CHAPTER 5

A single origin of numeral classifiers in Asia and the Pacific

A hypothesis

One-Soon Her[1,2] & Bing-Tsiong Li[2]
[1] Tunghai University | [2] National Chengchi University

A world map of more than 700 numeral classifier languages shows a pattern of distribution due to diffusion, where the classifier feature radiates outward from a clear center of clustering in Southeast Asia and classifier languages gradually thin out and with less intensive use of classifiers (Gil 2013; Her et al. 2015). A hypothesis of a single origin seems appealing to account for this particular pattern of distribution. We propose that in Asia and the Pacific numeral classifiers developed indigenously in one language group initially and all other classifier languages acquired this feature via language contact. The hypothesis thus excludes Europe, Africa, Papua New Guinea, and the Americas, pending further research. Based on the evidence available, we argue that, between Tai-Kadai and Sinitic, the latter is more likely the single origin for Asia and the Pacific, and we also consider a possible cause of the rise of classifiers in Sinitic.

Keywords: numeral classifier, measure word, single origin, Sinitic, Tai-Kadai

1. Introduction

Numeral classifiers typically appear with a numeral quantifier. This feature crucially distinguishes them from noun classes, or genders, which form an integral part of a noun, having nothing to do with the presence or absence of a numeral quantifier. Mandarin Chinese is a good example of a language with numeral classifiers. As shown in (1a–b), when a noun is quantified by a numeral, an additional element, known as 'numeral classifier', is required, whereas the English counterparts are formed by just a numeral and a noun.

https://doi.org/10.1075/cilt.362.05her

(1) Numeral classifiers in Mandarin Chinese
   a. *san ge* *nanren*
      3 CLF.GENERAL man
      "three men"
   b. *san ke* *xigua*
      3 CLF.ROUND watermelon
      "three watermelons"

In most classifier languages, the position of the classifier (CLF) can also be filled by a measure word (MENS), as in (2a–b). Her & Lai (2012) distinguish close to one hundred classifiers in Mandarin from a vast number of MENSs. The latter provide extra information in terms of the quantity of the head noun, while a CLF's semantic content overlaps with that of the head noun. A classifier contributes no additional meaning that the noun does not already have; instead, it serves to profile, or highlight, certain intrinsic semantic aspects of the noun (Her 2012: 1673).

(2) Measure words in Mandarin Chinese
   a. *san da* *pingguo*
      3 MENS.DOZEN apple
      "three dozens of apples"
   b. *san xiang* *pingguo*
      3 MENS.BOX apple
      "three boxes of apples"

The concept of measure words in the semantic sense is rather ordinary in all languages. However, in non-classifier languages such measure words are typically nouns and do not form a separate syntactic category. In classifier languages such as Chinese, however, such measure words belong to a distinct syntactic category with classifiers. The two subcategories, classifiers and measure words, are also known as 'sortal classifiers' and 'mensural classifiers' (e.g., Huang 2013; Saikia & Allassonnière-Tang, this volume), and 'classifiers' and 'massifiers' (e.g., Cheng & Sybesma 1998), respectively, among various other lesser-known terms. In this chapter we shall refer to the two as classifiers and measure words, and refer to the syntactic category formed by the two subcategories as C/M in short and 'numeral classifiers' in full.

The use of numeral classifiers, or C/Ms, has long been considered a prominent areal feature of languages in East and Southeast Asia (e.g., Greenberg 1990 [1974]; Bisang 1999; Aikhenvald 2000: 121). Jones (1970) and a number of other scholars have suggested that this areal feature is due to areal diffusion with Tai as the original center, though the dominant view among linguists in China is that C/Ms developed independently first in Sinitic languages (e.g., Wang 1994). Such characterizations have often led to an earlier misconception that C/Ms do

not occur in European languages. For instance, Rijkhoff (1998: 328) states "sortal classifiers are generally absent in the European languages". However, as several surveys demonstrate – most notably Greenberg (1990 [1972]), Aikhenvald (2000), and Gil (2013), each covering 103, 116, and 140 classifier languages, respectively – classifier languages also exist in the Pacific islands, Europe, and all other populated continents, except Australia.[1,2] Our first goal is to introduce a geographic information system (GIS) database of 713 numeral classifier languages in the world. Based on the distribution pattern of these 713 languages on a world map, a pattern emerges that is indeed rather similar to those in a diffusion scenario. Our second, more important, goal is to explore a bold and yet appealing hypothesis that there is a single origin of numeral classifiers in Asia and the Pacific.[3] The idea of a single origin is not entirely new, as it has been hinted at in the literature for the classifier languages in a particular area of the Asian continent. Janhunen (2000: 705) is rather explicit in making such a proposal.

> A feature like the numeral classifiers in East Asia can hardly have originated separately in several adjacent languages and language families. Rather, there was a single primary innovation where the principle was first created, and from where it radiated to all over the region. [...] In any case, there is no specific reason to assume that it was Chinese that was the first class language in Asia. For this particular sphere of linguistic phenomena, like for some others, the role of Chinese may well have been that of a transmitter, rather than that of an innovator.

While Janhunen talks only about East Asia, we expand the scope to Asia and the Pacific. Furthermore, though Janhunen (2000: 705) sees no specific reason to favor Sinitic as the original innovator of numeral classifiers in Asia, he offers

---

1. For European languages in our own database, numeral classifiers are found in these Indo-European languages: Breton, Bulgarian, Irish Gaelic, Polish, Russian, Standard German, and Swabian, as well as Hungarian (Uralic) and Crimean Tatar (Altaic); however, we are fully aware that the inclusion of some of these languages, e.g., Russian, German, and Polish, is contentious. For example, Polish is said to have only one numeral classifier, i.e., *sztuka* "piece" according to Sussex & Cubberley (2006: 314–315).

2. There is in fact one Australian language identified as a numeral classifier language, Anindilyakwa of the Northern Territories (van Egmond 2008), whose classifiers are likely due to contact with Makasar, an Austronesian language of South Sulawesi.

3. Note that this hypothesis does not apply to sign languages. Though it is well-documented that many sign languages have 'classifiers', to the best of our knowledge, no sign languages have 'numeral classifiers' per se, which is the focus of this hypothesis and also the current paper.

no better alternatives.[4] We take the view that this grammatical feature most likely originated in either Tai-Kadai or Sinitic and further argue that Sinitic enjoys a slight advantage given the evidence available. Other language groups, including Tai-Kadai, acquired this feature via contact. We should point out that this single origin hypothesis is meant to be just that, a hypothesis, one that competes with a number of other possible hypotheses where the rise of numeral classifiers in the language or language groups in question is due to factors independent of language contact. For example, while some groups may have acquired the feature by contact, others may have developed it internally without external influence. Therefore, the purpose of this chapter is not necessarily to *prove* the hypothesis; rather, the primary goal is to demonstrate that the hypothesis does offer a viable scenario for the rise of numeral classifiers in languages in the Asian continent and the Pacific.[5]

The organization of this chapter is as follows. Section 2 introduces a GIS database of 713 numeral classifier languages and the language cluster of SMATTI (an acronym for Sinitic, Miao-Yao, Austroasiatic, Tai-Kadai, Tibeto-Burman, and Indo-Aryan) as a hotbed of numeral classifiers, and presents our motivations for proposing a single origin of numeral classifiers from within SMATTI. We then discuss in Section 3 how several major language families or groups acquired this feature via contact. In Section 4, we deliberate between two candidates for this single origin, Sinitic and Tai-Kadai, and ultimately argue for Sinitic as the more likely original source, and further propose a scenario for the rise of numeral classifiers in Tai-Kadai under the influence of Sinitic. Section 5 consists of some concluding remarks.

2. SMATTI and the single origin hypothesis

In this section, we demonstrate that the proposal for a single origin of numeral classifiers in Asia and the Pacific is motivated by the pattern of distribution of classifier languages in the region. Our current knowledge would rule out genetic relationships as the historical explanation and also disfavor convergent multiple

4. We would certainly have no objection to Janhunen's skepticism towards Sinitic as the innovator if "this particular sphere of linguistic phenomena" in the quote refers to nominal classification in general and not numeral classifiers in particular.

5. In that spirit, in future research we should further look into other structural features besides classifiers and see if some of them can be attributed to a similar contact scenario. Given the intimate relation between numerals and numeral classifiers, in the present paper we will touch upon the numeral systems where such information is available as evidence for the hypothesis.

origins; we contend that diffusion, including stimulus diffusion, best accounts for this regional distribution.

2.1 Distribution of classifier languages in the world

To our knowledge, the earliest attempt at a survey of the world's classifier languages is Greenberg (1990 [1972]), where 103 classifier languages were investigated. A more recent survey by Gil (2013) identifies 140 classifier languages among 400 languages, available online at WALS (*The World Atlas of Language Structures*, http://wals.info), and is the largest published database for classifier languages so far. A significant advantage of Gil's (2013) database over that of Greenberg (1990 [1972]) is that it is incorporated in the geographic information system in WALS and can easily display the global distribution of these languages; see Figure 1. One can already recognize that rather than being randomly distributed, classifier languages cluster in several areas in the world. The most crowded area is in East and Southeast Asia, with classifier languages extending into the West Pacific, and other smaller clusters are found in West Africa, the Middle East, and the Americas.

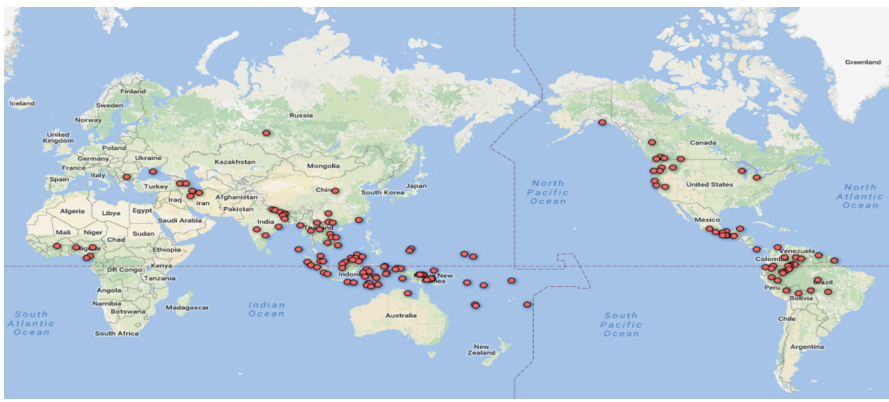


Figure 1. 140 classifier languages in the world (Gil 2013)[6]

In order to obtain better insight into the typology of C/M word orders in the nominal construction consisting of a numeral (Num), C/M, and N, the first author assembled a research team in 2012 and has since been constructing a database of classifier languages in the world. Entries of languages and genetic relations of the languages in this database are based on *Ethnologue* (Lewis et al. 2009). *Eth-*

6. Figures 1, 4, and 5 were generated with Google Fusion Table (Halevy & Shapley 2009).

nologue was chosen because of its completeness and the geographic information available in its World Language Mapping System (WLMS) (Lewis et al. 2016).

Data for numeral classifiers as well as numerals in each language were then collected via a survey of existing literature. A language is marked as a classifier language if an author explicitly states that it is a numeral classifier language or certain elements in their data are clearly and appropriately marked as numeral classifiers. There are of course cases of disagreement in the literature, where we must exercise our own discretion.[7] Neither the size of the classifier inventory nor the issue of whether classifiers are optional or obligatory is an essential consideration. An attested classifier language must have Cs, or sortal classifiers, which are the additional classificatory words or morphemes in numerical enumeration of nouns of high countability (Aikhenvald 2000; Seifart 2010; Gil 2013). In addition, as we will soon discuss in Section 2.2, such elements must form a multiplicative relation with the adjacent numeral and represent the value of *1* as the multiplicand (Her 2012). We first looked into the three major surveys mentioned above, i.e., Greenberg (1990 [1972]), Aikhenvald (2000), and Gil (2013) for classifier languages and Chan (2017) for numeral systems in languages. A sweep from language family to language family was then performed until each of the 7,000 plus languages listed in *Ethnologue* (Lewis et al. 2009) was looked into at least twice.[8] An ongoing project, the database currently has data on 713 classifier languages, the largest such database so far, to our knowledge.[9] With information for each language, including its coordinates, provided by WLMS, we are also able to show the distribution of the 713 classifier languages on a world map from Google.

An even clearer picture emerges from Figure 2, showing the dense clustering of classifier languages in East and Southeast Asia, where the majority of such languages belong to six language groups: **S**initic, **M**iao-Yao (Hmong-Mien), **A**ustroasiatic, **T**ai-Kadai, **T**ibeto-Burman, and **I**ndo-Aryan languages, hereafter referred to as SMATTI, an acronym proposed by Her et al. (2015). Figure 3 shows the distribution of SMATTI languages in *Ethnologue* (Lewis et al. 2009).

In our database, 310 SMATTI classifier languages have been identified, which account for nearly half of the classifier languages in the entire database. SMATTI is undoubtedly the hotbed of numeral classifiers in Eurasia, and indeed the logical place where we start our journey in search for an origin of numeral classifiers.

---

7. For example, Her (2017b) demonstrates that the several putative classifiers in Ejagham cited in Watters (1981) are in fact nouns.

8. We have since moved to the Glottolog database (Hammarström et al. 2021).

9. The database has been named 'The World Atlas of Classifier Languages', WACL in short. See Her et al. (2022) for an overview and its online release.

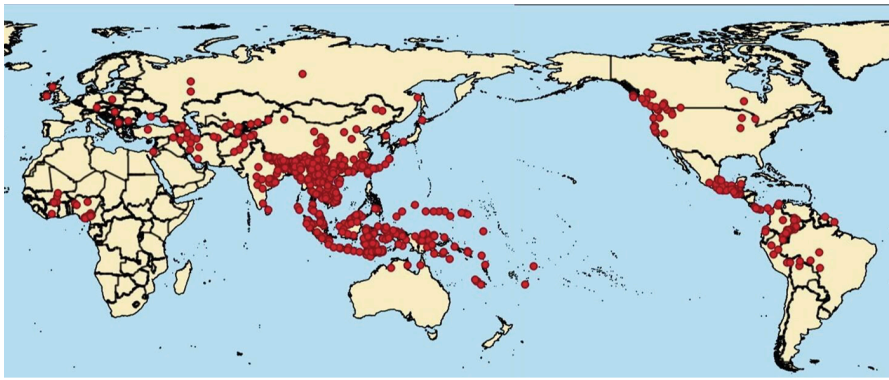


Figure 2. 713 classifier languages in the world[10]

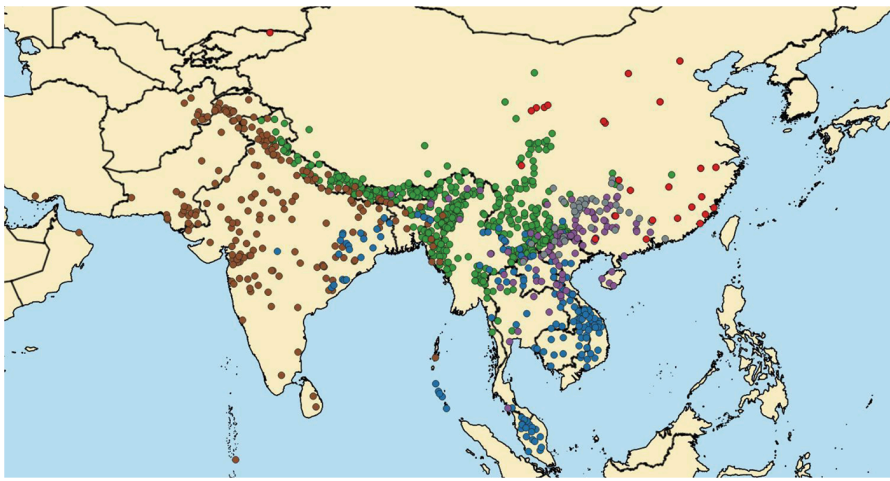


Figure 3. Distribution of SMATTI languages (red dots for Sinitic, gray for Miao-Yao, blue for Austroasiatic, purple for Tai-Kadai, green for Tibeto-Burman, and brown for Indo-Aryan)

According to *Glottolog*, SMATTI has a total number of 1001 languages, about 12.4% of the world's languages. See Table 1 for a summary.

Beyond SMATTI, classifier languages and the use of numeral classifiers thin out in all directions. Classifier languages are found in Western Austronesian, in Archipelago Southeast Asia, Micronesia, and Melanesia. However, there are no signs of classifier languages further eastward into the remote Pacific islands. The

10. Figures 2, 3, 6, 7, 8, 9, and 11 were generated with QGIS; see https://www.qgis.org/en/site/.

Table 1. Numbers of languages and classifier languages in SMATTI

Language group	Number of languages	Number of classifier languages	Percentage of classifier languages
Sinitic	26	19	73.1%
Miao-Yao	39	14	35.9%
Austroasiatic	159	58	36.5%
Tai-Kadai	94	50	53.2%
Tibeto-Burman	465	144	31.0%
Indo-Aryan	218	25	11.5%
Total	1001	310	31.0%

boundary of classifier languages in the Pacific seems to coincide with the International Date Line. To the north of SMATTI, there are classifier languages such as Korean and Japonic languages in the east, but in the west, only a few Altaic classifier languages can be found within the borders of China, but not further into Siberia.[11] In the south, some Dravidian classifier languages are found, but they are certainly not as abundant as their neighbors in the north. Classifier languages in the west also exhibit this pattern of distribution. While Indo-Aryan is one of the SMATTI language groups, its classifier languages are mostly found in the east of the subcontinent. Further to the west, the number and density of classifier languages decrease drastically in the Middle East and finally into Europe, where classifier languages are at best sporadic, and with very limited inventories of classifiers.

This distribution of classifier languages in Asia and the Pacific fits the scenario of diffusion. When an innovative feature occurs and continues to develop, it often results in a higher degree of variation and more intensive use of the feature. During its development, it may also diffuse to neighboring languages, and the intensity and diversity drop down gradually due to partial borrowing or shorter periods of development (e.g., Trudgill 1974; Aikhenvald & Dixon 2001; Cathcart et al. 2018). The distribution of classifier languages seems to match such a diffusion scenario; therefore, we propose a hypothesis of a single origin, somewhere in the center of this hotbed, of the rise of numeral classifiers within the languages in this area.

11. Given the controversy over whether the five language groups, Mongolic, Tungusic, Turkic, Japonic, and Koreanic, form a unified language family, we use the term 'Altaic' rather loosely and make no commitment to the language family.

2.2 A single origin within SMATTI

Another motivation for proposing a single origin from inside SMATTI is the clear signs of contact-induced classifiers in Tibeto-Burman (TB). Before introducing the Tibeto-Burman phenomenon, we need to introduce a multiplicative theory as a preliminary, which unites C/M with numeral bases in the numeral system from a mathematical perspective (Her 2012, 2017a, 2017b). Consider a numeral system with both addition and multiplication, for which Comrie (2006) gives a concise formulation, as in (3).

(3) General Pattern of Numeral Expressions (Comrie 2006)
$(n \times base) + m$, where $m < base$

A more detailed description is offered in Comrie (2013):

> By the "**base**" of a numeral system we mean the value *n* such that numeral expressions are constructed according to the pattern ... $xn+y$, i.e., some numeral *x* multiplied by the base plus some other numeral. (The order of elements is irrelevant, as are the particular conventions used in individual languages to indicate multiplication and addition). A simple example is provided by Mandarin, with base 10, in which the numeral 26 is expressed as in (1).
>
> (1) Mandarin
> *èr-shí-liù*
> two-ten-six
>
> In Mandarin, the convention is that the numeral before the word for 10 is to be multiplied by 10, while that after the word for 10 is to be added to this product ($[2\times10]+6$). (Comrie 2013)

Since the relation between the *n* and the *base*, e.g., *èr* "two" and *shí* "ten" in Mandarin is multiplication, the linear order between the two is irrelevant and does indeed vary across languages.

As for numeral classifiers, as mentioned above, the position where a classifier appears can also be filled by a measure word. How to distinguish between classifiers and measure words has been a contentious issue (e.g., Hsieh 2008; Her & Hsieh 2010). Her (2012) proposes a precise multiplication-based account for this distinction, synthesizing the insights gained from Landman (2004), Au Yeung (2005, 2007), and Yi (2009), i.e., [Num C/M] can be seen mathematically as [*multiplier* × *multiplicand*], exactly like the [*n base*] relation in a numeral such as *èr-shí* (two-ten) "twenty", as demonstrated in (4a–b) for classifiers and (5a–b) for measure words.

(4) Multiplication in a Mandarin Chinese classifier construction
   a. *san ge* *pingguo* (Num × CLF = 3 × 1)
      3 CLF.GENERAL apple
      "three apples"
   b. *san ke* *pingguo* (Num × CLF = 3 × 1)
      3 CLF.ROUND apple
      "three apples"

(5) Multiplication in a Mandarin Chinese measure word construction
   a. *san da* *pingguo* (Num × MENS.dozen = 3 × 12)
      3 MENS.DOZEN apple
      "three dozen apples"
   b. *san xiang* *pingguo* (Num × MENS.box = 3 × box)
      3 MENS.BOX apple
      "three boxes of apples"

From this mathematical perspective, classifiers and measure words are unified under the function as multiplicands and also distinguished in terms of their respective values, i.e., a classifier must have the precise numerical value of *1*, but a measure word must not. Measure words can have any value, numerical or non-numerical, except *1* (Her 2012). In other words, a measure word always implies an aggregate rather than an individual. Classifiers and measure words thus form the two subcategories of a single syntactic category, C/M. Given the fact that the mathematical function of [Num C/M] is exactly the same as [*n* × *base*], i.e., [*multiplier* × *multiplicand*], it is to be expected that, within the same language, the two should behave the same in terms of their relative order with the multiplier. This effect is called 'harmonization' by Greenberg (1990 [1978]: 293). Extending Greenberg's initial observation, Her et al. (2015), Her (2017a, 2017b), and Her et al. (2019) propose explicitly that there is harmonization in word order between C/M and *base* within a nominal phrase composed of Num, C/M, and N, as stated in (6), where C/M-final means [Num C/M], and C/M-initial, [C/M Num]; likewise, *base*-final means [*n base*], and *base*-initial, [*base n*]. C/M and *base* are seen as heads due to their function as the multiplicand.

(6) Harmonization in word order between C/M and *base* (Her 2017a: 43)
   a. C/M-final ⇒ *base*-final
   b. C/M-initial ⇒ *base*-initial

The insight is that a numeral with an internal [*n base*] order must precede C/M, or [Num C/M], and a [*base n*] numeral must follow C/M, or [C/M Num]. The consequence is that, cross-linguistically, among the three elements, *n, base*, and C/M, there can be only two expressions: [*n base* C/M] and its mirror image, [C/M *base n*]; the four other mathematically possible orders are all predicted to be ill-

formed (Her 2017b). This also explains why N does not intervene between Num and C/M.

Her et al. (2015) demonstrate that such *base*-C/M harmonization is indeed substantiated in SMATTI, i.e., out of 232 classifier languages, of which we obtained data for 218 languages, harmonization is observed in 213 languages, about 91.8% (213/232); see Figure 4 for their distribution. Specifically, 186 languages have *base*-final and C/M-final word order, and 27 others, *base*-initial and C/M-initial.

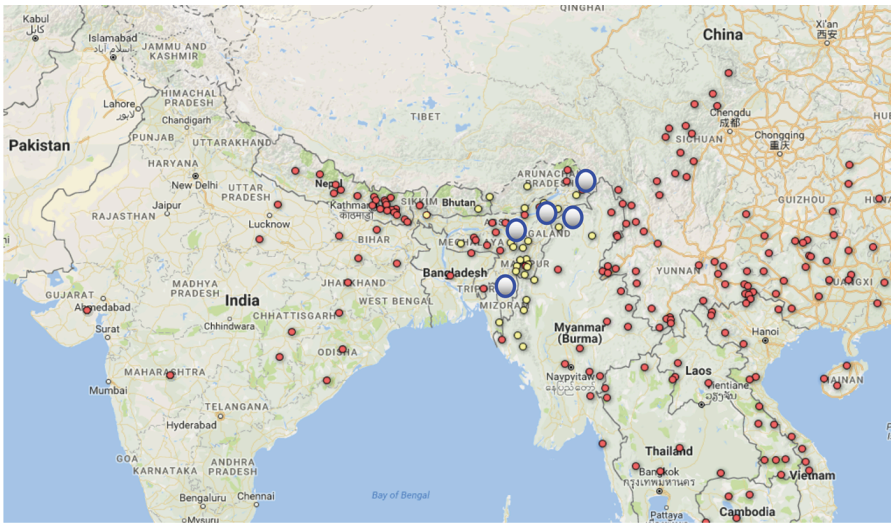


Figure 4. Classifier languages in SMATTI (red dots for *base*-final, yellow dots for *base*-initial, blue circles for violations)

Interestingly, the 27 *base*-initial, C/M-initial languages, along with the five exceptions, are all Tibeto-Burman. Yet, far more Tibeto-Burman languages are found to be *base*-final and C/M-final; 66 to be exact. The fact that TB is the only language group in SMATTI where *base*-initial numerals are found suggests quite strongly that such numerals are remnants from Proto-Tibeto-Burman (PTB). Indeed, Kiryu (2009: 64–66) hypothesizes that PTB has a numeral system that is *base*-initial, i.e., [*base n*],[12] while C/Ms as well as the [*n base*] order were due to

12. Note that on p. 65 of Kiryu (2009), the second of the three assumptions of (11) states: "The Proto-Tibeto-Burman order of numeral and quantifier was Num-Q." Num-Q here is a typo, which is [*n base*] in our terminology. It has been confirmed with the author via personal communication that the intended order was Q-Num, i.e., [*base n*], instead. This makes better sense since on p. 64 he correctly states that CLS-Num is the default order in Tibeto-Burman.

influence from neighboring language groups with C/Ms, which are all *base*-final and also C/M-final. This view is consistent with Fu's (2015: 50) position that the default order between N and Num in TB is [N Num], especially with Num seen as a complex unit composed of a C/M and a numeral.

Currently, however, roughly 19.8% (92/465) of TB languages are *base*-initial, while 37.6% (175/465) are *base*-final, and 4.9% (23/465) utilize both word orders. Although we still lack data on numerals for 37.2% (173/435) of TB languages, the biased distribution already indicates strong influence from contact with *base*-final languages. Now consider that out of 465 TB languages, only roughly 144, i.e., 31%, are found to have classifiers. This suggests areal diffusion even within TB. However, the borrowing of C/Ms into a *base*-initial TB language from a *base*-final, C/M-final non-TB neighbor is not a straightforward matter, given the *base*-C/M harmonization requirement. The concept of stimulus diffusion is thus necessary for the development of a classifier system in terms of areal diffusion in TB (Weidert 1984; Kiryu 2009: 66), where the idea of a system, but not necessarily its concrete content, is accepted by the recipient culture (Kroeber 1940: 1). Therefore, the *base*-initial TB languages must have adopted the idea of classifiers upon contact with *base*-final and C/M-final languages, and developed a C/M-initial system at first without taking on the word order in these languages. The change of word order of base and C/M from initial to final must come at a later stage of development.

Further issues related to Tibeto-Burman will be discussed in greater detail in Section 3.3. For now, evidence from Tibeto-Burman indicates that a grammatical device such as classifiers can be borrowed, even without its word order. This is a hint of the possibility that if we examine each language group within SMATTI, we may end up finding chains of borrowing, all linked to a single origin within. However, before we examine each language group in SMATTI and beyond and dig into their history of language contact, a few caveats should be noted about this hypothesis.

2.3 Current scope of the single origin hypothesis

As noted already, we limit our scope to Asia and the Pacific and exclude languages in the Americas, Africa, Europe, and Papua New Guinea from this single origin hypothesis. Classifier languages in the Americas are mainly distributed along the west coast, with only a few exceptions around the Great Lakes and in the Amazon basin. Such a distribution matches the migration and dispersal route out of Asia across the Bering Strait, southward through North and Central America, eventually into South America. The correspondence between the distribution of classifier languages and the migration route across the Americas does seem to suggest

that classifiers on this continent are genetically passed down since early migration into North America. Nichols (1992: 251) made the observation based on the data of only nine classifier languages of the New World:

> The distribution of numeral classifiers, a hotbed phenomenon for which there is only one clear hotbed and it lies along the circum-Pacific migration route, strongly suggests that *entrants to the New World came from a single population characterized by high frequency of numeral classifiers* (except apparently, for the earliest entrants, whose descendants are now mostly in southern South America and eastern North America, where numeral classifiers are rare).
>
> (Nichols 1992: 251, emphasis added)

This is certainly an attractive theory, especially from our standpoint. However, it is difficult at present to integrate our hypothesis of a single origin from within SMATTI into Nichols' theory. The most obvious difficulty lies in the time depth of Asian migration into, and dispersion across, the Americas, which is estimated to be around 16500–13000 years B.P. (e.g., Goebel et al. 2008). For Nichols' theory to stand, the people from Siberia who were ancestors of Native Americans should already have classifiers in their languages, but we simply have no evidence for that. We do not have any knowledge of the languages in Asia that early; furthermore, our survey has not turned up any classifier languages in that region. We thus must leave out the Americas from the single origin hypothesis for now.

Classifier languages in Africa are mainly Niger-Congo languages in and around Nigeria, with some Afro-Asiatic languages nearby. These classifier languages are too far apart from numeral classifier languages of other parts of the world. Also, in between these languages and SMATTI is a great number of Afro-Asiatic, Nilo-Saharan, and Niger-Congo languages without classifiers. It is therefore quite improbable that contact with Asian languages has played any meaningful role in the rise of numeral classifiers in Africa. On the other hand, Niger-Congo languages feature the use of noun classes, where a certain prefix is attached to the noun stem according to the class of the noun, regardless of whether numerals are used. It is much more likely that numeral classifiers in African languages arise by themselves from the noun class system.

Another region that we leave out is Papua New Guinea, though there is clear evidence suggesting that at least some PNG languages developed classifiers under the influence of Austronesian classifier languages (Klamer 2014: 160), which may be in turn the result of contact with SMATTI. If that is indeed the case, PNG languages are then in the circle of influence of our proposed single origin. Although classifiers in Austronesian will be discussed in Section 3, we rule out PNG languages partly because our data for PNG classifier languages are still rudimentary, but mainly because studies regarding linguistic prehistory of this area are

relatively scarce, and currently no inference regarding borrowing of classifiers can be drawn.

As for Europe, even though there are many classifier languages within Indo-European, especially in the Indo-Aryan and Iranian branches, the handful identified that are located in the European continent remain controversial. These include Breton, Bulgarian, Irish Gaelic, Polish, Russian, Standard German, and Swabian within Indo-European and Crimean Tatar (Turkic (Altaic)) and Hungarian (Uralic). Further research is required to firmly establish the proper status of the putative classifiers in each of these languages before we can look into the history of genuine classifiers in Europe.

2.4 Competing candidates for the single origin

The two most likely candidates for a single origin are Tai-Kadai and Sinitic, the two groups that enjoy the highest percentage of classifier languages, also the only two candidates debated in the literature as being the origin of the areal diffusion of the classifier feature on the Asian continent,[13] though the reconstruction of a linguistic feature to the proto-language does not require the presence of the feature in all extant languages in the language family. Sinitic languages are all attested classifier languages, although the range of syntactic functions that classifiers serve varies from one language to another within the group, and so do the sizes of the inventories of classifiers. In general, classifiers in southern Chinese languages have a wider range of syntactic functions than those in the north like Mandarin. See examples (7–9):

(7) Classifier in a bare classifier phrase [CLF N] to indicate definiteness
  a. *zek gau soeng gwo maalou* (Cantonese, Cheng & Sybesma 2005: 9)
  CLF dog want cross road
  b. *(na) zhi gou xiang guo malu (Mandarin)
  DEM CLF dog want cross road
  "The dog wants to cross the road."

(8) Classifier with an attributive to form a nominal phrase
  a. *sang ke* (Hakka, Sixian)
  raw CLF
  "raw thing(s)"

13. A language group with this feature in its proto-language is of course a better candidate for the origin, and a language group can have this feature reconstructed to the proto level if all its member languages share this feature.

b. *sheng* *ge/de (Mandarin)
raw *CLF/DE[14]
"raw thing(s)"

c. *qidga ke* (Hakka, Sixian)
self CLF
"thing(s) of oneself"

d. *ziji* *ge/de (Mandarin)
self *CLF/DE
"thing(s) of oneself"

(9) Classifier as a modifier marker

a. *louzisam tiu simzoeng* (Cantonese, You 1982: 38)
L. CLF Buddhist-stick

b. *luzhishen* *ge/de *chanzhang* (Mandarin)
L. *CLF/DE Buddhist-stick
"Luzhishen's Buddhist stick"

Larger inventories of classifiers are also found in the south than in the north. Besides the percentage, Chinese also shows signs of possible C/Ms as early as 3500 B.P. in its earliest written records, 甲骨文 *Jiaguwen* (oracle bone inscriptions) (e.g., Wang 1994). This is the oldest written record of a language with signs of C/Ms and this time depth gives Sinitic a practical advantage in being considered as the origin of C/Ms. However, we need to acknowledge that this is not in itself evidence for origin.

On the other hand, at least 50 out of 94 Tai-Kadai languages are attested to have classifiers; for the remaining 44 languages there is insufficient data for us to determine whether they have classifiers. Tai, with 76 languages, is the largest branch in Tai-Kadai and has currently 39 attested classifier languages in our database. While the percentages, 51% (39/76) in Tai, 61% (11/18) in Kadai, and 53% (50/94) overall, seem much lower in comparison with Sinitic, we believe it is due to the lack of data or our lack of access to some of the published data of these languages. In all the literature we have surveyed, not one Tai-Kadai language is referred to as not having C/Ms, and we are fairly confident that further investigation will reveal that languages in Tai-Kadai, like Sinitic, all have C/Ms. Morev (2000: 76, 81) in fact claims that all Tai languages have numeral classifiers, a view confirmed by Somsonge Burusphat (p.c.), and a similar claim for Kadai based on the current 65% already identified is certainly not far-fetched.[15] Tai-Kadai lan-

14. *De* is a functional element marking the preceding phrase, e.g., an adjective or a relative clause, as a modifier.

15. We do not expect to have the same issue in the data on other language groups, given what is already well-known and well-established in the literature, even though there are surely classifier

guages also have the most intensive use of classifiers, very similar to the southern Chinese languages (e.g., You 1982). They also have large inventories of more than forty classifiers, a feature that also characterizes southern Chinese languages (e.g., Peyraube 1991; Yue 2003; Cao 2008). Thai and Northern Zhuang, for example, have 80–90 (Haas 1942) and 52 classifiers, respectively (estimated from Burusphat & Qin 2006), meanwhile southern Chinese languages like Southern Min and Hakka have 83 (Chen 2013: 65) and 64 (Liao 2014: 63), respectively.

Regarding the time of emergence of Tai-Kadai classifiers, the earliest written records of the Tai-Kadai people date back to the 13th century. Though C/Ms were already an established grammatical feature in those scripts (Huffman 1973: 505), there is no way from written records alone to demonstrate a time of emergence of Tai-Kadai classifiers. However, considerations of genetic relationships seem to favor Sinitic over Tai-Kadai, though the issue is not yet settled, as indicated in the following quote:

> If we follow the predominant point of view that the Tai languages originally belong to the Austronesian family, then it would look more probable that Tais were recipients and Chinese donors. It comes from the fact that in modern [Austronesian] languages there is nothing that speaks for the existence in former times of a well developed category of classifiers so that they were able to share classifiers with Chinese. On the other hand, if we still agree with the former views about Chinese-Tai relationship, the question who is the donor and who is the recipient no longer arises. So, the issue is still under investigation. (Morev 2000: 78)

We will return to this issue in Section 4 and further compare the two candidates, ultimately arguing for Sinitic, but we should first demonstrate that it is a viable claim that other language groups besides Tai-Kadai and Sinitic in and around SMATTI acquired numeral classifiers due to direct or indirect contact with these two language groups.

3. Evidence of language phyla borrowing classifiers

In this section we consider other language families in the world and demonstrate that none of them is likely to have inherited the classifier feature genetically from an earlier ancestral language, and a much more plausible explanation is indeed areal diffusion via language contact, directly or indirectly, with Tai-Kadai and/or Sinitic. First we examine language families in SMATTI in the order of the

languages in other language groups that are yet to be identified, and conversely, some classifier languages identified may turn out to be misidentified.

acronym, that is, Miao-Yao, Austroasiatic, Tibeto-Burman, Indo-Aryan, before which we will briefly discuss the relevance of Dravidian. Then we will look at Austronesian, also rich in classifier languages and close to SMATTI, before other languages, including Japonic, Korean, and Altaic languages.

3.1 Miao-Yao (Hmong-Mien)

There are 39 Miao-Yao languages, mostly spoken in the mountainous areas of southern China, northern Laos, and northern Vietnam. Situated between Sino-Tibetan, Tai-Kadai, and Austroasiatic families, the distribution of Miao-Yao languages is rather sporadic, but can still be categorized into Miao (Hmongic) in the west and Yao (Mienic) in the east, with the She language (畲語) being an outlier. So far, we have confirmed that at least 14 are classifier languages. Again, we suspect that many more will be identified upon further research, as the accepted view is that classifier systems are a common Miao-Yao feature. Bu (2011a) summarizes the grammatical functions of classifiers and the size of classifier inventory, claiming specifically that the intensity of classifier use is the greatest in Tai, where a large inventory of classifiers serves many syntactic functions, is less in Miao-Yao, whose classifiers have a comparable size of inventory but a more restricted syntactic position, and ultimately is the least in Tibeto-Burman, where classifiers are very limited in number and use. What is crucial to our hypothesis is the fact that no classifiers have been reconstructed for Proto-Miao-Yao (Wang & Mao 1995; Ratliff 2010) and the borrowing of classifiers from Sinitic languages has been proposed (Ratliff 2000).

However, as the history of Miao-Yao involves a long period of contacts with various language families, to identify the sources of the influence is difficult. Historically, Miao-Yao has had intensive contact with Sinitic languages. Based on genetic and linguistic evidence, the homeland of the Miao-Yao people may have been from somewhere north of their current territories to the south of Yangtze River for more than 2000 years (Wen et al. 2004; Ratliff 2009: 643). During the history of Chinese expansion, Miao-Yao people have been forced to migrate to the south into modern Laos and Thailand (Blench 2004). Language contacts between Sinitic languages and Tai-Kadai, the two candidates for our single origin, are inevitable.

As for the evidence for the source of the influence, classifiers in Miao-Yao languages share features with both Sinitic and Tai-Kadai languages, e.g., Cantonese and Tai. For example, a bare classifier phrase [CLF N] in Miao-Yao languages can mark definiteness, which is a feature also found in southern Sinitic languages, but not in Mandarin (Bisang 1993; Li & Bisang 2012). In terms of word order, Tai-Kadai languages to the south have an N-initial order (i.e., [N Num CLF]), while

Sinitic languages, as well as Tai-Kadai languages in the north, have an N-final order (i.e., [Num CLF N]). Miao-Yao languages are like Sinitic languages in this regard (Enfield 2001: 265).

To resolve the issue, we may resort again to distribution. Haudricourt (1954) observed that Miao-Yao forms an intermediary group between Mon-Khmer and Tibeto-Burman. This observation forms the basis for Bu's (2011a) 'concentric model' of classifier diffusion. Based on the summary mentioned above, Bu (2011a) claims that the three language families form a concentric circle, with Tai-Kadai in the innermost circle, Miao-Yao the intermediate circle, and Tibeto-Burman the outermost circle. The strength of classifiers becomes weaker and weaker further away from the inner circle, implying a scenario of diffusion. We thus surmise that Miao-Yao classifiers are not indigenous and have been influenced more by Tai-Kadai than by Chinese.

3.2 Austroasiatic

The Austroasiatic family consists of 159 languages in *Glottolog*, in countries located between China and Indonesia, with two main groups, i.e., Mon-Khmer (137 languages, 54 of which are identified in our database as a classifier language) and Munda (22 languages, 4 classifier languages). Genetic, epigraphic, and ethnographic evidence suggests that Austroasiatic people, or at least the genetic ancestors of Austroasiatic-speaking people, were the oldest inhabitants in peninsular Southeast Asia from 4000 years B.P. onwards (Lertrit et al. 2008). While the issue of the homeland and expansion of Austroasiatic languages remains contentious, evidence of language contact with other language families in SMATTI has been proposed. A series of words in Old Chinese have been proposed to have been borrowed from Austroasiatic (Norman & Mei 1976; Schuessler 2007). Long after the time of this borrowing in the era of Old Chinese, the expansion of Tai people into the peninsula during the early second millennium brought forth another wave of language contact. Lastly, the Munda branch, as an especially isolated branch in South Asia, is certainly in contact with and influenced by the Indo-Aryan languages on the subcontinent.

Adams (1991), following Jones (1970), hypothesizes that the classifier system is not native to Austroasiatic but is borrowed from neighboring languages such as Thai, Burmese, and Chinese. Huffman (1973) compares the syntax, including classifiers, of Thai and Cambodian and comes to support such a claim. Compared to Thai, Cambodian has fewer classifiers, and they are optional in colloquial forms. This less intensive use of classifiers in Cambodian suggests a borrowed feature. The claim is further supported by the absence of classifiers in other Mon-Khmer

languages and the relatively strong and wide-spread classifier use in Tai languages (Huffman 1973).

In addition to Tai languages, there are other possible sources for classifiers in Austroasiatic languages, one of which being Sinitic. Vietnamese classifiers have been heavily influenced by, or borrowed from, Chinese, though some uses are innovations (e.g., Benedict 1947; Alves 2016). Munda languages, on the other hand, are influenced by Indo-Aryan in many ways, including classifier systems as well (Emeneau 1956). Though more studies are needed to draw a fuller picture of sources of classifiers in various Austroasiatic languages, we surmise that classifiers are not indigenous in this language family and have been influenced by the surrounding Indo-Aryan, Tai-Kadai, and Sinitic languages.

3.3 Tibeto-Burman

As shown in Section 2.2, within SMATTI, only Tibeto-Burman has *base*-initial, C/M-initial languages, where the only five exceptions to the *base*-C/M harmonization are also located. The overall geographical view of SMATTI in Figure 4 reveals a clear picture of TB sandwiched between *base*-final, C/M-final languages on its flanks. Figure 5 is a satellite picture of Figure 4 and gives a better view of the landscape of SMATTI classifier languages.

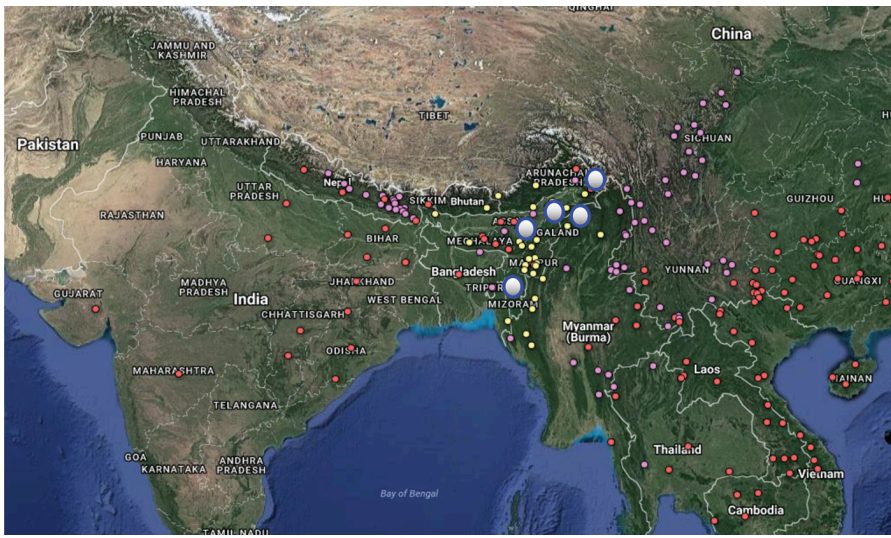


Figure 5. Satellite Map of SMATTI (yellow: TB initial, pink: TB final, blue circle: TB exceptions, red: SMATTI final)

Language contact not only accounts for classifiers in TB, but also the mixed patterns of numeral-internal orders, i.e., both *base*-initial systems and *base*-final systems, the former being indigenous, the latter borrowed. This fact, taken together with the strong tendency to harmonize *base* and C/M in word order, constitutes the key to the complex variation of *base* and C/M word orders in TB, as shown in (10), which is derived from data presented in Table 2.

(10) Variation of numeral systems and classifier systems in Tibeto-Burman
  A. *base*-initial, no C/M confirmed (56 languages)
  B. *base*-initial, C/M-initial (32 languages)
  C. *base*-initial, C/M-final (2 languages)
  D. *base*-final, C/M-initial (6 languages)
  E. *base*-final, no C/M confirmed (91 languages)
  F. *base*-final, C/M-final (76 languages)

Table 2. Types of numeral systems and classifier systems in TB

	***Base*-final**	***Base*-initial**	**Data on base not available**	**Total**
C/M-final	**76**	**2**	11	89
C/M-initial	**6**	**32**	9	47
C/M order not known	2	2	0	4
No attested data for C/M	**91**	**56**	176	323
Total	175	92	196	**463**[16]

As the current data of the first two columns in Table 2 show, far more TB languages now use base-final numerals. Given the indigenous *base*-initial order in PTB, type A languages are most conservative in maintaining the native *base*-initial numerals and rejecting classifiers. Type B languages maintain the native *base*-initial order, which dictates the word order of the borrowed classifiers. Type E languages lost their native *base*-initial system to a borrowed *base*-final system without classifiers. Type F languages have given in on both fronts, having switched to a *base*-final system with classifiers following the numeral. Both type C and type D languages are in an unstable transient state. Type C retains the native *base*-initial numerals but adopted classifiers with a foreign word order. Type D is in a more advanced stage of borrowing, where the numerals have started to give

16. The total number of Tibeto-Burman languages is 465, but Sunwar and Rabha are excluded from this table since these languages use two numeral systems, one indigenous and *base*-initial and the other, *base*-final.

in to a borrowed *base*-final order, while the classifiers borrowed earlier still follow an earlier *base*-initial order.

Languages of type C and D appear to be in a transitory state, which accounts for their scarcity. The overall variation unique in TB is thus the result of varying degrees of language contact with neighboring classifier languages, where different sociolinguistic factors (e.g., geographical, economic, political, and cultural, etc.) must also play a profound role. For example, Figure 5 seems to reveal a geographical factor at work, i.e., most of the *base*-initial, C/M-initial languages, together with the five languages of type C and D, aggregate along mountainous regions that constitute barriers to outside influence, as shown in dark green on the map.

While much more research is surely needed to have a better understanding of this pattern of distribution, the crucial point we wish to make from the scenario depicted above is that the TB variation of *base* and C/M orders can only be accounted for if we assume that C/M as a grammatical concept, like the *base*-final numeral, was borrowed. Note, however, that in Matisoff's (2003) *Handbook of Proto-Tibeto-Burman*, a single classifier for humans, *ra, is found, on p. 609, listed in the Index of Proto-Forms. Yet, closer examination of related data in the book indicates that *ra in PTB is likely a noun, not a classifier. The only basis for considering *ra a classifier in PTB is given on p. 170, where *k-ra is a PTB form meaning "strength/win". Matisoff (2003: 170, fn. c) explains why this form *k-ra and its meaning suggest that *ra is a classifier: "This seems certainly to be the same root as **ra* 'humans (classifier)', above 3.4.2. Cf. English expressions like '20 men strong'."

However, in the English expression *20 men strong* or *20-man strong, men* or *man* is neither a classifier nor a measure word under a proper definition of numeral classifiers. Also, on p. 43, in § 3.4.2, which the above quote refers to, *ra is listed as a form in Proto-Lolo-Burmese (PLB) meaning "CLF for humans", but no corresponding form in PTB is given, while the entry right below is "snake" *m-əy in PLB and its PTB form *s-b-rul is given. These observations, coupled with the fact that by far the majority of TB languages do not have classifiers, suggest strongly that no classifiers should be reconstructed for PTB.

In Figure 6, we show all 465 languages in TB, with green dots being the classifier languages and blue dots for the languages for which we have no information on their classifiers. Note the pattern of distribution of the green dots. First, the classifier languages are mostly located in or near lower altitude and possibly more agricultural places, where contacts are more likely to occur. Second, the classifier languages are mostly located either outside of, or on the fringe of, language clusters. This means that these languages are more likely to be influenced by neighboring non-TB classifier languages than those clustered tightly together. Far more research is needed to confirm these initial observations.

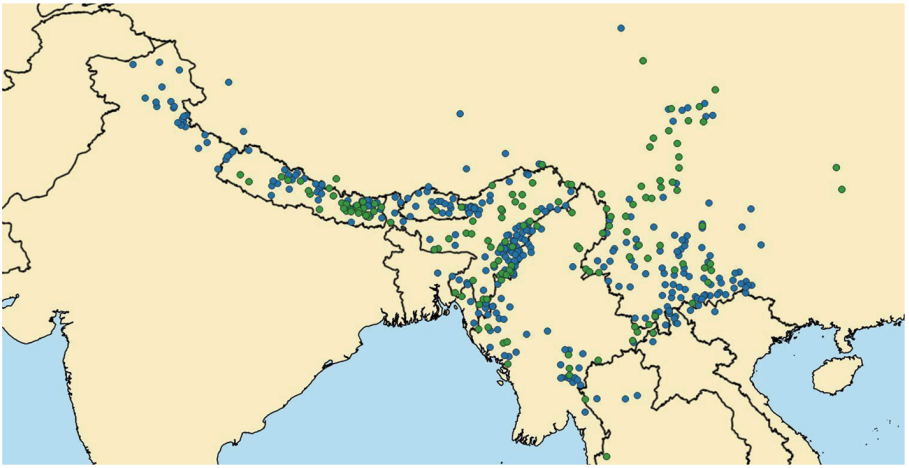


Figure 6. Attested classifier languages and other languages in Tibeto-Burman (green dots for classifier languages, blue dots for non-classifier languages)

Incidentally, word orders of numeral modifiers with respect to nouns may also shed light on our hypothesis. Dryer (2008: 49) suggests that languages with numeral classifiers may exhibit a correlation between Num/N word order and Num/CLF word order. But he also admits that there is no strong evidence for such a claim. It is indeed our hypothesis that cross-linguistically there is a correlation among N-parameter (N-NumP or NumP-N), *base*-parameter (*base-n* or *n-base*), and C/M-parameter (C/M-Num or Num-C/M). Note that NumP in classifier languages consists of Num and a C/M. We are now testing this hypothesis using Dryer's data on N-parameter in Tibeto-Burman and our data on *base*-parameter and C/M-parameter in Tibeto-Burman. In other words, we should find statistical significance in [N C/M *base-n*] and [*n-base* C/M N] over other attested word orders. The hypothesis has been shown to be valid in Allassonnière-Tang & Her (2020), which adds considerable weight to our claim regarding the status of classifiers in Tibeto-Burman.

3.4 Dravidian

The next language group to be considered is Dravidian, a group located mainly in the south of SMATTI, especially south of Indo-Aryan. Dravidian languages are spoken in southern India, on the continent and on the islands of Lakshadweep, and on the island of Sri Lanka. They are also spoken sporadically in other parts of the Indian subcontinent, including Brahui, a language spoken in Pakistan and Afghanistan. In our database, 11 Dravidian languages are found to have classifiers,

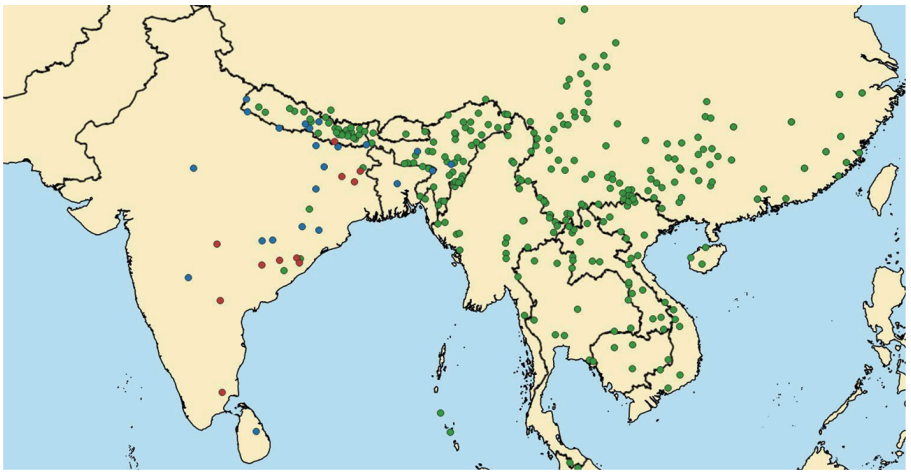


Figure 7. Classifier languages in Dravidian and other language families (red dots for Dravidian, blue dots for Indo-Aryan, and green dots for others)

out of the 85 listed in *Glottolog*. There is no indication that classifiers are an indigenous feature.

Except Tamil, all other classifier languages, i.e., Duruwa (Parji), Kurux (Kurukh), Kui, Kuvi, Kolami, Malto, and Telugu, are distributed in the eastern part of the subcontinent, on the fringes between Dravidian and Indo-Aryan speaking areas. This is a hint that classifiers in Dravidian are the result of language contact. Emeneau (1956) finds that many of these languages borrowed some numerals from the neighboring Indo-Aryan languages, and some also borrowed the classifier systems; as a result, classifiers are often used only with the borrowed numerals, e.g., in Duruwa, Kui, Kuvi, and Kolami. Duruwa, for example, having possibly borrowed its numerals from '6' on from Halbi, an Indo-Aryan language, uses the classifier *jan* for humans and *goṭa* for certain objects only with native numerals '5' and under. Furthermore, the similarities that *jan* and *goṭa* bear with their counterpart in the Indo-Aryan languages nearby further imply this borrowing relation (cf. *zɔn* in Assamese, *jan* in Halbi, *jana* in Bengali (Barz & Diller 1985) and *ʈo/ʈa* in Nepali (Pokharel 2010)).[17] Although few claims are made regarding Telugu and Malto, it is also conjectured in Emeneau (1956) that they developed classifier systems under the influence of Indo-Aryan classifier languages.

17. We thank an anonymous reviewer for pointing out this connection.

3.5 Indo-Aryan

Although Indo-Aryan is part of SMATTI, we have left it until now for a reason. It seems clear from the above discussion that classifiers in Dravidian came from those in Indo-Aryan. Languages such as Duruwa and Kolami borrowed not only the grammatical device but also the form of classifiers, while languages such as Malto and Telugu possibly only acquired the concept of classifiers from Indo-Aryan, but not the word forms, as in a scenario of stimulus diffusion. The origin of Indo-Aryan classifiers is the key to link Indo-Aryan and Dravidian to the single origin of Tai-Kadai or Sinitic. Currently, there are 25 attested Indo-Aryan classifier languages in our database. Since they are distributed in several primary subbranches of Indo-Aryan, there seems to be no genetic bias. However, there does seem to be a geographical bias; as shown in Figure 8, most of the Indo-Aryan classifier languages are located toward the center of SMATTI.

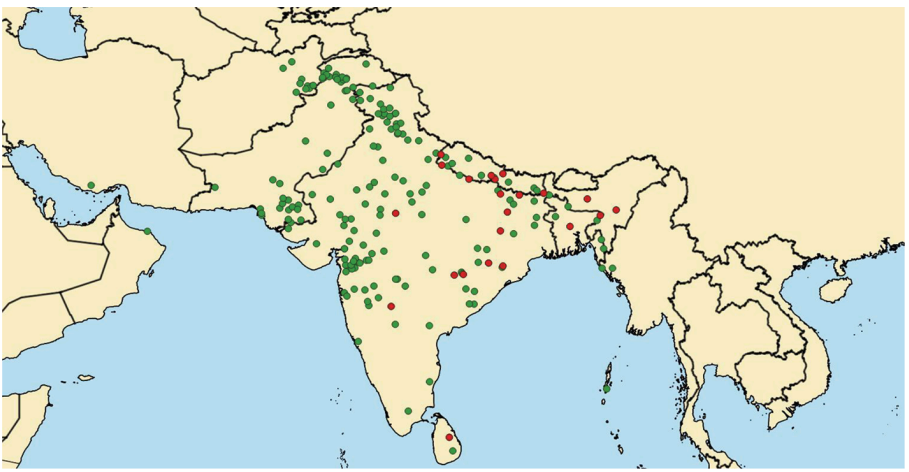


Figure 8. Indo-Aryan classifier languages (red: classifier languages, green: others)

Emeneau (1956) suggests that classifiers did not appear until the end of the Middle Indo-Aryan period, which spans from the end of Old Indo-Aryan to about 1500 A.D. Tibeto-Burman is one possible source of classifier influence for Indo-Aryan (Moral 1997). The Ahom kingdom presents another, perhaps better, possibility. The Ahom kingdom was established in the 13th century by Ahom people, a Tai-speaking ethnic group, in what is today the Assam state of India (Moral 1997). Ahom is a Tai language with a rich C/M-final classifier system (Phukan 1978; Moral 1997). Weidert (1984) and Kiryu (2009: 63) specifically suggest that numeral classifiers in the politically strong language Assamese came from Ahom, while Saikia & Allassonnière-Tang (this volume) cite different sources claim-

ing that Assamese classifiers were borrowed from its Tibeto-Burman and Austroasiatic neighbors. It is of course quite possible that both Ahom and other neighboring classifier languages all have played a role in the rise of classifiers in Indo-Aryan languages. Given that Dravidian classifiers were due to Indo-Aryan, the time that classifiers first appeared in Indo-Aryan languages in northeastern India should come before that in Dravidian. Further research is needed to confirm this conjecture.

3.6 Austronesian

Austronesian classifier languages seem to cluster in the eastern hemisphere, and virtually no classifier languages are found in Polynesia (see Figure 9).

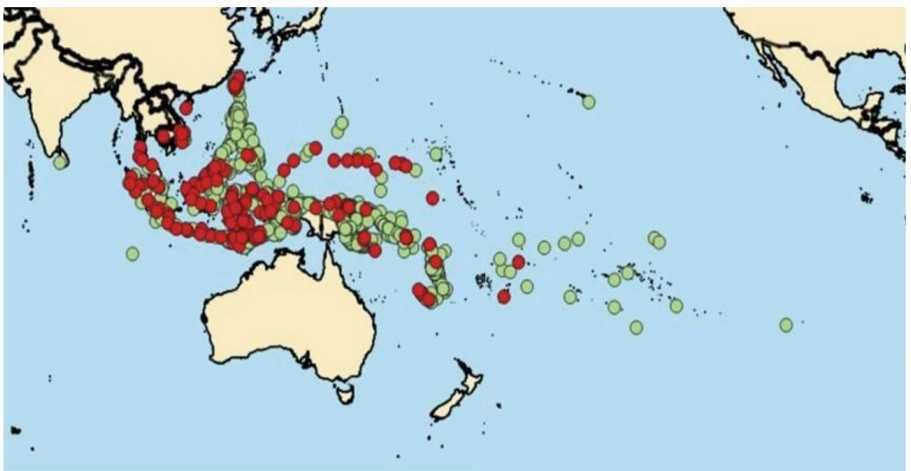


Figure 9. Distribution of classifier languages in Austronesian (red dots: classifier languages, green: other languages)

According to the Out-of-Taiwan model of the Austronesian people, a large scale of Austronesian migration from mainland China to Taiwan began around 5000–2500 B.C. and some research on radiocarbon dates suggests such a migration as late as 4000 B.C. (Kun 2006), depicted in Figure 10, which is based on the *Atlas historique des migrations* (Chaliand et al. 1999) and *The Austronesian Basic Vocabulary Database* (Greenhill et al. 2008). The two maps again display a 'thinning out' scenario among classifier languages, as many Micronesian languages in the west have classifiers, but virtually none in the Polynesian groups in the east (Bender & Beller 2006).

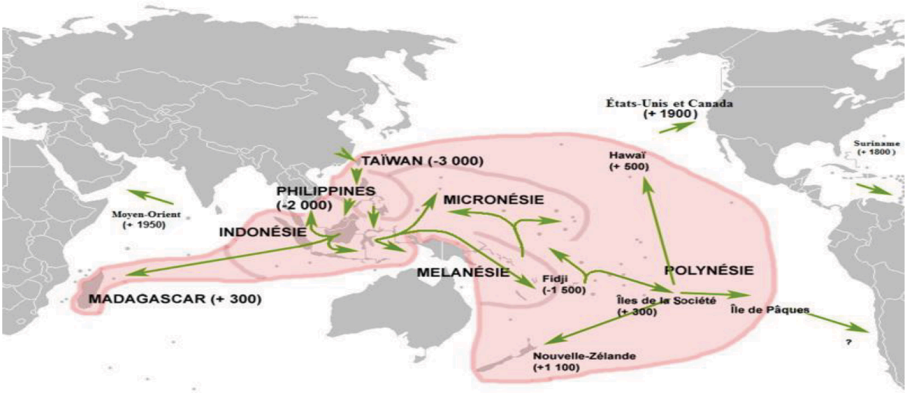


Figure 10. Map of expansion of Austronesian languages[18]

Under the Out-of-Taiwan model, if classifiers in Austronesian languages outside of Taiwan are an indigenous Austronesian feature, then they must have been inherited from Formosan languages. There are indeed six Formosan classifier languages in our database. Tang (2004) claims that Paiwan, Bunun, Kavalan, Amis, and Tsou are so-called 'poor' classifier languages with only two classifiers, one for humans, and the other, non-humans, and Zeitoun (2007) claims that Rukai has four classifiers. However, following a host of other Austronesianists, e.g., Himmelmann (2005: 173), Blust (2013: 292), Klamer (2014: 161), and Gil (2015: 286), Her & Ye (2017) specifically argue that there are no numeral classifiers in Formosan languages by demonstrating that the two putative classifiers in Tang (2004) are part of a rudimentary object-specific numeral system, which is common in Polynesian languages (Bender & Beller 2006), and the four putative classifiers in Zeitoun (2007) are better analyzed as nouns.

One piece of strong, though indirect, evidence against these putative classifiers in Formosan offered by Her & Ye (2017) is the fact that all Formosan languages employ *base*-final numerals. Yet, all these putative Formosan classifiers precede the numeral, violating the *base*-C/M harmonization. It is highly improbable, given the long history of *base*-final numerals since Proto-Austronesian, for classifiers that have passed down from Proto-Austronesian to be in a conflicting word order with numerals at the present time. After all, there has not been any outside C/M-initial influence. We should therefore properly reclassify Formosan languages as attested non-classifier languages.

18. Image obtained from http://en.wikipedia.org/wiki/File:Migraciones_austronesias.png; based on Chaliand et al. (1999) and Greenhill et al. (2008).

Benedict (1975) and Sagart (2004, 2005) propose an Austro-Tai hypothesis, arguing that Proto-Tai is a branch of Proto-Austronesian (PAN). Yet, as argued by Thurgood (1994) and Blust (1988), Proto-Tai also originated in southern China, and the correspondences between Tai-Kadai and Proto-Austronesian are the result of language contact in early times. Empirically, classifier languages are a minority in Austronesian and do not seem to be an indigenous grammatical feature. Jones (1970) proposes that some Austronesian languages like Javanese, Indonesian, and Cebuano have developed classifiers under Tai influence. Conklin (1981) further claims that the classifier systems in Eastern Austronesian, from Philippine to Polynesia, and Western Austronesian, from Indonesia to Madagascar, reflect Tai influence.

3.7 Other non-SMATTI languages

Currently in our database other classifier languages outside of SMATTI include Japonic languages, Korean, and Altaic, while further in Europe, classifier languages are also found in Indo-European and other languages spoken in Europe, e.g., Uralic.

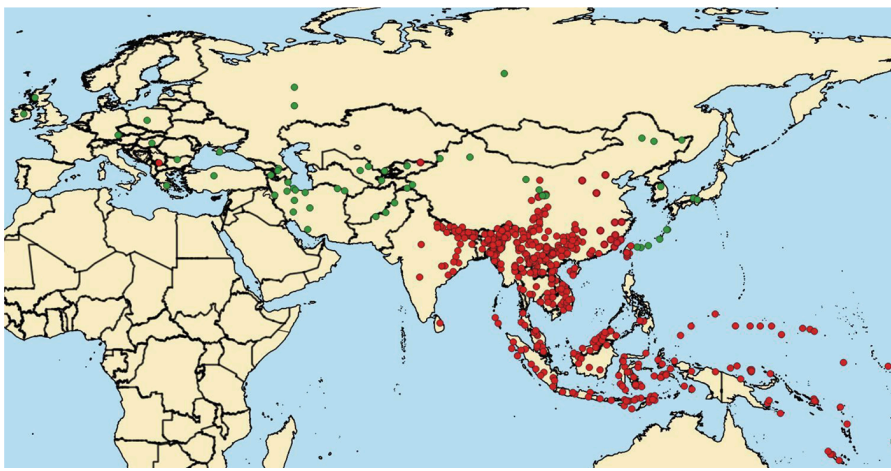


Figure 11. Classifier languages covered so far (red dots) and those covered in Section 3.7 (green dots)

One thing that is clear about these classifier languages is that most of them do not have a well-developed classifier system. Korean and the Japonic languages, which include Japanese and various Ryukyu languages, are nonetheless typical classifier languages, with large inventories of classifiers like the nearby Sinitic.

They likely have imported the system through a long period of contact with Chinese. This is clearly the case for Korean and Japanese, where the classifier system, along with a loan numeral system, is part of an identifiable Sinitic component in their respective grammars (e.g., Janhunen 2000: 693).

We also conjecture that Altaic acquired classifiers from other language families, since classifier languages in this family are rather sparse and sporadic, with only 22 in our database out of a total of 74 languages. Mangghuer, a Mongolic language, for example, has acquired some classifiers and numerals from Mandarin due to recent contact (Sandman & Di Garbo, this volume). While we will leave these languages to further research, for the purpose of our single origin hypothesis, suffice it to say for now that we have not seen any work on any of these classifier languages which claims indigenous classifiers for the language(s) in question.

4. The single origin: Sinitic or Tai-Kadai?

We have thus far argued that all language groups with numeral classifiers in and beyond SMATTI in Asia and the Pacific acquired this grammatical feature via language contact, which can be directly or indirectly linked to Tai-Kadai and Sinitic. Given their close proximity and constant interaction throughout history, one could easily have acquired the classifier feature under the influence of the other (e.g., DeLancey 2011). However, which between the two groups developed classifiers first, or who borrowed from whom, has been a hotly-debated issue for many years.

Tai-Kadai supporters argue that classifiers in Chinese were historically a borrowing from Tai into Chinese dialects in southern China and then diffused to dialects in the north (e.g., Jones 1970; Hashimoto 1977; Erbaugh 1986; Adams 1991; Peyraube 1991; Li 2000). However, many linguists disagree, e.g., Huang (1964), Zhang (1978), Liang (1983), Wang (1994), Morev (2000), Behr (2009), and Bu (2011b), and maintain that the classifier feature is indigenous to Chinese.

The various arguments that classifiers in Chinese were a borrowing from Tai are best summarized in Behr (2009: 40) and Huang (2013: 184). The arguments that Huang (2013: 184) lists are (1) lack of classifiers in Archaic Chinese, (2) lack of animal classifiers, (3) structural indeterminacy of classifier position in historical times, (4) absence of reconstructible classifier system in Tibeto-Burmese, (5) skewed distribution of classifiers in Chinese dialects, i.e., striking morphosyntactic parallels to Tai-Kadai classifiers in southern dialects are not shared by dialects further north, and (6) Tai-Kadai has the most developed system of all Southeast Asian languages. Behr (2009: 40) lists one more possible argument, i.e., Tai-Kadai is historically located in the center of other classifier language groups. We shall go

over these seven arguments and demonstrate that none of them can be substantiated as a solid argument against indigenous classifiers in Sinitic.

4.1 Classifiers in Archaic Chinese

First of all, classifiers are a well-formed category in the earliest manuscript in Tai, which dates back to the 13th century, roughly a thousand years later than the time when Chinese classifiers are likewise firmly established. Thus, even if it were true that there are no classifiers in the oracle bone inscriptions and bronze inscriptions from the Shang period (16th–11th c. B.C.), it does not add to the advantage of Tai-Kadai.

The use of measure words in the oracle bone and bronze inscriptions of Archaic Chinese is not in question (e.g., Wang 1994: 71; Lin 2006); the question is about classifiers. And the fact is that it is no longer in dispute that classifiers in the proper grammatical sense do exist in Archaic Chinese (e.g., Lin 2006; Jiang 2009: 138). Two things must be pointed out first: many classifier languages use classifiers optionally (e.g., Gil 2013) and many classifier languages have as few as one or two or three classifiers, which are usually optional. Such languages are nonetheless widely accepted as classifier languages, albeit in their incipient stage. Table 3 shows six Kiranti languages, a group of Tibeto-Burman languages spoken mainly in the east of Nepal, each with three classifiers or less (Ebert 1994: 79–80, cited in Kiryu 2009: 57).

Table 3. Sortal numeral classifiers in Kiranti languages (Ebert 1994)

	Limbu	**Athpare**	**Bantawa**	**Camling**	**Thulung**	**Khaling**
Human	*(-phu)*	*-paŋ*	*phaŋ, phop*	*-po*	*-ŋ*	*-pu, -ba*
Round				*-li*		
General	*-si*		*-tat, -tak*	*-ra*	*-le*	*-le*

Likewise, classifiers are both scarce and optional in Archaic Chinese, where echo or repeater classifiers are also used (e.g., Wang 1994: 71; Behr 2009; Jiang 2009: 138–140). It is well-recognized that a classifier system was in its incipient stage during this time in Archaic Chinese. Here are two examples of sortal classifiers from the bronze inscriptions.

(11) Classifiers from bronze inscriptions

a. 馬 四 匹 (Zhang 2001 (3.754))
ma si pi
horse 4 CLF
'4 horses'

b. 俘 車 三十 輛 (Zhang 2001 (5.2839))
fu *che* *sanshi liang*
capture chariot 30 CLF
'captured 30 chariots'

4.2 Animal classifiers in Chinese

A general observation on the rise of classifiers in languages is that designated classifiers for animate/inanimate and human/non-human distinctions are often among the first to appear and mature classifier systems seldom lack such specific classifiers. The classifier systems in Southeast Asian languages are known for their lack of animal classifiers (Adams 1991). Erbaugh (1986: 401) claims that Chinese likewise lacks classifiers for the distinction between animate/inanimate and human/non-human, which is taken as further support for Chinese as a borrower. This can be quite easily challenged on two fronts. First, even if it were true that Chinese lacks animal classifiers, it does not tell us anything about the direction of influence; in other words, it could be the other way around, which is indeed what we will argue for. Second, *contra* Erbaugh (1986: 401), Chinese does have animal classifiers.

It is true, however, that southern dialects and languages in Sinitic have fewer classifiers for animals than Mandarin. Southern Min, for example, has only one classifier, *bué*, that is exclusive for animals, fish, to be specific, while its more general animal classifier *tsiah* can also be used for vehicles and furniture with legs (Chen 2013: 60). In Mandarin, though certain animal classifiers, e.g., *tiao* and *kou*, are likewise not restricted to animals, the general animal classifier *zhi* is much more restricted than its Southern Min cognate *tsiah* and there are at least three classifiers exclusive for animals, i.e., *wei* for fish, *pi* for horses and other animals used for transport, and *tou* for larger mammals.

What is interesting is that the animal classifier *tua* in Thai is rather similar to the general animal classifier *tsiah* in Southern Min, in that its extended use includes shirts, tables, problems, and university courses, among other things (Lu 2012: ix). In Maonan, this same animal classifier can be used for young children (Lu 2012: 118). This similarity between Sinitic and Tai is likely due to borrowing from Chinese. The Chinese animal classifier 頭 *tou*, meaning "head" as a noun and still used as such today, came into use during the Han Dynasty (206 B.C.–220 A.D.) (Wang 1994: 109–110). Cognitively, it makes sense for languages to use a noun with the meaning "head" to refer to animals; after all, all animals have a head, hence the expression *three head of cattle* in English. It is therefore not surprising that Sinitic and Tai-Kadai both use "head" as an animal classifier, but the fact that *tou* as a noun is indigenous to Chinese favors it as the source of diffusion.

Manomaivibool (1976) and Alves (2015) propose that it is during Han and Weijin Nanbeichao when *tou* and a number of other Chinese classifiers were borrowed into Tai, together with some Chinese numerals.

A piece of indirect evidence favoring Chinese comes from the Japanese animal classifier *tou*, no doubt also a loan from Chinese. However, unlike Tai-Kadai and Sinitic, Japanese classifiers do not cross the boundary between animals and non-animals, and even *ko*, the general classifier, cannot be used for animals (Saalbach & Imai 2011: 386). Given that many classifiers in the earliest Japanese records were borrowed from Chinese before the 8th century (Downing 1996: 48), the fact that *tou* in Chinese, like *tou* in Japanese and unlike *tua* in Tai, never refers to non-animals indicates the early influence that Chinese had on its neighbors. The issue of animal classifiers turns out to be in favor of Chinese.

4.3 Chinese classifier positions in historical times

Proponents of the borrowing hypothesis also cite the structural indeterminacy of classifier position as evidence for the unstable state of classifiers as a syntactic category. Behr (2009: 11–20), for example, lists 11 attested types gathered from the literature on classifiers from Shang (13th–11th c. B.C.).

(12) Indeterminacy of classifier position in Shang inscriptions (Behr 2009: 11–20)
1. [N]
2. [Num-N]
3. [N-Num]
4. [Num-N-*yòu*-Num]
5. [Num-N-*yòu*-Num-N]
6. [Num-N-Num]
7. [N-Num-MENS]
8. [N-Num-N_{CLF}] (repeater classifier)
9. [N-Num-CLF]
10. [(N)-Num-CLF-*yòu*-Num] => context related elision[19]
11. [(N)-Num-CLF] => discourse related elision

We can dismiss (12.1), as it has nothing to do with C/Ms. (12.4) is an instance of (12.2) with a conjunct where N is elided; likewise, (12.10) is merely an instance of (12.9). (12.11) is the same as (12.9) with N ellipsis. What is left is exactly the seven types in (13) gathered by Peyraube (1998):

19. Behr (2009: 20) lists this type as [(N)-Num-CLF-*yòu*-N]; note the final N is a typo, it should be Num.

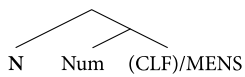
(13) C/M's seven word orders in Chinese history (Peyraube 1998)
1. [Num-N] (一牛 *yi niu* 1 ox)
2. [N-Num] (虎一 *hu yi* tiger 1)
3. [N_1-Num-N_2] (羌十人 *qiang shi ren* Qiang 10 person)
4. [N-Num-MENS] (貝廿朋 *bei ershi peng* shell 20 MENS-set)
5. [N-Num-CLF] (馬三匹 *ma san pi* horse 3 CLF)
6. [Num-MENS-N] (一杯羹 *yi bei geng* 1 MENS-cup soup)
7. [Num-CLF-N] (一株松 *yi zhu song* 1 CLF pine-tree)

Her (2017a: 54) further generalizes the seven types into four, as in (14), by collapsing classifiers and measure words into one syntactic category and taking repeater nouns and the N_2 in (13.3) to be classifiers.

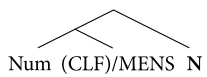
(14) C/M's word orders in Chinese history (Her 2017a: 54)
1. [Num-N]
2. [N-Num]
3. [N-Num-C/M]
4. [Num-C/M-N]

Her (2017a) further demonstrates that once we recognize that a classifier, but not a measure word, is optional and that N's position involves a head parameter, the four types in (14) can then be reduced to merely two, as in (15).

(15) C/M's word orders in Chinese history (Her 2017a: 54–55)
a.



b.



Wu et al. (2006) argue convincingly that in Archaic Chinese the N-initial (15a) and the N-final (15b) orders are different syntactically, semantically, and pragmatically. N-initial order (15a) has a small clause structure of subject and predicate, where [Num C/M] functions as a predicative measure, and (15b) is a nominal structure, where [Num C/M] functions as an attributive modifier. Tang (1996) likewise argues that in Mandarin today the two constructions are not derivationally related and should be treated as two distinct constructions. The difference between Chinese and Thai is simply that Thai has only the construction (15a) available, which some argue is derived by movement (e.g., Jenks 2011).

The crucial point here is that the position of C/M remains the same in the two constructions. In fact, for more than 3,000 years Chinese has been a perfect example of the *base*-C/M harmonization, with a consistent *base*-final [*n base*] numeral system and a consistent C/M-final [Num C/M] classifier system. Only two things have changed somewhat: classifiers have become increasingly obligatory and the N-final order has become much more dominant. In short, there is virtually no indeterminacy of C/M position to speak of.

4.4 Absence of classifiers in Proto-Tibeto-Burman

As demonstrated in Section 3.3, we also take the firm position that there are no classifiers in PTB. Following the conventional assumption that TB is a major branch under Sino-Tibetan, then it is reasonable that numeral classifiers are not reconstructable to Proto-Sino-Tibetan (PST) either (LaPolla 2002: 27). However, given what we know about classifiers in Sinitic, there is little doubt that Proto-Sinitic in oracle bone inscriptions already has a rudimentary system of C/Ms.

Now consider Proto-Tai-Kadai (PTK). According to DeLancey (1986), classifiers are not generally reconstructable for PTK, though there are seven classifiers appearing in all three branches of Tai. The fact that among the seven the animal classifier $*^{A}$tue, is a cognate of the Chinese animal classifier *tou*, and $*^{B}$phɛn (for rigid two-dimensional objects), is also a loanword, indicates that there was already borrowing from Chinese in Proto-Tai. It is therefore likely that TK acquired classifiers under Sinitic influence. In the following quote, Lee & Clontz (2012: 30) depict a scenario of Tai-Kadai history, which can easily accommodate a thesis that Tai-Kadai acquired classifiers from Sinitic.

> [...] it is a consensus among linguists and historians that Tai-Kadai languages are descents [sic] of ancient Baiyue languages and are heavily influenced by neighboring languages in a contact relationship such as Sino-Tibetan and Hmong-Mien, borrowing many loanwords and being sinicized. (Lee & Clontz 2012: 30)

For Chinese to have acquired classifiers from Tai-Kadai, PTK would have to have a classifier system mature enough to influence Chinese, and, more importantly, its date has to be earlier than that of Proto-Sinitic. In short, given our inadequate knowledge of PTK, the absence of classifiers in PTB and PST is not relevant to the debate in question.

4.5 Skewed distribution of classifiers in Chinese dialects

It is often assumed that southern Sinitic languages and dialects have more classifiers than the northern dialects of Mandarin (e.g., Yue 2003: 85). For example,

Dungan, a Gansu dialect of Mandarin spoken in Central Asia, has only the general classifier 個 *ge*. Such a tendency to use a single cover classifier can also be found in its related dialects and other northern Chinese languages from Shanxi, Shandong, Gansu, to Shanghai (Yue 2003). Cao (2008) conducts a thorough survey of the dialectal distribution of four classifiers: 個 *ge*, 頭 *tou*, 隻 *zhi*, 棵 *ke*, and likewise finds a clear pattern where the use of the general classifier is preferred in the north, while the use of the specific classifiers is much more frequent in the south. Such a skewed distribution of richer classifier systems in the south and poorer systems in the north therefore seems to favor a scenario of south-to-north diffusion pattern, which in turn would favor a Tai-to-Chinese direction of classifier development. However, we will argue that classifiers originated in Archaic Chinese in northern China and the northern dialects continued to develop very rich classifiers and there is a good explanation for their later decline.

It is important to distinguish between classifiers and measure words when considering this issue, as measure words are an open class, while classifiers are not. Here we cite three works, where classifiers and measure words are carefully distinguished, i.e., Her & Lai (2012) on Taiwan Mandarin, Chen (2013) on Taiwan Southern Min, and Liao (2014) on Taiwan Hakka. The respective counts of classifiers are 97, 83, and 64, with Mandarin the winner. However, such text-based inventories should be taken with a grain of salt, as Mandarin enjoys the advantage of having been the official language for centuries, the primary medium of formal education for the last century, as well as the most dominant written language for millennia.

It is however quite true that classifiers tend to be optional in northern varieties of Mandarin, while generally required in the southern varieties and other Sinitic languages, especially when the spoken language is considered. For example, in Taiwan Mandarin, Southern Min, and Hakka speech, classifiers are in general required, but in Beijing Mandarin they are much more easily dropped. In the dialogues of the movie 非誠勿擾 *Fei Cheng Wu Rao* [If you are the one] and its sequel, nearly a quarter of the classifiers are dropped (Her & Chen 2013). Her & Wu (2016) further observe that, in the movie 老炮兒 *Laopaoer* "old hooligan" [English title: *Mr. Six*], where the setting is the *hutong*, or alleyways, of Beijing, classifiers are routinely dropped.

(16) Quote from *Fei Cheng Wu Rao*
辦垮　過 三 公司。
bankua *guo san gongsi*
bankrupt ASP 3 company
"(He) bankrupted three companies."

(17) Quote from *Laopaoer*
頤和園 後邊 有 一 野湖。
yihe-yuan houbian you yi yehu
Yihe-Garden back have 1 wild-lake
"There is a wild lake behind Yihe Garden."

Another observation is also true, that there are striking morphosyntactic parallels to Tai-Kadai classifiers in southern Sinitic languages not shared by those further north, as already shown in (7–9). You (1982: 36–37) listed several syntactic functions that are not allowed in Mandarin but found in southern Sinitic and Tai languages. For example, classifiers can stand alone with the noun to express definiteness, with a demonstrative, adjective, verb, noun, pronoun or phrase to form nominals, or to serve as a modifier marker. None of the above uses is allowed in Mandarin, while found in southern Sinitic languages. Examples (7–9) are repeated here with parallels from Tai-Kadai languages.

(18) Bare classifier with a noun to mark definiteness
a. *zek gau soeng gwo maalou* (Cantonese, Cheng & Sybesma 2005: 9)
CLF dog want cross road
"The dog wants to cross the road."
b. *ko:ŋ ha:k ʔeu te pai ham nai* (Wuming Zhuang, You 1982: 36)
CLF officer tell him go night this
"The officer asks him to go tonight."

(19) Classifier with an attributive to form a nominal (Note that word order is reverse in Sinitic and Tai-Kadai.)
a. *sang ke* (Hakka, Sixian)
raw CLF
"raw thing(s)"
b. *qidga ke* (Hakka, Sixian)
self CLF
"thing(s) of oneself"
c. *luuk leklek* (Thai, Carpenter 1986: 15)
CLF small.REDUP
"little ones"
d. *khon nii* (Thai, Carpenter 1986: 15)
CLF.HUMAN here
"this person"

(20) Classifier as a modifier marker (Note that word order is [Mod. CLF N] in Sinitic but [N CLF Mod.] in Tai-Kadai.)

a. *louzisam tiu simzoeng* (Cantonese, You 1982: 38)
L. CLF Buddhist-stick
"Luzhishen's Buddhist stick"

b. *kun ko jɔm* (Xishuang Tai, You 1982: 37)
person CLF thin
"thin person"

Such a skewed pattern of distribution can be explained by south-to-north diffusion if the rise of Sinitic classifiers was due to a southern influence from Tai. However, we contend that there is a better explanation, i.e., the massive Altaicization and simplification of Northern Mandarin in the last two millennia and the intensive interaction between other Sinitic languages and Tai languages during the same period. This Altaicization thesis is succinctly summarized by LaPolla (2010: 6862):

> While many of the invaders assimilated to Chinese language and culture, they also had an effect on the language and culture of China, particularly in the north. Mantaro Hashimoto (e.g., 1976, 1980, 1986) has talked about this as 'the Altaicization of Northern Chinese', and has argued that a continuum of features from north to south, such as the northern dialects having fewer tones, less complex classifier systems, and an inclusive/exclusive distinction in the 1pl pronoun, while the southern dialects have more tones, more complex classifier systems, and other features similar to the Tai-Kadai and Hmong-Mien languages (see You 1982, 1995; Zhou & You 1986; Wang Jun 1991), is due to Altaic influence in the north, and Tai/Hmong-Mien influence in the south.

Beijing itself was largely the political center of Altaic peoples for the last thousand years, except for the three centuries of Chinese rule in the Ming Dynasty. However, besides the strong transfer effects of Altaicization, McWhorter (2007: Chapter 5) offers a complementary, and perhaps even more compelling, thesis, i.e., the simplification of Northern Chinese due to Altaic settlers' speaking the language as a foreign language.

> [...] while Altaic languages did have a decisive effect upon Mandarin, transfer was a relatively minor factor, while reduction due to nonnative competence was comprehensively transformative. The evidence also suggests that the crucial locus of transformation was not, as is often supposed, the occupations by Genghis Khan (thirteenth through fourteenth centuries) or the Manchus (1644–1911), but the widespread resettlement of conquered and dispossessed peoples amid Han Chinese on the northern Chinese frontier from the 600s through the 800s during native Chinese rule under the Tang dynasty. (McWhorter 2007: 105)

The two theses jointly form a powerful theory that accounts for the systematic skewed distribution of classifiers as well as a multitude of linguistic features in Sinitic languages and dialects in the south and those in the north.

Wu & Li's (2016) survey of the literature shows that the Chinese inventory of classifiers was on a path of steady growth since the Archaic Chinese of oracle bone inscriptions and reached full maturity during Weijin Nanbeichao, as widely recognized, with more than 100 classifiers, according to Liu (1965).[20] A recent study by Huang (2018) indicates a very rich classifier system in northern Chinese texts from the Yuan dynasty. The theory of Altaicization and simplification nicely explains the decline of classifiers in Northern and Northwestern Chinese in the last three hundred years, but these facts are difficult to be explained under the south-to-north diffusion hypothesis.

4.6 The most developed system of all Southeast Asian languages

The claim that Tai has the most developed classifier system may be hasty. Given the large inventories of C/Ms in southern Sinitic languages and also the striking morphosyntactic parallels of their classifiers to those in Tai, the classifier systems in southern Sinitic languages are just as developed as those in Tai.

However, there is one particular feature about the current Tai systems that is worth considering, i.e., the extensive use of echo or repeater classifiers. Jenks (2011: 94) states that as many as five hundred nouns in Thai can serve as their own classifiers in the repeater construction [N Num CLF]. It is well-known that in Tibeto-Burman classifier languages as well as in the historical development of Chinese classifiers, repeater classifiers are often a sign of a developing classifier system (e.g., Jiang 2009). A mature classifier system represents a conceptual classification system of nouns of high countability. The presence of productive repeater classifiers indicates that many entities form individual ad hoc classes yet to be incorporated into a more general classification scheme; it can be seen as a sign of the language still on its way to a full-fledged classification scheme.

Furthermore, repeaters can only appear in a [N Num CLF], or an N-initial, order, for the obvious reason that repeaters would be confused as nominal reduplication in an N-final, or [Num CLF N] order. In the historical development of classifiers, Chinese first had the N-initial order and also echo classifiers and then changed to the N-final order, where repeaters are not possible. Thus, the existence of repeaters and the dominant N-initial order in Tai also indicate that its classifier systems resemble an earlier stage of the classifier development. Thus, the fact that

20. Wu & Li (2016) counted 122 classifiers in Liu (1965), while Wang (1994: 112) counted 119.

none of the modern Sinitic languages uses repeater classifiers and the dominant word order is [Num CLF N] may be taken as a sign of greater maturity.

4.7 Historical center of classifier language groups

It is of little significance to claim either Tai-Kadai or Sinitic as the geographical center of classifier languages in SMATTI and beyond, given that to the east of China lies the vast Pacific Ocean and to the north of China proper lies the Great Wall. Land size, population, and many other sociolinguistic factors surely favor Sinitic as the origin, but the number of distinct languages favors Tai-Kadai.

4.8 Other arguments favoring Sinitic

The discussions in this section so far have seen three factors that seem to favor Sinitic over Tai-Kadai, i.e., the time depth of written records of Sinitic classifiers (Section 4.1), the animal classifier *tou* in Chinese and its cognates in Thai and Japanese (Section 4.2), and the existence of repeater classifiers in Tai-Kadai (Section 4.6). And we should also point out that, though the two language groups have interacted with, and borrowed from, each other, the direction of borrowing in terms of linguistic elements in general, including classifiers, favors Sinitic. Wang (1994: 164) claims that whether the classifiers in a language are derived from its own lexicon is a key point in determining whether the classifier feature is native to the language, and he painstakingly demonstrates that the several classifiers found in oracle bone inscriptions were all derived from nouns indigenously.

Out of the linguistic elements borrowed into Tai-Kadai from Chinese, one thing interests us in particular, i.e., numerals. Numerals are of significant importance to numeral classifiers. Given the *base*-C/M harmonization, a language simply cannot develop a classifier system without multiplicative bases in its numeral system. The scenario depicted below by Morev (2000: 81) thus seems reasonable.

> Historico-comparative investigations of Tai and Chinese testify that Tai-Chinese contacts were very intensive in the first half of the first millennium A.D. (see S. Starostin [1989], S. Yakhontov 1977, Downer 1963 e.a.). This comes from the fact that supposedly cognate Sino-Tai words, including numerals, can be traced back to the aforementioned period. As it has been said earlier, the first centuries A.D. were the period of consolidation of N Num Clf construction in Chinese. So, it gives us enough grounds to hypothesize that precisely at that period the Southwestern Tai had adopted the system of Chinese classifiers together with Chinese numerals.

In terms of the stimulus diffusion of linguistic features from one language to another, Weidert (1984: 205–206) takes into account two factors, an internal factor, i.e., the linguistic structure of the recipient language, and an external factor, i.e., the strength of the political, economic, and cultural contacts. Recall that Tai-Kadai numeral system is consistently decimal and base-final like that of Chinese, and the clear evidence of borrowing of some numerals from Chinese to Tai demonstrates an internal factor favoring the same direction of the borrowing of classifiers. In terms of the external factor, Wang (1994) argues against Erbaugh's (1986: 406) suggestion that Chinese borrowed classifiers from Tai through historical contacts such as exchange of tribute and goods in the Han dynasty. Historical facts related to the external factor undoubtedly likewise favor the direction of borrowing classifiers from Chinese into Tai.

From the discussion in this section, we conclude that no evidence currently available can give us a definitive answer as to whether Tai-Kadai or Sinitic is the innovator of the classifier feature in grammar, but it should also be clear that the evidence available to us at this time does seem to favor Sinitic.

5. Conclusion

A single origin hypothesis is proposed for numeral classifiers in the languages of Asia and the Pacific, excluding for now Europe, the Americas, Papua New Guinea, and Africa, based on the distribution of 713 classifier languages in the world. From the foremost hotbed of classifier languages, which we call SMATTI, an acronym for Sinitic, Miao-Yao, Austroasiatic, Tai-Kadai, Tibeto-Burman, and Indo-Aryan, we find links amongst the language groups in Asia and the Pacific. For some language groups, there is clear evidence in the literature indicating that classifiers were acquired through contact, e.g., Austroasiatic. Evidence for some other groups is inferred, e.g., Miao-Yao. There are still other language groups, e.g., Austronesian, for which we can exclude the possibility that classifiers are indigenous, but specific details of how they were acquired are lacking. Overall, although some links may require much further research, it seems plausible that all classifier languages in Asia and the Pacific acquired this feature through contact, except for those cases where classifiers developed spontaneously.

As for the origin, we have considered the two most likely candidates, Tai-Kadai and Sinitic, weighing various factors. At the end, we conclude tentatively that the evidence available at the present time gives Sinitic a slight edge.

Acknowledgements

The authors thank the anonymous reviewers of the paper and the editors of this volume for their constructive comments and suggestions. O.-S. Her gratefully acknowledges the following grants by Taiwan's National Science and Technology Council (NSTC): 101-2410-H-004-184-MY3, 102-2811-H-004-023, 103-2811-H-004-003, 103-2633-H-004-001, 103-2410-H-004-136-MY3, 104-2811-H-004-004, 104-2633-H-004-001, 104-2410-H-004-164-MY3, 106-2410-H-029-077-MY3, 107-2811-H-004-517, 108-2811-H-004-521, 108-2410-H-029-062-MY3, 109-2811-H-004-522, 111-2811-H-029-003, 111-2410-H-029-009-MY3. He also expresses his heartfelt gratitude to the graduate students and researchers for their help in building the database, including part-time RAs: Hsieh, Chen-tien; Lai, Wan-Chun; Chen, Meng-Ying; Wang, Wei; Chen, Ching-Perng; Chen, Yun-Ju; Liao, Jia-Yu; Li, Bing-Tsiong; Chia,Cheng-Pin; Allassonnière-Tang, Marc; Lin, Kun-Han; Huang, Yu-Min; Chen, Chia-Chi; Yeh, Chu-Hsien; Yang, Wen-Chi; Huang, Tsung-Chia; Chen, Shen-An; Jheng, Jhih Siou; Liang, Yu-Ting; Gao, Zhong-Liang; Cao, Zi-Yun; Hsu, Hung-Hsin; Kuo, Hui-Ting; Liang, Yung-Ping; Lo, I-Chieh; Chen, Yi-Ju; Chen, Wei-You; Cheng, Yu-Ching; full-time RAs: Chen, Ying-Chun; Lin, Yen-Tse; Ho, Pei-Hsuan; and post-docs: Hsieh, Fu-Tsai; Tsai, Hui-Chin; Hsiao, Pei-Yi; Hsu, Chi-Pin.

References

Adams, Karen L. 1991. The influence of non-Austroasiatic languages on numeral classification in Austroasiatic. *Journal of the American Oriental Society* 111(1). 62–81.

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

Aikhenvald, Alexandra Y. & R. M. W. Dixon (eds.). 2001. *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press.

Allassonnière-Tang, Marc & One-Soon Her. 2020. Numeral base, numeral classifier, and noun: Word order harmonization. *Language and Linguistics* 21(4). 511–556.

Alves, Mark. 2015. Grammatical Sino-Tai vocabulary and implications for ancient Sino-Tai sociolinguistic contact. Paper presented at the 48th International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL48), UC Santa Barbara, August 21–22, 2015.

Alves, Mark. 2016. Preliminary steps towards and challenges in reconstructing Proto-Vietic syntax. Paper presented at the International Conference of Austroasiatic Linguistics (ICAAL), Chiangmai University, Thailand, September 5–7, 2016.

Au Yeung, Ben Wai Hoo. 2005. An interface program for parameterization of classifiers in Chinese. Ph.D. dissertation, Hong Kong University of Science and Technology.

Au Yeung, Ben Wai Hoo. 2007. Multiplication basis of emergence of classifiers. *Language and Linguistics* 8(4). 835–861.

Barz, R[ichard] K. & A[nthony] V. N. Diller. 1985. Classifiers and standardisation: Some South and South-East Asian comparisons. In David Bradley (ed.), *Language policy, language planning and sociolinguistics in South-East Asia*, 155–184. Canberra: Pacific Linguistics.

Behr, Wolfgang. 2009. Classifiers, lexical tone, determinatives: A look at their emergence and diachrony in Old Chinese and beyond. Paper given at the workshop 'The Diachrony of Classification Systems', NIAS, Wassenaar, 12–14 March 2009.

Bender, Andrea & Sieghard Beller. 2006. Numeral classifiers and counting systems in Polynesian and Micronesian languages: Common roots and cultural adaptations. *Oceanic Linguistics* 45(2). 380–403.

Benedict, Paul K. 1947. An analysis of Annamese kinship terms. *Southwestern Journal of Anthropology* 3(4). 371–392. https://www.jstor.org/stable/3628527

Benedict, Paul K. 1975. *Austro-Tai: Language and culture, with a glossary of roots*. New Haven, CT: HRAF Press.

Bisang, Walter. 1993. Classifiers, quantifiers and class nouns in Hmong. *Studies in Language* 17. 1–51.

Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond". In Jadranka Gvozdanovic (ed.), *Numeral types and changes worldwide*, 113–185. Berlin: Mouton de Gruyter.

Blench, Roger. 2004. Stratification in the peopling of China: How far does the linguistic evidence match genetics and archaeology?. Paper presented at the Symposium of Human Migrations in Continental East Asia and Taiwan: Genetic Linguistic and Archaeological Evidence, Université de Genève, June 10–13, 2004.

Blust, Robert. 1988. The Austronesian homeland: A linguistic perspective". *Asian Perspectives* 26(1). 45–67. https://www.jstor.org/stable/42928105

Blust, Robert. 2013. *The Austronesian languages*. (Asia-Pacific Linguistics.) Canberra: Australian National University.

Bu, Lianzeng. 2011a. Nanfang minzu yuyan de mingcifenleici he shucifenleici [Nominal and numeral classifiers in the southern branch]. *Minzu Yuwen* 1. 48–54.

Bu, Lianzeng. 2011b. Nanfang hanyu zhong liangci dingzhi xianxiang laiyuan chutan: yi guinanpinghua wei li [The source of classifier definiteness: A case study of Guinan Pinghua]. *Yuyan Yanjiu* 3. 99–106.

Burusphat, Somsonge & Xiaohang Qin. 2006. *Northern Zhuang Chinese-Thai-English dictionary*. Nakhonpathom, Thailand: Mahidol University.

Cao, Zhiyun. 2008. Hanyu fangyan dituji: Cihuijuan [Linguistic atlas of Chinese dialects: Lexicon]. Beijing: The Commercial Press.

Carpenter, Kathie. 1986. Productivity and pragmatics of Thai classifiers. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 14–25.

Cathcart, Chundra, Gerd Carling, Filip Larsson, Niklas Johansson & Erich Round. 2018. Areal pressure in grammatical evolution: An Indo-European case study. *Diachronica* 35(1). 1–34.

Chaliand, Gérard, Michel Jan & Jean-Pierre Rageau. 1999. *Atlas historique des migrations*. Paris: Editions du Seuil.

Chan, Eugene. 2017. *Numeral systems of the world's languages*. (https://lingweb.eva.mpg.de/channumerals/). Accessed: 3 March 2017.

Chen, Meng-Ying. 2013. Classifiers in Taiwan Southern Min. MA thesis, National Chengchi University, Taipei.

Cheng, Lisa L.-S. & Rint Sybesma. 1998. *Yi-wan tang, yi-ge tang*: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies* 28(3). 385–412.

doi Cheng, Lisa L.-S. & Rint Sybesma. 2005. Classifiers in four varieties of Chinese. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 259–292. Oxford: Oxford University Press.

Comrie, Bernard. 2006. Numbers, language, and culture. Paper presented at the 16th Jyväskylä Summer School, Jyväskylä: University of Jyväskylä, July 24-August 11, 2006.

Comrie, Bernard. 2013. Numeral bases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/131). Accessed: 7 November 2016.

Conklin, Nancy Faires. 1981. The semantics and syntax of numeral classification in Tai and Austronesian. Ph.D. dissertation, University of Michigan, Ann Arbor.

doi DeLancey, Scott. 1986. Toward a history of Tai classifier systems. In Colette G. Craig (ed.), *Noun classes and categorization*, 437–453. Amsterdam: John Benjamins.

DeLancey, Scott. 2011. On the origins of Sinitic. In Zhuo Jing-Schmidt (ed.), Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23), vol. 1, 51–64. Eugene, OR: University of Oregon.

Downer, G.B. 1963. Chinese, Thai and Miao, Yao. In Harry Leonard Shorto (ed.), *Linguistic comparison in Southeast Asia and the Pacific*, 133–139. London: University of London, School of Oriental and African Studies.

doi Downing, Pamela A. 1996. *Numeral classifier systems: The case of Japanese*. Amsterdam: John Benjamins.

Dryer, Matthew S. 2008. Word order in Tibeto-Burman languages. *Linguistics of the Tibeto-Burman Area* 31(1). 1–83.

Ebert, Karen H. 1994. *The structure of Kiranti languages: Comparative grammar and texts*. Zurich: Seminar für Allgemeine Sprachwissenschaft, Universität Zürich.

doi Emeneau, M[urray] B[arnson]. 1956. India as a linguistic area. *Language* 32(1). 3–16.

Enfield, Nicholas J. 2001. On genetic and areal linguistics in mainland South-East Asia: Parallel polyfunctionality of 'acquire'. In Alexandra Y. Aikhenvald & R.M.W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 255–290. Oxford: Oxford University Press.

doi Erbaugh, Mary S. 1986. Taking stock: The development of Chinese noun classifiers historically and in young children. In Colette G. Craig (ed.), *Noun classes and categorization*, 399–436. Amsterdam: John Benjamins.

doi Fu, Jingqi. 2015. The status of classifiers in Tibeto-Burman languages. In Dan Xu & Jingqi Fu (eds.), *Space and quantification in languages of China*, 37–56. Dordrecht: Springer.

Gil, David. 2013. Numeral classifiers. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/55). Accessed: 7 November 2016.

doi Gil, David. 2015. The Mekong-Mamberamo linguistic area. In N.J. Enfield & Bernard Comrie (eds.), *The languages of Mainland Southeast Asia: The state of the art*, 266–355. Berlin: De Gruyter Mouton.

doi Goebel, Ted, Michael R. Waters & Dennis H. O'Rourke. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319(5869). 1497–1502.

Greenberg, Joseph H. 1990 [1972]. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. In Keith Denning & Suzanne Kemmer (eds.), *On language: Selected writings of Joseph H. Greenberg*, 166–193. Stanford, CA: Stanford University Press. [Originally published in *Working Papers on Language Universals* 9. 1–39, 1972.]

Greenberg, Joseph H. 1990 [1974]. Studies in numerical systems I: Double numeral systems. In Keith Denning & Suzanne Kemmer (eds.), *On language: Selected writings of Joseph H. Greenberg*, 194–206. Stanford, CA: Stanford University Press. [Originally published in *Working Papers on Language Universals* 14. 75–89, 1974.]

Greenberg, Joseph H. 1990 [1978]. Generalizations about numeral systems. In Keith Denning & Suzanne Kemmer (eds.), *On language: Selected writings of Joseph H. Greenberg*, 271–309. Stanford, CA: Stanford University Press. [Originally published in Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, vol. 3: Word structure*, 249–295. Stanford, CA: Stanford University Press, 1978]

doi Greenhill, Simon J., Robert Blust & Russell D. Gray. 2008. The Austronesian Basic Vocabulary Database: From bioinformatics to lexomics. *Evolutionary Bioinformatics* 4. 271–283.

doi Haas, Mary R. 1942. The use of numeral classifiers in Thai. *Language* 18(3). 201–205.

Halevy, Alon & Rebecca Shapley. 2009. Google fusion tables. (http://googleresearch.blogspot.com/2009/06/google-fusion-tables.html).

doi Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2021. *Glottolog 4.4*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://glottolog.org, accessed: 22 Oct. 2021.)

Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent. *Computational Analyses of Asian and African Languages* 3. 49–63.

Hashimoto, Mantaro. 1977. The genealogy and the role of the classifier in Sino-Tibetan. *Computational Analyses of Asian and African Languages* 7. 69–78.

Haudricourt, André G. 1954. Introduction à la phonologie historique des langues miao-yao. *Bulletin de l'École française d'Extrême-Orient* 44(2). 555–576.

doi Her, One-Soon. 2012. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122(14). 1668–1691.

doi Her, One-Soon. 2017a. Structure of numerals and numeral classifiers in Chinese: Historical and typological perspectives and cross-linguistic implications. *Language and Linguistics* 18(1). 27–73.

doi Her, One-Soon. 2017b. Deriving classifier word order typology, or Greenberg's Universal 20A and Universal 20. *Linguistics* 50(2). 265–303.

Her, One-Soon & Yun-Ru Chen. 2013. Unification of numeral classifiers and plural markers: Empirical facts and implications. Paper presented at the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27), Nov 22–24, 2013, National Chengchi University, Taipei, Taiwan.

doi Her, One-Soon, Harald Hammarström & Marc Allassonnière-Tang. 2022. Defining numeral classifiers and identifying classifier languages of the world. *Linguistics Vanguard* 8(1). 151–164.

Her, One-Soon & Chen-Tien Hsieh. 2010. On the semantic distinction between classifiers and measure words in Chinese. *Language and Linguistics* 11(3). 527–551.

Her, One-Soon & Wan-Jun Lai. 2012. Classifiers: The many ways to profile 'one', a case study of Taiwan Mandarin. *International Journal of Computer Processing of Oriental Languages* 24(1). 79–94.

Her, One-Soon, Marc Tang & Bing-Tsiong Li. 2019. Word order of numeral classifiers and numeral bases: Harmonization by multiplication. *STUF – Language Typology and Universals* 72(3). 421–452.

Her, One-Soon, Hui-Chin Tsai, Kun-Han Lin, Marc Tang & Meng-Chang Lee. 2015. Unification of numeral bases and numeral classifiers: Evidence from SMATTI. Paper presented at the Workshop on the Syntax and Semantics of Numerals, 48th Annual Meeting of the Societas Linguistica Europaea, Leiden.

Her, One-Soon & Jiun-Shiung Wu. 2016. Taxonomy of numeral classifiers and measure words: A formal semantic proposal. Unpublished manuscript, National Chengchi University, Taipei.

Her, One-Soon & Chu-Hsian Ye. 2017. Are there numeral classifiers in Formosan languages? Unpublished manuscript, National Chengchi University, Taipei.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London: Routledge.

Hsieh, Miao-Ling. 2008. *The internal structure of noun phrases in Chinese*. Taipei: Crane.

Huang, Shuanfan. 2013. *Chinese grammar at work*. Amsterdam: John Benjamins.

Huang, Tsung-Chia. 2018. Beifang hanyu fenleici shuliang zhi yanbian dui fenleici aiyuan zhi yihan [The development of Chinese numeral classifiers in Northern China: Implications for the origin of numeral classifiers]. MA thesis, National Chengchi University, Taipei.

Huang, Zaijun. 1964. Cong jiaguwen, jinwen liangci de yingyong kaocha hanyu liangci de qiyuan yu fazhan [Origin and development of Chinese classifiers in accordance with materials of inscriptions on bones and turtle shells and old manuscripts up to Jin period]. *Zhongguo Yuwen* 6. 432–441.

Huffman, Franklin E. 1973. Thai and Cambodian: A case of syntactic borrowing? *Journal of the American Oriental Society* 93(4). 488–509.

Janhunen, Juha. 2000. Grammatical gender from East to West. In Barbara Unterbeck & Matti Rissanen (eds.), *Gender in grammar and cognition*, 689–797. Berlin: Mouton de Gruyter.

Jenks, Peter. 2011. The hidden structure of Thai noun phrases. Ph.D. dissertation, Harvard University.

Jiang, Ying. 2009. Hanzang yuxi yuyan mingliangci bijiao yanjiu [A comparative study of classifiers in Sino-Tibetan languages]. Beijing: Minzu Press.

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

Kiryu, Kazuyuki. 2009. On the rise of the classifier system in Newar. In Yasuhiko Nagano (ed.), *Issues in Tibeto-Burman historical linguistics*, 51–69. (Senri Ethnological Studies 75.) Osaka: National Museum of Ethnology.

Klamer, Marian. 2014. Numeral classifiers in the Papuan languages of Alor and Pantar: A comparative perspective. In Margaretha Anna Flora Klamer & František Kratochvil (eds.), *Number and quantity in East Nusantara*, 103–122. Canberra: Australian National University.

Kroeber, A[lfred] L. 1940. Stimulus diffusion. *American Anthropologist* 42(1). 1–20.

Kun, Ho Chuan. 2006. On the origins of Taiwan Austronesians. In K. R. Howe (ed.), *Vaka Moana: Voyages of the ancestors*, 3rd edn., 92–93. Honolulu, HI: University of Hawai'i Press.

Landman, Fred. 2004. *Indefinites and the type of sets*. Malden, MA: Blackwell.

LaPolla, Randy J. 2002. Overview of Sino-Tibetan morphosyntax. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 22–42. London: Routledge.

LaPolla, Randy J. 2010. Language contact and language change in the history of the Sinitic languages. *Procedia Social and Behavioral Sciences* 2(5). 6858–6868.

Lee, Hugo Yu-Hsiu & Jack Clontz. 2012. Reviewing the prehistoric linguistic relationships of the Tai-Kadai language family and its putative linguistic affiliations: A meta-analysis study and abbreviation edition. *Journal of Liberal Arts Prince of Songkla University Hat Yai Campus* 4(1). 20–57.

Lertrit, Patcharee, Samerchai Poolsuwan, Rachanie Thosarat, Thitima Sanpachudayan, Hathaichanoke Boonyarit, Chatchai Chinpaisal & Bhoom Suktitipat. 2008. Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *American Journal of Physical Anthropology* 137(4). 425–440.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2009. *Ethnologue: Languages of the world*, 16th edn. Dallas, TX: SIL International. (http://www.ethnologue.com).

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2016, *Ethnologue global dataset*, 19th edn. Harvard Dataverse, V1.

Li, Ruohui. 2000. Yindai liangci chutan [An initial exploration of classifiers in Yin Dynasty]. *Guhanyu Yanjiu* 2. 79–84.

Li, XuPing & Walter Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua* 122. 335–355.

Liao, Jia-Yu. 2014. Qufen keyu zhong zhi fenleici [Identifying classifiers in Hakka]. MA thesis, National Chengchi University, Taipei.

Liang, Min. 1983. Zhuangtong yuzu liangci de chansheng he fazhan [Emergence and development of classifiers in Tai-Kadai]. *Minzu Yuwen* 3. 8–16.

Lin, Wan-jung. 2006. Yinzhou jinwen shuliangci yanjiu [A study of classifiers in the bronze inscriptions in the Yin-Zhou dynasty]. MA thesis, Soochow University, Suzhou.

Liu, Shiru. 1965. Weijin nanbeichao liangci yanjiu [On the classifiers in Weijin Nanbeichao]. Beijing: Chunghwa Book Company.

Lu, Tian-Qiao. 2012. *Classifiers in Kam-Tai languages: A cognitive and cultural perspective*. Boca Raton, FL: Universal Publishers.

Manomaivibool, Prapin. 1976. Layers of Chinese loanwords in Thai". In Thomas Wilson Gething, Jimmy G. Harris & Pranee Kullavanijaya (eds.), *Tai linguistics in honor of Fang Kuei Li*, 179–184. Bangkok: Chulalongkorn University Press.

Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley, CA: University of California Press.

McWhorter, John. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.

Moral, Dipankar. 1997. North-East India as a linguistic area. *Mon-Khmer Studies* 27. 43–53.

Morev, Lev. 2000. Some afterthoughts on classifiers in the Tai languages. *Mon-Khmer Studies* 30. 75–82.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.

Norman, Jerry & Tsu-lin Mei. 1976. The Austroasiatics in ancient South China: Some lexical evidence. *Monumenta Serica* 32. 274–301. https://www.jstor.org/stable/40726203

Peyraube, Alain. 1991. Some remarks on the history of Chinese classifiers. *Santa Barbara Papers in Linguistics* 3. 106–126.

Peyraube, Alain. 1998. On the history of classifiers in Archaic and Medieval Chinese. In Benjamin K. T'sou (ed.), *Studia linguistica serica*, 131–145. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

Pokharel, Madhav P. 2010. Noun class agreement in Nepali. *Kobe Papers in Linguistics* 7. 40–59.

Phukan, J.N. 1978. *An introductory primer & grammar of Ahom (Tai) language*. Paris: Lacito-CNRS.

Ratliff, Martha. 2000. Numeral classifiers and classifying nominal prefixes in Hmong-Mien: One more story of contact and its grammatical consequences. Paper presented at The Tenth Annual Meeting of the Southeast Asian Linguistics Society, Madison, WI.

Ratliff, Martha. 2009. Loanwords in White Hmong. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 638–658. Berlin: De Gruyter Mouton.

Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.

Rijkhoff, Jan. 1998. Order in the noun phrase of the languages of Europe. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 321–382. Berlin: Mouton de Gruyter.

Saalbach, Henrik & Mutsumi Imai. 2011. The relation between linguistic categories and cognition: The case of numeral classifiers. *Language and Cognitive Processes* 27(3). 381–428.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–440.

Sagart, Laurent. 2005. Tai-Kadai as a subgroup of Austronesian. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*, 166–181. London: Routledge.

Schuessler, Axel. 2007. *ABC etymological dictionary of Old Chinese*. Honolulu, HI: University of Hawai'i Press.

Seifart, Frank. 2010. Nominal classification. *Language and Linguistics Compass* 4(8). 719–736.

Starostin, Sergej A. 1989. *Rekonstrukcija drevsnekitajskoj fonologicheskoj sistemy* [Reconstruction of the Old Chinese phonological system]. Moscow: Nauka.

Sussex, Roland & Paul Cubberley. 2006. *The Slavic languages*. Cambridge: Cambridge University Press.

Tang, Chih-Chen Jane. 1996. *Ta mai-le bi shizhi* and Chinese phrase structure. *The Bulletin of the Institute of History and Philology* 67(3). 445–502.

Tang, Chih-Chen Jane. 2004. Two types of classifier languages: A typological study of classification markers in Paiwan noun phrases. *Language and Linguistics* 5(2). 377–407.

Thurgood, Graham. 1994. Tai-Kadai and Austronesian: The nature of the relationship. *Oceanic Linguistics* 33(2). 345–368.

Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3(2). 215–246.

van Egmond, Marie-Elaine. 2008. Incorporated adjunct classifiers in Anindilyakwa: An empirical challenge to LFG. In Miriam Butt & Tracy Holloway King (eds.), Proceedings of the LFG08 Conference. Stanford, CA: CSLI Publications. (http://web.stanford.edu/group/cslipublications/cslipublications/LFG/13/papers/lfg08vanegmond.pdf)

Wang, Fushi & Zongwu Mao. 1995. Miaoyaoyu guyin gouni [Phonological reconstruction of Miao-Yao]. Beijing: China Social Sciences Press.

Wang, Jun. 1991. Language interaction in China. In William S.-Y. Yang (ed.), *Language and dialects of China*, 159–184. (Journal of Chinese Linguistics Monograph Series 3.) Hong Kong: The Chinese University of Hong Kong Press.

Wang, Lianqing. 1994. Origin and development of classifiers in Chinese. Ph.D. dissertation, Ohio State University, Columbus, OH.

Watters, John R. 1981. A phonology and morphology of Ejagham – With notes on dialect variation. Ph.D. dissertation, University of California, Los Angeles.

Weidert, Alfonso. 1984. The classifier construction of Newari and its historical Southeast Asian background. *Kailash: A Journal of Himalayan Studies* 11(3/4). 185–210.

Wen, Bo, Hui Li, Song Gao, Xianyun Mao, Yang Gao, Feng Li, Feng Zhang, Yungang He, Yongli Dong, Youjun Zhang, Wenju Huang, Jianzhong Jin, Chunjie Xiao, Daru Lu, Ranajit Chakraborty, Bing Su, Ranjan Deka & Li Jin. 2004. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution* 22(3). 725–734.

Wu, Fuxiang, Shengli Feng & C.T. James Huang. 2006. Hanyu 'shu + liang + ming' geshi de laiyuan [On the origin of the construction of 'numeral + classifier + noun' in Chinese]. *Zhongguo yuwen* 5. 387–400.

Wu, Yicheng & Yanzhi Li. 2016. Zhongguo yuwen lun liangci duanyu de youding jiedu yu liangci de lishi fazhan [On the nature of the definite interpretation of [Cl + N] expressions in Chinese and the historical development of classifiers]. Unpublished manuscript, Zhejiang University, Hangzhou.

Yakhontov, Sergej E. 1977. Jazyki Vostočnoj i Jugo-Vostočnoj Azii v IV-I tysjačeletijax do n.è. [The languages of East and Southeast Asia in IV-I millennium B.C]. In Nikolaj Čeboksarov (ed.), *Rannjaja ètničeskaja istorija narodov Vostočnoj Azii* [Early ethnic history of peoples of East Asia], 98–108. Moscow: Nauka.

Yi, Byeong-uk. 2009. Chinese classifiers and count nouns. *Journal of Cognitive Science* 10. 209–225.

You, Rujie. 1982. Lun Taiyu liangci zai Hanyu nanfang fangyan zhong de diceng de yicun [On the substrata and remnants of Tai classifiers in Chinese southern dialects]. *Minzu Yuwen* 2. 33–48.

You, Rujie. 1995. Zhongguo nanfang yuyan li de niao chong lei mingci citou ji xiangguan wenti [The prefixes of the names of birds and insects in the languages of south China and related questions]. In William S-Y. Wang (ed.), *The ancestry of the Chinese language*, 253–268. (Journal of Chinese Linguistics Monograph Series 8.) Berkeley, CA: Project on Linguistic Analysis.

Yue, Anne O. 2003. Chinese dialects: Grammar. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 84–125. London: Routledge.

Zeitoun, Elizabeth. 2007. *A grammar of Mantauran (Rukai)*. (Language and Linguistics Monograph Series A4-2.) Taipei: Institute of Linguistics, Academia Sinica.

Zhang, Gongjin. 1978. Lun hanyu ji zhuangdong yuzu zhu yuyan zhong de danweici [About the markers of singularity in Chinese and Chuang-Dong languages]. *Journal of the Minzu University of China (Philosophy and Social Sciences Edition)* 4. 14–30.

Zhang, Yachu (ed.). 2001. Yinzhou jinwen jicheng yinde [Index of the collection of inscriptions from the Yin-Zhou period]. Beijing: Zhonghua Book Company.

Zhou, Zhenhu & Rujie You. 1986. Fangyan yu zhongguo wenhua [Dialects and Chinese culture]. Shanghai: Shanghai Renmin Chubanshe.