Journal of Liberal Arts, Thammasat University
วารสารศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์
Journal homepage: https://so03.tci-thaijo.org/index.php/liberalarts/index

Shift of Markedness of Relative Clause Markers /thiː3/, /sɯŋ3/, and /ʔan1/ in Thai

Sathon Thaiklang*

Faculty of Arts, Chulalongkorn University, Thailand

Article Info

Research Article

Article History:

Received 18 May 2025

Revised 9 September 2025

Accepted 3 October 2025

Keywords:

shift of markedness,
relative clause markers,
syntactic change

* Corresponding author
E-mail address:
sathon.thaiklang@gmail.com

Abstract

This study investigates the shift of markedness of relative clause markers /thiː3/, /sɯŋ3/, and /ʔan1/ in Thai across historical periods, using 3,879 texts in prose and verse across various genres from TNHC2 (Pittayaporn et al., 2023) and the Diachronic Thai Text Corpus (Thanyawong, 2025). Together, the corpus spans about 700 years from Sukhothai to the present. The findings reveal that /ʔan1/ had the highest token frequency in Sukhothai and Early Ayutthaya (14th–15th centuries) and Late Ayutthaya (18th century), functioning as the unmarked form from the 14th to the 18th centuries. By contrast, /sɯŋ3/ peaked in Middle Ayutthaya (16th century) but has never attained unmarked status because its use did not broaden across contexts. In the 19th century, /thiː3/ emerged as the unmarked form, consistent with Kullavanijaya (2008). Marker choice is further related to semantic, syntactic, and text-level factors. Semantically, /thiː3/ shifted from a locative-restricted marker to one that modifies a wider range of head noun domains. Syntactically, /ʔan1/ was once widely attested in direct-object and oblique positions but has become increasingly concentrated in subject positions, especially those with adjectival predicates. At the text level, /ʔan1/ is favored in fiction, legal, and non-academic genres and in content categories such as beliefs and ideas, social sciences, imaginative topics, and world events and history.

1. Introduction

Shift of markedness represents a mechanism of diachronic change whereby variants that were once frequent and broadly distributed may narrow to specific contexts and become marked, while formerly rarer, context-restricted variants may generalize and become unmarked (Haspelmath, 2006, pp. 25–30). This dynamic process shows how changes in frequency and distributional context jointly reshape the grammatical status of linguistic forms. The system of Thai relative clause markers provides an ideal case for examining this phenomenon, as marker choice has changed in both frequency and distribution across time.

A relative clause is a subordinate clause that modifies a noun phrase within a complex sentence. The modified noun phrase, often called the head noun or relativized noun, co-refers with a position inside the clause that is realized either as a gap or as a resumptive pronoun. A clause with a gap is structurally incomplete, whereas a clause with a resumptive pronoun is structurally complete because all arguments are overt (Kullavanijaya, 2006, p. 20).

In present-day Thai, relative clauses are introduced by /thi:3/, /sɯŋ3/, or /ʔan1/, with /thi:3/ being the most frequent marker (Prompapakorn, 1996, pp. 48–50). The alternation between /thi:3/ and /sɯŋ3/ has often been linked to differences in restrictiveness (Kullavanijaya, 2006, p. 49), as illustrated in examples (1)–(2), with /thi:3/ restrictive and /sɯŋ3/ non-restrictive.

(1) สัตว์<u>ที่</u>อาศัยอยู่ในทะเลบางตัวเป็นสัตว์เลี้ยงลูกด้วยนม

sàt tʰîː ʔaː.sǎj-jùː nai tʰa.leː baːŋ-tuːa pen
animal REL reside LOC sea some-CLF COP
sàt-líaŋ-lûːk-dûaj-nom
animal-feed-offspring-COM-milk
‘some of the animals that live in the sea are mammals’

(Kullavanijaya, 2006)

(2) สัตว์<u>ซึ่ง</u>อาศัยอยู่ในทะเลบางตัวเป็นสัตว์เลี้ยงลูกด้วยนม

sàt sɯ̂ŋ ʔaː.sǎj-jùː nai tʰa.leː baːŋ-tuːa pen
animal REL reside LOC sea some-CLF COP
sàt-líaŋ-lûːk-dûaj-nom
animal-feed-offspring-COM-milk
‘some of the animals, which live in the sea, are mammals’

(Kullavanijaya, 2006)

For /ʔan1/, marker choice is more closely associated with the syntactic function of the relativized position. In present-day Thai, /ʔan1/ is unattested as a relative clause marker when the relativized position functions as an oblique, as illustrated in (3) (Hongthong, 2018, p. 311). Although structurally parallel to relative clauses with /thi:3/ or /sɯŋ3/, the sentence in (3) is ungrammatical in present-day Thai.

(3) *โฮลเดนก็พาแซลลีไปเซนทรัลพาร์ค**อัน**เคยนั่งดูเป็ดด้วยกันสมัยยังเด็ก

ho:l.den kɔ̂ʔ pʰa: sɛ:n.li: paj sen.tʰan pʰa:k ʔan kʰɤ:j nâŋ
Holden PTCL take Sally go Central.Park REL EXP sit
du: pèt dûaj-kan sà.mǎj jaŋ dèk
watch duck COM-RECP period still child
'Holden then took Sally to Central Park where they used to sit and watch ducks together when they were still children.'

(adapted from Hongthong, 2018)

However, this was not always the case. In earlier Thai, as shown in (4), /ʔan1/ introduces a relative clause in which the relativized position functions as an oblique. This broader earlier distribution illustrates how /ʔan1/ was once unmarked for oblique contexts, demonstrating the diachronic narrowing predicted by shift of markedness.

(4) ปี**อัน**สถาปนาพระมหาธาตุ

pi: ʔan sà.tʰă:.pà.na: phrá-ma.hă:-tʰâ:t
year REL establish sacred-great-relic
'the year that the Great Relic was established

(Nakhonchum Inscription, 1900 BE)

All three markers show changes in frequency and usage patterns over time. In Sukhothai, /thi:3/ occurred exclusively with locative nouns and was the least frequent of the three markers. From Middle Rattanakosin onward, it generalized beyond locatives (Kullavanijaya, 2008, p. 457) and increased steadily, becoming the most common relative clause marker in present-day Thai (Prompapakorn, 1996, pp. 48–50). Early usage shows /ʔan1/ to be the most frequent relative clause marker in Sukhothai, compared to other markers and to its other functions, such as a classifier (Meesri & Wongwattana, 2020, p. 175). Finally, /sɯŋ3/ was attested as a relative clause marker from Sukhothai onward, and this use increased compared to its roles as a topic marker, an end-focus marker, and a complementizer (Yaowapat, 2008, p. 122).

While studies to date describe changes in the grammatical functions of /thi:3/, /sɯŋ3/, and /ʔan1/ individually, interactions and direct comparisons among these forms as relative clause markers remain underexplored. Building on the notion of shift of markedness, the present study addresses this gap by investigating the shift of markedness among /thi:3/, /sɯŋ3/, and /ʔan1/ in Thai. Changes in frequency and distribution are traced from Sukhothai to the present, and the analysis tests how marker choice varies with the semantic domain of the head noun, the syntactic function of the relativized position, and the text genre, content, and writing style in which the markers appear. Marker choice is modeled using mixed-effects multinomial logistic regression with period, semantic domain, syntactic function, and text-level factors as predictors.

The research questions are 1) How have the markedness relations among the relative clause markers /thi:3/, /sɯŋ3/, and /ʔan1/ in Thai changed from the Sukhothai period to the present? and 2) How have the factors related to the choice among /thi:3/, /sɯŋ3/, and /ʔan1/ changed from the Sukhothai period to the present? The hypotheses are that /ʔan1/, showing the highest token frequency and broad distribution, was unmarked in Sukhothai and Early Ayutthaya; /sɯŋ3/ was unmarked from Middle to Late Ayutthaya; and beginning in 2301–2350 BE, /thi:3/ became unmarked. Semantically, /thi:3/ is hypothesized to generalize from a locative-restricted use to modification of a wider range of head noun domains. Syntactically, /ʔan1/ is hypothesized to narrow diachronically, having once been widely attested in object and oblique positions but becoming increasingly restricted over time.

2. Literature Review

2.1 The Structure of Thai Relative Clauses

A relative clause is a subordinate clause that modifies a noun phrase in the matrix clause. Following Kullavanijaya (2006, p. 20), Thai relative clauses may be incomplete when the relativized position is a gap or complete when the relativized position contains a resumptive pronoun. Typological work classifies relative clauses by syntactic structure, head position, and the presence or absence of overt relative clause markers (Comrie, 1989, pp. 145–146). Thai falls into the category of externally headed, post-nominal relative clauses (Warotamasikkhadit, 1972, p. 48, as cited in Dryer, 2013a), similar to other Mainland Southeast Asian languages such as Lao, Vietnamese, and Khmer (Yaowapat & Prasithrathsint, 2008, p. 11). This configuration is also common in SVO languages, including English (Dryer, 2013b).

In externally headed, post-nominal relative clauses, the head noun is on the left and outside the relative clause. This configuration means the head noun is syntactically part of the

matrix clause while being semantically modified by the subordinate clause. The Thai pattern can be represented as **[NP Head [RC REL Clause]]**. In this construction, the relative clause marker introduces the relative clause, which dominates the clause structure organized as a TP (tense phrase). (5) illustrates this pattern. The head noun *ฝ่ายศัตรู* 'enemies' precedes the relative clause *ที่ยืนรายล้อมเขาอยู่* 'who were standing around him' and lies outside it, so the construction is post-nominal and externally headed.

(5) มานั่งประจันหน้ากับ [ฝ่ายศัตรู [ที่_ยืนรายล้อมเขาอยู่]]
ma: nâŋ prà.tɕan-nâ: kàp [fâ:j-sàt.tru: [tʰîː jɯ:n ra:j-ɦɔ́:m khǎwjù]]
come sit confront-face COM side-enemy REL stand surround 3SG PROG
'(They) came and sat face-to-face with the enemies who were standing around him.'
(Kusumawanna 2: The Literary Style, 2554 BE)

Another crucial component of Thai relative clauses is that the head noun in the matrix clause must be co-referential with either a gap or a resumptive pronoun inside the relative clause (Kullavanijaya, 2006, p. 20). Gap-based constructions are more common in Thai, whereas pronoun-based constructions occur mostly in colloquial speech and are rare in written texts (Yaowapat, 2005, p. 129). Above, example (5) illustrates a gap pattern in which the gap inside the relative clause co-refers with the head noun *ฝ่ายศัตรู*, which functions as the subject of the relative clause. Example (6) shows a pronoun-based pattern. The overt pronoun *เขา* 'they' fills the subject position and co-refers with the head noun *ครู* 'teacher' in the matrix clause. These examples highlight the flexibility of Thai relative clause constructions in allowing both gap-based and pronoun-based co-reference.

(6) **ครู**ที่**เขา**สอนภาษาอังกฤษ
kʰru: tʰîː kʰǎw sɔ̌:n pʰa:.sǎ:-ʔaŋ.krìt
teacher REL 3SG teach language-English
'the teacher who teaches English'
(Yaowapat, 2005)

However, Thai also has a clause type known as the pseudo relative clause, which is structurally similar to a relative clause. Like a relative clause, it follows and modifies a noun in the

matrix clause; by contrast, it lacks co-reference between the modified noun and any element inside the modifying clause (Wongwattana, 2005, as cited in Meesri & Wongwattana, 2020, p. 170). As illustrated in example (7), introduced by /ʔan1/, the clause *เขาได้ให้ยาแก่ผู้หญิงอันมีครรภ์นั้นกิน* ‘he gave medicine to that pregnant woman to take’ modifies the noun *บาปกรรม* ‘sin’. However, the modified noun does not co-refer with any element in the clause, so the construction is analyzed as a pseudo relative clause rather than a relative clause.

(7) บาปกรรม**อัน**เขาได้ให้ยาแก่ผู้หญิงอันมีครรภ์นั้นกิน

bà:p-kam ʔan kʰăw dâj hâj ja: kɛ̀:
sin-karma SUBORD 3SG PFV give medicine DAT
pʰû:-jĭŋ ʔan mi: kʰan nán kin
NMLZ.AG-woman REL have pregnancy DEM.DIST ingest
‘the sin which he gave medicine to that pregnant woman to take’
(Traibhumikatha, 2543 BE, as cited in Meesri & Wongwattana, 2020)

In addition, synchronic work has examined marker choice. Even in similar syntactic environments, Thai relative clause markers can alternate, but they differ in restrictiveness. The choice between /thi:3/ and /sɯŋ3/ reflects that distinction. According to Kullavanijaya (2006, p. 50), a restrictive relative clause narrows the reference of the head noun and is typically introduced by /thi:3/, whereas a non-restrictive relative clause provides additional, non-essential information without altering the referent and is typically introduced by /sɯŋ3/.

Although /thi:3/ and /sɯŋ3/ occur in the same syntactic position, in (8), *ที่เป็นโรงรถขนาดใหญ่* ‘that is a large garage’ restricts the referent to the large garage rather than garages of other sizes, whereas in (9), *ซึ่งเป็นโรงรถขนาดใหญ่* ‘which is a large garage’ supplies additional information about the garage’s size (Panthumetha, 1983, as cited in Kullavanijaya, 2006, p. 17).

(8) โรงรถของบ้าน**ที่**เป็นโรงรถขนาดใหญ่อยู่ทางด้านข้างของบ้าน

ro:ŋ.rót kʰɔ̌:ŋ bâ:n tʰî: pen ro:ŋ.rót kʰà.nà:t jàj jù:
garage GEN house REL COP garage size big be
tʰa:ŋ dâ:n-kʰâ:ŋ kʰɔ̌:ŋ bâ:n
direction side GEN house
‘The house’s garage that is a large garage is at the side of the house.’
(Panthumetha, 1983, as cited in Kullavanijaya, 2006)

(9) โรงรถของบ้าน**ซึ่ง**เป็นโรงรถขนาดใหญ่อยู่ทางด้านข้างของบ้าน

ro:ŋ.rót kʰɔ̌:ŋ bâ:n sŵŋ pen ro:ŋ.rót kʰà.nà:t jàj jù:
garage GEN house REL COP garage size big be
tʰa:ŋ dâ:n-kʰâ:ŋ kʰɔ̌:ŋ bâ:n
direction side GEN house
'The house's garage, which is a large garage, is at the side of the house.'

(Panthumetha,1983, as cited in Kullavanijaya, 2006)

In describing Thai relative clause structure, an important overlap must be noted. Thai permits a noun–classifier–adjective sequence of the form [N CLF ADJ] (Iwasaki & Ingkaphirom, 2005, as cited in Singhapreecha, 2007, p. 6). Because /ʔan1/ can function both as a relative clause marker and as a classifier, strings of the form [N /ʔan1/ ADJ] can be parsed in two ways. One parse is a classifier construction that forms a noun phrase. The other parse is a relative clause in which /ʔan1/ introduces a clause with an adjectival predicate, and the relativized position serves as the subject. In present-day Thai, this is the relative clause construction in which /ʔan1/ is widely attested (Sornhiran, 1981, p. 52). As a result, some tokens are ambiguous between these two analyses. Consider the following contrast, Example (10), which consists of (10a) and (10b), use the same words in the same order but allow different interpretations.

For the [N CLF ADJ] structure in (10a), /ʔan1/ functions as a classifier. The structure is a noun phrase. The noun *พระพุทธรูป* 'Buddha sculpture' is followed by the classifier /ʔan1/, and then the adjective *ใหญ่* 'big'. The interpretation is 'a big Buddha sculpture'.

(10a) พระพุทธรูป**อัน**ใหญ่

pʰrá-pʰút.tʰa-rû:p ʔan jàj
Buddha-sculpture CLF big
'A big Buddha sculpture'

(Wat Srichum Inscription, 1910 BE)

For the [NP Head [RC REL [ADJP ADJ]]] structure in (10b), /ʔan1/ functions as a relative clause marker. The head noun *พระพุทธรูป* is relativized as the subject of the adjectival predicate *ใหญ่*. The interpretation is 'the Buddha sculpture that is big'.

(10b) พระพุทธรูป**อัน**ใหญ่
pʰrá-pʰút.tʰa-rûːp ʔan jàj
Buddha-sculpture REL be.big
'The Buddha sculpture that is big'
(Wat Srichum Inscription, 1910 BE)

Although (10a) and (10b) can yield the same truth conditions, they differ in morphosyntax. In (10a), /ʔan1/ is a classifier in an NP of the form [N CLF ADJ]; the adjective is an NP-internal modifier. In (10b), /ʔan1/ is a relative clause marker introducing a clause [NP Head [RC REL [ADJP ADJ]]], where *ใหญ่* is the predicate of the relative clause. These examples reveal a structural ambiguity: the same string can be parsed as [N CLF ADJ] with /ʔan1/ as a classifier, or as [NP Head [RC REL [ADJP ADJ]]] with /ʔan1/ as a relative clause marker. This ambiguity motivates an explicit coding policy stated in the Methodology section.

2.2 Noun Phrase Accessibility Hierarchy

As noted in the introduction about the contextual constraint of /ʔan1/, it seems that the syntactic function of the co-referent argument within a relative clause is related to the choice of relative clause markers. In other words, the likelihood that a particular marker will be used depends on the syntactic function of the relativized position. This idea appears to be related to the Noun Phrase Accessibility Hierarchy (NPAH) (Keenan & Comrie, 1977, p. 66), which proposes a universal scale of relativizability. On this scale, subjects are the most accessible and therefore the easiest to relativize, while objects of comparison are the least accessible, as illustrated in (11).

(11) *Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of Comparison*

Examples (12)–(17) illustrate Thai head nouns relativized at successive positions on the NPAH (Keenan & Comrie, 1977, p. 66): (12) subject, (13) direct object, (14) indirect object, (15) oblique, (16) genitive, and (17) object of comparison. These examples demonstrate how the NPAH applies to Thai relative clauses.

(12) ผู้**ที่**ตกอยู่ในห้วงรัก
pʰûː tʰîː tòk jùː nai hûaŋ-rák
NMLZ.AG REL fall PROG LOC abyss-love
'the one who falls in love'
(Principles and Philosophy of Rhetoric, 2551 BE)

(13) ผลกรรม**ที่**ตนกระทำ

pʰŏn-kam tʰîː ton kra.tʰam

result-karma REL REFL do

'the karmic result that one brings about'

(A Brief History of Thai Literature: Traibhumi Story, Paraphrased Version, 2555 BE)

(14) คน**ที่**ข้าพเจ้าปลูกให้

kʰon tʰîː kʰâː.pʰá.tɕâw plùːk hâj

person REL 1SG inoculate DAT

'the person to whom I gave a vaccination'

(Vaccination Text, 2388 BE)

(15) ทาง**ที่**ชาวเมืองจะมาตักน้ำ

tʰaːŋ tʰîː tɕʰaːw-mɯaŋ tɕà maː tàk nâːm

Path REL town-people FUT come fetch water

'the path where the townspeople will come to fetch water'

(Tang Han, 2505 BE)

(16) ฉันรู้จักคน**ที่**เมียป่วย

tɕʰăn rúː.tɕàk kʰon tʰîː mia pùaj

1SG know person REL wife sick

'I know the person whose wife is sick.'

(Yaowapat, 2005)

(17) นักกรีฑา**ที่**ฉันวิ่งเร็วกว่า

nák-kriːtʰaː tʰîː tɕʰăn wîŋ rew kwàː

NMLZ.AG-athletics REL 1SG run fast COMP

'the athlete whom I run faster than'

(Hongthong, 2018)

In sum, all positions are attested in Thai, though in this corpus, the lower positions vary across periods. The distribution of Thai relative clauses across these positions follows predictable patterns. Relative clauses cluster at higher positions such as subject and direct object, and are rarer at lower positions such as indirect object, oblique, and genitive. Because accessibility differs

by function, marker choice tends to align with the relativized position; sustained spread into lower positions may signal functional expansion, whereas retreat from them may indicate restriction.

2.3 Markedness and Its Shift

Markedness is an idea pertaining to asymmetry between linguistic forms. Within a system, one variant tends to be common and broadly distributed, while another is rarer and more restricted. The former is *unmarked*, and the latter is *marked*. Markedness is linked to frequency and contextual range, so higher frequency and broader distribution imply an unmarked status, whereas lower frequency and narrower distribution align with a marked status (Greenberg, 1966, p. 14).

Relations between marked and unmarked forms can change over time. Shift of markedness occurs when a previously marked form increases in frequency, spreads to wider contexts, and consequently becomes unmarked, or when a previously unmarked form decreases in frequency, becomes restricted, and therefore turns marked (Haspelmath, 2006, pp. 25–30). Such shifts are documented in phonology (Rice, 2007), morphology (Brown & Witkowski, 1983), and syntax (Givón, 1977, as cited in Haspelmath, 2006, p. 29).

Shift of markedness may arise from several factors. One pathway that can trigger such shift is grammaticalization. In Thai, /thi:3/ originated as a noun meaning ‘place’. During the Sukhothai period, it appeared exclusively with locative nouns (Kullavanijaya, 2008, p. 453), as in (18), where /thi:3/ functions as a noun in apposition to the locative noun *ฝั่งน้ำมาวลิกคงคา* ‘the bank of Mavalik Ganga’.

(18) **<u>ฝั่งน้ำมาวลิกคงคา</u>**ที่พระพุทธเป็นเจ้าปูจรามขันประดิษฐานพระเกศาธาตุ

fàŋ-náːm maː.wá.lík-koŋ.kʰaː tʰîː pʰrá-pʰút.pen-câw
river.bank PN place Buddha
puː.tɕaː.raːm.kʰăn-prà.dìt.sà.tʰăːn pʰrá-keː.săː-tʰâːt
enshrine HON-hair-relic
‘the bank of Mavalik Ganga, the place the Buddha enshrined the Hair Relic’
(Wat Srichum Inscription, 1910 BE)

/thi:3/ subsequently grammaticalized into a relative clause marker by the Middle Rattanakosin period. Its frequency increased, and its distribution broadened beyond locative nouns (Kullavanijaya, 2008, p. 457) as seen in (19), where /thi:3/ modifies the animate noun *ท่าน* ‘person’, in contrast to its Sukhothai restriction to locatives.

(19) **ท่าน**ที่เป็นเจ้าของเดิม

tʰâːn tʰîː pen tɕâw-kʰɔ̌ːŋ dəːm

person REL COP owner original

'the person who is the original owner'

(The Royal Chronicle of King Rama V's Journey to the Malay Peninsula, 2433–2455 BE)

Viewed through a markedness lens, this grammaticalization process represents a clear shift: /thiː3/ evolved from a marked form restricted to locative nouns to an unmarked form that modifies a wider range of nouns.

Markedness can also shift through external pressures such as cultural salience. In Tzeltal, *čih* originally meant 'deer'. After sheep were introduced, they were first called *tunim čih* 'cotton deer'. As sheep gained cultural importance, *čih* came to denote 'sheep', and 'deer' then required the modifier *teʔtikil* 'wild', yielding *teʔtikil čih* (Haspelmath, 2006, pp. 28–29).

Frequency of use provides another window into shift of markedness. Regarding the word orders of Biblical Hebrew, VSO was once unmarked, while SVO was marked and restricted to specific contexts. As SVO spread into general use, it became unmarked, while VSO became restricted and marked (Givón, 1977, as cited in Haspelmath, 2006, p. 29).

Operationally, frequency is informative when increases persist across adjacent periods and coincide with the relative decline of a competitor. Taken together, persistence over time, relative decline of a competitor, and presence across contexts indicate shift of markedness (Haspelmath, 2006, pp. 25–30). These principles provide a framework for analyzing the diachronic development of relative clause markers in Thai.

3. Data and Methodology

Data are drawn from two sources. The first is TNHC2, a historical corpus of prose and verse spanning around 700 years, including fiction, legal documents, academic and non-academic texts, and newspapers, sourced from the Vajirayana Library website (Pittayaporn et al., 2023). The second is the Diachronic Thai Text Corpus, a multi-period corpus that augments early coverage with Sukhothai and Early Ayutthaya inscriptions and, for Middle Ayutthaya, diplomatic correspondence, trade records, and travelogues drawn from the Thai Inscription Database (Thanyawong, 2025). Together, these two corpora comprise 3,879 texts.

The Diachronic Thai Text Corpus is combined with TNHC2 to improve coverage of the early periods, since TNHC2 underrepresents Sukhothai and Early Ayutthaya inscriptions. Following TNHC2, texts are grouped into nine historical periods to achieve a balanced distribution across

time (Pittayaporn et al., 2023). The combined corpus constitutes the dataset for the analyses that follow.

1.	Around 1857–2072 BE	Sukhothai and Early Ayutthaya
2.	2133–2231 BE	Middle Ayutthaya
3.	2232–2300 BE	Late Ayutthaya
4.	Years 2301–2350 BE	Thonburi and Early Rattanakosin
5.	Years 2351–2400 BE	EarlyRattanakosin
6.	Years 2401–2450 BE	Middle Rattanakosin to the present
7.	Years 2451–2500 BE	
8.	Years 2501–2550 BE	
9.	Years 2551 BE–present	

In this study, token and type frequencies of the relative clause markers are analyzed to trace shift of markedness. Token frequency, the total number of occurrences of each marker within a period, tracks changes in relative use over time. Type frequency, the number of distinct antecedent types with which a marker occurs across periods, tracks how broadly a marker applies and whether it expands into new contexts or becomes constrained. Taken together, the two measures provide complementary evidence for shift of markedness.

Occurrences of /thi:3/, /sɯŋ3/, and /ʔan1/ were retrieved with the corpus-launcher software, AntConc version 4.3.1 (Anthony, 2024). Only tokens that function as relative clause markers were retained. This study follows the definition of a relative clause marker set out in the literature review. A token counts as a relative clause marker only when it introduces a noun-modifying clause that contains an argument co-referential with the head noun. Constructions without such co-reference, such as pseudo relative clauses, are excluded from the analysis.

Sampling procedures differ between early and later periods due to data availability. For Sukhothai and Early Ayutthaya, all attestations were included to ensure a complete picture of early usage, yielding 122 samples because the amount of data at this stage is substantially smaller than in later periods. For each of the remaining eight periods, a stratified random sample of 125 attestations was drawn, resulting in 1,122 total samples across all periods. Within each period, attestations of /thi:3/, /sɯŋ3/, and /ʔan1/ were stratified by two dimensions: the semantic domain of the head noun and the syntactic function of the relativized position. This stratification is necessary because the first appearance of a particular semantic domain or syntactic function is crucial for tracing the development of the marker system. The balanced design improves comparability across periods and stabilizes maximum likelihood estimates and their standard errors.

Marker choice is modeled with multinomial logistic regression, which is appropriate for a nominal outcome with more than two categories and for making comparisons among them. The method fits the research design because the dependent variable, relative clause marker, has three categories, namely /thi:3/, /sɯŋ3/, and /ʔan1/, and pairwise differences among the categories are examined. The independent variables are period; semantic domain of the head noun; syntactic function of the relativized position; and text-level factors including genre, content, and writing style. Interactions are specified between period and semantic domain and between period and syntactic function. A mixed-effects specification is used. Semantic domain and syntactic function are modeled as fixed effects because they are observed across periods and can be controlled explicitly. Text-level factors are modeled as random effects to accommodate their uneven distributions across time. Metadata for these uncontrolled variables are provided in TNHC2 and the Diachronic Thai Text Corpus (Pittayaporn et al., 2023; Thanyawong, 2025). All statistical analyses were conducted using R.

The Concordance (keyword-in-context, KWIC) module was then used to inspect the left and right contexts of each hit. The left context is analyzed to determine the semantic domain of the head noun. Four semantic domains were categorized: concrete nouns (e.g., ‘house’, ‘book’), abstract nouns (e.g., ‘idea’, ‘truth’), locative nouns (e.g., ‘city’, ‘village’), and temporal nouns (e.g., ‘day’, ‘era’). This classification was based on the meaning of the noun that the relative clause modifies. The right context is analyzed to determine the syntactic function of the co-referential argument inside the clause. Functions are classified into five types: subject, direct object, indirect object, oblique, and genitive, following the NPAH (Keenan & Comrie, 1977, p. 66). Example (20) illustrates the procedure. To the left of the marker, the head noun *พนักงาน* ‘employee’ is categorized as concrete because it denotes a tangible person. To the right, the co-referential gap occupies the subject position in the relative clause *ที่ทำแพและตกแต่งเรือ* ‘who makes rafts and decorates boats’, identifying the agent of these actions. The verbs *ทำ* ‘make’ and *ตกแต่ง* ‘decorate’ take the gap as their subject, which co-refers to the head noun *พนักงาน*.

(20) พนักงาน**ที่** ___ทำแพและตกแต่งเรือ
pʰa.nák.ŋa:n tʰî: tʰam pʰɛ: lɛ́ tòk.tɛ̀ŋ rɯa
employee REL make raft and decorate boat
‘the employee who makes rafts and decorates boats’
(The Chronicle of the Reception of Phraya Settakunchon, the First White Elephant Obtained in the Reign of King Rama II, 2355 BE)

This study also examines relative clauses with non-core arguments. These arguments function as obliques, indicating location, time, or instrument (Hongthong, 2018, p. 20). In many cases, the oblique is not expressed inside the relative clause. There is neither a gap nor a resumptive pronoun, and the relation is inferred from the predicate. Example (21) illustrates this pattern. In the relative clause *ข้าพเจ้าอยู่* 'I live', no overt element inside the clause co-refers with the head noun *ตึก* 'building'. The head is interpreted as the location associated with *อยู่* 'live', so the relativized position is analyzed as oblique.

(21) **ตึก**ที่ข้าพเจ้าอยู่
tɯ̀k tʰîː kʰâː.pʰá.tɕâw jùː
building REL 1SG be
'the building where I live'
(Kosapan's Memoirs, 2228 BE)

All instances of /thiː3/, /sɯŋ3/, and /ʔan1/ in the corpus were manually examined. For each token, the head noun was identified, its semantic domain was classified, and the syntactic function of the relativized position was determined from its relation to the verb and surrounding constituents. This procedure produced categorical annotations used in the statistical analysis to test how marker choice varies with semantic and syntactic factors across periods.

As mentioned in the literature review, in sequences where /ʔan1/ occurs in the patterns of [N CLF ADJ] or [NP Head [RC REL [ADJP ADJ]]] and the function of /ʔan1/ is ambiguous, the tokens are annotated as a relative clause. Because the corpus does not reliably distinguish /ʔan1/ used as a relative clause marker from /ʔan1/ used as a classifier in these strings, ambiguous cases are uniformly coded as relative clause markers. Excluding ambiguous tokens would remove the environments that are informative for tracking shift of markedness.

4. Findings

4.1 Change in Relative Frequency

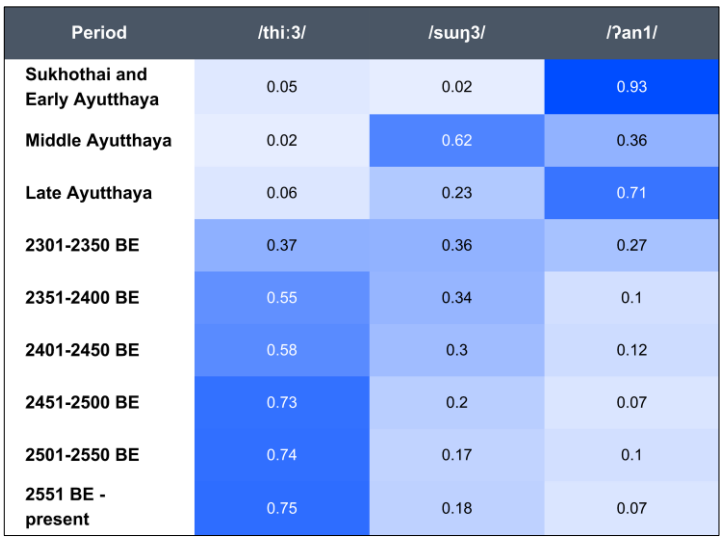
The shift of markedness among Thai relative clause markers is traced through token and type frequencies across historical periods and through modeled probabilities. Together these measures show how the relative prominence of /thiː3/, /sɯŋ3/, and /ʔan1/ changes and how the default choice reorganizes over time.

Figure 1 visualizes predicted probabilities by period. Darker blue indicates a higher probability that a marker is selected in that period, and lighter blue indicates a lower probability. The data reflect ongoing competition among the markers. The pattern reveals successive dominance.

Early periods favor /ʔan1/, Middle Ayutthaya favors /sɯŋ3/, and modern periods favor /thi:3/. This sequence aligns with the idea of shift of markedness, in which the unmarked option changes as the frequencies and distributions reconfigure across time.

Figure 1

Relative frequencies of relative clause markers in Thai over time.



To further examine and corroborate the patterns in Figure 1, this study fits a multinomial logistic regression. The model estimates the effect of historical period on relative clause marker choice. Table 1 reports coefficients for /ʔan1/ and /sɯŋ3/ with /thi:3/ as the reference category, together with significance levels. Positive coefficients indicate higher odds of choosing that marker rather than /thi:3/. Negative coefficients indicate lower odds.

Table 1

Coefficients for /ʔan1/ and /sɯŋ3/ relative to /thi:3/ across periods.

Periods	Raw Counts			Coefficients	
	/thi:3/	/sɯŋ3/	/ʔan1/	/ʔan1/ to /thi:3/	/sɯŋ3/ to /thi:3/
Sukhothai and Early Ayutthaya	6	2	114	2.87***	-0.96
Middle Ayutthaya	3	77	45	2.56***	3.10***
Late Ayutthaya	7	29	89	2.48***	1.37***
2301–2350 BE	46	45	34	-0.30	-0.02

Table 1

Coefficients for /ʔan1/ and /sɯŋ3/ relative to /thi:3/ across periods. (Cont.)

Periods	Raw Counts			Coefficients	
	/thi:3/	/sɯŋ3/	/ʔan1/	/ʔan1/ to /thi:3/	/sɯŋ3/ to /thi:3/
2351–2400 BE	69	43	13	-1.64***	-0.47**
2401–2450 BE	73	37	15	-1.56***	-0.67***
2451–2500 BE	91	25	9	-2.27***	-1.28***
2501–2550 BE	92	21	12	-2.00***	-1.46***
2551 BE–present	94	22	9	-2.30***	-1.44***

*** = $p < 0.001$

4.1.1 /ʔan1/: Dominant Marker in Early Periods

Figure 1 and Table 1 show that the frequencies of /ʔan1/ and /thi:3/ move in opposite directions over time. The statistical analysis indicates that /ʔan1/ exhibits the clearest overall decline among the three relative clause markers, with significant decreases in nearly every period. This pattern traces a trajectory from being the primary choice to a marker with a more limited role within the Thai relative clause system.

In Sukhothai and Early Ayutthaya, /ʔan1/ was by far the most frequently used marker, with a significantly positive coefficient over /thi:3/ equal to 2.87 and $p < 0.001$, corresponding to roughly 17 times the odds of /thi:3/. /sɯŋ3/ and /thi:3/ were much less common. The /sɯŋ3/ versus /thi:3/ contrast is -0.96 and not significant. This predominance suggests that /ʔan1/ was almost the sole option for introducing relative clauses during this interval, with little competition from the other two markers.

By Middle Ayutthaya, /ʔan1/ was overtaken by /sɯŋ3/. The /sɯŋ3/-to-/thi:3/ coefficient is 3.10 with $p < 0.001$, exceeding the /ʔan1/-to-/thi:3/ coefficient of 2.56 with $p < 0.001$. In Late Ayutthaya, /ʔan1/ returned as the most used marker. The /ʔan1/-to-/thi:3/ coefficient is 2.48 with $p < 0.001$, compared with 1.37 with $p < 0.001$ for /sɯŋ3/-to-/thi:3/. After that, the use of /ʔan1/ gradually declined, with consistently negative and significant coefficients in subsequent periods.

Apart from its decreasing frequency, /ʔan1/ also shows semantic and syntactic constraints. Semantically, /ʔan1/ shows a narrowing constraint with concrete nouns. Since 2301–2350 BE, when /thi:3/ started to become the dominant marker, /ʔan1/ has been mostly found relativizing concrete nouns related to holiness or divinity. This pattern suggests a narrowing toward more specific semantic contexts for /ʔan1/.

Example (22) shows /ʔan1/ modifying a concrete noun associated with divinity or sacred status, *พระเจ้าละโว้* ‘King of Lavo’, a historical figure widely regarded as sacred or divine in Thai history. Although the referent is concrete, its association with holiness or royal sacrality pulls /ʔan1/ toward more specialized contexts.

(22) **พระเจ้าละโว้**อันอยู่ในเมืองลพบุรี
pʰrá-tɕâw lá-wó: ʔan jù: nai mɯaŋ lóp-bù-ri:
lord PN REL be LOC city PN
‘the king of Lavo who is in the city of Lopburi’
(Thapraiwan Thonburi, 2314 BE)

Moreover, syntactically, /ʔan1/ has rarely been attested in direct-object relative clauses from 2301–2350 BE onward, when /thi:3/, the competing marker, expanded into that position. Usage of /ʔan1/ in the direct-object position has declined quickly thereafter, as its use has become more focused on the subject position (for statistical evidence and discussion, see Section 4.2.2).

Along with its subject concentration, /ʔan1/ also shows increasing specialization in terms of predicate types. From 2301–2350 BE onward, /ʔan1/ has commonly occurred with adjectival predicates. In (23), /ʔan1/ introduces a clause whose relativized position is the subject of the adjectival predicate *สูงส่ง* ‘exalted’. This pattern is consistent with Sornhiran (1981, p. 52), who notes that /ʔan1/ usually introduces relative clauses containing adjectival predicates in present-day Thai.

(23) ผลสำเร็จอัน**สูงส่ง**
pʰŏn-săm.rèt ʔan sŭ:ŋ-sòŋ
result-success REL high-exalted
‘the success that is exalted’
(Wiwatchaiyanuson, 2504 BE)

Furthermore, from the 2301–2350 BE period onward, /ʔan1/ has no longer been attested in oblique relative clauses. This restriction, however, represents a later development. Example (24) shows an earlier use in Sukhothai and Early Ayutthaya in which *ปี* ‘year’ functions as the head noun, and /ʔan1/ serves as the relative clause marker introducing *จักสิ้นศาสนาพระพุทธ* ‘the Buddhism will come to an end’. This example indicates that the relativized position corresponding to *ปี* is an oblique, specifying the time at which the event occurs.

(24) ปีอันจักสิ้นศาสนาพระพุทธ

pi: ʔan tɕàk sîn sà:t.sà.nă: pʰrá-phút

year REL FUT end religion Buddha

‘the year when the Buddhism will come to an end’

(Nakhonchum Inscription, 1900 BE)

These findings demonstrate that /ʔan1/ has become less flexible across syntactic functions. The marker has shifted from appearing with relativized positions across many syntactic functions to occurring in a narrower set of contexts. This result is consistent with Hongthong (2018, p. 311), who reports that in present-day Thai, /ʔan1/ does not introduce relative clauses whose internal co-referential element is a non-core argument. Although that study did not explicitly address relativized positions functioning as obliques, such relative clauses involve non-core arguments and typically do not show a clear gap position and lack a resumptive pronoun.

4.1.2 /sɯŋ3/: Marker of Choice during Middle Ayutthaya

The statistical analysis for /sɯŋ3/ reveals a distinctive trajectory that differs from the other markers. The pattern shows a parabolic profile with a peak in Middle Ayutthaya, followed by a decline in later periods. This rise and fall suggest a temporary preference tied to that period.

Table 1 shows the coefficient for /sɯŋ3/ relative to /thi:3/ of 3.10 with $p < 0.001$ in Middle Ayutthaya, indicating a sharp increase in the use of /sɯŋ3/ during this period. In Late Ayutthaya, the coefficient remains positive and significant at 1.37 with $p < 0.001$, yet it is clearly reduced relative to Middle Ayutthaya and remains below the /ʔan1/-to-/thi:3/ value of 2.48 with $p < 0.001$, signaling the start of a downturn. In 2301–2350 BE, the /sɯŋ3/-to-/thi:3/ coefficient turns slightly negative at -0.02 and is not statistically significant, matching the near parity in their counts during this period, with 45 tokens of /sɯŋ3/ and 46 tokens of /thi:3/. From 2351–2400 BE onward, the coefficients for /sɯŋ3/ become progressively negative and statistically significant, demonstrating the declining role after its peak in Middle Ayutthaya.

However, the frequency of /sɯŋ3/ may not reflect system-level markedness. Unlike /ʔan1/ and /thi:3/, whose usage is determined by the syntactic and semantic contexts, /sɯŋ3/ seems to be shaped by author-specific stylistic choices. An inspection of Middle Ayutthaya texts reveals text-by-text preferences in marker selection, often between /sɯŋ3/ and /thi:3/ separately. For example, in *The Letter of Okprajommuang Srirajakosa*, *The Thai and French Contract in the Reign of King Narai*, and *The Letter from Sithammaraj to Monsieur Pontchartrain*, only /sɯŋ3/ and /thi:3/ were attested, with no instances of /ʔan1/. However, in *Kosapan’s Memoir*, *The Letter from Kosapan to Monsieur le Marquis de Lagny*, and *The Letter from Kosapan to Monsieur le Marquis de Seignelay*—all written by the same author—/sɯŋ3/ was used most frequently, followed by /ʔan1/,

with /thiː3/ least frequent, indicating a clear authorial preference for /sɯŋ3/ even in texts where multiple markers occur. In contrast, *The Letter to King Don Philip of Portugal* relied primarily on /ʔan1/, with only limited use of /thiː3/ and no /sɯŋ3/. Because, in this data, texts that preferentially use /sɯŋ3/ are more numerous than those favoring /ʔan1/, the overall Middle Ayutthaya corpus shows higher frequencies of /sɯŋ3/ than /ʔan1/.

These comparisons suggest that some texts consistently select one marker and avoid others, while another set of texts makes the opposite choice. Consequently, the higher overall frequency of /sɯŋ3/ relative to /ʔan1/ in Middle Ayutthaya does not necessarily index system-wide markedness on its own. This Middle Ayutthaya rise aligns with how that period's corpus is built. There is a concentration of authors whose texts prefer /sɯŋ3/, and the period contains a higher proportion of those texts. These texts contribute a large share of tokens, so when the period includes more of them, the counts for /sɯŋ3/ rise even if other texts do not show the same preference. There is little evidence of competition for new contexts. /sɯŋ3/ did not displace /thiː3/ nor consistently occupy positions vacated by /ʔan1/ across syntactic functions or semantic domains, and the increases did not carry over into adjacent periods.

4.1.3 /thiː3/: Increasingly Frequent Marker from 2301–2350 BE to the Present

The statistical profile for /thiː3/ shows an S-curve growth pattern from a low-frequency option to a high-frequency marker. Although /thiː3/ serves as the reference category in the model, its ascent is clear when the other markers are compared against it.

As noted in the findings of /ʔan1/ and /sɯŋ3/, /thiː3/ was comparatively infrequent from Sukhothai and Early Ayutthaya through Late Ayutthaya. During these periods the coefficients for /ʔan1/ and /sɯŋ3/ relative to /thiː3/ are positive, indicating that /thiː3/ was chosen less often than its competitors. Beginning in 2301–2350 BE, /thiː3/ grew steadily; however, as stated, the coefficient of /sɯŋ3/ relative to /thiː3/ is not statistically significant, consistent with near parity in counts. Then /thiː3/ rose sharply in relative use by 2351–2400 BE, with the coefficient of /sɯŋ3/ relative to /thiː3/ becoming significantly negative. As this trend continues to the present, /thiː3/ has become the most frequently used relative clause marker. Its coefficients remain comparatively higher than those of /ʔan1/ and /sɯŋ3/, reflecting its consistent growth in usage relative to the other two markers.

Taken together, these results point to a historical trend in which /ʔan1/ dominated in Sukhothai and Early Ayutthaya, /sɯŋ3/ peaked in Middle Ayutthaya, and /thiː3/ expanded gradually to become the most widely used form from 2301–2350 BE to the present. These quantitative trends set the stage for examining the semantic and syntactic factors that contributed to these developments. Beyond overall frequency, Table 2 shows a clear broadening of /thiː3/'s semantic range over time, which aligns with its status as the present-day default.

Table 2

Expansion of /thi:3/ across semantic domains.

Period	Location	Concrete	Abstract	Time	Total	Trend
Sukhothai and Early Ayutthaya	6 (100%)	0 (0%)	0 (0%)	0 (0%)	6	Restricted to locative domain
Middle Ayutthaya	2 (66%)	0 (0%)	1 (33%)	0 (0%)	3	Expands to abstract domain
Late Ayutthaya	0 (0%)	7 (100%)	0 (0%)	0 (0%)	7	Expands to concrete domain
2301–2350 BE	4 (8%)	39 (83%)	3 (6%)	0 (0%)	46	Established in concrete, abstract, and locative domains
2351–2400 BE	4 (5%)	59 (85%)	6 (8%)	0 (0%)	69	Established in concrete, abstract, and locative domains
2401–2450 BE	8 (11%)	51 (69%)	8 (11%)	6 (8%)	73	Expands to temporal domain; coverage in all domains
2451–2500 BE	6 (6%)	52 (57%)	26 (28%)	7 (7%)	91	Across all domains
2501–2550 BE	8 (8%)	65 (70%)	15 (16%)	4 (4%)	92	Across all domains
2551 BE–present	4 (4%)	46 (48%)	37 (39%)	7 (7%)	94	Across all domains

In Sukhothai and Early Ayutthaya, /thi:3/ modified only locative nouns. Then, Middle Ayutthaya introduced the first use of /thi:3/ with an abstract noun, as shown in example (25), where /thi:3/ modifies the head noun *สิ่ง* ‘thing’, which in this context refers to anything used as a simile or comparator to *พระกรุณา* ‘royal benevolence’. This study analyzes *พระกรุณา* as an abstract noun because it is an intangible concept; accordingly, the comparator *สิ่ง* should likewise be treated as belonging to the abstract domain.

(25) มี**พระกรุณา** ... หา**สิ่ง**ที่จะอุปมาบมิได้

mi: phrá-karu-na: hăː sìŋ tʰîː tɕà ʔù.pà.ma: bɔ̀.mí dâj

have royal-benevolence seek thing REL FUT compare NEG ABLE

'has royal benevolence ... cannot find the thing that can be compared'

(The Letter to King Don Philip of Portugal, 2161 BE)

This finding appears to conflict with Kullavanijaya (2008). Whereas the present study attests /thiː3/ modifying nouns beyond the locative domain as early as the Middle Ayutthaya period, Kullavanijaya (2008) reported the first such uses only in the Middle Rattanakosin. This is likely due to differences in corpus size and coverage. This study draws on TNHC2 and the Diachronic Thai Text Corpus, whereas Kullavanijaya's dataset was much smaller. For Middle Ayutthaya, it consists only of *Kosapan's Memoir*. By contrast, the Middle Ayutthaya sample in this study consists of eleven texts, including *The Letter to King Don Philip of Portugal*, which contains example (25) and is absent from the corpus used in Kullavanijaya (2008). This difference in corpus size plausibly accounts for the different results.

After that, Late Ayutthaya showed a shift to concrete nouns. The first example appeared in (26), where /thiː3/ modifies a head noun in the concrete domain. In this case, the head noun is *หลวงวรวา* 'Luang Worawa', a person. This example illustrates the ongoing expansion of the semantic range of /thiː3/, which increases across successive periods.

(26) **หลวงวรวา**ที่มีบรรดาศักดิ์ดาบทอง

lŭaŋ-wɔː.ra.waː tʰîː miː banˑdaː.sàk dàːp-tʰɔːŋ

HON-PN REL have noble.rank sword-gold

'the Luang Worrawa who has the noble rank of Dap Thong'

(The Letter from Sithammaraj to Monsieur Pontchartrain, 2236 BE)

From 2301–2350 BE onward, /thiː3/ has been attested across multiple domains, with a strong concentration in concrete nouns. By 2401–2450 BE, it also appeared with temporal nouns and was thus found in all domains, as shown in examples (27)–(30): in (27), *บรรดาคน* 'people', a concrete noun; in (28), *ธรรมะ* 'dhamma', an abstract noun; in (29), *เขตแดน* 'boundary', a locative noun; and in (30), *เวลา* 'time', a temporal noun.

(27) **บรรดาคน**ที่ลงมาทำงาน

ban.da: kʰon tʰîː loŋ maː tʰam-ŋaːn
all person REL descend come do-work
‘the people who came down to work’

(Javanese Letters, 2439 BE)

(28) **ธรรมะ**ที่ดี

tʰam.má tʰîː diː
dharma REL be.good
‘the dharma that is good’

(The Royal Ceremonies of the Twelve Months, 2431 BE)

(29) **เขตแดน**ที่เป็นของขึ้นแก่เมืองเชียงแสน

kʰèt-dɛːn tʰîː pen kʰɔ̌ːŋ-kʰûn kɛ̀ː mɯaŋ tɕʰiaŋ.sɛ̌ːn
territory REL COP property-subordinate DAT city PN
‘the territory that is subordinate to the city of Chiang Saen’

(On the Chronicle of Phra Phuttha Chinnarat,
Phra Phuttha Chinnasri, and Phra Srisatsada, 2409 BE)

(30) **เวลา**ที่ควรจะลงไปข้างใต้

weːlaː tʰîː kʰuan tɕà loŋ paj kʰâːŋ-tâːj
time REL should FUT descend go south-side
‘the time when one should go down to the south’

(Royal Letter No. 39 in Klai Ban II, 2450 BE)

In the most recent period, from 2551 BE to the present, the distribution remains broad. Concrete uses remain frequent, with a notable rise in abstract and temporal uses, indicating sustained expansion of /thiː3/ beyond its early locative restriction.

It should be noted that this study recognizes /thiː3/ as a relative clause marker earlier than reported by Kullavanijaya (2008) because the definitions differ. Before Middle Rattanakosin, Kullavanijaya (2008, p. 457) classified /thiː3/ as a noun denoting land rather than a relative clause marker, since it appeared only with locative nouns. According to that study, /thiː3/ was not considered a relative clause marker until Middle Rattanakosin, when it was found beginning to modify nouns in a wider range of semantic domains. In contrast, the present study treats /thiː3/ as a relative clause marker throughout whenever it introduces a noun-modifying clause that contains an argument co-referential with the head noun, regardless of the head noun’s semantic domain. This uniform syntactic criterion keeps the marker set constant across periods and improves comparability.

4.2 Factors Related to Shift of Markedness

Several factors are related to the shift of markedness, including the semantic domain of the head noun, the syntactic function of the relativized position, and the text genre, content, and writing style in which the markers appear. Analyzing these elements clarifies how linguistic factors shape changes in relative clause markers in Thai.

4.2.1 Semantic Domains of Head Nouns

The analysis of interactions between semantic domains of head nouns and historical periods using a multinomial logistic regression model provides a systematic picture of how the choice of relative clause markers in Thai has shifted over time. The heatmaps in Figures 2 and 3 visualize coefficients comparing /ʔan1/ and /sɯŋ3/ against /thi:3/ by semantic domain and period. The earliest period, Sukhothai and Early Ayutthaya, serves as the baseline, /thi:3/ is the reference marker, and the abstract domain serves as the reference level for semantic domain. These heatmaps highlight whether the use of a marker increased or decreased relative to /thi:3/ in each semantic domain across historical periods. Overall, the semantic evidence indicates that /thi:3/ progressively overtakes both /ʔan1/ and /sɯŋ3/.

Figure 2

Probability of /ʔan1/ selection relative to /thi:3/ across semantic domains by period.

Period	Concrete	Abstract	Location	Time
Sukhothai and Early Ayutthaya	Reference	Reference	Reference	Reference
Middle Ayutthaya	-2.07	-0.77	-1.38*	-0.97
Late Ayutthaya	-2.27*	-0.96	-1.57**	-1.17
2301-2350 BE	-2.54**	-1.23	-1.86**	-1.46
2351-2400 BE	-2.86**	-1.54	-2.19**	-1.77
2401-2450 BE	-3.11*	-1.79	-2.45*	-2.02
2451-2500 BE	-3.49*	-2.18	-2.83**	-2.41
2501-2550 BE	-3.96*	-2.65	-3.30**	-2.88
2551 BE - present	-4.21**	-2.89	-3.55**	-3.13

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Figure 2 shows modeled coefficients for /ʔan1/ relative to /thi:3/ by period and semantic domain. For /ʔan1/, coefficients decline steadily across all domains, indicating a long-term decrease relative to /thi:3/. The coefficients first reach statistical significance in the Middle Ayutthaya period

for locatives at -1.38 with $p < 0.05$. The coefficients remain significant, and the effect becomes stronger over time. For locatives, significance begins in Middle Ayutthaya, and the effect grows to -3.55 with $p < 0.01$ by 2551 BE to the present. For concrete nouns, significance begins in Late Ayutthaya at -2.27 with $p < 0.05$ and becomes more negative, reaching -4.21 with $p < 0.01$ by 2551 BE to the present. In the abstract domain, coefficients calculated from raw counts are consistently negative, ranging from -0.77 in Middle Ayutthaya to -2.89 in the present. In the temporal domain, they range from -0.97 to -3.13 over the same periods. Although negative across periods, none of these coefficients reach statistical significance. Taken together, these declines show /ʔan1/ receding from domains where it had previously been frequent while /thiː3/ gradually rose to occupy the space of a broadly distributed marker.

Figure 3

Probability of /sɯŋ3/ selection relative to /thiː3/ across semantic domains by period.

Period	Concrete	Abstract	Location	Time
Sukhothai and Early Ayutthaya	Reference	Reference	Reference	Reference
Middle Ayutthaya	0.29	-0.17	0.82**	n/a
Late Ayutthaya	0.18	-0.22	0.88**	n/a
2301-2350 BE	0.12**	-0.29	0.91***	n/a
2351-2400 BE	0.05**	-0.36	1.02**	n/a
2401-2450 BE	-0.08**	-0.42	1.15**	n/a
2451-2500 BE	-0.19**	-0.51	1.27**	n/a
2501-2550 BE	-0.28**	-0.6	1.36***	n/a
2551 BE - present	-0.34**	-0.66	1.44***	n/a

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Figure 3 shows modeled coefficients for /sɯŋ3/ relative to /thiː3/ by period and semantic domain. In the locative domain, /sɯŋ3/ is consistently preferred from Middle Ayutthaya onward, with coefficients that are positive and grow over time: 0.82 with $p < 0.01$ in Middle Ayutthaya, increasing to 1.44 with $p < 0.001$ by 2551 BE to the present. In the concrete domain, coefficients shift from positive to negative, moving from 0.12 with $p < 0.01$ in 2301–2350 BE to -0.34 with $p < 0.01$ by 2551 BE to the present. In the abstract domain, coefficients calculated from raw counts are negative but not statistically significant throughout, for example, -0.17 in Middle Ayutthaya and -0.66 in 2551 BE to the present. /sɯŋ3/ is not attested in the temporal domain in this data set.

In summary, the association between markers and semantic domains helps explain the shift of markedness. While /ʔan1/ initially had the flexibility to relativize nouns across various domains, its role became increasingly restricted, particularly as /thi:3/ expanded beyond an early locative restriction to cover a broad range of semantic domains. /sɯŋ3/ remained relatively stable, never becoming the dominant marker and maintaining a neutral role across periods. Unlike /ʔan1/ and /thi:3/, it did not undergo any restriction or expansion in distribution.

4.2.2 Syntactic Functions of Relativized Positions

The relative clause markers /thi:3/, /sɯŋ3/, and /ʔan1/ in Thai relativize a range of syntactic functions, including subjects, direct objects, indirect objects, obliques, and genitives. Their changing distributions over time reflect the shift of markedness. The following analyses track how each marker's role changes across periods.

To deepen the account of how Thai relative clause marker choice has changed, the interaction between historical periods and syntactic functions of relativized positions is analyzed. The results show that changes across periods do not occur independently of syntactic context; rather, the two factors interact in a statistically meaningful way. Figures 4 and 5 plot coefficients for marker choice by syntactic function and period, evaluated against the reference group, the period of Sukhothai and Early Ayutthaya, with /thi:3/ as the reference marker and direct objects as the reference level for syntactic function. These figures provide a fine-grained view of function-specific developments and make explicit the role of syntactic context in the historical redistribution of the Thai relative clause marker system.

Figure 4 shows modeled coefficients for /ʔan1/ relative to /thi:3/ by period and syntactic function. In the subject position, the coefficients for /ʔan1/ are significantly negative in Middle Ayutthaya, at -17.0 with $p < 0.001$, then shift to very large, consistently significant positive coefficients, increasing from 5.38 with $p < 0.001$ in Late Ayutthaya to 23.0 with $p < 0.001$ in 2551 BE to the present. This pattern indicates a transition from earlier broad use to later concentration on subjects, compared to other syntactic functions.

As /ʔan1/ concentrates in subjects, its presence in direct objects wanes. Although direct model coefficients are unavailable because direct objects serve as the reference level for syntactic function, coefficients calculated from the raw counts show a consistent shift from positive to negative, moving from about 1.61 in Middle Ayutthaya to -2.4 in 2301–2350 BE and -2.83 by 2551 BE to the present. The raw counts confirm this trend. /ʔan1/ occurs 19 times with direct objects in Sukhothai and Early Ayutthaya but drops to only 2 tokens in Middle Ayutthaya, 4 in Late Ayutthaya, and then virtually disappears, with just 1 token each in 2351–2400 BE, 2401–2450 BE, and 2501–2550 BE, and none attested in the remaining periods. This trajectory is consistent with /thi:3/ gradually replacing /ʔan1/ in the direct-object position.

Evidence for indirect objects is sparse, with coefficients only for 2401–2450 BE and 2501–2550 BE at -1.36 and -0.88, respectively, neither of which is significant because each period has a single token, both instances using /thi:3/. These values indicate no meaningful difference between /ʔan1/ and /thi:3/ for this function.

For obliques, coefficients for /ʔan1/ are negative and statistically significant whenever it is not the most frequent marker, particularly after Late Ayutthaya, from -11.0 with $p < 0.001$ in 2301–2350 BE to -6.67 with $p < 0.001$ at the present, which indicates lower selection than /thi:3/. Notably, /ʔan1/ is not attested in oblique relativization after Late Ayutthaya in this dataset. Genitives show a significantly negative effect in Middle Ayutthaya, at -3.11 with $p < 0.001$, with no later attestation. Taken together, these patterns indicate a diachronic narrowing of /ʔan1/ toward the subject domain alongside the broader expansion of /thi:3/ in object and oblique contexts.

Figure 4

Probability of /ʔan1/ selection relative to /thi:3/ across syntactic functions by period.

Period	Subject	Direct Object	Indirect Object	Oblique	Genitive
Sukhothai and Early Ayutthaya	Reference	Reference	Reference	Reference	Reference
Middle Ayutthaya	-17.0***	1.61	n/a	-3.21***	-3.11***
Late Ayutthaya	5.38***	1.1	n/a	36.2***	n/a
2301-2350 BE	30.1***	-2.4	n/a	-11.0***	n/a
2351-2400 BE	17.7***	-1.95	n/a	-14.6***	n/a
2401-2450 BE	18.4***	-2.61	-1.36	-3.61***	n/a
2451-2500 BE	24.8***	-4.2	n/a	-2.05***	n/a
2501-2550 BE	21.3***	-3.08	-0.88	-6.68	n/a
2551 BE - present	23.0***	-2.83	n/a	-6.67***	n/a

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Figure 5 shows modeled coefficients for /sɯŋ3/ relative to /thi:3/ by period and syntactic function. In the subject position, coefficients are significantly negative in Middle and Late Ayutthaya, -4.89 with $p < 0.001$ and -1.32 with $p < 0.001$, respectively, then turn positive from 2301–2350 BE onward. They remain strongly significant through the present, rising from 1.27 with $p < 0.001$ to 4.53 with $p < 0.001$, though never reaching the magnitudes observed for /ʔan1/.

Direct objects serve as the reference level, so no model coefficient is available. Derived values range from 1.44 in Middle Ayutthaya to -2.34 at the present, indicating a decrease in the use of /sɯŋ3/ as /thi:3/ increases. For obliques, the pattern shifts from positive to negative. Middle

Ayutthaya through 2351–2400 BE show advantages for /sɯŋ3/, peaking at 2.10 with $p < 0.001$ in 2301–2350 BE. From 2401–2450 BE to the present, the contrasts are negative, with -1.45 and $p < 0.05$ at present. For indirect objects, all available estimates are negative and not statistically significant, for example, -1.72 in 2401–2450 BE and -1.29 in 2501–2550 BE, reflecting sparse data. For genitives, /sɯŋ3/ is attested only once in Middle Ayutthaya, with a small negative value of -0.54, which is not interpretable statistically.

Figure 5

Probability of /sɯŋ3/ selection relative to /thi:3/ across syntactic functions by period.

Period	Subject	Direct Object	Indirect Object	Oblique	Genitive
Sukhothai and Early Ayutthaya	Reference	Reference	Reference	Reference	Reference
Middle Ayutthaya	-4.89***	1.44	n/a	0.45	-0.54
Late Ayutthaya	-1.32***	1.52	n/a	1.41	n/a
2301-2350 BE	1.27***	-1.14	n/a	2.10***	n/a
2351-2400 BE	3.59***	-1.08	n/a	1.21	n/a
2401-2450 BE	4.73***	-2.18	-1.72	-0.54	n/a
2451-2500 BE	4.27***	-2.42	n/a	-1.22*	n/a
2501-2550 BE	4.23***	-2.27	-1.29	-0.41	n/a
2551 BE - present	4.53***	-2.34	n/a	-1.45*	n/a

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Taken together, these results point to a diachronic redistribution consistent with shift of markedness. Initially, /ʔan1/ began as the most versatile marker, relativizing subjects, direct objects, and obliques, then narrowed as /sɯŋ3/ and later /thi:3/ took over these roles. /sɯŋ3/ gained prominence in the subject position mid-timeline but did not generalize across syntactic functions. /thi:3/ eventually became the most flexible and even relativizes indirect objects. Genitive relativization remains rare throughout all periods.

4.2.3 Text-Level Factors

Beyond semantic domain and syntactic function, additional factors that may correlate with marker choice were examined at the text level. These include text genre, content, and writing style.

Across genres, /ʔan1/ is chosen significantly more often than /thi:3/ in fiction, legal, and non-academic texts, with confidence intervals that exclude zero. /sɯŋ3/ does not differ reliably from /thi:3/ in any genre. Estimates for newspapers are uninterpretable because the sample consists of only two tokens, both introduced by /thi:3/.

For content, higher odds for /ʔan1/ over /thi:3/ are observed in beliefs and ideas, social sciences, imaginative topics, and world events and history, whereas applied sciences, natural or pure sciences, and arts show no reliable differences. Again, /sɯŋ3/ shows no significant contrasts with /thi:3/ in any content category.

By writing style, coefficients for /ʔan1/ and /sɯŋ3/ do not differ significantly from those for /thi:3/, indicating no reliable effect of prose versus verse on marker choice.

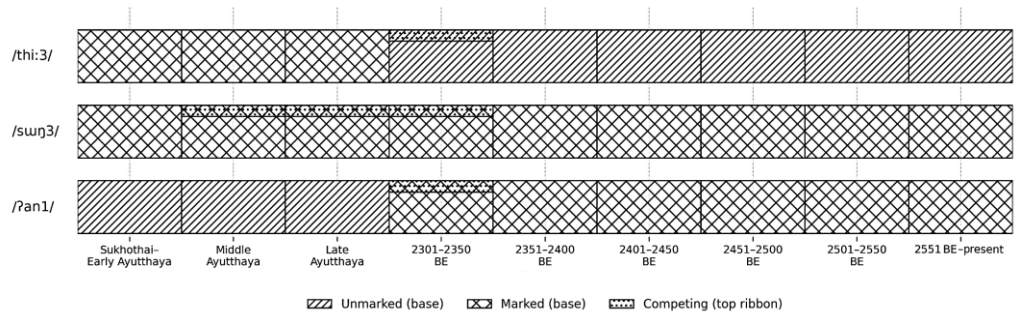
Overall, text-level factors show uneven effects. Genre and content maintain contexts where /ʔan1/ remains more common, while style is neutral. Nevertheless, across periods, /thi:3/ has steadily increased in frequency, eventually surpassing /ʔan1/ in overall usage.

5. Conclusion and Discussion

This study examined the diachronic development of the relative clause markers /thi:3/, /sɯŋ3/, and /ʔan1/ in Thai to explain how their usage and markedness changed over time. Unlike prior studies that focused on individual markers in isolation, this analysis considered all three markers as a system and traced how they interacted, competed, and shifted in markedness status over a span of more than 700 years. Figure 6 presents the timeline that summarizes these developments.

Figure 6

The timeline of shift of markedness of /thi:3/, /sɯŋ3/, and /ʔan1/



In Sukhothai and Early Ayutthaya, Middle Ayutthaya, and Late Ayutthaya, /ʔan1/ was unmarked. It functioned as the default marker, capable of relativizing a wide range of semantic domains and syntactic functions. From 2301–2350 BE onward, it has become marked, showing significant restriction. It has no longer been attested in oblique functions and has rarely occurred in direct-object relative clauses. This indicates contraction toward more specific contexts, notably concrete head nouns associated with holiness and divinity, and the subject position with an adjectival predicate.

Across periods, /sɯŋ3/ has remained marked. Its apparent rise occurred in Middle Ayutthaya; however, the higher counts in that period likely reflect author- or text-specific preferences and that period's corpus composition, specifically a larger share of tokens from a small set of authors and several documents that favored /sɯŋ3/, rather than expansion across syntactic functions or semantic domains. The increase was not sustained in Late Ayutthaya or later periods. /sɯŋ3/ neither displaced /thiː3/ nor filled contexts vacated by /ʔan1/. In this dataset, gains are confined to Middle Ayutthaya, and thus /sɯŋ3/ has never attained unmarked status.

/thiː3/ began as a marked option, largely restricted to locative nouns. This restriction is consistent with its lexical source as a noun meaning 'place'. It then expanded steadily in both semantic and syntactic range. Although its relative frequency was only marginally higher than that of /sɯŋ3/ in 2301–2350 BE (hence the competing ribbon in Figure 6), /thiː3/ became unmarked during this period, where it replaced /ʔan1/ in the semantic domains and syntactic functions that /ʔan1/ vacated. /thiː3/ has remained unmarked to the present, combining the highest overall frequency with the broadest coverage across domains and functions.

Text-level effects are modest and uneven. Relative to /thiː3/, /ʔan1/ is more likely in fiction, legal, and non-academic genres. It also appears more frequently in content areas such as beliefs and ideas, social sciences, imaginative topics, and world events and history. However, these few /ʔan1/ cases do not alter the overall rise of /thiː3/. /sɯŋ3/ never differs significantly from /thiː3/, and writing style has no effect.

Taken together, the evidence points to an internal change, namely the grammaticalization of /thiː3/, as the principal driver of the shift of markedness. Diachronically, /thiː3/ began as a lexical noun meaning 'place' and was used only with locative nouns (e.g., (18) in Sukhothai). Over time, it underwent the familiar grammaticalization trajectory that included semantic bleaching (Hopper & Traugott, 2003, p. 94) from concrete 'place' to a relational linker, decategorialization with loss of nominal behavior (Hopper & Traugott, 2003, p. 106), and, crucially, distributional expansion from locative heads to a broader set of semantic domains and syntactic functions (e.g., (19) in Middle Rattanakosin; Kullavanijaya, 2008). In the terms used here, this expansion shifted /thiː3/ from a marked, context-limited option to the unmarked relative clause marker. It rose in token frequency and, more importantly, acquired the widest coverage across semantic domains and syntactic functions. This pattern accords with the view that forms occurring more frequently in texts tend to be further along the path of grammaticalization (Hopper & Traugott, 2003, p. 106). As /thiː3/ increased in frequency and broadened its distribution, the rise both signaled and facilitated its reanalysis as the primary relative clause marker, helping explain its emergence as the unmarked option. The corpus results indicate that this unmarking is in place by 2301–2350 BE, when /thiː3/ replaced /ʔan1/ in precisely those syntactic and semantic environments that /ʔan1/ began to vacate.

These patterns align with a common trajectory in language change. A form undergoing grammaticalization broadens its semantic and syntactic range and stabilizes as the unmarked default, while the prior default narrows and becomes marked. Frequency shifts mirror this restructuring. Forms that broaden their scope increase in frequency, while those that narrow their scope decrease (Hopper & Traugott, 2003, pp. 127–129), which is consistent with links between frequency and grammaticalization. The result is a reorganized paradigm with one general marker and residual marked alternatives.

Ethics Statement

The author used an AI language assistant, ChatGPT by OpenAI, to help with proofreading and language editing. Assistance was limited to improving grammar, clarity, wording, and paragraph grouping. The tool did not generate research ideas, analyze data, create figures or tables, or fabricate or select references. All edits suggested by the tool were reviewed and approved by the author, and the final text is the author's own. No confidential or sensitive information beyond the manuscript text was provided to the tool. The author is fully responsible for the accuracy and integrity of the content.

Abbreviations

1SG	first person singular	HON	honorific
3SG	third person singular	LOC	locative
ABLE	ability	NMLZ.AG	agentive nominalizer
CLF	classifier	NEG	negation
COM	comitative	PFV	perfective
COMP	comparative	PN	proper name
COP	copula	PROG	progressive
DAT	dative/benefactive	PTCL	particle
DEM.DIST	distal demonstrative	RECP	reciprocal
EXP	experiential	REFL	reflexive
FUT	future	REL	relative clause marker
GEN	genitive	SUBORD	subordinator

References

Anthony, L. (2024). *AntConc* (Version 4.3.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/AntConc

Brown, C. H., & Witkowski, S. R. (1983). Polysemy, lexical change and cultural importance. *Man*, *18*(1), 72–89. https://doi.org/10.2307/2801765

Comrie, B. (1989). *Language universals and linguistic typology* (2nd ed.). University of Chicago Press.

Dryer, M. S. (2013a). Order of relative clause and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Version 2020.4). Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.13950591

Dryer, M. S. (2013b). Relationship between the order of object and verb and the order of relative clause and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Version 2020.4). Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.13950591

Greenberg, J. H. (1966). *Language universals: With special reference to feature hierarchies.* MIT Press.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*(1), 25–70. https://doi.org/10.1017/S0022226705003683

Honthong, K. (2018). *Noun-modifying clause constructions in Thai* [Doctoral dissertation, Chulalongkorn University]. Chula DigiVerse. https://doi.org/10.58837/CHULA.THE.2018.1048

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139165525

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8*(1), 63–99. http://www.jstor.org/stable/4177973

Kullavanijaya, P. (2006). Noun-modifying clauses: Relative clauses and noun complement clauses. In A. Prasithrathsint (Ed.), *Controversial constructions in Thai grammar: relative clause constructions, complement clause constructions, serial verb constructions, and passive constructions* (pp. 7–65). Chulalongkorn University Press.

Kullavanijaya, P. (2008). A historical study of /thîi/ in Thai. In A. V. N. Diller, J. A. Edmondson, & Y. Luo (Eds.), *The Tai-Kadai languages* (pp. 445–467). Routledge.

Meesri, W., & Wongwattana, U. (2020). Grammatical functions and semantic aspects of “ʔan” in Sukhothai period. *Humanities and Social Sciences Nakhonsawan Rajabhat University Academic Journal*, *6*(2), 156–179. https://so05.tci-thaijo.org/index.php/hssnsru/article/view/241616

Pittayaporn, P., Kanjanoläre, M., Thanyawong, S., & Nuchprahan, K. (2023). *TNHC2 dataset* [Data set]. ChulaSEAL. https://www.arts.chula.ac.th/chulaseal/tnhc2/

Prompapakorn, P. (1996). *Variation and change in relative clauses in Thai during the Ratanakosin period* [Master's thesis, Chulalongkorn University]. Chula DigiVerse. https://doi.org/10.58837/CHULA.THE.1996.2116

Rice, K. (2007). Markedness in phonology. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 79–98). Cambridge University Press.

Singhapreecha, P. (2007). A reference grammar of Thai: S. Iwasaki, P. Ingkaphirom, Cambridge University Press, Cambridge, 2005, vii + 392 pp. *Lingua*, *117*(8), 1497–1512. https://doi.org/10.1016/j.lingua.2006.09.005

Sornhiran, P. (1981). /thiː3/, /sɯŋ3/, /ʔan1/ in relative clauses. *Journal of Letters*, *13*(1), 50–59. https://doi.org/10.58837/CHULA.JLETTERS.13.1.3

Thanyawong, S. (2025). *Diachronic Thai text corpus* [Data set]. Thai Language Research Center. https://osf.io/wzx48/

Warotamasikkhadit, U. (2017). *Thai syntax: An outline.* De Gruyter. https://books.google.co.th/books?id=sZNsDwAAQBAJ

Yaowapat, N. (2005). Pronoun retention in Khmer and Thai relative clauses. In *SEALS XV: Papers from the 15th Annual Meeting of the Southeast Asian Linguistics Society* (pp. 121–132). Pacific Linguistics.

Yaowapat, N., & Prasithrathsint, A. (2008). A typology of relative clauses in mainland Southeast Asian languages. *Mon-Khmer Studies, 38,* 1–23.

Yaowaphat, N. (2008). *The development of the multiple functions of /sin/ in Thai* [Doctoral dissertation, Chulalongkorn University]. Chula DigiVerse. https://doi.org/10.58837/CHULA.THE.2008.1884