

Alves, M. J. (2007). Sino-Vietnamese grammatical borrowing: An overview. In Y. Matras, & J. Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 343–362). Mouton de Gruyter. [CrossRef]

Alves, M. J. (2018). Early Sino-Vietnamese lexical data and the relative chronology of tonogenesis in Chinese and Vietnamese. *Bulletin of Chinese Linguistics*, *11*(1–2), 3–33. [CrossRef]

Alves, M. J. (2022). Vietic and early Chinese grammatical vocabulary in Vietnamese: Native Vietic and Austroasiatic etyma versus early Chinese loanwords. In M. J. Alves, & P. Sidwell (Eds.), *Papers from the 30th meeting of the Southeast Asian linguistics society (2021)* (JSEALS Special Publication No. 8, pp. 1–24). University of Hawai'i Press. Available online: http://hdl.handle.net/10524/52498 (accessed on 4 August 2026).

Alves, M. J. (2025). Vowels in early Chinese loanwords in Vietnamese with comparative data from Vietic. In P. Sidwell (Ed.), *Austroasiatic linguistics: In honour of Michel Ferlus (1935–2024)* (pp. 19–44). Myanmar Center, Faculty of Humanities, Chiang Mai University.

Alves, M. J., & Vũ, Đ. N. (2025). Cần xem xét lại nguồn gốc, sự hình thành và quan niệm về những từ được gọi là "từ Hán Việt Việt hoá". *Ngôn Ngữ*, *1*, 3–21.

Barðdal, J., Enghels, R., Feltgen, Q., Van Hulle, S., & Lauwers, P. (2026). Productivity in diachrony. In A. Ledgeway, E. Aldridge, A. Breitbarth, K. É. Kiss, J. Salmons, & A. Simonenko (Eds.), *The Wiley Blackwell companion to diachronic and historical linguistics*. Wiley Blackwell. [CrossRef]

Bisang, W. (1999). Classifiers in East and Southeast Asian languages: Counting and beyond. In J. Gvozdanović (Ed.), *Numeral types and changes worldwide* (pp. 113–186). Mouton de Gruyter. [CrossRef]

Bisang, W. (2015). Problems with primary vs. secondary grammaticalization: The case of East and Mainland Southeast Asian languages. *Language Sciences*, *47*, 132–147. [CrossRef]

Bisang, W., & Quang, K. N. (2020). (In)definiteness and Vietnamese classifiers. In K. Balogh, A. Latrouite, & R. D. Van Valin Jr. (Eds.), *Nominal anchoring: Specificity, definiteness and article systems across languages* (pp. 15–49). Language Science Press. [CrossRef]

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]

Diewald, G. (2002). A model for relevant types of contexts in grammaticalization. In I. Wischer, & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 103–120). John Benjamins. [CrossRef]

Đoàn, Q. N. T., Everaert, M. B. H., & Reuland, E. J. (2019). (In)definiteness of Vietnamese noun phrases. In N. Duffield, T. Phan, & T. Trinh (Eds.), *Interdisciplinary perspectives on Vietnamese linguistics* (pp. 155–180). John Benjamins. [CrossRef]

Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer, & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 83–101). John Benjamins. [CrossRef]

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

Her, O.-S., Hammarström, H., & Allassonnière-Tang, M. (2022). Defining numeral classifiers and identifying classifier languages of the world. *Linguistics Vanguard*, *8*(1), 151–164. [CrossRef]

Hopper, P. J. (1991). On some principles of grammaticization. In E. C. Traugott, & B. Heine (Eds.), *Approaches to grammaticalization* (Vol. 1, pp. 17–36). John Benjamins. [CrossRef]

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge University Press.

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

Li, Y.-H. A. (1999). Plurality in a classifier language. *Journal of East Asian Linguistics*, *8*(1), 75–99. [CrossRef]

Liu, J. (2024). Cognitive, linguistic, and socio-cultural perspectives on classifier usage: A diachronic and synchronic semantic study of the classifier 只 (zhī) in Modern Mandarin. *Asian Languages and Linguistics*, *5*(2), 310–336. [CrossRef]

Matras, Y., & Sakel, J. (Eds.). (2007). *Grammatical borrowing in cross-linguistic perspective*. Mouton de Gruyter. [CrossRef]

Nguyen, D. H. (1957). Classifiers in Vietnamese. *Word*, *13*(1), 124–152. [CrossRef]

Nguyen, T. H. (2004). *The structure of the Vietnamese noun phrase* [Doctoral dissertation, Boston University].

Nguyễn, P. P. (1995). *Questions de linguistique vietnamienne: Les classificateurs et les déictiques*. École française d'Extrême-Orient.

Nguyễn, T. C. (1975). *Từ loại danh từ trong tiếng Việt hiện đại*. Khoa học Xã hội.

Peyraube, A. (1998). On the history of classifiers in Archaic and Medieval Chinese. In B. K. T'sou (Ed.), *Studia linguistica serica* (pp. 131–145). City University Press.

Phan, J. 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].

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

Phan, T., & Chierchia, G. (2022). Identifying (in)definiteness in the Vietnamese noun phrase. *Journal of the Southeast Asian Linguistics Society*, *15*(2), 27–49. Available online: http://hdl.handle.net/10524/52502 (accessed on 4 August 2026).

Phan, T., & Chierchia, G. (2026). Putting three pieces of a puzzle into place: Classifiers–plurals–articles in Vietnamese. In A. Simpson (Ed.), *New insights into theoretical syntax from Asian languages: Studies in honor of C.-T. James Huang* (pp. 436–458). John Benjamins. [CrossRef]

Phan, T., Nguyen, T.-C., & Shimizu, M. (2024). The rise of negative markers in Vietnamese: Evidence from primary Nôm texts. In T. Phan, T.-C. Nguyen, & M. Shimizu (Eds.), *Studies in Vietnamese historical linguistics* (pp. 29–50). Springer. [CrossRef]

Shimizu, M. (2020). Sino-Vietnamese initials reflected in the phonetic components of 15th-century Nôm characters. *Journal of Chinese Writing Systems*, *4*(3), 183–195. [CrossRef]

Simpson, A. (2012). Classifiers and DP structure in Southeast Asia. In G. Cinque, & R. S. Kayne (Eds.), *The Oxford handbook of comparative syntax* (pp. 806–838). Oxford University Press. [CrossRef]

Simpson, A., & Ngo, B. (2018). Classifier syntax in Vietnamese. *Journal of East Asian Linguistics*, *27*, 211–246. [CrossRef]

Taavitsainen, I., Włodarczyk, M., & Tyrkkö, J. (2026). Genre variation and change across registers of writing. In A. Ledgeway, E. Aldridge, A. Breitbarth, K. É. Kiss, J. Salmons, & A. Simonenko (Eds.), *The Wiley Blackwell companion to diachronic and historical linguistics*. Wiley Blackwell. [CrossRef]

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.

Tran, H. T. T. (2021). *A corpus-based study of inanimate classifiers in Vietnamese* [Doctoral dissertation, University of Manitoba]. Available online: http://hdl.handle.net/1993/35426 (accessed on 4 August 2026).

Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford University Press.

Trinh, T. (2011). Nominal reference in two classifier languages. In I. Reich, E. Horch, & D. Pauly (Eds.), *Proceedings of Sinn und Bedeutung 15* (pp. 629–644). Universaar.

Vũ, Đ. N. (2023). Do so-called classifiers in Vietnamese categorize nouns? 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. 339–378). Geolinguistic Society of Japan. [CrossRef]

Wang, L. (1948). Hàn-Yuè yǔ yánjiū [A study of Sino-Vietnamese]. *Lingnan Journal*, *9*(1), 1–96.

Whitt, R. J. (Ed.). (2018). *Diachronic corpora, genre, and language change*. John Benjamins. [CrossRef]

Yan, N., Suraratdecha, S., & Yurayong, C. (2026). Generational variation in language convergence: Lexical and syntactic change in Dai Lue under Chinese influence. *Languages*, *11*(1), 3. [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.