Verbal numeral classifiers in languages of Eastern Eurasia : A typological survey

Honkasalo, Sami

de Gruyter
2024-09-26

Honkasalo, S & Yurayong, C 2024, 'Verbal numeral classifiers in languages of Eastern Eurasia : A typological survey', Open Linguistics, vol. 10, no. 1, 20240021. https://doi.org/10.1515/opli-2024-0021

http://hdl.handle.net/10138/586339
10.1515/opli-2024-0021

Research Article

Sami Honkasalo* and Chingduang Yurayong

Verbal numeral classifiers in languages of Eastern Eurasia: A typological survey

https://doi.org/10.1515/opli-2024-0021
received July 19, 2024; accepted July 22, 2024

Abstract: This article presents a typological survey of verbal numeral classifiers (VNCs) in languages of Eastern Eurasia. As classifiers of occurrence (e.g., to V once, to V twice), VNCs are prevalent in languages of East Asia and Southeast Asia, yet the phenomenon has been largely overlooked in typological studies that have overwhelmingly focused on the adnominal use of numeral classifiers. Analyzing a sample of 543 linguistic varieties, the study categorizes the language families of Eastern Eurasia into three groups based on their use of VNCs: 1) VNCs consistently present in all languages (e.g., Tai-Kadai); 2) VNCs present in some languages (e.g., Sino-Tibetan); and 3) VNCs consistently absent (e.g., Yukaghir). Additionally, the meanings of VNCs are shown to follow a non-random distribution by centering on certain semantic fields, such as quantifying hits with 'verbs of violence'. The study aims to offer the first step toward an areal typological study of VNCs, highlighting the importance of including this phenomenon in future descriptive work.

Keywords: verbal numeral classifiers, classifiers, event quantification, Eastern Eurasia, areal typology

1 Introduction

The present study offers an areal typological survey of verbal numeral classifiers (VNCs), namely classifiers of occurrence, in languages of Eastern Eurasia. While nominal numeral classifiers (NNCs) have been subject to extensive previous research, VNCs that frequently occur in East and South-East Asian languages have not been studied to the same extent from a comparative-typological viewpoint. Previous studies addressing the phenomenon have been mostly limited to the systems present in the most dominant individual languages of the region, Mandarin Chinese in particular. It is no exaggeration to say that rarely has such a regionally ubiquitous phenomenon been neglected in typological research literature. To address this gap, the present study offers the first typological study of VNCs in Eastern Eurasia, utilizing a representative sample of surveyed languages from all 23 major language families and language isolates (i.e., one-language families) of the region discussed further in Section 1.3.

The study has the following three primary goals: First, based on the surveyed languages, we formulate a more precise and typologically oriented definition for VNCs and differentiate the phenomenon from related linguistic concepts, since the term has been used inconsistently in various differing and ill-defined ways in earlier research, including its application to morphosyntactic phenomena that should not be considered prototypical numeral quantification.

Second, the study examines the distribution of VNCs among the 23 surveyed language families of Eastern Eurasia and demonstrates how languages and language families of the region can be grouped into three main

*** Corresponding author: Sami Honkasalo,** Department of Languages, Faculty of Arts, University of Helsinki, Helsinki, Finland; Research Institute for Languages and Cultures of Asia, Mahidol University, Nakhon Pathom, Thailand, e-mail: samihonkasalo@gmail.com
Chingduang Yurayong: Department of Languages, Faculty of Arts, University of Helsinki, Helsinki, Finland; Research Institute for Languages and Cultures of Asia, Mahidol University, Nakhon Pathom, Thailand

categories based on whether VNCs can be identified. The distribution of prototypical VNC systems centers around Southeast and East Asia. In language families, such as Tai-Kadai and Hmong-Mien, VNCs are ubiquitous and show great semantic breadth. In contrast, many of the surveyed language families contain languages, such as Mangghuer (Mongolic) that have simple systems of verbal classifiers that in several cases result from language contact while many other languages in such families lack verbal classifiers. Finally, when moving outside the core distribution area of verbal classifiers, languages and language families, such as Uralic Yukaghir, with no systems of VNCs become dominant.

As the third primary goal, the study investigates the semantics of VNCs in the surveyed languages. It demonstrates that the meanings of VNCs in the surveyed languages across language families manifest non-random distribution by clustering around specific semantic core fields. For instance, in addition to a generic VNC present in every language with a grammaticalized system of such classifiers, languages with full VNC systems often include one or several classifiers to quantify and classify hits with 'verbs of violence'.

1.1 Terminology

Three terminological points need clarification. To begin with, 'Eastern Eurasia' refers to the Eastern reaches of the Eurasian continent that comprises East Asia, Central Asia, and Mainland Southeast Asia, regions where language contact has historically been frequent.[1] In particular, Mainland Southeast Asia and East Asia form the geographic core area for the distribution of both NNC and VNC systems. While the main island groups next to the Eurasian mainland and in historical contact with it, such as the Japanese archipelago, Taiwan, and Hainan, are included, the Eastern maritime regions, such as Maritime Southeast Asia and the Pacific islands, remain outside the scope of investigation. Individual languages from the Indian subcontinent are also included and all language families indigenous to India are surveyed, but the Indian subcontinent remains outside the focus of the present study. In the territory corresponding to Russia, we include languages and language families spoken to the east of the Ural Mountains, a traditional geographic boundary between Europe and Asia. While this artificial boundary does not necessarily correspond to any linguistic reality, it is nevertheless necessary for delimiting the scope of the study. Also, we exclude languages of Europe and Southwest Asia. It should be noted that the selection aims not to deny the possible existence of VNCs outside the investigated region.[2] For a universal typology, further scholarship is consequently needed to confirm the global distribution of VNC systems in languages of the world.

Second, the present study addresses a phenomenon for which we coin the term 'exact and adjustable event quantification', since VNCs constitute one possible device for this purpose. Exact and adjustable quantification is a subtype of quantification that fulfills two conditions, namely 1) the attribution of an exact numerical value, and 2) the possibility to adjust such value. Languages of the world possess a wide range of means for event quantification, but not all of these attribute an exact and adjustable numerical value to the event, such as the English adverb *sometimes* in *She sometimes goes out for a walk in the evening.* Such non-exact and non-adjustable devices of event quantification remain outside the scope of the study. They have been widely discussed from the viewpoint of aspect, pluractionality, reduplication, and adverbs in previous research. On the other hand, VNCs constitute a grammatical device for exact and adjustable event quantification, since an exact numerical value is attributable to event occurrence and the value can also be adjusted, as

1 The study is limited to Eastern Eurasia only. This is not intended as a statement concerning the global distribution of verbal numeral classifiers that may exist also in some languages of the Americas, for instance. Rather, the geographic boundaries of the study resulting in an areal convenience sample are due to its pilot nature: We wished to focus on East and Southeast Asia that form the distributional core for verbal numeral classifiers as identified in earlier research and to further investigate the spread of the phenomenon in the adjacent areas.

2 As shown by typological studies of classifiers, such as Gil (2013) and Her and Li (2023), among the regions adjacent to Eastern Eurasia, nominal numeral classifiers can be identified in languages of Maritime Southeast Asia, and the Pacific, although more sparsely than in their core distributional center investigated in the present study. As discussed in Section 2.2, verbal numeral classifiers may exist in languages with nominal numeral classifiers and not vice versa. Consequently, a possibility for discovering verbal numeral classifiers in languages of the regions excluded in this study cannot be ruled out.

with the Mandarin Chinese classifier *cì* ‘time’ in *qù sān cì* 去三次 (go three CLF.time) ‘to go three times’ where the numeral *sān* ‘three’ can be productively replaced by *sì* 四 ‘four’, for instance.

In addition to VNCs, other devices for exact and adjustable event quantification exist in languages of Eastern Eurasia. These devices, found in Eastern Eurasian languages that lack VNCs, are not covered in the present study but are briefly introduced below (refer to Yurayong et al. 2024, Section 4, for more details). The main types include counted nouns, iterative numerals, and verbalized numerals. Counted nouns refer to nouns that can be counted directly as a device of exact and adjustable quantification, such as English ‘time’ and its Written Tibetan counterpart *thengs* ‘time’ (1).

(1) **Counted nouns: Written Tibetan (Sino-Tibetan)**
nga rgyagar=la ***thengs*** *gsum ‘gro myong yod.*
1SG India=DAT **time** three go EXP.PRF AUX.EGO
‘I have been to India three times.’ (constructed)

In turn, iterative numerals are a denumeral category (Fradin 2015) used for indicating exact and adjustable quantification. For instance, Alytor derives iterative numerals by means of the suffix *-saŋ* attaching to a numeral, as in (2).

(2) **Iterative numerals: Alytor (Chukotko-Kamchatkan)**
ŋəra-saŋ *jat-ə-tkə* *nedelʲa-k.*
three-ITER come-E-IMPF:2SG.S week-LOC
‘You come three times a week.’ (Nagayama 2003, 81)

Finally, as a rare strategy in languages of Eastern Eurasia, verbalized numerals are attested in the Formosan Austronesian languages that attach verbalizing prefixes to numerals for deriving verbs that indicate frequency. For instance, Thao uses verbalizing prefixes, such as *mu-* (3).

(3) **Verbalized numerals: Thao (Austronesian)**
Mu-turu-z *iza* *yaku* *mu-tusi*
VBLZ-three-SUFF already 1SG.NOM VBLZ-there
Lipún m-riqaz nak a azazak.
Japan AF-see 1SG.POSS LNK child
‘I have been to Japan three times to visit my daughter.’ (Blust 2003, 1021–2)

The phenomenon termed ‘verbal numeral classifiers’ in the present study is known by several names in previous research. To begin with, many specialists of Sino-Tibetan and Southeast Asian languages have used the term ‘verbal classifier’ to refer to classifiers of occurrence (e.g., Aikhenvald 2000, 9, Paris 2013, Zhang 2017a), illustrated by Thai ครั้ง *kʰráŋ* ‘time’ in expressions such as สองครั้ง *sɔ̌ːŋ kʰráŋ* ‘twice, two times’ and สามครั้ง *sǎːm kʰráŋ* ‘three times’, respectively. In addition, the research literature contains particularly frequent mentions of ‘verbal action classifiers’ (e.g., Bai 2019, 295–6, Chirkova 2009, 35, Daudey 2014, 146, LaPolla and Huang 2003, 68, Tunzhi 2019, 238–9, Zhang 2013, Zheng 2016, 116). Other similar terms include ‘activity classifiers’ (Ahrens and Huang 2016, 174–5, Massupong 1982, 115) and ‘measures for verbs of action’ (Chao 1968, 627).

The term ‘verbal action classifiers’ is commonly used by Chinese scholars or other scholars operating primarily in the linguistic context of China. Together with ‘activity classifier’, it reflects a Chinese language term for the phenomenon. The Chinese research tradition frequently uses the term *dòngliàngcí* 动量词 translatable as ‘action classifier’ or ‘verbal classifier’. It contrasts with *wùliàngcí* 物量词 ‘object classifier’ used for NNCs. The term *Dòngliàngcí* 动量词 can be found in most Chinese publications dealing with VNCs,

such as in multiple volumes of the influential series of descriptive grammars, *Zhōngguó xīn fāxiàn yǔyán yánjiū cóngshū* 中国新发现语言研究丛书 *Newly Found Minority Languages in China Series.*

Less frequently, the term '(verbal) event classifier' is also attested (Kobuchi-Philip 2021). Finally, highlighting a functional perspective, the term 'adverbial classifiers' appears infrequently in previous studies (Thurgood et al. 2014, Donlay 2015, 235).

Many of the used terms present certain challenges. For instance, while the term 'verbal action classifier' avoids the polysemic confusion of the term 'verbal classifier', not all instances of use are limited to action verbs only. Also, as will be discussed in the following section, the term 'verbal classifier' is polysemous and already used differently in the established research tradition of native American languages. Consequently, based on the established term 'numeral classifier' (e.g., Aikhenvald 2000), we adopt the term 'verbal numeral classifier' to specify numeral classifiers that occur in the verbal context.[3] Similarly, we use the term 'nominal numeral classifier' in the nominal contexts. The terminological choice also reflects our theoretical interpretation that numeral classifiers branch into two main subcategories: nominal and verbal. While 'verbal numeral classifiers' as a term has not been in frequent use, it is not an innovation coined here, but can also be found in previous research, such as Bisang (2018, 276) and Zhang (2017a).

1.2 Previous studies of VNCs in languages of Eastern Eurasia

VNCs are often omitted in descriptive studies. For instance, Iwasaki and Ingkaphirom's (2005) representative grammar of Thai lacks any mentions of VNCs, even though Thai possesses a rich system of VNCs. Even when the phenomenon is explicitly identified, it is often given superficial treatment with few if any examples of use. Also, verbal classifiers may be present in the examples of descriptive grammars without receiving any dedicated discussion, which results in documentation without description. Consequently, the present study identifies VNCs not only from direct mentions but also indirectly, despite the risk of misinterpretation that is mitigated by the specialist survey.

Verbal classifier languages have been subject to highly uneven attention regarding the phenomenon. Of the surveyed languages, the most substantial research concerning VNCs has focused on Mandarin Chinese. Previous studies, such as that of Chao (1968, 627–30), identify tens of different VNCs in the language. Other representative studies of the phenomenon include the pioneering work of Bhaskararao and Joshi (1985) on Newar (Sino-Tibetan, Newaric) and Gerner (2009) on Kam (Tai-Kadai, Kam-Sui). Among recent studies, Li (2020, 166–232, 369–81) not only offers a particularly detailed study concerning the NNCs of Zauzou (Sino-Tibetan, Loloish), but also discusses the VNC system at great length, showing the way and providing a standard for future grammar writing.

The uneven and often insufficient treatment of VNCs has direct relevance to the limitations of this study. To enhance our understanding of this frequent yet paradoxically neglected grammatical phenomenon in languages of Eastern Eurasia and beyond, more fieldwork-based in-depth research from lesser-researched languages is needed. The authors hope that the present study with an areal typological approach lays the groundwork for such research.

3 An anonymous reviewer proposed the term 'event quantification classifier' as an alternative. The term is informative as a label describing the function of the classifiers. Following a similar functional approach in terminology, nominal numeral classifiers could be termed 'entity classifiers' to maintain consistency in the principles of nomenclature. In this study, we nevertheless prefer the terms 'verbal' and 'nominal numeral classifiers' that focus on syntactic relations. Also, the choice of the term 'verbal numeral classifiers' highlights the classifier type's close relationship with 'nominal numeral classifiers', the classifier type better known from earlier research.

1.3 Methodology and structure of the article[4]

The study is based on an analysis of 543 linguistic varieties[5] from Eastern Eurasia. It aims to include languages from all major language families of the region and from diverse branches of the families. The following 23 language families are included in the study: Ainuic, Amuric, Austroasiatic, Austronesian, Burushaski, Chukotko-Kamchatkan, Dravidian, Great Andamanese, Hmong-Mien, Indo-European (Indic and Tocharian branches), Japonic, Koreanic, Kusunda, Mongolic, Nihali, Ongan, Sino-Tibetan, Tai-Kadai, Tungusic, Turkic, Uralic, Yeniseian, and Yukaghir. In addition, available data from the regional creole languages are also taken into account.

Rather than following random sampling suited for formulating statistical generalizations, the sampling of languages included herein is pragmatic based on the availability of descriptive data. Several closely related languages in genealogically relatively compact language families were investigated since distinct patterns of language contact may have yielded different results in those languages regarding the presence of VNCs. A case in point is the Mongolic language family where Khalkha lacks a grammaticalized classifier system, yet such a system with both NNCs and VNCs exists in Mangghuer due to Chinese influence. Finally, reflecting Dahl's (2001, 1456) insight that areal typology manifests both a synchronic and a diachronic side, the survey also includes historical languages of Eastern Eurasia in addition to contemporary languages of the region, since they provide valuable information concerning the diachronic development of VNC systems and explaining the diffusion of the phenomenon.[6]

Methodologically, the data collection and analysis rely on three approaches. First, as the cornerstone of this study, we surveyed existing descriptive grammars and text collections from among the regional languages. Since VNCs are often neglected to varying degrees in grammar writing, this step involved developing hypotheses and dealing with uncertainties regarding the analysis of several phenomena.[7] As Section 3 will show, due to the fragmentary and limited nature of the existing data, some of such uncertainties remain to be answered

4 In the discussion that follows, verbal numeral classifiers are named descriptively with the format CLF.X where X briefly characterizes the nature of the event, as in CLF.time (the generic undifferentiated verbal numeral classifier) and CLF.step (steps taken with a movement verb as the predicate). When more semantic detail is necessary, further distinctions are included with the colon CLF.X:Y, as in CLF.hit:knife (stabs with a knife with a verb of violence as the predicate), as in Stau *œ-ndə ɲcʰə* (one- CLF.hit:knife hit) 'to stab once'. Due to the nature of the phenomenon and semantic domains associated with it, some examples unavoidably include instances of aggressive and violent behavior. Finally, the paper contains examples from three kinds of sources identifiable as follows: Examples from previous published research are attributed to their sources while examples received via the survey or an interviewed native speaker of a language are attributed to the contributor, unless they have been published previously in another context. Examples with the label 'constructed' originate from the authors of the study. To harmonize data from various sources with different research traditions, we have implemented slight adjustments to the glossing conventions followed by each contributor and the surveyed original sources.

5 The term 'linguistic varieties' is used *in lieu* of 'language' here, since some of the surveyed varieties can be argued to be dialects of a language. While the issue is universal in linguistics, in the Eastern Eurasian context, the Greater Himalayan region contains many linguistic cases where drawing the line between a language and a dialect is often challenging. Two main approaches to distinguish languages from dialects have been attempted to draw a line between languages and dialects, namely mutual intelligibility and 'counting sameness' when measuring different domains of the linguistic system (Wichmann 2019). As part of the latter, Wichmann uses a computational method based on the Normalized Levenshtein Distance, reaching the conclusion that the language-dialect dichotomy has an observable basis. Some scholars nevertheless advocate the interpretation of languages and dialects as socially constructed entities without clear boundaries; refer to Weber and Horner (2012, 27–38) and Romaine (2000, 10) for the critique of the notion of countable distinct languages.

6 When offering a visual geographic overview of the language sample and the presence of verbal numeral classifiers from the viewpoint of the present time, only contemporary linguistic varieties are included in Figures 1 and 2. In turn, a brief diachronic overview is offered in Section 5.

7 The topic of verbal numeral classifiers is often not discussed directly in descriptive studies. Even when they are claimed to exist, grammars frequently omit examples of practical use that could serve for identifying the morphosyntactic behavior of the proposed verbal numeral classifiers further. This has required reading the descriptive grammars interpretively as source data and analyzing whether certain descriptions and examples are interpretable as instances of verbal numeral classifiers. Therefore, the classifications and analyses of individual language varieties in this study should not be taken universally as claims by the original authors, but in many instances as our interpretations.

in future work. Following, we conducted a VNC questionnaire distributed among linguists researching languages of Eastern Eurasia. The brief questionnaire serves the function of supplementing the data obtained from written research outputs. It thus served as a supplementary tool for clearing points of doubt, not as a primary research method. Finally, as an additional step to disambiguate the collected data, we informally interviewed several native speakers of regional languages who shared their native intuition regarding VNCs and event quantification in their respective native languages. The remaining challenges are further discussed in Section 3.

The results from the three steps above were collected into an event quantification marker database (see the supplementary file). In the database, each language variety was assigned a language family and geographical coordinates. As the first step, we identified whether the language possessed verbal and nominal classifiers. When the answer to both was positive, to the extent that this was possible, we investigated the semantic domains where the VNCs occur. For languages with no VNCs, we aimed to determine the strategy followed for exact and adjustable quantification. Finally, the completed database was used for drawing typological generalizations.

The present study follows the following structure: In Section 2, we will offer a definition of VNCs and equally illustrate several morphosyntactic phenomena that should be excluded from the definition to keep it meaningful. Following, Section 3 introduces the language sample serving as the major data source and provides an overview of both nominal and verbal classification in the surveyed languages. Analyzing the data, Section 4 shows that typologically, the semantics of VNCs in the surveyed languages follow a non-random distribution by centering around certain core semantic domains discussed in detail. Section 5 briefly delves into the diachronic aspects of VNCs in languages of Eastern Eurasia and discusses the role of language contact in their diffusion. We propose that both NNCs and VNCs originate from the Tai-Kadai languages and have spread widely via language contact. Finally, Section 6 with a conclusion summarizes the central findings of the study and discusses its limitations.

2 Defining VNCs

This section contrasts VNCs with their nominal counterparts and thus starts with a brief discussion on grammaticalized nominal classification devices in Section 2.1. Then, in Section 2.2 it presents a detailed definition of VNCs since the term has been used polysemously in earlier studies. It is argued that not all instances that have been described as verbal or event classifiers in earlier studies should be included inside the scope of the definition to keep it typologically meaningful.

In brief, VNCs are classifiers of occurrence. In other words, they indicate the exact and adjustable frequency of an event expressed by the predicate verb, e.g., *once, twice, thrice,* as *-li* ‘time’ in (4) from Geshiza. As discussed in detail in Section 4, in addition to a generic VNC ‘time’ as illustrated in the example, languages with VNCs may make additional different degrees of semantic distinctions regarding the occurrence of an event, e.g., ‘steps’ or ‘hits with the fist’.

(4) **Geshiza (Sino-Tibetan)**
sop^{h}o *œ-**li*** *dœ-ɛ^{h}oŋ* *zda.*
Suopo one-**CLF.time** PFV-go.PST.1 EXP
‘I have been to Suopo once.’ (Honkasalo 2019, 330)

2.1 Grammaticalized nominal classification devices

Nominal classification and NNCs have received extensive typological interest. Research on VNCs builds on this research tradition. Consequently, before proceeding further and offering a more detailed definition of VNCs, it

is necessary to contrast them with NNCs, a prominent noun categorization device. Furthermore, in what follows, we will briefly investigate other common related noun categorization devices.[8]

Some grammatical devices for categorizing nouns and nominals exist in almost all languages (Aikhenvald 2006, 463). Gender systems in languages, such as Arabic, German, and Spanish, categorize nominal referents into grammatical genders based on various parameters, including animacy and sex. Such systems are characterized by grammatical agreement, exemplified by Spanish where both the indefinite article *un/una* and the adjective *caro/cara* 'expensive' agree in gender with the masculine head noun *apartamento* 'apartment' (5a) and the feminine head noun *casa* 'house' (5b).

(5a) **Spanish (Indo-European)**
un *apartamento* *car-**o***
ART.INDF.M apartment.M.SG expensive-**M.SG**
'an expensive apartment' (constructed)

(5b) ***una*** *casa* *car-**a***
ART.INDF.F house.F.SG expensive-**F.SG**
'an expensive house' (constructed)

Many linguists see gender as a subtype of noun classes with few available categories. Swahili provides an illustrative example of an extensive noun class system and the resulting agreement. The noun *ki-tabu* 'book' belongs to the seventh noun class along with many other tools. Both the adjectival modifier and the predicate verb must agree with the noun by means of the agreement prefix *ki-*, as in (6a). In contrast, *m-toto* 'child' is a noun of the first class for people, and the agreement surfaces as *m-* for the adjectival modifier and as *a-* for the subject marker in the verb, illustrated in (6b).

(6a) **Swahili (Niger–Congo)**
***Ki**-tabu* ***ki**-zuri* ***ki**-li-kuwa* *meza-ni.*
7-book **7**-good **7**-PST-EXV table-LOC
'The good book was on the table.' (constructed)

(6b) ***M**-toto* ***m**-zuri* ***a**-li-kuwa* *nyumba-ni.*
1-child **1**-good **1**-PST-EXV home-LOC
'The good child was at home.' (constructed)

Finally, NNCs are a noun categorization device that co-occurs with accompanying numerals and/or quantifiers and classifies nominal referents into different semantic categories. The classificatory criteria may be based on animacy, shape, and other inherent properties (Aikhenvald 2006, 466). NNCs form a system at a midway point between lexical (e.g., English two *head* of cattle) and grammatical (e.g., gender and noun classes) means of nominal classification (Grinevald 1999, 2000). The two main categories of NNCs are sortal and mensural classifiers, although the latter is sometimes treated separately and the criteria distinguishing the two have been debated. Sortal numeral classifiers categorize nouns in terms of their inherent properties, such as animacy and shape (Aikhenvald 2000, 115). For instance, in Xong, the classifier *ngonl* marks the referent, namely *mioul* 'fish', as an animate entity (7), unlike the classifier *hnanf,* for instance, which is used for inanimate referents in the semantic field of clothing (Sposato 2015, 258–9).

8 The treatment here is not intended as exhaustive, but merely to give sufficient background information placing numeral classifiers into a larger context and to illustrate how verbal numeral classifiers differ from nominal classification devices. In addition, noun classifiers, classifiers in possessive constructions (relational, possessed, and possessor classifiers), locative classifiers, and deictic classifiers exist in languages of the world (Aikhenvald 2000, 2006, 2019).

(7) **Xong (Hmong-Mien)**

Wel	*naond*	*aod-**ngonl***	*mioul*	*neind*	*liox*	*guaot!*
1SG	ASSOC	one-**CLF.animate**	fish	DEM.PROX	big	pass

'This fish of mine is really big!' (Sposato 2015, 229)

Mensural numeral classifiers, also known as measure words, individuate entities in terms of quantity (Lyons 1977, 464). In contrast to sortal classifiers that categorize nouns based on their inherent properties, they categorize based on temporal states. In other words, they "contain information about how the referent is measured" (Aikhenvald 2006, 466). In (8), the Xong mensural numeral classifier *zhet* 'CLF.bowl' indicates that the cooked rice that the speaker ate was in a bowl and the speaker ate two bowls of such rice.

(8) **Xong (Hmong-Mien)**

Tat-hnef	*wel*	*nonx*	*oub-**zhet***	*hlit.*
this-CLF.day	1SG	eat	two-**CLF.bowl**	cooked.rice

'I ate two bowls of rice today.' (Sposato 2015, 235)

Since the present study focuses on VNCs, rather than NNCs, we adopt the following broad criterion for the presence of NNCs: A language is defined as having grammaticalized NNCs if it has at least one sortal numeral classifier and at least some nouns in the language are typically counted with the help of classifiers. The size of the classifier inventory does not matter, and classifiers need not be mandatory, but a prototypical classifier inventory contains several classifiers so that it can serve the function of classifying nouns. When counting, NNC insertion may be compulsory (e.g., Mandarin Chinese) or optional (e.g., Khmer), especially under language-specific conditions, such as in the context of large numbers. Systems that deviate from the prototype above and yet show similarity with it are defined as limited classifier systems. For instance, a language might be in the process of classifier genesis, as Kurtöp (Hyslop 2017, 148–9). Alternatively, it might possess counters for the subset of specific nouns, such as humans, which serve the function of noun categorization, as in Ainu (Tamura 2000, 255–9). Systems with mensural quantifiers for entities of low countability, such as English *bottles* and *ounces* in *three bottles of Coke* and *five ounces of gold dust,* respectively, occur in most or all languages (Gil 2013) and are excluded here. Like many other linguistic phenomena, however, NNCs form a continuum from prototypical to limited or emerging systems, which occasionally makes it difficult to establish whether a language is a 'classifier language' or not in borderline cases.

NNCs show strikingly uneven geographic distribution with a core area in East and Southeast Asia extending both eastwards and westwards (Gil 2013). At the time of writing, the research database of Her and Li (2023) qualifies as the most extensive cross-linguistic source of information concerning NNCs, and the authors identify over 700 classifier languages from among languages of the world, also confirming the earlier findings of strikingly uneven areal distribution. Hölzl and Cathcart (2019) show how classifiers become less frequent when moving from East to West in Eurasia. In all, as confirmed by the present study, the distribution of NNCs greatly overlaps with that of VNCs. The two are nevertheless not identical, since only a subset of languages with NNCs also possess VNCs. Consequently, verbal numeral classifier languages form a smaller subset of all classifier languages.

2.2 Defining VNCs

The notion of verbal classifiers has been used in various ways in different linguistic approaches. For instance, in the research tradition of indigenous languages of the Americas and Papua, verbal classifiers refer to "morphemes which occur on the verb and characterize a core argument in terms of its shape, form, consistency, and other semantic properties" (Aikhenvald and Dixon 2011, 157). To illustrate, Imonda, a Papuan language, possesses a wide range of verbal classifiers that attach prefixes to the verb. In (8), the classifier *u-* indicates that the referent *tõbtõ* 'fish' belongs to the group of small animals, such as frogs and fish (Seiler 1985, 121). Since such devices do not mark event frequency, they are not discussed in the present study. Rather, they constitute yet another form

of noun categorization device with agreement (Aikhenvald 2000, 151). Against this backdrop, we propose a terminological distinction, namely 'verbal classifiers' to be used in the meaning illustrated above and 'verbal numeral classifiers' for grammaticalized classifier devices for counting and classifying events.

(9) **Imonda (Border/Upper Tami)**
tõbtõ *ka-m* ***u**-ai-h-u.*
fish 1-GL **CLF.small.animal**-give-REC-IMP
'Give me the fish!' (Seiler 1985, 120)

To avoid terminological confusion, it is necessary to define VNCs in a fashion that succinctly, yet accurately characterizes their behavior from a cross-linguistic viewpoint in Eastern Eurasia. Below, we offer three criteria for defining VNCs. The first criterion defines the classificatory device as dependent on a grammaticalized NNC system. Following, criteria two and three address the syntax and semantics of the classifiers, respectively. In addition to showing what VNCs are, the criteria also illustrate what they are not. Multiple linguistic phenomena have either been included as instances of VNCs or could potentially be conceptually conflated with it. The end of the section summarizes the discussed criteria.

2.2.1 Criterion 1: Dependence on nominal numeral classification

VNCs constitute a grammaticalized classifier system and thus differ from lexical quantification devices, such as counted nouns (e.g., English 'time' in *She saw the movie three times*). To qualify as a VNC, a morpheme needs to fulfill the basic properties of classifiers in a language. The definition should not be taken as a requirement for fully identical morphosyntactic behavior, an impossibility due to among other things the different syntactic distribution of NNCs and VNCs. Rather, the emphasis lies on similarities between the two. In other words, both VNCs and NNCs are subclasses of the word class 'classifiers'.

The criteria of classifierhood must be established on a language-specific basis and as an example, commonly include (at least moderately) free co-occurrence with a numeral. To illustrate further, Indonesian has NNCs, such as *ekor* used for animals. The numeral *satu* 'one' appears in a prefixed form *se-* attached to a classifier, including both the sortal and mensural subtypes. Consequently, *tendangan* 'kick' fails to qualify as a VNC in the language, since it cannot host the prefixal form of the numeral and should thus be considered a counted noun (10), unlike the genuine VNC *kali* 'time' that behaves like other classifiers in the language by hosting the numeral prefix, as in *se-kali* 'one time, once'.

(10) **Indonesian (Austronesian)**

Atlet	*memberikan*	***satu/*se-***	*tendangan*	*kepada*	*lawan-nya.*
athlete	give	**one**	kick	to	opponent-3SG

'The athlete gave a kick to his/her opponent.' (Syafruddin)

The classifier system of a language may have NNCs and lack VNCs, but no opposite systems where VNCs exist while NNCs are lacking were identified in this study (see also Vittrant and Tang 2021). To illustrate, English has no grammaticalized system of nominal numeral classification, and 'time' does not qualify for a VNC. Rather, in terms of its word class properties, it behaves like other nouns of the language and thus constitutes a lexical means for achieving the same function, namely to indicate the frequency of an event. At the same time, Japanese in example (11) contains a VNC *-do* 'time', since the language possesses a system of NNCs and in terms of its word class, *-do* shares the basic properties and behavior of classifiers in the language, rather than that of prototypical nouns.

(11) **Standard Japanese (Japonic)**

彼女は	東京を	三度	訪れた	ことが	ある。
kanojo=wa	*tokyo=wo*	*san-**do***	*otozure-ta*	*koto=ga*	*aru.*
she=TOP	Tokyo=ACC	three-**CLF.time**	visit-PST	NMLZ=NOM	EXV

'She has visited Tokyo three times.' (constructed)

Importantly, the requirement above signifies that similar to prototypical NNCs, VNCs must be compatible with a wide numeral range, not only with the number one, for instance. The choice of the term 'wide' is intentional: A wide numeral range does not necessarily mean the same as a limitless numeral range, since among others, pragmatic and cultural reasons may limit the numeral range a certain VNC typically co-occurs with. Stau possesses both NNCs and VNCs as closed word classes, *-ŋɖʐə* 'time' illustrating the latter (12). In addition, many verbs may undergo semelfactive-like derivation in which the number prefix *œ-* 'one' attaches to the unconjugated verb stem followed by the light verb *v-rœ* (13). Unlike in the case of VNCs with productively adjustable numeral value, however, such semelfactives that closely resemble classifiers cannot replace the numeral prefix *œ-* by any other number. Consequently, *ɣnə-* 'two' in (14) results in an ungrammatical formulation, while in (12), the numeral prefix can be freely adjusted, as in *ɣnə-ŋɖʐə* 'two times'. Ergo, classifier-like semelfactive formulations of the type (13), *sensu stricto,* fall outside the scope of the prototypical verbal numeral classification.

(12) **Stau (Sino-Tibetan)**

ŋa	*kapəla*	*gɛ-ʁjɛ*	*də-ŋoŋ.*	*stɛwu*	***œ-ŋɖʐə***	*ʝaŋ.*
1S	forehead	ADJZ-good	PFV-COP.1	Stau	**one-CLF.time**	come.PST.1

'I was lucky. I came (i.e., I was able to come back) to Daofu once.' (personal fieldwork)

(13)

tʰe=qʰe	*ŋe*	*vdʐə=gi*	***œ-skʰəri***	*də-re.*
DEM=LOC	1SG.GEN	friend=DAT	**SEM-shout**	IMP-LV.2SG

'Call to my friend over there!' (personal fieldwork)

(14)

**tʰe=qʰe*	*ŋe*	*vdʐə=gi*	***ɣnə-skʰəri***	*də-re.*
DEM=LOC	1SG.GEN	friend=DAT	**two-shout**	IMP-LV.2SG

Intended meaning: 'Call twice to my friend over there!' (rejected)

Paris (2013, 270) divides VNCs into 'weak' and 'strong'. Cases restricted to co-occurrence with the number 'one' belong to 'weak classifiers' while 'strong classifiers' co-occur with the full range of numbers. From this, we can see that Paris's 'weak' classifiers correspond to the semelfactive quasi-classifiers of this article. They do not quantify the times of occurrence with the possibility of adjustability in the number value and are thus excluded from the present study, an issue which will be discussed in more detail in criterion 3. Classifier-like semelfactives, however, exist in several languages of Eastern Eurasia and would thus form a fruitful ground for a separate typology study. Example (15) from Japanese illustrates what Paris's categorization labels 'weak verbal classifiers' and what the present study calls semelfactive quasi-classifiers. The word *koe* 声 literally means 'voice' and it is used idiomatically in the expression *hito-koe wo kakeru* 'to call (someone)'. The numeral is fixed and cannot be replaced. Interestingly, the number 'one' occurring in such expressions may play the role of marking infiniteness, rather than counting.

(15)

koe 声	'voice'
hito-koe 一声	'one call'
**futa-koe* 二声	intended: 'two calls'
**hyaku-koe* 百声	intended: 'a hundred calls'

Finally, the similarity between NNCs and VNCs implies that like the former, the latter are generally compatible with interrogation by an interrogative pro-form that often carries the basic meaning 'how

Table 1: Compatibility of both NNCs and VNCs with interrogative pro-forms

Language	Interrogative	Nominal example	Verbal example
Japanese	*nan(i)*	*nan-biki*	*nan-po*
	'what'	'how many small animals'	'how many steps'
Korean	*myŏt*	*myŏt-myŏng*	*myŏt-bŏn*
	'how many'	'how many people'	'how many times'
Mandarin Chinese	*jǐ*	*jǐ běn*	*jǐ quán*
	'how many'	'how many volumes of books'	'how many hits with a fist'
Geshiza	*xazi*	*xazi-q^{h}a*	*xazi-rəu*
	'how many'	'how many stick-like objects'	'how many hits with an axe'
Thai	*kìː*	*kìː k^{h}an*	*kìː kâːw*
	'how many'	'how many cars'	'how many steps'
Vietnamese	*bao nhiêu*	*bao nhiêu người*	*bao nhiêu lần*
	'how many'	'how many people'	'how many times'
Malay	*berapa*	*berapa orang*	*berapa kali*
	'how many'	'how many people'	'how many times'

much, how many'. The two typically share at least one interrogative pro-form, as in Thai: กี่ *kìː* 'how many', กี่คน *kìː k^{h}on* 'how many people' vs. กี่ครั้ง *kìː k^{h}ráŋ* 'how many times'. In other words, interrogatability constitutes a shared feature between NNCs and VNCs, illustrated further in Table 1. It follows that incompatibility with an interrogative generally indicates that a form is not a VNC. This is connected to VNCs' semantic functions of quantifying events discussed in criterion 3. However, since the alternative devices for VNCs, such as counted nouns and iterative numerals, may also be compatible with interrogation, co-occurrence interrogation does not necessarily confirm the status of a morpheme as a VNC.

To conclude the first criterion and to inspect the similarity and overlap in VNCs and NNCs, three distributional possibilities exist for a numeral classifier regarding its co-occurrence (or lack thereof) in nominal and verbal contexts. A numeral classifier may be dedicated for either nominal or verbal use only or it may be compatible with both. Matthews and Leung (2004) illustrate how Cantonese Chinese exhibits all these three patterns. For instance, *seng^{1}* 声 'voice' is used exclusively for vocalizations, such as calls and instances of cough. On the other hand, *faai^{3}* 塊 'slice' shows exclusive nominal use only for counting slices of objects, such as a cake. Finally, numeral classifiers, such as *caan^{1}* 餐 'meal' show compatibility with both nominal and verbal use.

2.2.2 Criterion 2: Syntactic link to the verb phrase

NNCs belong syntactically to noun phrases (NPs) where they may either modify a head noun or substitute for it, depending on the language in question. In (17) from Thai, the classifier เล่ม *lêm* used for books modifies the head noun หนังสือ *nǎŋsŭːˑ* 'book' together with which they form a classifier phrase (CLFP), a subunit of NP.

(16) **Central Thai (Tai-Kadai)**

[[หนังสือ	สาม	เล่ม]$_{\textbf{CLFP}}$	นั้น]$_{\text{NP}}$	หนัก	มาก
[[***nǎŋsŭː***	***sǎːm***	***lêm***] $_{\textbf{CLFP}}$	***nán***]$_{\text{NP}}$	*nàk*	*mâːk.*
book	**three**	**CLF.book**	**DEM.DST**	be.heavy	very

'Those three books are very heavy.' (constructed)

In contrast, VNCs manifest a syntactic connection with the predicate verb. They combine with a numeral to form a verb CLFP that serves the function of a complement to the verb. Regarding their syntactic function, VNCs resemble adverbs at least in some of the investigated languages, which is the reason for the term 'adverbial classifiers' occasionally met in the research literature. In Stau, the generic VNC *-ŋqʐə* 'time' that attaches to its numeral host *ɣsu* 'three' directly precedes the verb in (17a). The adverb *vɛrvi* 'slowly' in (17b) shows identical syntactic placement by directly preceding the verb.

(17a) **Stau (Sino-Tibetan)**
vdzi *ɣsu-ʁe* *ɣsu-**ŋɖʐə*** *nə-ɕə-sə.*
person three-CLF.person three-**CLF.time** PFV-go.3-IFR
'Three people went three times.' (Gates 2021, 225, transcription harmonized)

(17b) ***vɛrvi*** *tə-ɕən.*
slowly IMP-go.2
'Go slowly!' (Gates 2021, 124)

It is important to distinguish VNCs from NNCs occurring in or as an object argument to a verb. As Gerner (2014, 288–9) demonstrates, Nuosu allows some 'container nouns', such as 'mouth', to be used as classifiers. Such classifiers, however, modify a noun that may be present or omitted, as in (18). While frequent omissions of the possible head noun complicate the quick identification of such cases, they must nevertheless be excluded from VNCs, since they do not modify the verb, rather than a head noun that may be present or omitted.

(18) **Nuosu (Sino-Tibetan)**
$ts^{h}i^{33}$ ***(dza^{33})*** $ts^{h}i^{21}$ ***ba^{33}ɬɔ^{21}*** *dzɯ^{33}.*
3SG **food** one **CLF.mouth** eat
'He ate two mouthfuls of food.' (Gerner 2014, 289)

Somewhat distinctly from the above, NNCs indicating frequency should not be conflated with VNCs that indicate occurrence. In Mandarin Chinese, the classifier *cì* 次 'time' functions as the generic VNC for iterations (19a). The same classifier, however, is polyfunctional by also occurring in NPs, as in (19b) where in *jǐ cì huìtán* 几次会谈 'some/several talks' *cì* indicates that the head noun 'talk' took place several times. In sum, the requirement for a syntactic link with the verb phrase rules out (19b) as an instance of VNC, yet (19a) qualifies. As stated earlier, this reflects how a single classifier may function both in the nominal and verbal domains.

(19a) **Mandarin Chinese (Sino-Tibetan)**
我 去 过 日本 两 次。
wǒ *qù* *guò* *Rìběn* *liǎng* ***cì***.
1SG go EXP Japan two **CLF.time**
'I have been to Japan twice.' (constructed)

(19b) 他们 进行 了 几 次 会谈。
tā-men *jìnxíng* *le* *jǐ* ***cì*** *huìtán.*
3-PL hold ASP some **CLF.time** talk
'They held several talks.' (Li and Thompson 1981, 110)

Also, the study omits what Gerner (2014, 289–91) calls 'verbal auto-classifiers', namely 'verbs which serve as their own phase and event counter'. Coined by Matisoff (1973, 89) for Lahu (Sino-Tibetan, Lolo-Burmese), auto-classifiers in the nominal context refer to scenarios where some nouns function as their own classifiers, such as *yè tê yè* 'one house' in Lahu. The term 'repeaters' is also used synonymously. In the verbal context, the term has been used for expressions, such as Mandarin Chinese *kàn (yí) kàn* 看(一)看 (look (one) look) 'to have, take a look'.[9] To distinguish the nominal and verbal domains, Liu (2018) labels the former 'auto nominal classifiers' and the latter 'auto verbal classifiers' with the 'reduplicants coding lexical or grammatical aspects'. In languages, such as Mandarin Chinese, auto-verbal classifiers do not quantify events and show compatibility only with the numeral *yī* 一 'one', which makes them functionally semelfactives (see criterion 1) and further highlights their difference from auto-nominal classifiers.

9 It has been widely discussed whether the presence of the numeral *yī* 一 'one' included in parenthesis changes the meaning of the expression. However, we will not explore this issue further, as it is tangential to the main arguments of the present study.

In some languages, auto-verbal classifiers may possibly be fully productive in the sense of being compatible with a broad base of verbs and serve the function of adjustable event quantification. Adjustability in event quantification, namely whether the accompanying numeral value can be modified or not, is typically not discussed in descriptions of VNCs. This issue thus requires more investigation and would benefit from a dedicated study.

2.2.3 Criterion 3: Semantic function of quantifying events

VNCs quantify events rather than referents, the domain of NNCs. Consequently, due to their distinct function, even when they show similarity with NNCs, they cannot be considered a nominal categorization device. Duratives, namely expressions indicating temporal lengths of time units, such as 'two days', 'three weeks', and 'five years' remain outside the scope of the definition, since they do not indicate frequency. In many of the surveyed languages, duratives are either temporal nouns or temporal adverbs. In others, they appear as 'self-referential' classifiers, for instance in Geshiza, Mandarin Chinese, and Thai (Table 2).

Self-referential classifiers count themselves (Konnerth 2014, 142), rather than quantifying events. Also called 'independent classifiers' by Haas (1942, 204–5) in the context of Thai, they are close to mensural nominal classifiers, since rather than serving the role of noun categorization device, they measure the quantity of time units, albeit self-referentially (see Honkasalo 2019, 328). The term 'time classifier' is also occasionally used for the classifier type (Zhang 2014, 901). In Karbi, self-referential classifiers, such as *arnì* 'day' and *jó* 'night', appear (20). Similarly, วัน *wan* 'day' in Thai counts itself, as in สองวัน *sɔ̌ːŋ wan* (two CLF.day) 'two days'. Repeated formulations, such as *วันสองวัน **wan sɔ̌ːŋ wan* (day two CLF.day), with the intended meaning 'two days' are not used.

(20) **Karbi (Sino-Tibetan)**
aphi atum ejo arnisi dokokta
a-phì *a-tūm* ***e-jó*** ***arnì-sī***[10] *dō-kòk=tā*
POSS-grandmother POSS-PL **one-CLF.night** **CLF.day-one** stay-back=ADD
"ehem chedamnang erit chedamnang […]"
e-hēm *chV-dām-nāng* *e-rīt* *chV-dām-nāng*
1PL.INCL-house RR-go-HORT 1PL.INCL-field RR-go-HORT
'The grandmother and grandfather stayed just one night and one day, and then (the old man) said: let's go home, let's go to our property […]' (Konnerth 2014, 143)

In addition to the semantic function of quantifying events with an exact and adjustable numerical value, VNCs also manifest a delimiting function by setting a boundary for an event (Matthews and Leung 2004, 453). To illustrate, in (21a), the bare verb *haang*^{A} 行 'to walk' describes the event as an unbounded event with no VNC present. In contrast, the introduction of the VNC *bou*^{6} 步 'CLF.step' in (21b) delimits the event and converts the

Table 2: Examples of self-referential classifiers in Geshiza, Mandarin Chinese, and Thai

Type	Geshiza	Mandarin Chinese	Thai
day	*-sni*	*tiān*	*wan*
night	*-rja*	*wǎn*	*kʰɯːn*
week	n/a	n/a	*ʔaːtʰít, sàpdaː*
month	*-slə*	n/a	*dɯan*
year	*-ko*	*nián*	*piː*

10 Konnerth (2014, 143) points out that the enumeration is irregular and surfaces instead of the expected **e-nì*.

act of walking into a bounded event by establishing a boundary to its occurrence. As the authors further argue, this also corresponds to transforming the event from an atelic into a telic one.

(21a) **Cantonese (Sino-Tibetan)**
行
haang^{4}
walk
'to walk' (Matthews and Leung 2004, 453, Chinese character added)

(21b) 行 兩 步
haang^{4} *loeng*^{5} ***bou***^{6}
walk two **CLF.step**
'to walk two steps' (Cantonese; Matthews and Leung 2004, 453, Chinese characters added)

2.2.4 Summary of the criteria for defining VNCs

In conclusion, the present study defines prototypical VNCs as follows: VNCs are a grammatical classificatory device for quantifying events in some languages that also have NNCs. VNCs manifest a strong syntactic link with the predicate verb, rather than with an NP that is the domain of NNCs. VNCs show productive compatibility with wide numeral ranges, rather than co-occurring with fixed numbers only. *Via negativa,* the following are excluded from the definition of prototypical VNCs: 1) formulations occurring with fixed numbers only, typically 'one'; 2) NNCs occurring as the object argument of a verb regardless whether a head noun is present or not; 3) verbal 'auto-classifiers' where a verb is repeated to serve as its own counter; and 4) duratives appearing typically as self-referential classifiers to indicate temporal lengths without quantifying events.

3 Language sample of the present study

This section offers an overview of the distribution of numeral classifier systems in the surveyed 23 language families. The 11 major phylogenetic units of Eurasian languages in the present study include Austroasiatic, Austronesian, Dravidian, Hmong-Mien, Indo-European (mostly Indo-Iranian), Mongolic, Sino-Tibetan (also known alternatively as Trans-Himalayan, Tibeto-Burman), Tai-Kadai, Tungusic, Turkic, and Uralic. Additionally, in the figures that follow, we include under an umbrella unit 'Paleo-Asiatic' the following 12 smaller genealogical units: Ainuic, Amuric, Burushaski, Chukotko-Kamchatkan, Great Andamanese, Japonic, Ongan, Koreanic, Kusunda, Nihali, Yeniseic, and Yukaghir, and also take the major creoles and pidgins of the region into account. In total, the sample contains 543 distinct contemporary linguistic varieties.

The analysis of the languages and the language families reveals that language families of Eastern Eurasia form three core groups regarding the presence and absence of VNCs. First, Hmong-Mien, Japonic, Koreanic, and Tai-Kadai are the only language families where VNCs are consistently present among all member languages of the respective families. Also, only languages of these language families, together with some branches of Sino-Tibetan (e.g., Qiangic), possess fully fledged VNC systems. In contrast, in the remaining language families, grammaticalized VNCs are generally rudimentary if they are present.

Second, in Austroasiatic, Austronesian, Indo-European, Mongolic, Sino-Tibetan, and Turkic, verbal classifiers are present only in some, but not in all languages of the respective language families. Finally, in Ainuic, Amuric, Burushaski, Chukotko-Kamchatkan, Dravidian, Great Andamanese, Kusunda, Nihali, Ongan, Tungusic, Uralic, Yeniseian, and Yukaghir, no systems of VNCs were identified.

Summarizing the findings, Table 3 provides an overview of the surveyed languages in terms of the presence or absence of VNCs.

Following, Figure 1 illustrates the geographic distribution of the surveyed languages. In turn, Figure 2 illustrates the distribution of classifier systems in Eastern Eurasia. Red dots on the map indicate the surveyed

Table 3: Languages of Eastern Eurasia and VNCs

Typ. group	Phylogenetic units	Remarks on distribution
Group 1	Hmong-Mien, Japonic, Koreanic, Tai-Kadai	VNCs consistently present across the units
Group 2	Austroasiatic, Austronesian, Creoles and pidgins, Indo-European, Mongolic, Sino-Tibetan, Turkic	VNCs present in some languages of the units
Group 3	Ainuic, Amuric, Burushaski, Chukotko-Kamchatkan, Dravidian, Great Andamanese, Kusunda, Nihali, Ongan, Tungusic, Uralic, Yeniseian, Yukaghir	VNCs consistently absent across the units

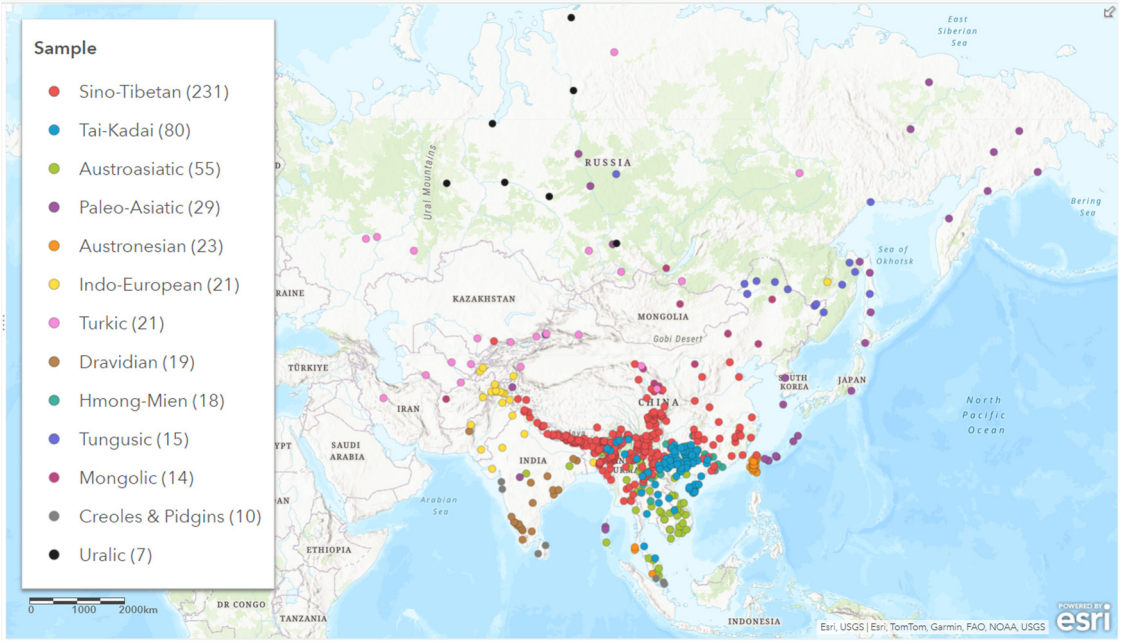

Figure 1: Geographical locations of the surveyed languages (map base © ArcGIS).

languages with both NNCs and VNCs present; blue dots mark languages with only NNCs being identified; and black dots code the languages with no classifiers of either kind.

Since there is also a significant internal variation in each phylogenetic unit as well as ambiguity observed in the published grammatical descriptions, we subsequently discuss three worthwhile observations from the data.

The first issue concerns the nature of data sources in historical languages. For instance, it is impossible to ascertain whether the absence of VNCs in the Shang Chinese oracle bone inscriptions in general results from a lacuna in the surviving inscriptions or from a genuine lack of this grammatical feature. At the same time, while the surviving materials from Old Korean are insufficient for determining how the language counted verbal events, the lack of NNCs helps us deduce that VNCs must also have been absent.

Second, drawing the line between devices of exact and adjustable quantification, particularly in the case of counted nouns and VNCs, is often challenging. While not unique in this respect, Modern Khmer presents such a problem for analysis, since *da:ŋ* 'time' is analyzable either as a VNC or as a counted noun, namely a noun that is counted directly by means of a numeral. Old Khmer with the oldest historical documents available shows incipient classifier genesis, but the language likely had no classifier system in its earlier stages that have not been recorded. Also, it is historically clear that Khmer has used the device of counted nouns for exact and adjustable event quantification. Against this backdrop and due to the scarcity of usable morphosyntactic clues,

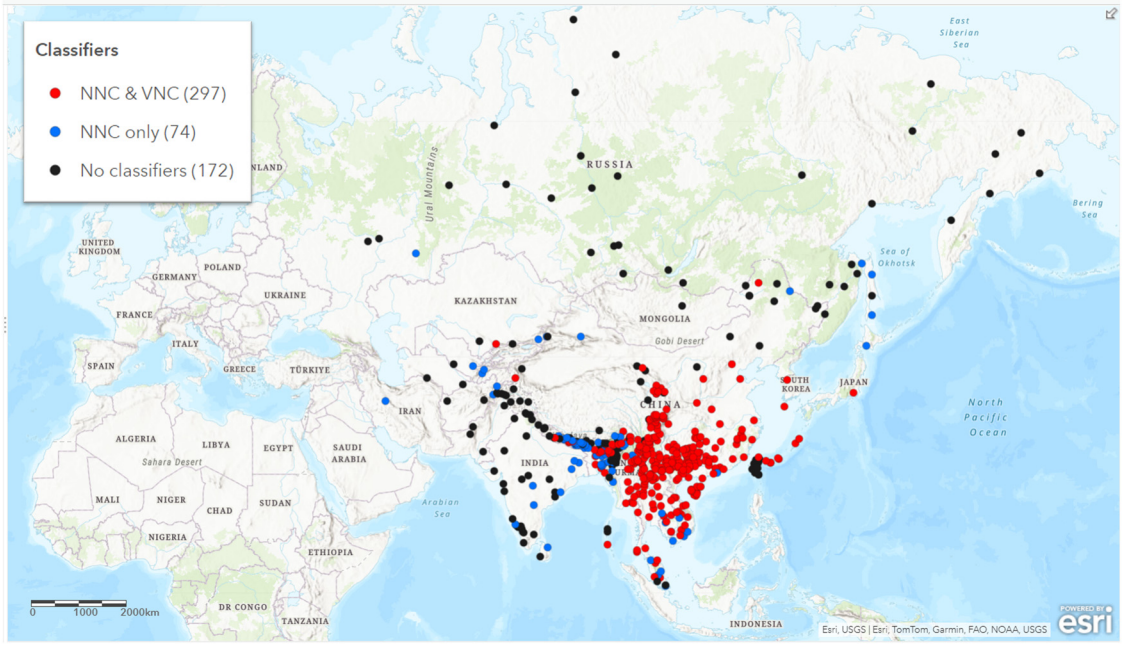

Figure 2: Core distribution of VNCs (map base © ArcGIS).

it remains challenging to determine whether *da:ŋ* has been reanalyzed as a VNC or whether it remains as a counted noun. In all, *da:ŋ* either continues as a counted noun or has become reanalyzed as a classifier sometime at the Middle Khmer stage or after when the classifier system emerged in the language. The issue requires further investigation.

As the mirror image of the above, some earlier research has used the label 'verbal numeral classifier' to devices of exact and adjustable quantification that the criteria established in the present study analyze differently. In Jingpaw for instance, *làŋ* is often described as a VNC, but the morpheme is likely a counted noun forming right-headed noun-noun compounds typical to the language; it also shows different word order (NUM-*làŋ*) when compared with NNCs (CLF-NUM; Kurabe, personal communication, July 25, 2022). Consequently, following the criteria for establishing VNCs adopted in this article (Section 2.2.1), a non-VNC interpretation is preferred.

The third area of potential confusion is what counts as a classifier system in contrast to sporadic counting devices which can be alternatively analyzed as counted nouns. For example, China Coast Pidgin shows limited use of a general classifier *piece(e)* that originates from Chinese influence (Ansaldo et al. 2010; Shi 1991, 20). This limited use of a counter is not considered a fully-fledged classifier system here. Likewise, in Sri Lankan Portuguese Creole, Smith (1977, 148–9, 2013) shows how *pesaam* 'person' occasionally occurs as a classifier for humans. This limited use of a counter is not interpreted as a fully-fledged classifier system in the present work.

We propose a two-part universal based on the respective sizes of the VNC and NNC inventories. First, complex NNC systems correlate with the presence of verbal numeral systems. The most complex VNC systems of Tai-Kadai, Hmong-Mien, and some branches of Sino-Tibetan, such as Qiangic and Sinitic, invariably occur in languages with complex NNCs. Conversely, very simple NNC systems correlate with the lack of VNCs. Second, in a language with both subtypes of classifiers present, the NNC subsystem typically shows richer formal complexity than its verbal counterpart. While the universal is not absolute, it is nevertheless a very strong tendency: The only identified counterexamples originate from Sinitic contact varieties that have greatly simplified their systems of NNCs, such as Kazakhstani Gansu Dungan (Honkasalo 2024), but at the same time, retain systems of VNCs.

4 Typological trends in lexical meanings of VNCs

The present study shows that rather than following random distribution, lexical meanings of VNCs in languages of Eastern Eurasia are restricted to particular semantic fields. This parallels many other grammatical phenomena. For instance, lexicalizations of negative senses center around a limited number of semantic domains (Veselinova 2013).

The semantic types discussed below are established with the help of two criteria. Primarily, we have formulated hypernymic categories that encompass a wide range of subtype-level manifestations of VNCs (apart from the category of 'generic') identified in the source materials. For instance, classifiers of hits with first, kicks with feet, and bites of mouth, etc. have been grouped under the semantic category 'instrumental A: body parts as classifiers with "verbs of violence"'. Secondarily, we have attempted to create categories for widely attested semantic types of VNCs present in the source materials.

Table 4 summarizes the findings. The ten language families with identified VNCs can be grouped further on the elaborateness of the respective systems. The widest semantic range of VNCs is found in Tai-Kadai, Hmong-Mien and Sino-Tibetan, particularly in the Qiangic and Sinitic branches. In contrast, the Japonic and Koreanic language families show less elaborate systems of VNCs. The VNC systems of Indo-European, Mongolic, and Turkic are of very simple nature, highlighting their origin due to language contact. Also, the systems in Austroasiatic and Austronesian include only an identified generic classifier and, in a few cases, additionally an identified locomotive classifier, typically for steps. Due to the nature of existing grammatical descriptions, many instances in these two language families remain opaque regarding whether a lexical item is best analyzed as a classifier or a counted noun. Consequently, Austroasiatic and Austronesian are clearly not core languages for VNCs, and further research may result in narrowing down the range of language with VNCs. These observations are also directly linked with the diachrony of VNCs, as will be discussed in Section 5.

We propose a hierarchy between the semantic fields for predicting the occurrence of a classifier type: onomatopoeic > instrumental > locomotive > generic. While this hierarchy should not be taken as an absolute universal, it illustrates the general tendencies in languages with verbal classifiers. If a language possesses the onomatopoeic classifier type, all other types are likely to be equally present (e.g., Thai). In turn, if onomatopoeic classifiers are lacking, but the instrumental type is present, a language likely also possesses locomotive and generic verbal nominal classifiers (e.g., Geshiza). Furthermore, if a language lacks both onomatopoeic and instrumental classifiers, yet possesses the locomotive type, it will additionally have one or several generic classifiers (e.g., Dungan).

4.1 Generic VNCs and other general classifiers

All languages with grammaticalized verbal numeral classification include a classifier with a broad functional range for quantifying times of occurrence. Such a classifier is termed 'generic verbal numeral classifier' in the

Table 4: Semantic ranges of the VNCs in the analyzed language families

Language family	Generic	Locomotive	Instrumental B	Instrumental A	Onomatopoeic
Tai-Kadai	✓	✓	✓	✓	✓
Hmong-Mien	✓	✓	✓	✓	✗
Sino-Tibetan	✓	✓	✓	✓	✗
Japonic	✓	✓	✓	✗	✗
Koreanic	✓	✓	✓	✗	✗
Austroasiatic	✓	✓	✗	✗	✗
Austronesian	✓	✓	✗	✗	✗
Indo-European	✓	✓	✗	✗	✗
Mongolic	✓	✓	✗	✗	✗
Turkic	✓	✓	✗	✗	✗

Table 5: Examples of generic VNCs in the surveyed languages

- ***Austroasiatic:*** *Aslian:* **Semelai** *tən; Bahnaric:* **Jruq/Loven** *ruoh,* **Stieng** *wət; Katuic:* **Kuy** *deːŋ,* **Pacoh** *kəntiʔ; Khasi-Palaungic:* **Mok** *pɔ́k; Khmuic:* **Mlabri** *thɯɯ; Pearic:* **Chong** *dɔːŋ; Vietic:* **Vietnamese** *lần*
- **Austronesian: Malay** *kali,* **Urak Lawoi'** *kali*
- ***Japonic:*** *Japanese:* **Old Japanese** 多妣 *-taNpyi,* **Standard Japanese** 回 *-kai; Ryukuan:* **Yuwan Amami** 回 *-kəi,* **Irabu** 回 *-kwai,* **Okinoerabu** 回 *-(k)koi*
- ***Koreanic:*** **Standard Korean:** 번 *-bŏn*
- ***Sino-Tibetan:*** *Gyalrongic:* **Geshiza** *-li,* **Stau** *-ɳɖʐə,* **Khroskyabs** *-sti,* **Jiaomuzu Situ Gyalrong** *cʰa; Karenic:* **Sgaw Karen** *blɔ1; Lolo-Burmese:* **Bisu** *tsaŋ*^{55}, **Cosao** *thɔ*^{55}, **Enu/Ximoluo** *phv̠*^{33}, **Khatso** *thɤ55,* **Lalo** *phuq,* **Lisu** *xua*^{35}, *tʂo*^{33}, *tʂɒ*^{33} (dialectal variation), **Yao'an Lolo** *bol; Newaric:* **Bhaktapur Newar** *kAlA/kA,* **Pahari Newar** *-pu; Qiangic:* **Yongning Na** *ʂɯ33,* **Pumi/Prinmi** *xɜjF; Sinitic:* **Gan Chinese** 次 *tshi213,* **Mandarin Chinese** 次 *cì,* **Shaowu Min Chinese** 回 *fei22,* **Southern Min Chinese** 擺 *pai3,* **Xiang Chinese** 回, 次 *fei13, tshɪ45; unclear affiliation:* **Tujia** *qie2*
- ***Tai-Kadai:*** *Kam-Sui:* **Kam** *tau*^{53}*; Kra:* **Lakhi** *ɕɔN*^{22}, **A'ou/Zoulei** *vie*^{31}, *Tai:* **Lao** ເທື່ອ *tʰʉa,* **Thai** ครั้ง *kʰráŋ,* **Neo-Ahom** *pak,* **Aiton** *thaa1,* **Lungchow** *baːt4,* **Nung** *bạt*

present work, and Table 5 lists some such classifiers in the surveyed languages. Moreover, in some languages, such as Jruq/Loven (Austroasiatic, Bahnaric) *ruoh* 'time', the generic VNC is the only identified VNC that exists in addition to more numerous NNCs. Due to its generic nature, the general VNC may replace other more specific classifiers in some languages, as in Bhaktapur Newar (Joshi 1984, 188).

The following examples illustrate generic VNCs in use. To begin with, Gan Chinese uses *tshi*^{213} 次 as a generic VNC (22).

(22) **Gan Chinese (Sino-Tibetan)**

我去过三次。

ŋo^{34}	*tɕhiɛ*^{44}*-kuo*	*san*^{34}	***tshi***^{213}.
1SG	go-EXP	three	**CLF.time**

'I've been there three times.' (Li 2018, 141)

In a similar fashion, Zoulei, the VNC *vie*^{31} counts the frequency of events (23). Stieng and Korean use *wət* (24) and *-bŏn* (25), respectively.

(23) **Zoulei (Tai-Kadai)**

vlei^{31}	*na*^{13}	*i*^{13}	*lɔ*^{55}	*vəɯ*^{13}	*ʁəɯ*^{31}	*əɯ*^{55}	***vie***^{31}.
year	DEM.PROX	1SG	return	go	home	one	**CLF.time**

'This year, I went home once.' (Li et al. 2014, 129)

(24) **Stieng (Austroasiatic)**

kənɔŋ	*di-cənam*	*ɓaːn*	*bəːh*	*ti*	*psaːr-snual*	*ɓaːn*	*di-**wət**.*
in	one-year	TAM:get	come	to	market-Suol	TAM:get	one-**CLF.time**

'I could come to Snuol market once a year.' (Bon 2014, 128)

(25) **Standard Korean (Koreanic)**

리더는	하루에	백	번	싸운다
lidŏ-nŭn	*haru-e*	*paek*	***bŏn***	*ssau-nda.*
leader-TOP	day-DAT	hundred	**CLF.time**	fight-PRS.DECL

'The leaders fight hundred times a day' (title of a book by Jo U-Sŏng 2019)

Even though the generic VNCs are often used for counting trips and spatial movement, they are not dedicated trip classifiers (Section 4.4), but typically co-occur with a wide set of predicates. In Jiaomuzu Situ Gyalrong, the generic classifier *cʰa* is attested as a counter of various kinds of events, such as marrying (26a) and killing (26b) in addition to trips (26c), highlighting its generic nature.

(26a) **Jiaomuzu Situ Gyalrong (Sino-Tibetan)**
stonmoŋ kəŋes ***cʰa*** *na-ˈa-va-w pəʒər nə-ŋanəŋka-jn.*
wedding two **CLF.time** PRF-NEV-do-3SG again PRF-divorce-3PL
'She married twice and divorced again.' (Prins 2016, 272)

(26b) *kə-**cʰa**=tʃe təwaʔm na-ˈa-sat-w.*
one-**CLF.time**=LOC bear PRF-NEV-kill-3SG
'He killed a bear once.' (Prins 2016, 271)

(26c) *wuɟo-ɲo lhase kə-**cʰa** ji-ˈa-tʃ ʰi-jn.*
3-PL Lhasa one-**CLF.time** PRF-NEV-go$_1$-3PL
'They went to Lhasa once.' (Prins 2016, 271)

A question arises whether a language may possess two or several generic VNCs. In languages such as Japanese, two classifiers show particularly wide usability. The classifiers *-kai* and *-do* show frequent interchangeability, yet a closer inspection reveals that in contrast to *-kai,* which is compatible with any numeral, the functional range of *-do* is mostly limited to small numerals (Kobuchi-Philip 2021, 182, see also Tao 2013 concerning the distribution of the two). Preferring the more generic over the more specific and restricted, *-kai* thus constitutes the generic VNC in Japanese. Similarly, while both ครั้ง *kʰráŋ* and หน *hŏn* (originally meaning 'path, way') can count a wide variety of events in Thai, the former far exceeds the latter in frequency and is consequently considered more basic. In sum, while it is theoretically possible that a language possesses more than one generic VNC, in practice, differences in functional range and frequency can be generally used to argue for the primacy of one form over the others. Due to the limited source materials, however, in some instances, it has not been possible to determine whether one of the several generic VNCs is more basic than the other(s). Detailed future studies of individual languages will cast more light on the issue and help to solve such uncertainties.

In some languages with NNCs, it is possible to specify the referential value of a noun with the choice of a classifier. For instance, Zhang (2013, 400) illustrates this from Ersu where the word *ndzŋ* 'buckwheat' is compatible with a wide array of classifiers, each of which results in foregrounding some distinct semantic aspects of the word, although broadly from the same semantic field. When used with the classifier *-pu* 'plants', *ndzŋ tə-pu* refers to 'a buckwheat plant'. Co-occurring with the classifier *-pa* 'roundish objects', however, *ndzŋ ta-pa* means 'a buckwheat seed'. Other examples by Zhang are offered in (27).

(27) no classifier: *ndzŋ* 'buckwheat'
-pu 'plants' *ndzŋ tə-pu* 'a buckwheat plant'
-pa 'roundish objects' *ndzŋ ta-pa* 'a buckwheat seed'
-tɕʰo 'bundles' *ndzŋ tə-tɕʰo* 'a bundle of (harvested) buckwheat'
-tɕʰo 'pieces of land' *ndzŋ tə-tʂʰo* 'piece of land where buckwheat grows'

A parallel phenomenon is identifiable in some languages with VNCs. In other words, the choice of a classifier with a single verb may specify the semantics of the verb. To illustrate, in Mandarin Chinese, the verb *dǎ* 打 'to hit' has a very general meaning, to the extent that it has grammaticalized into a light verb, as in *dǎ diànhuà* 打电话 'to call', literally 'to hit a phone'. The verb shows compatibility with a multitude of VNCs, each of which foregrounds a certain semantic value from the wide semantic field of the verb. With the generic VNC *cì* 次 'time', as in *dǎ yí cì* 打一次 'to hit once' the meaning remains general. With the classifier *quán* 拳 'hits with fist', *dǎ yì quán* 打一拳 acquires the meaning 'to hit with a fist once'. Additional examples are listed in (28).

(28) no classifier: *dǎ* 打 ‘to hit’
cì 次 ‘generic’ *dǎ yí cì* 打一次 ‘to hit once’ (general)
quán 拳 ‘hits with the fist’ *dǎ yì quán* 打一拳 ‘to hit with a fist once’
qiāng 枪 ‘shots with a gun’ *dǎ yì qiāng* 枪打一枪 ‘to shoot once’
bāzhang 巴掌 ‘slaps with the palm’ *dǎ yì bāzhang* 打一巴掌 ‘to give a slap with the palm’

Finally, in the non-generic VNCs discussed below, a selectional relation holds between the predicate verb and its accompanying classifier. In other words, the use of many non-generic VNCs is controlled by the predicate verb, similarly to the selectional relation between an NNC and the head noun. To illustrate, in Geshiza, the classifier *-ba(r)* ‘CLF.step’ requires a predicate that is a verb of movement, such as *v-tɕə* ‘to take steps’. In contrast, the classifier *-bi* ‘CLF.hit:stick’ only accepts the light verb *v-ra* ‘to hit’ as the predicate.

4.2 Instrumental classifiers type A: body parts as classifiers with ‘verbs of violence’

In many languages of East and Southeast Asia, VNCs exist for counting hits, either with body parts used instrumentally or with an instrument proper. In the first type, a classifier has been formed for hits with various body parts predicated by verbs of violence. This semantic subtype of VNCs can be identified in Tai-Kadai, Hmong-Mien, and Sino-Tibetan. Languages show major variation regarding the subdivisions in this category. To illustrate, Pahari Newar (Sino-Tibetan, Newaric) applies the versatile classifier *-thu* for violence perpetrated with body parts, such as blows, kicks, and scratches (Shrestha 2010, 218). In contrast, Geshiza with a wide spectrum of subdivisions for instrumental VNCs possesses the dedicated classifiers *-rgu* ‘hits fists’, *-rtsʰo* kicks with the feet, *-ltə* ‘hits with the head’, and *-wʈʂœ* ‘clawings with claws’ (Honkasalo 2019, 329).

To illustrate the use of the first subtype of instrumental classifiers, Wadu Pumi counts strikes with the fist involving vertical movement with the VNC *-bi* and slaps with *-l̥á* (Daudey 2014, 146–7) (29).

(29) **Wadu Pumi (Sino-Tibetan)**

tsáw=sə́	*tsàw*	*ʑìŋ*	*wèŋ*		*kʰì=bù,*
pound=CONTR.TOP	pound	can	CUST.EXCL		time=TOP
*sòŋ-**bí**=nòŋ=bù*		*sòŋ-**l̥á***	*tɕə̀*	*qʰù*	*mədzə̀*
three-**CLF.pummel**=COORD=TOP		three-**CLF.slap**	do	need	GNOMIC

‘(…) You can pound, but you will have to thump three times and slap three times (…)’ (Daudey 2014, 147)

Similarly, Khroskyabs counts hits with the palm with the classifier, as in *təlpʰœʁ* (30).

(30) **Khroskyabs (Sino-Tibetan)**

nœ-ɕôŋ=ɕəvœ	*gəvâ=tə*	*cə̂mɲəd*	*ə̂-**təlpʰœʁ***	*o-tsʰóŋ.*
PST-go.II.1SG=CONJ	wife=DEF	like.this	one-**CLF.hit:palm**	PST-hit.II.1SG

‘When I left, I hit my wife with the palm like this.’ (Lai)

Finally, Khatso uses the classifier *tɕo*^{24} to mark the frequency of kicks with the feet (Donlay 2015, 237) (31).

(31) **Khatso (Sino-Tibetan)**

i^{33}	*ŋa*^{33}	*tɛi*^{31}	***tɕo***^{24}	*tɕhɛ*^{55}	*wa*^{323}.
3SG	1SG	one	**CLF.kick**	Kick	PFV

‘He kicked me once.’ (Donlay 2015, 237)

4.3 Instrumental classifiers type B: tools and objects with ‘verbs of violence’

In the second instrumental type, the VNC counts hits by tools and objects. This group resembles group A closely, but the hits occur by means of an instrument proper, not a body part used as an instrument. Examples of common semantic domains include hits or cuts with a knife, strokes with a rod, and gunshots. Stabbing or cutting with a knife in particular constitutes a common type of VNC contained in this semantic category. In Zhaozhuang Bai, the classifier *ta*^{35} counts the number of cuts with a knife (32).

(32) **Zhaozhuang Bai (Sino-Tibetan)**
sɛ^{44} *sa*^{55} ***ta*^{35}**
cut three **CLF.cut:knife**
‘to cut three times with a knife’ (Zhao 2012, 75–88)

In turn, Stau uses the classifier *-ndə* for counting hits with a knife (33). Importantly, while the instrumental classifiers often derive from instrument nouns cross-linguistically, the classifier bears no direct connection with the noun *pərzi* ‘knife’ and remains synchronically opaque.

(33) **Stau (Sino-Tibetan)**
pərzi *a-q*^{h}*a=k*^{h}*œ* ***œ-ndə*** *nə-ɲc*^{h}*ə-sə.*
knife one-CLF.stick=INSTR one-**CLF.hit:stab** PFV-LV.3-IFR
‘S/he hit it with a knife once.’ (personal fieldwork)

Also, gunshots constitute a cross-linguistically common semantic field for the instrumental VNCs. In Xong, the classifier *put* is used for counting the number of gunshots (34). Similarly, Mandarin Chinese uses the classifier *qiāng* 槍 for counting gunshots (35).

(34) **Xong (Hmong-Mien)**
Beul *baond* *aod-ngonl* *nbat-doub* *dox* *bub-**put**,* *beul*
3 shoot one-CLF:animate pig-earth that three-**CLF:gun** **3**
deit *jix* *daos.*
still NEG$_1$ die
‘He shot the wild pig three times (with a gun), but it still didn’t die.’ (Sposato 2015, 278)

(35) **Mandarin Chinese (Sino-Tibetan)**
大林 對著 小 鳥 打 了 四 槍。
Dàlí *duìzhe* *xiǎo* *niǎo* *dǎ* *le* *sì* ***qiāng.***
Dàlín to small bird hit ASP four **CLF.gunshot**
‘Dàlín shot the small bird four times.’ (Zhang 2017b, 629)

Since firearms are a relatively new technology, these classifiers, at least with their current meanings, lack great antiquity. Also, while firearms have largely replaced the bow and arrow as the primary trajectory weapon, we may speculate that historically, classifiers for counting the arrows shot may also have existed, although this remains speculative and not confirmed by any data at the moment.

4.4 Locomotive type

The locomotive type of VNCs measures the frequency of events that involve movement. By far, the most frequently attested classifier measures the number of steps taken when walking, as in (36) with *ʔja:m*^{5} ‘CLF.step’ from Bouyei. In many languages, it constitutes the only instance of a locomotive VNC. In addition,

the frequency of jumps, runs, rounds around an object, trips, and round trips, among others are measured with VNCs of this semantic category.

(36) **Bouyei (Tai-Kadai)**
ku[1] *na:i*[5] *tɯ*[2]*ða:i*[4] ***ʔja:m***[5] *lo*[1] *tu*[3] *pia:i*[3] *mi*[2] *ʔdai*[4] *leu*[4].
1SG tired really **CLF.step** one all walk NEG can already
'I am so tired that I cannot move one more step.' (Burusphat et al. 2001, 559)

When counting trips, languages may make a distinction between simple trips with no regard to return and return trips where the start and end points are identical. Offering an illustration, simple one-way trips are counted with the classifier *thɔŋ*[5] in Jianghua Mien (37a). In contrast, the language applies a distinct classifier *dzuŋ*[6] for return trips (37b).

(37a) **Jianghua Mien (Hmong-Mien)**
miŋ[2] *ji*[2] ***thɔŋ***[5] *ta*[2] *ji*[2] ***thɔŋ***[5]
go one **CLF.one.way.trip** come one **CLF.one.way.trip**
'to go and to come once' (Zheng 2011, 162)

(37b) *ming*[2] *i*[1] ***dzuŋ***[6] *mai*[3] *lɔ*[4] *pwə*[8]
go two **CLF:round.trip** NEG find meet
'to go and come back twice without finding (the person one was looking for)' (Zheng 2011, 162)

4.5 Onomatopoeic type

Onomatopoeic VNCs imitate the sound that results from the action or event of a verb. They exist at least in some languages of the Tai-Kadai family, but among all semantic types of VNCs, they have received by far least attention. In Thai, Noss (1964, 108–9) identifies this group of morphemes as classifiers, and they are also mentioned in other previous research (e.g., Lu 2012, 229–30) for Thai. To illustrate, sounds of laughter in Thai can be counted with the onomatopoeic VNC ฮาก *hâ:k* (38a) while instances of flatulence receive a dedicated onomatopoeic classifier ป้าด *pâ:t* (38b).

(38a) **Standard Thai (Tai-Kadai)**
หัวเราะ สอง ฮาก
hŭarɔ́ *sɔ̌:ŋ* ***hâ:k***
laugh two **CLF.ONOM:laughter**
'to laugh twice' (Lu 2012, 229–30)

(38b) ตด สอง ป้าด
tòt *sɔ̌:ŋ* ***pâ:t***
fart two **CLF.ONOM:fart**
'to fart twice' (constructed)

Some other Thai onomatopoeic words commonly observed in the adjustable event quantification structure are, for instance, เฮือก *hûak* for breathing and attempting to accomplish something, ฟอด *fɔ̂:t* for pecking someone on the cheek, and กรึ๊บ *krúp* for sipping a drink.

4.6 Others

Like in the case of NNCs (Aikhenvald 2000, 273), languages may have specific VNCs attested less frequently in other languages. Such classifiers are placed in the category of 'other' in the present work. However, applying the criteria described at the beginning of this section may warrant establishing new semantic categories as more data becomes available in the future.

To illustrate, Rabha possesses a dedicated classifier *sákho-* for the frequency of bearing offspring in humans and *batri-* for animals (39).

(39) **Rabha (Sino-Tibetan)**
***batri**-sa* *sábra* *pra-a.*
CLF.**bear.offspring.animal**-one child give.birth-PRS
'to bear offspring (animal) once' (Joseph 2007, 462)

Vocalizations, such as calls, shouts, and other productions of noise, frequently appear subject to counting, which may even justify treating them as a semantic subtype of VNCs when more data become available. Vocalizations are illustrated in Khatso (40), Kam (41), and Namuzi (42).

(40) **Khatso (Sino-Tibetan)**
$za^{31}ni^{24}za^{31}$ jo^{35} $ŋ^{31}$ ***tɕi*^{55}** v^{24} wa^{323}.
girl CLF two **CLF.shout** call PFV
'(He) shouted twice (at) the girl.' (Donlay 2015, 237)

(41) **Kam (Tai-Kadai)**
mau^{33} pan^{55} sam^{35} ***so*^{33}**.
3SG call three **CLF.voice**
'He called three times. (lit. He called three voices).' (Gerner 2009, 729)

(42) **Namuzi (Sino-Tibetan)**
Âmî-m û sô-qo luó-ndzhuò, sô-kì ló-gô dzhuò.
ʔa˥mi˥-mu˥ *so˥-**qo**˦* *lu̯o˦˥-ndzu̯o˥˦* *so˥-ki˥˦* *lʷo˦˥ go˥* *dzu̯o˦˧.*
mother-SG three-**CLF.voice** up-call three-CLF.time up-turn hither
'Mother has called [him] three times, [so he] three times turned back.' (Pavlík 2017, 125)

5 Brief diachronic remarks

It remains outside the scope of this pilot study to delve deeply into the history of VNCs. The following illustrates the authors' general interpretation of the collected data.

The question regarding the origin of VNCs connects closely to that of NNCs since the former are dependent on the latter, as explained earlier in this article. In addition, based on the proposed NNC primacy, we argue that NNCs predate VNCs that emerge from the former through analogy. In other words, an existing classificatory device is extended from the nominal into the verbal domain. A hypothesis of polygenesis would postulate that verbal and numeral classifiers emerged in languages of Eastern Eurasia independently from each other. However, as illustrated in Section 4, the great degree of cross-linguistic similarity in the semantics of VNCs hints that such systems result from language contact and have spread through diffusion. Consequently, a hypothesis of monogenesis cannot be ruled out, and it has been offered in earlier research, e.g., Her and Li (2023) and Janhunen (2000).

Classifiers are likely a later innovation in the Sino-Tibetan language family. Shang Chinese, the Sino-Tibetan language with earliest written records as the oracle bone inscriptions, did not have compulsory sortal

numeral classifiers while some 'proto-classifiers', such as 人 'person', and some mensural classifiers (Wang 1994, 77–8, 171–2), including 玉 'a pair of jade', 朋 'a pair of shell', and 丙 'pair of horses' (see also Takashima 1984-1985, 231–3) are identifiable in the language. Consequently, two possible sources emerge for classifier genesis: Hmong-Mien and Tai-Kadai. Classifiers are universal in languages of both language families the languages of which contain both VNCs and NNCs. Classifiers are postulated for Proto-Tai-Kadai (DeLancey 1986). Moreover, in his pioneering comparative work of classifiers, Jones (1970, 11) proposes that Tai languages are the possible original source of the areal phenomenon. Complex classifier systems, however, are also universally attested in Hmong-Mien languages, and the language family thus cannot be ruled out as a possible source of classifier systems without further evidence.

In both language families, VNCs show a wide range of semantic subtypes, but only Tai-Kadai contains onomatopoeic classifiers that were not identified in the analyzed Hmong-Mien languages. The wider semantic range may indicate Tai-Kadai primacy in classifier genesis. It may also, however, result from limited descriptions, since onomatopoeic phenomena that require a high level of mastery in a language are often neglected to various extents in descriptive grammar. The issue of classifier origins becomes further entangled with that of the origin of complex number systems, since classifier systems are dependent on developed numeral systems. Hmong-Mien numerals are borrowed from Sino-Tibetan with the native system comprising only 'two', 'three', and possibly 'one' (Ratliff 2010, 214). While it is possible that borrowing resulted in the replacement of earlier native numerals, Proto-Hmong-Mien may also have been a language with a simple number system (Ratliff 2010, 214). In Tai-Kadai languages, Chinese numerals have been borrowed in the Kam-Tai branch whereas Hlai and Kra maintain the (more) original numerals. It thus seems that proto-Tai-Kadai possessed a more evolved number system fertile for the development of a classifier system.

While VNCs likely originate from Tai-Kadai, Chinese has played a major role in spreading VNCs in languages in Eastern Eurasia in later times. Starting with the Japonic languages, the native VNC 多妣 *taNpyi* 'time' has fallen into disuse save in some fixed expressions, such as *hito-tabi* 'once'. All productive VNCs originate as Chinese loans, such as Standard Japanese 回 *-kai* 'time', 度 *-do* 'time', and 步 *-ho* 'step', corresponding to 回 *huí,* 度 *dù,* and 步 *bù* in Modern Mandarin, respectively. Likewise, the Vietnamese generic *lần* 'times' traces its origins to the Chinese *lún* 轮 'round, turn' (Alves 2007, 223) and the Korean generic 번 *-bŏn* 'time' originates from the Chinese *fān* 番 'turn, time'.

Along the northern and north-western contact zone of Sinitic and Altaic-type languages, particularly the ones spoken in the Qinghai and Gansu areas, some non-Sinitic languages have borrowed their VNCs from Chinese (further discussion on the contact aspects of classifiers in Yurayong et al. 2024). Even Sinitic varieties with generally simple classifier systems, for instance, Wutun (Sino-Tibetan, divergent) with *tang* 'time' *cf.* Mandarin Chinese *tàng* 趟 'CLF.trips', have retained VNCs, as discussed earlier in this study. As for the Altaic-type languages, classifier systems are generally not common, but the counted noun 'time(s)' could have existed among the earliest counting devices. While the function of the limited classifiers across Altaic-type languages is primarily dedicated to counting entities or general verbal events 'time(s)', classifiers for counting more versatile types of verbal events are also observed in a limited fashion, particularly in those languages spoken in the Gansu-Qinghai areas. Salar with limited NNCs may have taken steps toward the emergence of VNCs (43a–43b) and Mangghuer has undoubtedly done so (44). Notice that while Salar uses native Turkic lexical items, the Mangghuer VNC *-zhuan* is a matter a borrowing from Chinese 转 *zhuàn* for counting rotations, and it has been borrowed together with a set of Chinese numerals, as can be seen from *yi-*, which is a Chinese numeral 'one.'

(43a) **Salar (Turkic)**

ebise	*hergune*	*bər*	*gun-de*	*baŋke*	*beɕ*	***kez***	*oqu-ba.*
1PL.EX	everyday	one	day-LOC	prayer	five	**CLF.time**	read-IPFV

'Everyday we read the prayer five times in one day.' (Simon 2016, 296)

(43b)

ergine	*sen*	*bu*	*mani*	*elaŋ-gu-sə*	*muni*	*uɕ*	***elen*** *…*
dawn	2SG	DEM	prayer_wheel	turn-NMLZ-3	DEM.ACC	three	**CLF.turn**

'At dawn, you will turn that prayer wheel three times.' (Simon 2016, 35)

(44) **Mangghuer (Mongolic)**

ni	*zhaler*	*yi-**zhuan***	*mergu*	*nuqi*	*ri-ku, …*
this	hired_farmhand	one-CLF.**circle**	kowtow	pass	come-IPFV

'When this hired farmhand came back (after) kowtowing for one circuit (around the temple), […]' (Slater 2003, 95–6)

In all, as a second great diffusion wave, language contact and borrowing from the Sinitic languages are responsible for the wide distribution of numeral classifiers in Eastern Eurasia, for the striking cross-linguistic semantic similarity in the classifiers, and for the word order change toward the Sinitic model NUM + CLF + N (Szeto and Yurayong 2022, 36, 42).

6 Conclusion

The present study offered a typologically oriented definition for VNCs and addressed the presence of the phenomenon in languages of Eastern Eurasia. VNCs concentrate in Southeast and East Asia, their distribution largely overlapping that of NNCs. The language families where all member languages possess VNCs are Hmong-Mien, Japonic, Koreanic, and Tai-Kadai. VNCs have also emerged in some languages more peripheral to the core distribution area through language contact, such as Mongolic. They are nevertheless absent in language families of the periphery, such as Dravidian and Yukaghir. In turn, the meanings of VNCs show non-random distribution by centering around certain semantic fields, such as a generic classifier for times of occurrence, classifiers for various kinds of hits with either a body part or a tool as an instrument, and classifiers for various types of locomotion, such as trips and steps. Since the Tai-Kadai languages show uniform attestations of VNCs and the existing semantic scopes of available classifiers are generally the widest among these languages, we traced the origin of VNCs to Tai-Kadai, rejecting the alternative hypothesis of Sino-Tibetan origin. Furthermore, it was shown that VNCs frequently spread in language contact situations.

Abbreviations

NNC	nominal numeral classifiers
VNC	verbal numeral classifiers
1	'first verb stem in Jiaomuzu Situ Gyalrong
1	first person; first noun class in Bantu languages
2	second person
II	second verb stem in Khroskyabs
3	third person
4	fourth person
7	seventh noun class in Bantu languages
ACC	accusative
ADD	additive focus
ADJZ	adjectivizer
ADV	adverbial
AF	actor focus
AFF	affiliative
ART	article
ASP	aspect
ASSOC	associative
AUX	auxiliary

CAUS	causative
CLF	classifier
CONTR.TOP	contrastive topic
COORD	coordinator
COP	copula
CUST.EXCL	exclusive customary
DECL	declarative
DEM	demonstrative
DISJ	disjunctive
DST	distal
E	epenthetic sound
EGO	egophoric
EX	exclusive
EXP	experiential perfect
EXV	existential verb
F	feminine
FIN	final
GL	goal
GNOMIC	gnomic
HORT	hortative
IFR	inferential evidential
IMP	imperative
INCL	inclusive
IND	individuator
INDF	indefinitie
INSTR	instrumental
IPFV	imperfective
IRR	irrealis
ITER	iterative
LNK	linking particle, ligature
LOC	locative
M	masculine
NEV	non-direct evidentiality marker
NMLZ	nominalizer
NOM	nominative
ONOM	onomatopoeic
PART	partitive
PFV	perfective
POSS	possessive
PL	plural
PRF	perfect
PROX	proximal
PRS	present
PRS.PERF	present perfect
PST	past
REC	recipient
REFL	reflexive
REP	repetitive
RR	reflexive/reciprocal
S	single participant of an intransitive clause
SEM	semelfactive

SENS	sensory evidential
SG	singular
SLEVL	same (topographic) level
SUFF	suffix
TOP	topic
VBLZ	verbalizer

Acknowledgements: We would like to thank the following scholars for their participation in the survey or for other contributions: Katia Chirkova (Lizu), Juha Janhunen (Mongolic languages), Masato Kobayashi (Kurukh, Malto), Keita Kurabe (Jinghpaw), Yunfan Lai (Khroskyabs), Erika Sandman (Wutun), Pui Yiu Szeto (Guangdian Iu Mien, Western Xiangxi Miao), Ye Xu (Old Korean, Old Japanese), and Juha Yliniemi (Denjongke). In addition, the authors thank the following individuals for sharing their native expertise in the following languages: Andi Juliastri Syafruddin (Indonesian), Hui He (Kam), Ibrahim Mamatazhiev (Uzbek), Ciya Manoj (Malayalam), M.A. Mueed (Bengali), and Mereinur Tabuldin (Kazakh). Finally, we extend our warm thanks and gratitude to Jia Wu (RILCA, Mahidol University) for her help in locating some of the less accessible grammatical descriptions. The authors are solely responsible for the analysis of the linguistic data and any possible mistakes therein.

Funding information: Sami Honkasalo was granted a RILCA fellowship in 2022 and spent approximately 2 months at Mahidol University carrying out the research project. Conferences and workshops where the authors have presented earlier versions of this work were supported by a University of Helsinki research grant for new faculty members ('starttiraha'). Open-access publication of the article is funded by the Helsinki University Library.

Author contributions: Both authors have accepted responsibility for the entire content of this manuscript and consented to its submission to the journal, reviewed all the results, and approved the final version of the manuscript. S.H. is primarily responsible for the theoretical foundations and the semantic typology of verbal numeral classifiers discussed in Chapters 1, 2, and 4. S.H. and C.Y. have jointly collected and analyzed the data presented in this article and discussed in Chapter 3. Based on the data collected by both authors, C.Y. created the database used in this study and employed GIS software for its visualization. Both S.H. and C.Y. contributed to the diachronic analysis presented in Chapter 5.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: All data generated and analysed during this study are included in this published article and its supplementary file.

Institutional review board statement: The study was exempted from ethical review by the Committee for Research Ethics (Social Sciences), Faculty of Social Sciences and Humanities, Mahidol University (2022/001.1701, January 17, 2022).

References

Ahrens, Kathleen and Chu-Ren Huang. 2016. "Classifiers." In *A Reference Grammar of Chinese*, edited by Huang Chu-Ren and Shi Dingxu, 169–98. Cambridge: Cambridge University Press.

Aikhenvald, Alexandra. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra. 2006. "Classifiers and Noun Classes: Semantics." In *Encyclopedia of Languages and Linguistics*, edited by Keith Brown, 463–71. Oxford: Elsevier.

Aikhenvald, Alexandra. 2019. "Noun Categorization Devices: A Cross-Linguistic Perspective." In *Genders and Classifiers: A Cross-Linguistic Typology*, edited by Alexandra Aikhenvald and Elena Mihas, 1–29. Oxford: Oxford University Press.

Aikhenvald, Alexandra and R. M. W. Dixon. 2011. "Non-Ergative Associations between S and O." In *Language at Large: Essays on Syntax and Semantics*, edited by Alexandra Aikhenvald and R. M. W. Dixon, 143–69. Leiden and Boston: Brill.

Alves, Mark. 2007. "Categories of Grammatical Sino-Vietnamese Vocabulary." *Mon-Khmer Studies* 37: 217–29.

Ansaldo, Umberto, Stephen Matthews, and Geoff Smith. 2010. "China Coast Pidgin: Texts and Contexts." *Journal of Pidgin and Creole Languages* 25 (1): 63–94.

Bai, Junwei. 2019. "Numeral classifiers in Munya, a Tibeto-Burman language." In *Genders and Classifiers: A Cross-linguistic Typology*, edited by Alexandra Aikhenvald and Elena Mihas, 282–98. Oxford: Oxford University Press.

Bhaskararao, Peri and Sunder Krishna Joshi. 1985. "A Study of Newari Classifiers." *Bulletin of the Deccan College Research Institute* 44: 17–31.

Bisang, Walter. 2018. "Nominal and Verbal Classification." In *The Diachrony of Classification Systems*, edited by William McGregor and Søren Wichmann, 241–81. Amsterdam: John Benjamins Publishing Company.

Blust, Robert. 2003. *Thao Dictionary*. Taipei: Academia Sinica.

Bon, Noëllie. 2014. *Une Grammaire de la langue stieng: Langue en danger du Cambodge et du Vietnam* [A grammar of the Stieng language: An endangered language of Cambodia and Vietnam]. PhD dissertation, Université Lumière Lyon 2.

Burusphat, Somsonge, Sumittra Suraratdecha, and Guoyan Zhou (eds.). 2001. *Bouyei-Chinese-English-Thai Dictionary*. Nakhon Pathom, Thailand: Institute of Language and Culture for Rural Development, Mahidol University.

Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

Chirkova, Ekaterina. 2009. "Shixing, a Sino-Tibetan Language of South-West China: A Grammatical Sketch with Two Appended Texts." *Linguistics of the Tibeto-Burman Area* 32 (1): 1–89.

Dahl, Östen. 2001. "Principles of Areal Typology." In *Language Typology and Language Universals: An International Handbook*, edited by Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, 1455–70. Berlin: De Gruyter.

Daudey, Gerdine Henriëtte. 2014. *A Grammar of Wadu Pumi*. PhD dissertation, La Trobe University.

DeLancey, Scott. 1986. "Toward a History of Tai Classifier Systems." In *Noun Classes and Categorization*, edited by Colette Craig, 437–52. Amsterdam: John Benjamins Publishing Company.

Donlay, Chris. 2015. *A Functional Grammar of Khatso*. PhD dissertation, University of California, Santa Barbara.

Fradin, Bernard. 2015. "Denumeral Categories." In *Word-Formation: An International Handbook of the Languages of Europe, Vol. 2*, edited by Peter Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, 1424–37. Berlin and New York: Mouton de Gruyter.

Gates, Jesse. 2021. *A Grammar of Mazur Stau*. PhD dissertation, Paris: Ecole doctorale de l'EHESS & Centre de recherche linguistiques sur l'Asie (CRLAO).

Gerner, Matthias. 2009. "Instruments as verb classifiers in Kam (Dong)." *Linguistics* 47 (3): 697–742.

Gerner, Matthias. 2014. "Verb Classifiers in East Asia." *Functions of Language* 21 (3): 267–96.

Gil, David. 2013. "Numeral Classifiers." In *The World Atlas of Language Structures Online*, edited by Matthew Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/55).

Grinevald, Colette. 1999. "Typologie des Systèmes de Classification Nominale." *Faits de langues* 14: 101–22.

Grinevald, Colette. 2000. "A Morphosyntactic Typology of Classifiers." In *Systems of Nominal Classification*, edited by Gunter Senft, 50–92. Cambridge and New York: Cambridge University Press.

Haas, Mary. 1942. "The Use of Numeral Classifiers in Thai." *Language* 18 (3): 203–5.

Her, One-Soon and Bing-Tsiong Li. 2023 "A Single Origin of Numeral Classifiers in Asia and Pacific: A Hypothesis." In *Nominal Classification in Asia and Oceania: Functional and Diachronic Perspectives*, edited by Marc Allassonnière-Tang and Marcin Kilarski, 113–160. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Honkasalo, Sami. 2019. *A Grammar of Eastern Geshiza: A Culturally Anchored Description*. PhD dissertation, Helsinki: University of Helsinki.

Honkasalo, Sami. 2024. "Kazakhstani Gansu Dungan: An Analysis of Language Contact with Russian." *Languages Special Issue "Typology of Chinese Languages: One Name, Many Languages"* 9 (2): 59. doi: 10.3390/languages9020059.

Hölzl, Andreas and Chundra Cathcart. 2019. "Sortal Numeral Classifiers in Central Asia." In *13th Conference of the Association for Linguistic Typology*. Pavia (Italy), 4–6 September 2019.

Hyslop, Gwendolyn. 2017. *A Grammar of Kurtöp*. Leiden and Boston: Brill.

Iwasaki, Shoichi and Preeya Ingkaphirom. 2005. *A Reference Grammar of Thai*. Cambridge and New York: Cambridge University Press.

Janhunen, Juha. 2000. "Grammatical Gender from East to West." In *Gender in Grammar and Cognition, Part II: Manifestations of Gender*, edited by Barbara Unterbeck, Matti Rissanen, Terttu Nevalainen and Mirja Saari, 689–797. Berlin & New York: Mouton de Gruyter.

Jones, Robert. 1970. "Classifier Constructions in Southeast Asia." *Journal of the American Oriental Society* 90 (1): 1–12.

Joseph, UV. 2007. *Rabha*. Leiden and Boston: Brill.

Joshi, Sunder Krishna. 1984. *A Descriptive Study of the Bhaktapur Dialect of Newari*. PhD dissertation, University of Poona.

Kobuchi-Philip, Mana. 2021. "Japanese Numeral Quantifiers that Count Events." In *Numeral Classifiers and Classifier Languages: Chinese, Japanese, and Korean*, edited by Lee Chungmin, Kim Young-Wha, and Yi Byeong-uk, 72–196. London and New York: Routledge.

Konnerth, Linda. 2014. *A Grammar of Karbi*. PhD dissertation, University of Oregon.

LaPolla, Randy and Chenglong Huang. 2003. *A Grammar of Qiang with Annotated Texts and Glossary*. Berlin and New York: Mouton de Gruyter.

Li, Charles and Sandra Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Li, Xia, Li Jinfang, and Luo Yongxian. 2014. *A Grammar of Zoulei, Southwest China*. Bern: Peter Lang AG, International Academic Publishers.

Li, Xuping. 2018. *A Grammar of Gan Chinese: The Yichun Language*. Boston and Berlin: De Gruyter Mouton.

Li, Yu. 2020. *A Grammar of Zauzou*. PhD dissertation, State University of New York.
Liu, Boyang. 2018. "Verbal Classifier Structures in the Wu Dialects of China." In *Syntax of the World's Languages 8*. Paris: Institut National des Langues et Civilisations Orientales. September 3–5, 2018.
Lu, Tian-Qiao. 2012. *Classifiers in Kam-Tai Languages: A Cognitive and Cultural Perspective*. Boca Raton: Universal Publishers.
Lyons, John. 1977. *Semantics*, Vol. 2. Cambridge: Cambridge University Press.
Massupong, Chantra. 1982. *A Grammar of the Tai Dialect of Lung-Chow*. Silpakorn University MA thesis.
Matisoff, James. 1973. *The Grammar of Lahu*. Berkeley: University of California Press.
Matthews, Stephen and T. T. C. Leung. 2004. "Verbal vs Nominal Classifier Constructions in Cantonese and Thai." In *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society*, edited by Somsonge Burusphat, 445–59. Tempe, Arizona: Arizona State University, Program for Southeast Asian Studies.
Nagayama, Yukari. 2003. *Очерк грамматики алюторского языка* [Grammatical outline of Alutor]. Osaka: Osaka Gakuin University.
Noss, Richard. 1964. *Thai Reference Grammar*. Washington, D.C.: Foreign Service Institute.
Paris, Marie-Claude. 2013. "Verbal Reduplication and Verbal Classifiers in Chinese." In *Breaking Down the Barriers: Interdisciplinary Studies in Chinese Linguistics and Beyond*, Vol. 1, edited by Cao Guangshun, Hilary Chappell, Redouane Djamouri, and Thekla Wiebusch, 257–78. Taipei: Academia Sinica.
Pavlík, Štěpán. 2017. *The Description of Namuzi Language/Popis jazyka Namuzi*. PhD dissertation, Charles University.
Prins, Maria. 2016. *A Grammar of rGyalrong, Jiǎomùzú (Kyom-kyo) Dialects: A Web of Relations*. Leiden: Brill.
Ratliff, Martha. 2010. *Hmong-Mien Language History*. Canberra: Pacific Linguistics.
Romaine, Suzanne. 2000. *Language in Society: An Introduction to Sociolinguistics*, 2nd ed. Oxford and New York: Oxford University Press.
Seiler, Walter. 1985. *Imonda, A Papuan Language*. Canberra: Pacific Linguistics.
Slater, Keith. 2003. *A Grammar of Mangghuer: A Mongolic Language of China's Qinghai-Gansu Sprachbund*. London and New York: Routledge Curzon.
Shi, Dingxu. 1991. "Chinese Pidgin English: Its Origin and Linguistics Features." *Journal of Chinese Linguistics* 19 (1): 1–41.
Shrestha, Omkareshwor. 2010. *A Grammar of Pahari*. PhD dissertation, Tribhuvan University.
Simon, Camille. 2016. *Morphosyntaxe et sémantique grammaticale du salar et du tibétain de l'Amdo: analyse d'un contact de langues*. PhD dissertation, Université de la Sorbonne Nouvelle (Paris 3).
Smith, Ian. 1977. *Sri Lanka Portuguese Creole Phonology*. PhD dissertation, Cornell University.
Smith, Ian. 2013. "Sri Lanka Portuguese Structure Dataset." In *Atlas of Pidgin and Creole Language Structures Online*, edited by Susanne Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/41).
Sposato, Adam. 2015. *A Grammar of Xong*. PhD dissertation, State University of New York.
Szeto, Pui Yiu and Chingduang Yurayong. 2022. "Establishing a Sprachbund in the Western Lingnan Region: Conceptual and Methodological Issues." *Folia Linguistica* 56 (1): 25–55.
Takashima, Ken-ichi. 1984-1985. "Noun Phrases in the Oracle Bone Inscriptions." *Monumenta Serica* 36: 229–302.
Tamura, Suzuko. 2000. *The Ainu Language*. Tokyo: Sanseido.
Tao, Ping (陶萍). 2013. "近代小説における助数詞「回」「度」「遍」の一考察 [A study concerning the counter words *kai*, *do*, and *hen* in modern (Japanese) novels]." *The Journal of Cultural Sciences* 630: 882–92.
Thurgood, Graham, Ela Thurgood, and Li Fengxiang. 2014. *A Grammatical Sketch of Hainan Cham: History, Contact, and Phonology*. Boston and Berlin: De Gruyter Mouton.
Tunzhi. 2019. *Outline of Bra'go variety of rTa'u (Horpa)*. PhD dissertation, Victoria, Australia: La Trobe University.
Veselinova, Ljuba. 2013. "Lexicalization of Negative Senses: A Cross-Linguistic Study." *Paper presented at The 10th Biennial Conference of the Association for Linguistic Typology (ALT 10)*. Leipzig University/Max Planck Institute for Evolutionary Anthropology, Leipzig, August 16, 2013.
Vittrant, Alice and Marc Tang. 2021. "Classifiers in Southeast Asian Languages." In *The Languages and Linguistics of Mainland Southeast Asia: A Comprehensive Guide*, edited by Paul Sidwell and Jenny Mathias, 733–72. Berlin and Boston: Walter de Gruyter.
Wang, Lianqing. 1994. *Origin and Development of Classifiers in Chinese*. PhD dissertation, Ohio State University.
Weber, Jean-Jacques and Kristine Horner. 2012. *Introducing Multilingualism: A Social Approach*. London and New York: Routledge.
Wichmann, Søren. 2019. "How to Distinguish Languages and Dialects." *Computational Linguistics* 45 (4): 823–31.
Yurayong, Chingduang, Pui Yiu Szeto, and Sami Honkasalo. 2024 (in press). "Areal Typology of Sinitic in the Temperate Climate Zone: Focusing on Adjustable Quantification and Negation Asymmetry." In *Chinese Language Contact and Typology*, edited by Dan Xu and Cong Wang. (Journal of Chinese Linguistics Monograph Series).
Zhang, Ning. 2017a. "The Syntax of Event-Internal and Event-External Verbal Classifiers." *Studia Linguistica* 71 (3): 266–300.
Zhang, Ning. 2017b. "Verbal Classifiers." In *Encyclopedia of Chinese Language and Linguistics: Volume 1 A–Dǎi*, edited by Rint Sybesma et al., 627–31. Leiden: Brill.
Zhang, Sihong. 2013. *A Reference Grammar of Ersu: A Tibeto-Burman Language of China*. PhD dissertation, James Cook University.
Zhang, Sihong. 2014. "Numeral Classifiers in Ersu." *Language and Linguistics* 15 (6): 883–915.
Zhao, Yanzhen (赵燕珍). 2012. 赵庄白语参考语法 [A reference grammar of Zhaozhuang Bai]. Beijing: China Social Sciences Press.
Zheng, Wuxi. 2016. *A Grammar of Longxi Qiang*. PhD dissertation, National University of Singapore.
Zheng, Zongze (郑宗泽). 2011. 江华勉语研究 [The Mien Language of Jianghua County]. Beijing: Publishing House of Minority Nationalities.