

# ADMIXTURE ASPECTS OF VIETNAMESE CLASSIFIERS

V. U. Nguyen

## ABSTRACT:

Characteristics of Vietnamese Classifiers are examined under the framework of the Admixture Theory in which the language is treated as a continuum in both space and time. It is found that for the type of Classifiers investigated, the original arrangement used to be in reverse order, with the numeral and classifier placed after the noun. Under that category of word order, as in a number of languages in the region, the classifier is not placed next to the noun, but separated by the numeral. As a result, the classifier is allowed to assume a word identical to the noun, having the same meaning. Through admixing processes, the final word order most commonly used today in the language is that having the numeral followed by classifier and then noun. As such, the classifier is placed right next to and in front of the noun, and accordingly, the classifier generally has to assume an equivalent but different morphology, sourced from various constituent languages of Vietnamese.

It is also found that the Classifier may have many other roles in its formative stages, some being a marker to make distinction between cardinal number and ordinal number, a repeater in a constituent dialect, a misnomer, and a word forming pairs in descriptive phrase to the noun or noun phrase in the front, to satisfy disyllabic features of some constituent languages in old times. Construction of the pair of “Classifier and Noun” is very similar to that of many Compound words in that one element of compound word is sourced from one constituent language and the other, from another language or dialect.

The study also looked into homophonic features of Vietnamese lexicon hitherto obscured under both Nôm and alphabetic quốc ngữ, and presented detailed etymology of some of the common Classifiers of the language.

.....

It is widely known that classifiers are intrinsic in many Asian languages, and their use is intimately linked with the nature of noun classes, the most common being referents in quantity or entity. Interest in studies of Vietnamese classifiers, among others, appears to have been on the rise in recent years (e.g. [1] [2] [3] [33] [34]), and yet a universal view on their characteristics and etymology is apparently still outstanding.

In this paper, results of a short study on classifiers, based on a new theory on the origin of the Vietnamese [4], are presented, noting that the two commonly

known *Cái* and *Con*, are not included and to be reported separately. Basically, the theory was developed, using cultural, historical and linguistic data among others, around the popular folklore about the union of lady *Âu Cơ*, representing the Fairy, and lord *Lạc Long Quân*, symbolizing the Dragon [27]. The couple had 100 children fertilized from 100 eggs. Due to differences in ethnic background, however, the couple was separated, soon after the birth of their children. One half of the children later followed the mother (*Âu*) to the mountainous regions, and the other half were relocated with the father (*Lạc*) to the sea and river plains. The eldest son, on lord *Lạc*'s side, later on became the first king, known as *Hùng Vương*, who passed on the reign to his descendants by 18 generations. The theory uncovered a number of similarities between ancient histories, mostly through folklore, of Vietnam and other regions in Asia, notably China, Thailand, and Korea [4]. In practical terms, the theory is intertwined with many premises relevant to studies in Vietnamese, the most important being Vietnamese is a historical and evolutionary merger or admixture of a large variety of languages. It has a substratum of Mon-Khmer mixed with Munda, Thai, and Polynesian, and overlain by many old languages of the *Bai-Yue* (*Bách Việt*) group in Southern China, and reinforced by the Hmong-Mien and Hakka combination. Of importance to note [4], are contributions to Vietnamese lexicon from two major migrant groups, one being the Hakka from the Shandong peninsula, represented by the Ly dynasty (1010-1225), and the other, the Fujian – Teochew groups, by the Tran dynasty (1225-1400).

The basic assertion above has been tested against analysis on the etymology of personal pronouns, body parts, and basic colors [5] [6] [7], and the results are encouraging, on the proviso of a list of desiderata. Central to the desiderata is the assertion that Vietnamese as a language has had more than 95% historical existence in both space and time under various spoken forms, and its current written form using the alphabet is by no means a truthful reflection of any of its spoken forms in the past.

Analysis of Vietnamese classifiers as shown in the following, will therefore treat the language as a continuum, in space and time, with porous or fuzzy boundaries. The lexicon illustrated will be checked for similarity with their own alternative “ancient” forms listed in earlier dictionaries ([9] [18] [43]), as well as correspondence in many different languages or dialects, prior to alphabetic codification known as *quốc-ngữ*, which in a paradox way, is intrinsically and arbitrarily dualist. Such dualist standardization and classification of the language, if taken as bedrock premises by and large would easily lead to discarding or burying many essential characteristics of the language in its long history. The search for cognates or kinwords presented herein however is not exhaustive, and to avoid confusion, may not include many ethnic dialects currently spoken inside Vietnam, as kinwords available

from languages contributing to Vietnamese, closely related to internal dialects, already proved to be adequate.

To illustrate just one aspect of the present “admixture approach”, also called the Tree-and-Soil approach, consider word in Burmese, [lu-dó], and in Munda, [mijDa] and [murDa?]. In Burmese [lu-dó] means “person”, not differentiating between singular and plural, and likely of Mon origin, which is also part of the substratum of Vietnamese [5][6][7]. In Vietnamese [lu-dó] under monosyllabic pressure was split into [lũ] and [đồ] denoting “people” in a derisive sense: *lũ thất học* (illiterate mob), *đồ mất dạy* (ill-educated person(s)). Munda [mijDa] is directly equivalent to Vietnamese “*Miếng đất*”, meaning Piece of Land, apparently with only difference in alphabetic phonetization. Similarly “rainy season” is called [MurDa?] in Munda and *Mùa Nước* (water season), or *Mùa Mưa* (rainy season), in Vietnamese. It can be seen that [Mur] (Munda) is *Mùa* (season), and [Da?] is “*Đác*” or “*Nác*” or “*Nước*” (water), all used in different places of Vietnam, with *Nước* as the most popular one. It is also noted that initial of both [Đác] and [Nác], i.e. [Đ] and [N], are alveolar sound, but one voiced and the other, nasal, respectively. In spoken languages, voiced and nasal alveolar were very likely undifferentiable.

On contributions from the *Bai Yue* {Hundred Yue (Việt)} group, the present theory has a core assertion stating the demarcation between *Nôm* and Sino-Vietnamese (*Hán Việt*) lexicon is at best a fuzzy one. This assertion, shown previously in [4] [5] [6] [7], will have many important implications, and hopefully will be formally presented on another occasion. Many aspects and relevant examples of this assertion, nevertheless, will become apparent in the present analysis of some common classifiers, as follows.

For illustrative purposes, consider the following classifiers: *Cánh*, *Đám*, *Quyển*, *Khuôn*, *Con*, *Chiếc*, *Đôi*, *Cặp*, *Bữa*, etc., and place focus on their etymology or range of kinwords, one by one.

### **Cánh** tay (CL + arm)

First it can be noted that in Thai, Burmese, Cambodian and Mandarin, in particular, spoken form of consonant [t] is somewhere between [d] and [t]. For example: Mao Ze Dong and Deng Xiao Ping in pinyin, used to be coded as Mao Tse Tung (*Dong ~ Tung*) and Teng Hsiao Ping (*Deng ~ Teng*), respectively, under the old Wade-Giles system. Thus [tay] (arm) in Vietnamese, during the pre-quốc-ngữ days might not be quite dissimilar to [day] in Khmer in some sections of the community of “Vietnamese” speakers. “Tay” also has cognate in [thay] of the *Muong* language.

*Cánh* on the other hand, again under spoken forms of the language during pre-quốc-ngữ times, could likely be related to: [kǎen] Thai, [kǎan] Lao,

[met-Kai] Tamil, and [kum-si] Gorum (Munda), all meaning “arm”. That is classifier “*Cánh*” appears to be derived from other languages, with meaning similar to “*Tay*” (arm), but not recognised in present-day Vietnamese under that meaning.

### **Cánh cửa** (CL + door)

*Cánh* is also used as classifier for “Door” (Cửa), where [cửa] has similar sound in [kuE] 關 of the Wu dialect, which principally had origin in the state of Yue and Wu of the Spring-Autumn period (about 722-481 BC) located in the regions of the two provinces Jiang Su and Zhe Jiang, today. It is noted that pronunciation of [Yue] is equivalent to “*Việt*” in Vietnamese and [Yue] is written with the same character 越, as Yue in Yue Nan (Vietnam). Wu pronunciation for “door”, i.e. [kuE], in fact corresponds to [guan] in Mandarin (Quan – Vietnamese), and [gan] in Hakka, which is very close to “*Cánh*” in *Cánh cửa*. Another etymon of “*Cánh*”, in both sound and meaning, is 閤 [kang]<sup>4</sup> Mandarin, with Hakka and Cantonese pronunciation as [kong] (= door). In Tay-Nung dialect, which is related to Zhuang in GuangXi – China, ending [ng] is sometimes equivalent to [nh] in Vietnamese: *Càng* => *Cành* (tree branch). Other words denoting “door” having one phoneme close to “*cánh*”, though not monosyllabic, may include: [kadavu] Tamil and [katuba] Fijian. It should be noted that by assigning 門 [men] (Môn – Vietnamese) as the official Chinese word for “door”, under a one-to-one regime, meaningful search for etymology or metaphorology of “*cánh*” and “*cửa*” has been obscured for a very long time. *Cánh* in *Cánh cửa*, as classifier, again has origin in other languages but similar meaning as “*Cửa*”, its own referent “Door”.

### **Cánh rừng** (CL + Forest)

*Cánh* can be a classifier for Forest, wood or jungle:

Hôm qua, tôi bị lạc trong **cánh rừng** kia

Yesterday, I was lost in **CL forest** that => Yesterday, I was lost in that forest

*Cánh* as classifier for “*Rừng*” (forest) is likely related to: [kAnreG] Sora, [kun] or [kunDa] Gorum and Remo (Munda), and [Glay] Champa. Of interest to note is [kAnreG] in Sora (Munda), having a semblance of “*Cánh rừng*”, in pronunciation after monosyllabic split, like [MijDa] for *Miếng đất* above.

*Cánh* as a noun on its own may refer to “Wing” (bird’s wing), and metaphorically may also cover “side” (eg. The West Wing), like classifier for “forest” above. “Wing” could also be metaphor for “arm”, and vice versa. For example, Gorum word [dena] denotes both “wing” and “arm”. Words meaning “wing” under disyllabic form that has one phoneme similar to “*cánh*” include: [Kapha] in Korku (Munda), [paKura] Korwa (Munda), [\*kapak] or [kampak], or [kələpa?] Mon Khmer. It is interesting to note that Turkish word for “flap” or “wing” or “arm” is [kanat] or [kol]. Similarly for Hungarian “Kar”, or Albanian

“Krah”.

*Nôm* characters of *Cánh* principally relied on 更 [geng] (canh) for sound, and were written in combination with other radicals like [kou] (khẩu) (口), [mu] (mộc) (木), [yu] (vũ) (羽), [ji] (kỷ) (几), etc. It is noted that basically *Nôm* spelling does not provide any insight on the admixture aspects in the present discussion, especially those related to homophonic features of “*Cánh*”.

**Đám Tang / Đám Cưới (CL + funeral, CL + wedding)**

The search for etymology of *Đám* is perhaps the most complex of all, since, like *Cánh* above, *Đám* has long been overshadowed by its own alphabetic form with hidden homophonic features, which in turn were also obscured by *Nôm* spelling. It will be shown asunder that *Đám* was likely to be sourced from languages or dialects belonging to both the lower bound (substratum) and upper bound (waves of migration).

*Đám* etymologically, and in accordance with the present theory, can have three different but interrelated meanings: (i) “place” or “location”; (ii) “crowd” or “gathering”, and (iii) “ceremony” or “to moderate or perform ceremony”.

In Mandarin, *đám* is equivalent to [chang] 场, being a classifier for both “funeral” and “wedding” like Vietnamese, meaning “an open space” or a “site or place for a special purpose” [7]. Under the meaning of “site” or “place”, *đám* is also equivalent to the Thai classifier [dti] (likely to be cognate of 地 [di] meaning “place”, “location”, etc.) also used as counting word for “funeral”, and “wedding”: [ngaan sohp sawng dti] = two funerals. Other Thai words relating to “place” or “area” or “village”, and having sound close to “Đám” include: [thaan], [dtang], and [khaam]. In the Munda substratum, there are lexicon that have one phoneme component similar to *Đám*, denoting or relating to “place” or “site”: [tan] Gutob, [da:l] Kharia, and [Da(b)ko] Sora. Similarly, word for “place” in Tamil is [iDam]. In some *Nôm* dictionaries (e.g. [8] [43]) “*đám*” is written as 埗, having a variety of meanings in Chinese, but with meaning as “location” or “plot of land”, or simply covering the whole range of “*Đám*” under *Nôm*, and similar sound [diam] in Hakka. Measure word for “funeral” and “wedding” in Cambodian is normally [dawng], meaning “time(s)”, but still having sound close to *đám*. Cambodian classifier [dawng] for “funeral” and “wedding” has similar use to Chinese [ci] 次 (thứ), meaning both “time / order” and “place” among others, also used as classifier for “funeral” and “wedding” in many Chinese dialects. Likewise, Japanese classifier 回 [kai], meaning “time / order”, is also used with “funeral” and “wedding”.

Under meaning of “crowd” or “gathering”, or an “ensemble” or “assortment”, *Đám* is normally used either in compound form: *đám đông* (crowd), *đình-đám*

(gathering at the village community hall), or as a classifier (see [9]): *đám ruộng* (paddy field), *đám cỏ* (grass), *đám mạ* (rice seedling), *đám mây* (cloud), etc.

In *đám-đông* (crowd), or *đám người đi chợ* (shoppers), *đám* as a classifier, showed that conception of “noise” and “crowd” was practically lumped together, and both “*đám*” and “*đông*” in “*đám đông*”, would very likely share common meaning (crowded / noisy) from a wide range of thesaurus or cognates. As such, “*đông*” for “crowd” in the Tay-Nung dialect is identically [đông]. Word for “crowd” in Cantonese: [deoi] 堆, with sound close to [đông] and radical [tu] 土 on the left denoting “place”, could be a cognate of “*đông*”. Another with one Cantonese pronunciation close to “*đông*” is 口囀, pronounced as [dung] or [hung], having the [kou] (mouth) radical (口) on the left, denoting “people”, with meaning “noise from a crowd”. “*Đông*” (crowded) with a change in tone will give “*động*” (ie. *đông* added with *Nặng*-accent), meaning “noise”, also under metonymic transformation. Thai word for “loud” or “noisy” is [dang] or [deung], very close to both “*đám*” and “*đông*”.

Each element of compound word “*Đình đám*”, on the other hand, originally was sourced from *Bai-Yue* dialects, denoting first “village community hall” or “worship hall”. *Đình* is related to 庭 [ting] with one meaning as a “Hall”, and *Đám* very close to 壇 [taan] Cantonese, meaning “altar” or “hall for meeting or ceremony”. Over time “*đình đám*” took on meaning of “being noisy for the sake of showing off” or “something sensational”. In the lower bound of Munda lexicon, word for one’s own village is [DaoD] (Gorum), also having similar sound to *Đám*.

As a classifier for Rice seedling, Rice field, Grass, Cloud, Land, etc., *Đám* takes on meaning of an ensemble, variety, assortment, multitude, etc, and as such, would most likely have a cognate in Thai word [dtaang], or [daaen], or [daan], meaning: land, region, area, field, etc. For the word “Cloud” in particular, its classifier *Đám* (in *Đám Mây*) sometimes can be replaced by another classifier “*Áng*” which is derived from Minnan (Fujian) pronunciation as [ung] or [aong] of 雲 [yun] Mandarin (meaning “cloud”) [52]. In *Nôm*, “*Áng*” is written as 𣎵 [ang-4] [43], with completely different meaning in Chinese.

Searching for a commonality among *Đám tang* (or *Đám Ma*) (funeral), *Đám cưới* (wedding), *Đám hỏi* (engagement), *Đình đám* (community hall or gathering for some ceremonial purpose), and *Đám giỗ* (kỵ) (anniversary of the dead) – with focus on metaphor or metonymy of *Đám*, it can be seen that *Đám* has much to do with some kind of “ceremony” or rituals associated with, or performing, ceremony. As such “*Đám*” is related to Thai [dahm], and possibly, also Hmong [ua tawm], meaning “performing ceremony”, and Cantonese [daam], or Hakka [tam] 𣎵 𣎵, meaning “sacrifice offered on anniversary”,

where radical [shi] 示 refers to “spirit”, coupling with [tan 覃] for sound. In *Đám giỗ* (or *Kỵ*) (anniversary of the dead, commonly ancestors), *giỗ* has cognate, under monosyllabic split, in [karJo] Sora (Munda), and “*Kỵ*” in [ki] 忌 Hakka, or [gi] Teochew (MinNan), or [gei] Cantonese. Words in Munda relating to “ceremony” that have sounds similar to “*Đám*” include: [kung-Dak] (Korwa), meaning “ceremony taking place 10 days after death”, [Tal] (Gorum) meaning sacrificial ceremony, and [ru-Da?a] (Gorum) or [du:l-Da:ʔ] (Korwa) denoting a marriage ceremony involving water [Daʔ] (*đác* or *nác* (nước) in Vietnamese) being poured over the heads of both the bride and the groom.

*Nôm* spellings of *Đám* as shown in various dictionaries [10] [11] [43], as expected, do not reveal a great deal about its etymology, in various contexts. One way of spelling *Đám* under *Nôm* is to join the earth radical [tu] (thổ) (土) with the character 店 [dian] (điếm) (store / inn): 土店. Another is {站}, from Chinese [dian] meaning “stand for goblets”. And yet another is to combine the grass radical [cao] (thảo) (艹), or field radical [tian] (điền) (田), or mouth radical [kou] (khẩu) (口), or hand radical [shou] (thủ) (扌), with either [dian] above, or the right half of character [dan] {đam} (耽). Nevertheless, interpretation of *Nôm* spellings for *Đám*, in some ways seems to be consistent with the present discussion.

It can be seen from the discussion above, that *Đám* is a fairly versatile classifier due to its various homophonic features, and while in some applications *Đám* still retains similar meaning as its referent (e.g. *đám giỗ* - anniversary of the dead; *đám ruộng* – rice field), and in others *Đám* takes on the role of referential salience, as presented by Daley [1], and reviewed by Thurgood [37]. In the main, however its function as measure word may not be discounted altogether. For example: *Trên trời hiện có vài đám (áng) mây đen*: In the sky now, there are some (+ CL) dark clouds. Under *đám tang* (*ma*) (funeral), *đám cưới* (wedding), *đám giỗ* (*kỵ* - anniversary of the deceased), meaning of *đám* as something to do with “ceremony” is still apparent through its usage in utterances such as: *Tháng tới, chúng tôi sẽ làm đám cưới cho đứa con trai*. (Next month, we will get/ have marriage ceremony performed for our son). In such utterance, the use of verb “*làm*” (make/do/get/have) preceding ‘*đám cưới*’ (CL+marriage) indicates strongly that the whole verbal phrase “*làm đám cưới*” would incorporate “*đám*” having the sense relating directly to “ceremony” or “moderating ceremony”.

### **Cuốn / Quyển:** Classifier for *Sách* (Book)

Under conventional view, *Sách* is often considered as a *Nôm* word, with Sino-Vietnamese equivalent “*Thư*”, standing for Chinese [shu] 書. Chinese classifier for [shu] is [ben] 本, which also means “book”. In fact, “*Sách*” is also of Chinese origin, according to *Vân Đài Loại Ngữ* [12], an encyclopaedic treatise by Lê Quý Đôn, one of old Annam’s foremost scholars (1726-1784),

and written as 策 [ce] Mandarin, with similar pronunciation: [chaak] in Cantonese, [tsak] in Hakka, [tshâ?] Wu, and [chhek] Minnan (Fujian) (Sino-Korean: [chayk]). “*Sách*” {[tsak] 策 or 冊}, in its original meaning, denotes a series of “bamboo pages” (called “*giản*” [jian] 簡), each containing about one line of characters, bound together by strings.

*Cuốn* and *Quyển* as classifiers for *Sách*, like [ben], in original meaning, may also refer to “book”, and are pronounced similarly in Mandarin [juan] 卷, Hakka [kian], Cantonese [gyun], and Minnan [koan]. Combination of Hakka [kian], Mandarin [juan], and Cantonese [gyun] with the assistance of *quốc-ngữ* codification would easily result in “*quyển*”, and Minnan [koan], in “*cuốn*”.

*Cuốn* and *Quyển*, as classifiers for Book (*Sách*), therefore demonstrated that both classifier and referent noun, shared the same meaning at some time in the past, and merged into the language by admixture process. It is noted that *Cuốn* has a cognate as *Cuộn*, meaning “a roll” exactly like 卷 [juan] Mandarin, [koan] Fujian, [gyun] Cantonese above, and which could serve also as a classifier: *Hai cuộn giấy fax* (Two fax-paper rolls).

#### **Bộ / Khuôn / Gương: Classifier for *Mặt* (Face)**

In the Champa language, words for ‘Face’ are [muta] and [bo?], with [bo?] related to ‘*bộ*’ often used in compound word ‘*bộ mặt*’ in Vietnamese. *Bộ* is also a classifier as in “*nhiều bộ mặt quen thuộc*” (lots of familiar faces), with application as counting word. In the Sora dialect of Munda, word for intimate friendship is [a-Bo-Mugka], literally meaning “one face”, involving syllables resembling both *Bộ* and *Mặt*.

*Khuôn* and *Gương* are also two other alternative classifiers for “*Mặt*” (face). As seen from [15], it is clear that *Khuôn* and *Gương* are different pronunciations of a common word denoting the “face”. *Khuôn* has similar sound to Japanese [kao], whilst *Gương*, to Korean [olGul]. A closer search [13] will reveal that there is another word for “face”, but specifically denoting “face of an elderly”, written as 耆 [gou] Mandarin, pronounced as [kou] Japanese, with equivalent [gung] or [gang] Hakka (very close to *Gương*), and [k<sup>h</sup>əŋ] Wu (*Khuôn*). *Khuôn* could also be connected with the second syllable of [muKhO] Juang (Munda) following monosyllabic separation, noting that in sound [muKhO] is similar to “*mặt khuôn*” or “*khuôn mặt*” in reverse order. The same can be said to be applicable to Chinese compound word for the face: 面孔 [mian-kong], where [kong] is likely to be connected with both “*gương*” and “*khuôn*”, in pre-alphabetic times.

All common classifiers for *Mặt* (Face), ie *Bộ*, *Khuôn*, and *Gương*, thus have meaning closely related, or identical, to *Mặt* itself, but were derived from different dialects (or languages) of the region.

**Trận / Con:** Classifier for *Bão* (Storm, hurricane), *Mưa* (Rain), *Lụt* (Flood), *Gió* (Wind), ...

• **Trận**

Consider:

*Trong tháng qua, thị xã này đã chịu 3 trận bão:*

During the last month, this town suffered 3 storms

In that *trận* is used as classifier, deriving directly from Chinese 陣 [zhen], meaning “a certain duration of time for an event - like storm, rain, flood, etc - to take place.” (e.g. [13]). Its Cantonese pronunciation is [zhan], and Hakka, [chin], showing similar sound correspondence between Vietnamese surname “Trần” (with tone difference) and Chinese [zhen] 陳, and Cantonese [chan].

*Trận* is thus a common classifier for many an event, generally catastrophic, taking place over some duration of time, like storm, wind, rain (heavy), flood, fire, etc. It is extended also to cover “earthquake” (*trận động đất*), of which Chinese equivalent is [di zhen] 地震 (địa chấn), with the second sound [zhen] being close to [zhen] 陣 above, being classifier. With origin from the North, *Trận* showed some sense of generalization, with meaning further distanced from those of referents.

• **Con**

Different from *Trận*, *Con* appears to have sound similarity with words from other languages or dialects related to “old” Vietnamese, giving similar meaning.

- For *lụt* (*lũ*) (flood): [gong] 潑 in Hakka, and [gun] 灌 in Cantonese, have similar sound to *Con*, also meaning “flood”.
- For *bão* (hurricane, cyclone): [geoi] 颶 or [kyun] 風工, both in Cantonese, meaning *bão*, but with similar sound to *Con*.
- For *gió* (wind): *Con* as classifier for *Gió* (Wind) came about, most likely, from monosyllabic split of [\*kazal] or [\*kanh'jal] (into: con + gió) (Mon-Khmer / Proto-Austic [14]), with Japanese equivalent: [kaze], or [khyo'l] in Cambodian. *Con* in *Con-gió* could also be related to the first syllable in [gor-Zey] Gorum, [koyo] Juang, [koyo?] Korku (all Munda), or came straight from [\*cal] (Proto-Mnong), [kial] Bahnar, or 風工 [kyun] Cantonese, all meaning “wind”.
- For *mưa* (rain): Like for *Gió* (wind) above, *con* for *mưa*, very likely could result from monosyllabic separation of [\*qami(a,q,s)] (P.Austic) [14], [qmia] Halang, [kum'a] Ruc (Mon-Khmer), or be related to [gAnur] Sora, [gur] Gorum, and [gum] Gutob, from Munda dialects.
- As a result of being in existence under spoken forms over a long period of time, *Con* sometimes can be used interchangeably with “*Con*” which is a general “animacy” classifier, to be dealt with in another paper. For example: High tide = *Con nước lớn* = *Con nước lớn*.

- For *hoả hoạn* (fire): *Con* could be related to 燿 [gwon] Hakka, [gun] Cantonese, 焜 [kun] Hakka, 火果 [gun] Cantonese, or remotely [ajCo] Korku (Munda), all meaning “fire”. Another word with similar sound is [guj] in Juang and Kharia (Munda), meaning “fire” or “to light a fire”. This [guj] in Munda is equivalent to 燬 [kui] Cantonese, and [củi] in Mường, meaning “fire”. “*Nhóm củi*” in Mường can be used interchangeably with “*nhóm lửa*” in Vietnamese (where *lửa* is “fire”), and over time “*củi*” took on meaning of “firewood” (used to light or to start a fire). Another classifier for “Fire” in Vietnamese, but used in collocation with “*cháy*”, is “*Đám*” (“a fire” is called *đám cháy*) discussed above, in the category applied to an ensemble like: Grass (*đám cỏ*), Cloud (*đám mây*), paddy-field (*đám ruộng*), etc., where “*cháy*” is likely a cognate of 焯 [tseoi] in Cantonese.

*Con* is also a classifier for words relating to sickness (*bệnh*) or an attack, like asthma attack, heart attack, seizure, etc., and in this case, *Con*, is likely related to [Kosu] Kharia (Munda), 瘡 [gon] Hakka, 瘵 [kwan] Hakka, all meaning “illness” or “sickness”. Sino-Japanese word equivalent to Hakka’s [kon] or [kwan] above is pronounced as [kan], and Cantonese pronunciation for “*suyễn*” (asthma) is [tşyn] 喘, corresponding to [chon] or [tson] in Hakka [13].

*Con* can also refer to Mood, like an Angry Mood (*Con giận*), and here *Con* has an equivalent in Khmer [keng] meaning “anger”. It is also interesting to note that each of the phonemes in *Con Giận* is connected to 愠 [gyun] Cantonese (*Con*), and [gian] Hakka (*Giận*), and strictly speaking, with no direct equivalent in Mandarin. Another compound word for “Anger” (or “to be angry”) is “*Cấm gan*” or “*cấm giận*”, with “*gan*” normally taken as “*Gan*” denoting the Liver. In fact, both morphemes of “*Cấm Gan*” are *quốc-ngữ* transliteration of lexicon among various Southern Chinese dialects denoting “anger”. *Cấm* could come from Hakka [kam] 鬪, and “*Gan*” or “*Giận*”, also from Hakka [gan] (Hailu, Siyan) or [gian] (Meixian) 愠, all meaning “angry” or “mood or temper arising from anger”, with sound almost identical to *Gan* and *Giận* in Vietnamese. Interestingly, connection between the Liver and human emotions is most prevalent in many Miao (Hmong) dialects, noting that in ancient times [51], Miao-Yao (Hmong-Mien) and Hakka nomadic groups were known to have mutual affinity, and their settlements, were very close to each other [4] [39] [40]. It is also reported that many of the lexicon are shared between the Miao and Hakka groups [39]. In Miao dialects, “happy” can be expressed as [zoo sab], literally meaning “good in the liver”, and “angry” is [chim sab], “agitated or sad in the liver”.

It should be stressed that the type of lexical transfer in “*Cấm gan*” or “*cấm giận*” (being angry), from Hakka dialects to Vietnamese illustrates “transfer as an ensemble” embodying different dialect pronunciations as well as deep structure of conception and metaphory. Examples abound regarding this type of

lexical transfer, from contributive dialects to the national language which was in the making, and some presented in [6] [7].

**Chiếc:** Classifier for *Chiếu* (mat), *Xe* (Car, cart), Singular object, etc.

Etymological analysis of Vietnamese classifiers or Vietnamese lexicon in general is normally carried out with an inherent “bedrock assumption” about a one-to-one régime between word and meaning, in the language. In fact, from examination of classifier lexicon thus far, it is seen that Vietnamese word may involve a multitude of homophones, where, unlike English and Chinese, most homophonic words conform to exactly the same standard spelling. Examples abound and include *Cánh*, *Đám*, *Con*, above, and *Chiếc*, as follows.

- As classifier for *Xe* (Cart, vehicle, car, ...), **Chiếc** was derived from [chiaa] (車), under Southern Fujian (Minnan) pronunciation [44], meaning: car, cart, bicycle, or car, etc., exactly like *Xe*, and corresponding to [tse] Cantonese, and [tsa] Hakka (Sino-Viet: *Xa*).
- *Chiếc* as classifier for Boat, Ship (*Chiếc thuyền*, *chiếc tàu*) could have cognates in Southern Chinese dialects, such as: [zaak] 舢, and [zoek] 棹 both Cantonese, having similar sound to “*Chiếc*” (*quốc-ngữ*), {[chel] or [chek] or [chayk] in Korean}, and meaning Boat or Ship. Other possible cognates may include: [tsit] Cantonese, written with [zhou] (boat) radical {舟} replacing [che] (vehicle) {車} in [che] or [zhe]: 轍; or [cha] 舢, all meaning Boat. *Nôm* character for *Chiếc* used for Boat is normally written as: 舟隻 being “Chu 舟 + Chích 隻”. Another *Nôm* character commonly used for *Chiếc* is 析 [sit] Hakka (or *chiết* in Vietnamese), meaning “to separate”, showing that *Nôm* writing relies much on sound copying, and could not reveal or uncover hidden conceptions in the lexicon examined.
- As classifier for *Chiếu* (sedge mat), **Chiếc** was sourced from Teochew (Southern Fujian) [chao chiat] 草席, ie. “*chiếc chiếu*” in reverse order, with [chiat] (*chiếc*) 席 {Hakka [tsit]}, being the main word for “mat”. “*Chiếu*”, pronounced in Minnan as [chao] 草, actually means “grass”, or “dried grass”, being material to make sedge mat. Vietnamese word (conventionally: Sino-Vietnamese) for “Grass” is “*thảo*” which bears strong resemblance with [tshɔ] in Wu (ie. Wu-Yue or Ngô Việt), and has a *Nôm* equivalent in “*Cỏ*”, having a cognate with [Co] in Fijian. “*Chiếu*” and [Chao], interestingly also follow an observation propounded in [4] which states that: “*If a word from one Chinese dialect has phoneme [ou] (or [au]), there is bound to be a corresponding word from another dialect, which will have phoneme [iu] in place of [ou] (or [au])*”. Example: [Zhou] 周 (as in Zhou En Lai), pronounced with [ou] in Mandarin, corresponds to [Zhiu] in Hakka, and [Chiu] in Teochew (Minnan). Tak Wa *Lau* (actor-singer Andy Lau), as pronounced in Cantonese, is called De Hua *Liu* 德華劉 in Mandarin. [Hu zhao] 護照 {Passport} in Mandarin, is [hu chiu] in Teochew (Minnan – Southern Fujian) and *Hộ chiếu* in Vietnamese.

Vietnamese word for breeze “*hiu-hiu*” could be said to be closely related to [hau-hau] in Rapanui and Maori. Difference in sounds between “Chiếu” and [Chao] (grass) can be thus explained from the established sound correspondence between [iu] and [ou]. Vietnamese *quốc-ngữ* in fact tried to compromise the two phonemes [au] and [iu] by devising the diphthong [ru] for both: [Lau] / [Liu] => [Lru]. Vestiges of “ancient” dialects still remain however with northern accent being close to [Liu], whereas southern, closer to [Lou]. In some, also, correspondence between [iu] and [au] is still retained intact: *kiêu ngạo* (vain) => *cao ngạo*.

- *Chiếc* in *Chiếc Chiếu* (sedge mat) may still appear to have some etymological complications, as the language had been in existence over a very long period of time under spoken forms. *Chiếc* could be an homonym of a word denoting “singleton” as opposed to a pair (*đôi / cặp*), normally in compound form, *đơn chiếc* (單隻) (Chaozhou: [duan-chek], Tay-Nung [tan chich]), meaning “single”, and used as a classifier for a singleton, such as: *chiếc đũa* (one chopstick), *chiếc bông tai* (an earring), *chiếc giày* (one shoe), *chiếc chiếu* (a sedge mat), *chiếc gối* (a pillow), etc. In fact, Chaozhou word for number 1 (one) can be [cheç] 蜀, which is likely another kinword of “*chiếc*” (singleton) [38]. In Chinese, as well as in Nôm (pronounced as *Chích*), this *Chiếc* is written as 隻, pronounced as [zhi] in Mandarin, and [chak] in some dialects of Hakka. In Chinese it is used as a classifier for hen, bird, pigeon, ox, goat, hand, foot, etc., and in Hakka dialect [chak] (or [zhak]) is a general classifier, equivalent to [ge]4 個 in Mandarin. Another reason for *Chiếu* (sedge mat) to employ a singleton classifier like *Chiếc* is because in olden times, *Chiếu* [17] mostly was made for individual use, ie. no queen-size but only single sedge mat. This is evidenced from one verse of a popular folklore where a male proposed to a female, by using metaphor:

Giúp em *đôi chiếu* em nằm

((I will) help you with *a pair of sedge mats* to lie on)

*Đôi* chăn em đắp *đôi* tằm em đeo [16]

(*A pair* of blankets to sleep with, and *a pair* of earrings to wear)

- *Chiếc* can also be used as classifier for Coat, Shirt, ... (*Chiếc Áo*), again as a singleton contrasting a suit (or *un complet*). In sound *Chiếc* (in *Chiếc Áo*) is likely related to [jas] in Malay and Indonesian, with meaning Shirt, Coat, which in turn may reflect word from a dialect in the past that was also present in Vietnamese.

### **Đôi / Cặp: Pair of...**

*Đôi* and *Cặp* belong to a different category of classifiers compared to those having the same meaning but originating from different dialects, like *Chiếc*, *Cơn*, *Cánh*, *Cuốn*, *Quyển*, *Bộ*, etc. above.

*Đôi* and *Cặp* can assume the role of a noun, like “pair” in English; for example:

*Đôi đũa* (pair of chopsticks), *đôi giày* (pair of shoes), *cặp chim* (two birds), *cặp vợ chồng* (a married couple), etc.

Both *Đôi* and *Cặp* are strongly connected to lexicon external to Vietnam, but intimately linked with old “Vietnamese” during formative stages. Cognates of *Đôi* include: [doh] Persian, Hindi, [dui] Bengali, [toh] Thai, [dui] Mandarin, [deoi] Cantonese, [dua] Malay, and [duih] Fujian, and remotely: “deux” (French), whereas *Cặp* on surface and under the meaning of “Pair” appears to be related to: [khuu] Thai, and [goo] Khmer. *Cặp* under close examination, however, may have connotation not merely as a “Pair”, but closer to a “Couple”, and akin to [kau] 媾 in Cantonese (*cặp vợ chồng* (a married couple)). The Huỳnh Tịnh Paulus Cửa dictionary [18] gave *Cặp* some declensional relationship with *Kẹp* (to clamp / to couple), and as such, *Cặp* was most likely derived from [kap] 合 Hakka and Cantonese, meaning: to combine, to join together; later extended to assume the role of a classifier or a noun (pair, couple), being measure word in application.

### **Tấm and Miếng: Piece**

*Tấm* is another classifier involving many homophones. Primarily it refers to (small) broken pieces, and is likely related to: [zam] or [tsaam] Cantonese, [tug] Sora (Munda), [tundu] Hindi, [dtawn] Tamil, and [dom] Khmer. *Tấm* can also denote some flat object, like Mat or Carpet (*tấm thảm*), and in that case *Tấm* has [diam] as a similar classifier in Hmong.

As classifier for Paper (*tấm giấy*), *Tấm* has kinword in 單 [dan] or [sam] Hakka, meaning “sheet of paper”, with sound correspondence [tE] in Wu, transliterated under alphabetic quốc-ngữ as *Tờ*, another classifier for Paper (*tờ giấy*) [41].

Although *Tấm thảm* (Mat or Carpet) often appears together as compound noun, *Tấm*, when analysed separately as a classifier, would have many cognates with similar sound and meaning (as Mat or Carpet), but in different but related languages or dialects: 簾 [tiam] Hakka, 毯 [tan] Mandarin, [taam] Cantonese, and 筴 [tan] Hakka. Character 筴, pronounced as [tim] in Cantonese, meaning “mat made of rushes or bamboo”, has another pronunciation as [daat] (Cantonese), having sound close to “*Đất*” in Vietnamese, meaning Land, Earth, Soil. In Cantonese, [daat] 筴 is a classifier for use with Places and Lands. This brings us to *Miếng* and *Mảnh*, being alternative classifiers for “Piece”.

As a Piece of Land, *Miếng Đất* was likely derived from disyllabic [Mij-Da] in Sora (Munda), referring to the same. *Miếng*, in use with Land, could also have cognate in 土眉 [Mi] Hakka, denoting Land allotment [13].

*Miếng*, in collocation with Cake (or Bread), *Miếng Bánh (Mi)*, could be related to 饅 [maan] (*mảnh*) in Cantonese, denoting Bread, or Buns. Another classifier

for Cake and Bread is *Mẫu*, having equivalent in 饜 [mo] in Mandarin, Hakka, and Cantonese, meaning Bread, and in 璫 [mau] Hakka, with original meaning as “piece of precious jade”. *Bánh* (cake or bread), on the other hand, is close to [Baan] 粿, Cantonese, or [bpang] Thai. Another classifier for Cake and Bread is ‘Ô’ (loaf), having cognate in [haaw] Thai, and [Olag] Juang (Munda).

*Mảnh* is another classifier denoting Piece: *Mảnh vải* (piece of cloth), *mảnh gỗ* (piece of wood), etc. In *Mảnh vải* (fabric), *Mảnh* could originate from [muaan] in Thai, being a classifier meaning Roll (of fabric), and [meu] Hakka, and [maau] Cantonese [13] meaning a thick raw silk fabric. *Mảnh* in *Mảnh gỗ* (timber / wood), on the other hand, may have sound correspondence in [maai] Thai, [mai] Lao, and [maram] Tamil.

### **Bữa (ăn): Classifier for Meal**

Bữa is a common classifier for Meal:

3 **bữa** ăn mỗi ngày (or: *Mỗi ngày 3 bữa ăn*): 3 meals a day

Bữa has a cognate in Juang (Munda): [Bua], meaning Paddy, Grains, Rice Plant. It could also come originally from [Satiag Bua] denoting a Meal of “roasted and fine grains”, and could be related to [Bi:] meaning “being full after eating” (Korwa – Munda).

### **Thức / Đồ / Món (ăn): Food, Dish**

*Thức* is an alphabetic and tonal declension form of *Thực*, meaning To Eat or Food. Closest form is Cantonese [sik] 食 or Hakka [shit]. *Đồ*, more frequently used in the southern dialects, could be related to: [don] in Gorum (Munda) meaning “to take or to accept food”, or 登 [den] Hakka ([tou] Japanese) denoting ancient vessels to contain food, or 餽 [dou] Mandarin for “food items on display” [13].

Another classifier used to denote a dish, or a type of food prepared, is *Món*, which seems to have kinwords with [mAnur] in Sora (Munda), or 饜 [mung] (dish filled with food) in Hakka and Cantonese, and 皿 [men] Hakka and [ming] Cantonese, meaning dish, plate, or saucer.

### **Toà (nhà cao tầng): Classifier for Building**

*Toà* can be used for Big house (*Toà Nhà Trắng (Bạch Ốc)*: the White House), or Building (*Toà cao ốc*), generally with more than 2 storeys. Its origin is that of a “misnomer” formed on lexical transfer, like *Ô* and *Dù* for Umbrella from [Yu] (rain) [29], or *Ly* (glass, cup) from [Bo Li Bei] (see [29] and Table I), as it came from ChaoZhou (southern Fujian) compound word: [Tua Lao] 大樓 (*Toà Lầu*), equivalent to Mandarin [Da Lou] (*Đại Lầu*), with [Da] (Big), and [Lou] (Multi-storey Building). *Toà* actually is a *quốc-ngữ* transliteration of spoken monosyllabic word [Tua] having Fujian origin, being merely an adjective, with original meaning: Big (*Lớn, To, Bự*). Strictly speaking, *Toà* is *To* (Big) in

Vietnamese ([dtoh] in Thai), and *Toà Lầu* or *Toà Nhà*, is just *Lầu To* or *Nhà To*, in reverse order.

An antonym of *Toà Lầu* is *Túp Lều* meaning a little Cottage or a Hut, with *Túp* a Thai word [dtuup], and *Lều* from 寮 [liu] Cantonese, [liau] Hakka, or 廬 [leoi] Cantonese, all sharing the same meaning “Hut”.

In *Toà Án* (Law Court) or *Toà Đô Chánh* (City Hall, Mayoral Office), *Toà* is a homophone of *Toà* in *Toà Cao Ốc* (building). However, the Taberd dictionary [43] gave Nôm spelling of *Toà* (in *Toà Án*) as 座 [zuo] (toạ), denoting a Seat, and thus *Toà Án* makes sense as “a seat for judgement” (tribunal). Under Nôm however *Toà* as “Big” above, or as “Seat”, has only one spelling {座} (toạ) [43], which would, as usual, completely obscure its role as an adjective for “Big” from the Teochew (Minnan) dialect.

### **Cuộc (Đua): Classifier for Race, Competition**

Classifier for Race, Competition, Contest, Wrestle, etc. is *Cuộc*, having kinword as [gok] Cantonese and Hakka, and [ko?] Wu, written as 較 or 角, pronounced as [jue] or [jiao] in Mandarin. In Chinese *dialects*, it has the same meaning as “*Đua*” (race, competition), equivalent to, and having similar sound with [du] Hakka {賭}, or [dyut] Cantonese, [duo] Mandarin {奪}, or [diu] Hakka, [dou] Mandarin {鬪}, which comes close to Vietnamese “*Đấu*”, having the same meaning. As a classifier *Cuộc* again falls into the category of alternative synonym with its referents, but from different dialects.

### **Cụm / Mớ (Tù, đồ ăn, hoa,...): Words, mouthful, collection, cluster, etc**

*Cụm* comes directly from Thai classifier [Cum] used for Words and Mouthful of food, and is also related to another classifier [gloom] in Thai, extended to cover Group, Assortment, Collection, Team, Cluster, etc. Sometimes *Cụm* is used with Cloud (*cụm mây*) as *Đám* and *Áng* above. As to Chinese dialects, *Cụm* could be related to [kiun] Hakka, and [kwan] Cantonese: {群}. In old Vietnamese (see [5] [9]) and among the southern dialects, *Cụm* could be used interchangeably with *Mớ*. *Mớ* has kinword in Khmer as [mat] and Thai as [muu]. Examples: *Mớ đồ ăn còn lại* (left-over food), *mớ kiến thức* (knowledge), *mớ chữ nghĩa* (literary training), *mớ quần áo dơ* (dirty clothes). *Mớ* in turn could be pronounced, or alphabetically coded, differently as *Máy*, or vice versa: *Máy người* (you people), *Máy cuốn sách trên bàn* (Those books on the table). *Mớ* and *Máy*, in many contexts, however could also be taken as Plural word.

### **Lời / Tiếng (Word, Speech): Classifier for Thanks, Statement, etc**

*Lời* and *Tiếng* are classifiers for Word, Speech or Statement, e.g. *đôi lời cảm tạ* (few words of thanks), *tiếng xin lỗi* (apologies), etc. *Tiếng* as classifier may be just a homophone of the other *Tiếng* denoting language (Tongue), as in *tiếng Anh* (English). As such *Tiếng* could have kinword in 稱 [tsih] Hakka, [tsing]

Cantonese, [ts<sup>h</sup>əŋ] Wu, meaning (to) Say, State, Commend, or associated nouns; or in 暢 [t'iong] Hakka, [tseong] Cantonese, meaning Statement, to Elaborate, to make clear. While not being exhaustive, *Tiếng* could be related to Thai [thaawy] meaning words, statement, or [dtaawng waa] denoting “something must be said”. *Lời* is more widely used than *Tiếng* as it could cover written words, like *Lời Mở Đầu* (ProLogue), *Lời giới thiệu* (Introduction), etc. Under the present theory, *Lời* and *Lưỡi* (Tongue, Langue (French)) are metaphorically related, and *Lời* could have kinword in Thai [glao] (say, declare), or [laa] (say goodbye), or [lin lohm] (eloquence ~ *lanh lợi*). From old *Bai-Yue* groups, *Lời* as in *Lời Cảm Tạ* (*Lời Cảm Ôn*) (Word of Thanks) could be derived from identical [Loi] 禮 in Chaozhou (i.e. Teochew – Southern Fujian) denoting etiquette, civility, in the phrase [gio gam loi sia] 叫感禮謝, equivalent to Vietnamese: *Gọi* (or *Gởi*) *cảm lời tạ* - or in a more standard way: *Gởi lời cảm tạ*. It should be noted that 禮 [Loi] in Viet and Chaozhou, is pronounced as [leo] in Cantonese, and [li] in Mandarin, same as [li] 利, meaning interest, advantage, or profit, showing an “arbitrary” change in Vietnamese tones: *Lợi* (advantage), *Lời* (interest, profit), which may reflect variation in tones among different sub-dialects in a single “Southern Chinese” dialect, like Fujian.

#### **Bàn (Ủi, Là, Chải, Tay): Classifier for Iron, Brush, Palm, Sole**

As a noun *Bàn* is *Cái Bàn* (Table), which could be a ChaoZhou word [bang] for Table, with Mandarin equivalent [ban] 板, meaning wooden board or plank.

*Bàn*, in another homophonic form, can also be classifier for Iron (for ironing clothes), i.e. *Bàn Ủi* / *Bàn Là*, deriving from [ban] Hakka 板, or [baan] Cantonese, referring to “an iron plate”. It is noted that *Ủi* in *Bàn Ủi* (iron) has sound correspondence in [Uui] 熨 Hakka, [ʔuht] Fujian, [ʔut] Khmer, [Uʔamea] and [ʔauli] Samoa, whereas *Là* in *Bàn Là* (iron) (northern) corresponds closely with [lek] in Thai, meaning “pressing clothes”. *Bàn Ủi* can be used as compound noun, and as such its classifier is the inanimate “*Cái*” (or at times *Chiếc*): *Cái bàn ủi (là)*.

*Bàn*, as classifier for Brush (*Bàn Chải*), has word with similar sound and meaning (Brush), as [Bpaang] in Lao, and [Biang] Hakka, whereas “*Chải*” (to brush) corresponds to: 擦 [chaat] Cantonese, [chal] Korean, and 漱 [chiu] Chaozhou (southern Fujian).

*Bàn* can be regarded as classifier for “Sewing Machine” in *Bàn Máy May* (sewing machine), which came directly from Cantonese [bang] 縫, meaning “Base for mounting sewing machine”. Like *bàn ủi* (iron) above, *bàn máy may* (sewing machine) can be used together as a noun, and in such case the classifier is the inanimate “*Cái*”.

*Bàn* in the old days, was also used in compound word *Bàn Tọa* [43], denoting

the Buttock (bottom), with focus on its use in sitting. In fact both *Bàn* {班} [ban] and *Tọa* {座} [zuo] may have the same meaning: Seat.

Homophonic characteristics of Vietnamese lexicon again can be seen through *Bàn* when it is used as classifier or part of compound noun for Palm (*bàn tay*) or Sole (*bàn chân*). As Palm, *Bàn* has kinwords as: [pal] Chamic, [palad] (Tagalog), [pang], [kopang] Mon-Khmer, [Pan] Mường, all denoting Hand in general. Word from Munda dialects [\*panlema] has the first syllable resembling *Bàn* with direct meaning “Palm of hand” [14]. Under polysemy, however *Bàn tay* or *bàn chân* could be used to denote Hand or Foot, entirely. Further distinction on Palm (or Sole) can be made by yet another classifier called “*Gan*” or “*Mu*”: *gan bàn tay* (palm), *mu bàn chân* (sole). In *Gan bàn tay*, “*Gan*” has kinword in [Gata] Remo (Munda), whereas *Mu* in *Mu bàn chân* has cognate in [pOdu**Mo**] Juang (Munda).

### **Lá (phối /gan): Classifier for Lung / Liver, and others**

*Lá* under *quốc-ngữ* appears to have only one principal meaning as “Leaf” (Leaves). As a classifier for use with body parts, *Lá* is often obscured by the missing link between its different etyma or cognates, normally disyllabic, during its long past in spoken forms, and that under *quốc-ngữ* codification (as *Lá*) within the last two centuries or so. Due to the missing link, the search for *Lá*’s origins, as classifier, like many others examined above, often proved to be futile or resulted in further misconceptions, relating mainly to *Lá* being used to denote thin flat objects, metaphorizing tree leaves, such as *lá thư* (letter), *lá đơn* (application form) [22]. This is reflected in Nôm characters for *Lá* coded only to denote “leaf”, e.g. 蕻 [10] and 淚 [8].

*Lá* in fact is a very common classifier for body parts. It can be used with Lung, Liver, Spleen, Tongue, Nose, compartments of ruminants’ stomach, such as: reticulum, omasum, abomasum.

As classifier for “lung” (*lá phổi*), *Lá* is used mainly to differentiate between the left and right lung, or as a general compound noun, and like many other classifiers presented here, it could be derived from many disyllabic sources within languages or dialects contributing to the formation of Vietnamese in the long past. *Phối* for Lung could be matched with [fo:y] Mường (Thanh Hoá) [14], [phoksa] in Kharia (Munda), [fai] Cantonese or [fei] 肺 Mandarin (phế), whereas *Lá* resembles one syllable of: [koLoz] Gutob (Munda), [aoLe] Juang (Munda), [buruk-Le] {lung / heart} Remo (Munda) [20], and [paLay] Alak (Mon-Khmer). It should be noted that like Vietnamese “bụng” and “lòng” (belly, heart, stomach, or abdomen, in general), most lexicon in Munda or Mon-Khmer relating to “lung” is also used to denote Liver, or Heart, and vice versa [24], within the ambit of “polysemy” (e.g. [32]). In the Munda dialects, however, ending [le] as in [bati-Le] (lungs) {Remo} is often used for the plural.

As classifier for “liver” (*lá gan*), *Lá* in the long past, could be related to one main syllable of [koLoz] or [koLoj] Gutob, [baLa-ku] Gta, [keLeJja] Korku, [piLhig] Korwa (Munda), signifying heart, liver, lung or the whole ensemble, noting also that Khmer word for Liver is [tlaum]. “Gan” for Liver, has cognate in [Gar] Sora (Munda), apart from [guan] Southern Fujian, [gan] 肝 Mandarin, and [cang] Sino-Vietnamese and Sino-Korean. *Lá* in *lá gan* could also result from monosyllabic split of [paLay] Alak, [paLai] Brou, [pLay] Stieng, [pLei] Bahnar, etc. of the Mon Khmer group [14], meaning the heart and/or liver [24] [25]. Metaphorically, liver, heart, or kidneys also look like fruits, and thus “*trái*” (fruit) relating (in sound) to [plei], [play], [palai] above, with equivalents in old Vietnamese and Mường as [blái] and [tlái] [9] [21], respectively, easily led to the use of “*trái*” to denote “fruit” and as a classifier for heart (*trái tim*), kidney (*trái cật*) (where *cật* means “backbone” [6]). Hayes [14] listed some cognates of [plei] (equivalent to *Trái*) as classifier for “round things” such as [play num] in Jeh meaning “lower abdomen”, [pli ho’nih] (muscle) Rengao, and as other body parts: [palâi] (stomach) Katu, [khyllái] (kidneys) Khasi, [phalhi] or [pli] (spleen) Mon.

The *Huỳnh Tịnh Của* dictionary [18], first published in the 19th century, listed also *Lá* as classifier for *Lưỡi* (tongue: *lá lưỡi*). This is quite feasible, as *lưỡi* (tongue) has a number of cognates in the Munda dialects [20], having pronunciation very close to *Lá*: [laʔag] or [lag] Gorum, [lag] Juang], [laG] Kharia, and [lan] Korku. Furthermore, the tongue has a two-dimensional shape quite akin to “leaf”.

*Lá lách* is a “pure *Nôm*” term for the Spleen, which is otherwise known as “*Tỳ tạng*” 脾臟 [pi zang]. *Lá* being classifier for Spleen, and *Lách* itself, could have some connection with, or result from monosyllabic split of, [piLa] in Ho (Munda), denoting both the Spleen and Liver. It is noted that the first syllable [pi] in [piLa] closely resembles Chinese word [pi] in [pi zang] above.

*Lá* in the past was used as classifier for *Mũi* (nose), under the form: *Lá Mía*. This is a good example to illustrate the missing link mentioned above, in that some dictionaries such as the *Khai Trí Tiển Đức* gave *Lá Mía* as the nose bridge or its bone, dispensing the use of *lá* as classifier for a body part in its totality. Under the present theoretical framework with basic premise stating that the substratum of Vietnamese would include Mon-Khmer, Munda, Thai and Polynesian, it will follow that *lá mía* under any form of old (spoken) Vietnamese could not be used to denote just one component (ridge bone) of a body part (nose), as per old dictionaries. It thus can be seen that *Lá* should fit perfectly as a classifier for some body parts, and *Mũi* (nose) should have some cognates in the substratum dialects that have sound similar to *Mía*. In fact *Mũi* could be alternatively pronounced very close to “*Mía*” in the Munda dialects:

[Mi?] Gutob, [mii] Remo, [aamiG] Korwa. Close to *Mũi*, on the other hand, in the Munda dialects are [mu~] Kharia, [mu?u] Sora, [mu:] Korku. In Thai, *Mũi* (nose) is [ja-muuk], and in Gorum as [le-bad], both of which can be said to be related to “*lá mía*” as “*mũi*” in disyllabic form. Under spoken form, *Lá* could be related as a declension to *lỗ* (hole), as in *lỗ mũi* (nostril), which has cognates in [plok-mu] Gta (Munda), [phlawng moo] Thai, and [pologka] Remo. *Lá* as a classifier for some body parts could also be related to [lub] (pronounced as /loo/) in the Hmong language for many body parts like liver, stomach, lung, heart, and nose.

*Lá* in the 19<sup>th</sup> century and in some Vietnamese dialects, could be used as classifier for various compartments of ruminants’ stomach, again as a result of metaphory. Etymologically, it can be seen that a general word for “stomach” in Korku (Munda) is [Laj], *extremely close* to *Lá*. First, according to the Huỳnh Tịnh Của dictionary [18], the rumen is called “*Lá dày*”, with *dày* meaning “thick” referring to the rumen’s thick lining, and *Dạ dày* being a term for the stomach. Lay-term for “rumen” is Paunch, close to [pung] in Mon-Khmer, [poong] in Thai, and “Bụng” in Vietnamese (see [6]). Term for Rumen in Mandarin is 瘤胃 [liu wei] (bumpy stomach) or 反刍 [fan chu] (recycled grass).

The second compartment, Reticulum, is known as “*Lá tàng ong*” (*Lá tổ ong*, or *Lá ong*, or *Dạ tổ ong* – beehive), similar to Mandarin [feng chao wei] 蜂巢胃, and English lay-term: “honey-comb”. The omasum is called “*Lá sách*” (book) or “*Lá chạch*”, noting that “*sách*” and “*chạch*” (策 or 冊) are Bai-Yue terms for Book, depending on dialect pronunciation (see “*Cuốn*” above). English lay-term for Omasum is the same: Book (Latin: Libri folium), whereas Mandarin equivalent is [zhong ban wei] 重瓣胃, with [zhong] being “heavy” or “lots of”, [ban] (flower) petals, and [wei] stomach. Its equivalent in Cantonese is [ngou paa(t) yip] 牛栟葉, meaning “leaves of hemp palm tree”, and possibly in sound: a hundred leaves (or Book). The fourth and last compartment of ruminants’ stomach is the Abomasum, known in Vietnamese as “*Dạ múi khế*” (star-fruit segment folds), or *Lá lau* (sậy, Reed) [18], with similar metaphor “Reed” in English. In Mandarin it is [zhou wei] 皱胃 referring to a compartment full of wrinkles or folds.

*Lá Mía*, as shown above, was used for the Nose (Mũi) [18]. According to the Huỳnh Tịnh Của dictionary [18], *Lá Mía* could denote some internal organ that is common in both pig and cow. The dictionary compiled by Lê Văn Đức and Lê Ngọc Trụ [26], possibly based on [18], gave *Lá Mía* as a term for Spleen of cows and buffaloes, as different from *Lá lách* applied to humans. Noting that basic Annamese lexicon for internal organs, as well as those in many other admixture dialects, before the 16<sup>th</sup> century mostly did not give clear distinction between various organs (see also [6]), and by simple reasoning taking into

account the difficult subject of anatomy, it can be seen that entries on *Lá Mía* from the two authoritative dictionaries above ([18] [26]) may involve some inaccuracy. *Lá Mía* under the framework of the present theory, was a term resulting from lexical transfer, used to denote both the Spleen and the Abomasum (i.e. Reed), and more likely the latter.

It can be seen from metaphory or metonymy as in the foregoing (see also [6] and [7]), words for various compartments of ruminants' digestive system, could result principally from frequent usage of some spoken form of *Lá* ([Laj] in Munda for Stomach), coupled with lexical transfer, most likely from European languages, involving advanced knowledge of anatomy. It can be recalled that East West contact in the country flourished from the 16<sup>th</sup> century, during the time of conflict between the Trịnh and Nguyễn Lords, culminating in a prolonged civil war, from 1627 to 1673. Noting also that in the *Đàng Ngoài* (Exterior Expanse) of the Trịnh Lords, influence of the Dutch was more pronounced than others [27], it is thus instructive to select Dutch lexicon for Spleen and Abomasum, to elucidate such lexical transfer through phonology and metaphory.

Dutch word for Spleen is “Milt” (German: Milz), sufficiently close to *Mía*, for *Lá Mía* to denote “spleen”. However Spleen (English) and Milt (Dutch) are used for both humans and animals, whereas the language already has *Lá Lách* for Spleen [28] (Latin: Lyenis). In English, lay-terms for Abomasum also include “Maw”, which is equivalent to “Muil” in Dutch, also very close to *Mía*. Another Dutch word, having 2 syllables, equivalent to Abomasum is “Lebmaag” (German: Labmagen), closely resembling *Lá Mía* [48]. More likely, therefore, *Lá Mía* was a term for ruminants' Abomasum.

In summary, general Munda word for “stomach” or Digestive system is [Laj]. *Lá phôi* and *Lá gan* are for the Lung and Liver, respectively. *Lá lưỡi* is for the Tongue, and *Lá mía* for the Nose. *Lá lách* is equivalent to the Spleen. For ruminants' stomach compartments, *Lá dày* is for the rumen, *Lá tàng ong* for the reticulum, *Lá sách* - the omasum, and *Lá lau* (or *Lá mía*, as argued above) - the abomasum.

In the pre-quốc-ngữ centuries, *Lá* was also used as classifier for Oar (*lá chèo*), used for rowing boat, as listed in the Alexandre de Rhodes dictionary [9].

Interestingly *Lá* in *Lá chèo* (oar) was later superseded by *Mái* (*mái chèo* – oar), and this will serve to illustrate the argument presented herein, since all of *Lá*, *Mái* and *Chèo* can be found from most Chinese dictionaries (eg. [13]) to be words of similar sounds, all denoting “Oar”. For *Lá* it is 槽 or 艫 or 舟魯 [lou] in Cantonese and [lu] Hakka; for *Mái*, it is 木尾 being [mei] Cantonese; and for *Chèo*, 櫂 or 棹 pronounced as [zhao] in Mandarin, [zaau] in Cantonese, and

TABLE I. List of some commonly-found classifiers

Classifier	Meaning (Classifier)	Usage (Referent)	Kinwords & Notes
BÌNH, LỌ (bình hoa)	Vase, Jar, Bottle, Jug	Flowers, Ink, Wine	[ping] 瓶 Mandarin, [lou] 婁瓦 Man., [lau] Cant., [leoi] 罌, [kalauk] Cham
BỘN (bộ ăn cướp)	Group, Gang	People (derisive)	<i>Bộ ăn cướp</i> (gang of robbers). Related to: [Baan] 班 Cantonese
CĂN, GIAN (gian phòng) (căn nhà)	House, Room	House, Room	間 [jian] Mandarin, [gan] Hakka: classifier for Room. Word for House [Jang] 棧 in Chaozhou. 廩 [king] Cant (room) 家 [ka] Wu 館 [gon] Hakka (house)
CHAI (chai rượu)	Bottle	Liquid, wine	缶垂 [Chui] Mandarin, [tseoi] Cant.
CỐC, LY (ly rượu)	Glass, Cup	Water, wine	[Kác] Bengali, [Bo Li bei] 玻璃杯 Mandarin, Kaca (Munda), Livan (Cham), [Li] Katuic
CỌNG, SỢI (sợi tóc)	Hair	Hair	[Konda] (Sinhalese), [sôk] Khmer
ĐOÁ (Hoa)	Flower	Flower	[Daawk] Thai
GIỐNG	Genus, Species	Race, plants Descendants	[Zhung] 種 Hakka
HẠT (HỘT) (hạt đậu) (hat nhân)	Kernel, Nut, Grain	Pea, Nut, Bean Grain (rice)	核豆 [he dou] Mandarin, [hat dau] Cant.. <i>Nhân</i> (kernel): 仁 [yan] Canton.. <i>Hạt thóc</i> ( <i>hột gạo</i> ): 禾 [he] Mandarin, [hu] Wu -Rice grain. [Oro] Korku (Munda)
LÁT (lát bánh)	Slice	Meat, Bread	[let] Chaozhou, 酉力 [laai] Cantonese, [laae] Thai
LOÀI	Species	Species	類 [Leoi] Cantonese
LỌNG, BÚI CHÙM, LÀN (tóc)	Tuft, Lock (hair)	Hair	[BpaawY], [bpuy] Thai (Búi), [jook] Thai, 撮 [chot] Chaozhou, [lou'ulu] Tongan, 監毛 [lan], 絡 [lau] Mandarin
NHÓM	Group, assembly	Group, gathering	[jOm], [rUn-Dai] Gutob, [run] Remo (Munda), [ruaam] Thai, [zon] 攞 Hakka
NIỀM	*	Hope, wish, belief	願 [nien] Hakka= hope, wish, desire. 佇 [ning] Cantonese= hope. 佞 [nin] Hakka= belief. [neuk wang] Thai= hope
NGÓN (ngón tay)	Finger, Toe	Finger, toe	[aGgul] Gorum (Munda), [niu gaawy] {little finger} Thai, [tuJgol] {index finger} Kharia (Munda),
NGỌN (ngọn núi/dao)	Summit, Sharp end	Mountain, Spear, knife	岩 [ngam] Hakka. 岳 [ngok] Cant.= mountain. 刃 [ngiun] Hakka // 鏢 [ngok] = knife edge // sword blade

almost identical [jaaeo] in Thai. Sometimes, *mái chèo* is used as a compound noun, and in the circumstances the inanimate *Cái* is used as classifier, *cái mái chèo*.

Some other commonly used classifiers are tabulated in Table I. It is clear from Table I and the foregoing, that a good majority of classifiers share their own basic meanings with, or are metaphorically related to, the referents, be they sourced from a variety of related languages or dialects. Some, like *Bình*, *Lọ* (vase, jar), *Chai* (bottle), *Niềm* (hope), may customarily assume roles other than that of classifiers, like that of a noun (*chai rượu* – bottle of wine), or part of a compound noun (*niềm hy vọng* – hope, *niềm tin* – belief). In compound nouns, the construction is generally identical to the “classifier + noun” pair, in that the two main elements (like *niềm & hy vọng* (hope), *cơn + giận* (anger), *mái + chèo* (oar)) have similar connotations, originally, but were sourced from different languages, dialects, or sub-dialects. The underlying construction or etymology of all classifiers is the same, being none of them is uniquely or stand-alone Vietnamese, but intimately related to, or derived from, lexicon of neighboring languages or dialects, which formed constituents of Vietnamese, in the long past.

### **Continuum in both space and time**

Examination of some of the most commonly found classifiers through illustrative examples above, treating the language as a continuum in both space and time in contrast with a discrete entity like *quốc-ngữ*, has again confirmed many important aspects of the admixture theory [4]-[7], those most relevant to the present discussion being:

- 1) Demarcation between pure *Nôm* and the so-called Sino-Vietnamese lexicon is at best a fuzzy one, especially when Southern Chinese dialects like Fujian, Cantonese, Hakka, etc are grouped into Chinese dialects or part of the Sino-Tibetan family, under conventional axioms. Some of the kinwords between Vietnamese and Southern Chinese dialects shown in the present discussion, such as *cấm gan* and *cấm giận* (angry); *lời cảm tạ* (words of thanks); *toà lầu* (multistorey building); *chiếc chiếu* (sedge mat); *chiếc đũa* (a chopstick), etc., are so strikingly similar that an absence of a genetic link between a large part of “nationalized” Vietnamese and “*Bai-Yue*” Chinese, would be most likely implausible.
- 2) Like results from Body Parts [7], Colors [6], and Personal Pronouns [5], contributions from the substratum languages, like Mon-Khmer, Munda and Thai, are also widespread in terms of classifiers. For example: *Cánh* (door, forest, arm), *Đám* (funeral, wedding, grass), *Cơn* (rain, storm, quake, fire, sickness), *Bữa* (meal), *Cọng*, *Sợi* (Hair), *Búi* (hair), *Đóa* (flower), *Cụm*

(words), *Lá* (body parts), *Túp Lều* (hut), etc.. Lexicon that has connection with Polynesian, shown above, included: [ʔauli] Samoa, for *Bàn ủi* (iron), and [Co] Fijian, for *Cỏ* (grass). It is noted that none of the lexicon used in the paper is uniquely Vietnamese, or found not to be related or “affiliated” to at least one language (or dialect) external to the land known as Vietnam, then and now. This has cast great doubt on the proposition that Vietnamese is an isolating language.

- 3) Origin of classifiers is mixed. Some originally assumed a different role, which could simply be an adjective from another migrant dialect (*Toà Lầu*: [Tua Lou] TC = Big Building), others vacillated between its role as an element of compound word, and that of classifier, e.g. *lá chèo*, *mái chèo*. Classifier could also originate from part of a disyllabic word in constituent languages, such as: [MijDa] for *Miếng* in *Miếng Đất* (piece of land), [kAnReG] for *Cánh* in *cánh rừng* (forest), [\*Kazal] or [\*Kanh'jal] for *Con* in *con gió* (wind). Configuration of the “classifier+noun” pair, is also strikingly similar to common structure of compound words (see [47]) in that one word is sourced from one language and the other from another, with both having the same meaning. For example: *Túp lều* (Hut) with *Túp* from [dtuup] Thai, and *Lều* from Southern Chinese dialects, such as Cantonese: [liu] or [leoi] (see above). *Sanh đẻ / Sinh đẻ* (birth, give birth, born) has *Sanh (Sinh)* from 生 [sang] Hakka, [sinh] Tay-Nung, and [deh] Hainan. Similarly, compound word *Đường Sá* (Road), has *Đường* (or *Đàng*) related to [taang] Thai, and *Sá* to [sala] Fijian, [salan] Champa, [seh] Thai, and [\*zalan] Mon-Khmer. In Thai, another word for Road is [ta'non] related to [tango:r] Munda, and either [seh] (ie. *Sá*) or [taang] can be used as classifier: *Ta'non sawng* (2) *seh*: 2 roads.
- 4) Examination of the classifiers above, typically through *Con*, *Cánh*, *Chiếc*, *Lá*, *Đám*, *Bàn*, *Toà*, has dispelled an old myth about the one-to-one régime between word and its meaning, and at the same time, revealed many homophonic features of the Vietnamese language inadvertently buried through conventional treatment of the language as a discrete entity (under both *Nôm* and *quốc-ngữ*), as opposed to a continuum in both space and time as in the present approach. It should be stressed that the discrete entity approach (particularly through alphabetic *quốc-ngữ*) sometimes may lead to abandonment of some age-old lexicon in favor of new replacement, resulting possibly from frustration in failure to decipher or to understand the traditional one. Consider as an example compound verb “*có thể*” (can, be able to, possible) commonly used in the past, now being replaced by “*có khả năng*”, with Chinese equivalent 有可能 [you ke neng] (*hữu khả năng*).

The problem with “*có thể*” is that Vietnamese is misled by its own alphabetic form (*quốc-ngữ*), where *Có* would be easily misunderstood as “to have” or “there is” at all times, and that would render “*thể*”, and also “*có thể*”, meaningless, at least in etymology. In fact *Có* in *Có thể* only lost

its true meaning through that alphabetic *quốc-ngữ*. It could be remedied by having recourse to the continuum theory, and noting that in the Lisu dialect in Thailand [35], [Ko] merely means “can, able to, possible”, like *Có* in *Có thể*, whereas in Tahitian, [tiʔa], with sound close to “*thể*”, is word for “possible”. Likewise Tibetan word for “can” is [thoob], also very close to “*thể*”. Furthermore, Hakka pronunciation for word equivalent to “possible” or “can” could be [kon-theux] (Taiwan), or [ko the] [13] [42], strikingly similar to *Có thể*, which in turn is equivalent to the other “Sino-Vietnamese” *Khả thi* (可施 [ke shi]).

- 5) Like *Lá* above, in regard to *Lá chèo* (oar) being transformed to *Mái chèo*, and *Lá Mía* from Dutch “Lebmaag” for Abomasum; *Toà Lầu* from [Tua Lou] Teochew; *Có thể* versus *Khả Thi* and Hakka dialects, the question of “Proto-Vietnamese” should be treated with great care as the language is a continuum in both space and time. As the present admixture theory is intimately linked with the continuum conception, access to a great amount of data will thus become available, and as the language “barriers” have been broken, a new theory on the origin of Vietnamese classifiers can now be developed by looking at characteristics of classifiers of other languages in the whole region.

### **Classifiers under the Admixture Theory**

One of the basic characteristics of a language evolving from admixture is that elements in lexicon for personal pronouns, or numbers from 1 to 10, should exhibit some rearrangement in reference of their counterparts from constituent languages or dialects [4] [5] [46]. For example, word for the second personal pronoun (you) in many related languages is [naw] (Lahu), [no] Korean, [nek] Khmer, [no] P’u Noi (in Myanmar), [nong] Shanghai, but was shifted to assume the third pronoun (he/she/it) in Vietnamese: *Nó*. Likewise, [nãm] is a common phoneme for number 6 (six) (e.g. Mon-Khmer, Champa), and yet in Vietnamese it denotes number 5. It should be remarked here that a link between the Vietnamese language and other “remote” languages like Korean, Japanese, and even Tibetan [50], is always possible through mutually shared constituent dialects – such as Hakka, Wu, Fujian, etc., or their “Proto”-counterparts – whose speakers known to inhabit along the East coast of China, stretching from the Shandong peninsula down to the present-day GuangDong (see [4]); or through the House-of-Cards effects, or a combination of both. This lexical rearrangement, as exemplified in the third personal pronoun and number 5, indicates strongly that some very important vestiges or evidence of admixing processes in the past may still be recognizable under the present form of the language, though customarily hidden under the discrete entity approach.

Accordingly, for analysis of classifiers using the admixture theory, it is propositioned that each type of classifiers was connected with a particular

constituent ethnicity (or sub-ethnicity) in the long past, and evolved through the millennia or centuries into Vietnamese classifiers (or compound nouns) as end products. More specifically, and as shown in another paper [45], it is suggested that the inanimate “*Cái*” and animate “*Con*” came from some subgroups of Mon-Khmer speakers, and its usage is customarily overlapped with that of “article”, called Type I.

Most of the classifiers presented above, originated from counting system with construction of “Noun + Number + Classifier” such as in the Burmese and Thai languages, which later on evolved into a Vietnamized re-arrangement of the type “Number + Classifier + Noun”, where either Classifier or Noun has to be sourced from different language or dialect, Type II. This Type II is largely mixed with compound words, both in construction and in usage. At times and originally, a classifier or an element of compound word could have originated as misnomer, or resulted from split of a multi- or di-syllabic word. For example: An adjective like *Toà* in *Toà Lầu* (multistorey building) turned into classifier or element of compound noun; a verb or qualifying noun like *Đám* in *Đám tang* (funeral) became mixed with other *Đám*’s as classifier; or monosyllabic split of [MijDa] for *Miếng Đất* (piece of land) gave rise to *Miếng*.

There is another category, Type III, involving noun classifiers assuming the role of measure words of the kind: *một hạt gạo* (a grain of rice), *nắm muối* (handful of salt), *đôi đũa* (pair of chopsticks), *chai rượu* (bottle of wine), *hộp đồ ăn trưa* (lunch box), etc. in which the measure word (eg. *Chai* - bottle) confers directly to a noun an entity, connoting generalization or a class of objects. The term “classifier” is considered to be most aptly applied for description of Type III.

As discussed above, however, many of the classifiers and Vietnamese lexicon in general, are homophonic, or originated as misnomer, or from disyllabic lexicon, and thus distinction between Type II and Type III classifiers sometimes may be blurred. For example: *Đám* in *Đám ruộng* (paddy field), *Chiếc* in *Chiếc Đũa* (a chopstick), and *Chai* in *Chai rượu* (bottle of wine) are Type III, whereas *Đám* in *Đám cưới* (wedding), *Chiếc* in *Chiếc Xe* (vehicle), *Toà* in *Toà Cao Ốc* (building), and *Niềm* in *Niềm Hy Vọng* (hope), belong to Type II. Likewise, *Chiếc* in *chiếc áo* (shirt, coat), where *Áo* is of Mon-Khmer origin [?Aav], in reference to a singleton could be said to belong to Type III, but in comparison with a kinword, with similarity in sound, like [jas] (Indonesian), it should be put under Type II. Due to limited space, the remainder of this discussion will focus only on Type II classifiers.

## Two Categories of “Measure Structure”

Measure words of Type III, in general are similar to their English counterparts,

which largely resulted from classification of nouns into count nouns (e.g. cars) and mass nouns (e.g. salt, sand). Studies of the characteristics of Vietnamese classifiers should therefore begin by analysing ancient mode of genesis of classifiers, now made possible through examination of constituent languages or dialects of the language.

Consider the following numeral usage [15] [35]:

- in Burmese: ‘ein *dta* ‘**ein** (one house – house-one-house), where *dta*= 1, and ‘**ein** (bold) is classifier.
- in Khmer: ptéah *bey* (three houses) {no **CL**}; gruu *búan* **néak** (four teachers)
- in Thai: rôt *sawng* **rôt**, Or: rôt *sawng* **kan** (two cars – car-three-car);  
pùu-yíng *săhm* **kon** (three women – women-three-human )
- in Sinhalese: gǎanu *dennek* (two women); pot *tunák* (three books) – {no **CL**}
- In Japanese: seito *san-nin* (three students); tegami *san-mai* (three letters).
- In Lahu: chaw *te*<sup>v</sup> **g’ă** (one person); person-one-CL(human)
- In Akha: ăji<sup>v</sup> *nyi*<sub>v</sub> **maw**<sup>v</sup> (two birds); bird-two-CL (animal)
- In Akha: pu *ti*<sub>v</sub> **pu** (one village); village-one-CL (village)

in which the numerals (*dta* (1), *bey* (3), *buan* (4), *sawng* (2), *săhm* (3), *dennek* (2), *tunák* (3), etc) are placed after the nouns. In the majority, the numerals are followed by the classifiers (in **bold** letters), called Category A (structure) for convenience.

The general structure is thus: Noun + *Numeral* + **CL**, Or: Noun + *Numeral*, Which is in contrast with the other counting structure found in Vietnamese, Chinese, Hmong, Malay, Tamil, Bengal, Hindi, Persian, etc. where the Numeral precedes the Noun, and in many, a Classifier is placed between the Numeral and the Noun, i.e. *Numeral* + **CL** + Noun, Or: *Numeral* + Noun, called Category B. For example:

- *Hai* **ngườì** đàn bà (two women) (Vietnamese). {CL: **ngườì** (human being)}
- *Dua* **orang** perempuan (two women) (Malay). {CL: **orang** (human being)}
- *Rendu* payi (two bags) (Tamil). {No CL}.
- *Dujon* chele (two boys) (Bengali). {CL-slot: **jon** (human being)}
- *Dos’ta* boi (ten books) (Bengali). {CL-slot: **ta** (thing)}
- *Do* **nafar** mard (two men). (Persian). {CL-slot: **nafar** (human being)}
- *Ib* **tug** dle (one dog) (Hmong). {CL for humans and animals: **tug**}

From these, and in reference to the admixture framework some observations can be made as follows:

- a) Under the “Noun + *Numeral* + **CL**” structure (Category A), typically represented by the Thai language (e.g. rôt *sawng* **rôt** / car-two-car => two cars), a kind of anaphora between noun and classifier is generally acceptable, though classifier has been gradually replaced by an equivalent

which is different from the noun, e.g. *rõt sawng rõt* => *rõt sawng kan*.

- b) In reference to Category A above, examination of the other measure structure, Category B, being “*Numeral + CL + Noun*”, as in Vietnamese, coupled with Type II classifiers above, will reveal that it is almost mandatory that under Category B, the classifier must be different in form to the noun followed, e.g. *Con Gió* (wind), *Cuốn Sách* (Book), in which *Con* is different from *Gió*, *Cuốn* (or *Quyển*) from *Sách*, at least in morphology..
- c) It is also seen that all the three Types of Classifiers (I, II, and III) may at some time in the past co-exist, interwoven with measure structures of Category A and B, in any state of Asia. And the final arrangement, in terms of proportion or predominance of any one Type, would depend on the success of admixing process, and the duration of time a particular group of ethnicity with prevalent affinity for a particular Type had control of the military, political and cultural agenda of the mixed community. For example, as shown in another article about *Cái* and *Con*, Vietnamese use of these inanimate and animate classifiers (Type I) has manifested a perfect match from the Khasi system, and the use of *Cái* and *Con* is more widespread in Vietnamese than their general classifier counterparts in Burmese, Thai, or languages in the North like Chinese, Korean and Japanese.
- d) In respect of Category of measure structure, it can be seen that Category B, being *Numeral* (with classifier) preceding *Noun*, by and large correlated well with Type I, involving animacy differentiation, although the reverse is not always true. For example, in Vietnamese there is a large amount of lexicon using classifiers *Cái* and *Con* (Type I), and at the same time, the measure structure is of Category B (*Numeral* preceding *Noun*). On the other hand, the Sinhalese language has animacy distinction, but through suffix *-ennek* (living) and *-ák* (non-living) attached to the numeral (see above), and its measure structure is of Category A. Most characteristic of Category B structure in Vietnamese, it should be stressed again, is word for Classifier being different in “morphology” to that of the *Noun* followed, e.g. *hai chiếc xe* (2 + CL + cars), where **Chiếc** (CL) is different from “Xe” (only in morphology and etymology).
- e) Competition between the two categories of measure structure (A and B) was settled in conjunction with that among the various Types, but normally not without compromises. One of these, for example, was shifting the position of the *Numeral* in Category A-Structure, from the back to the front, to precede the *Noun* as in Category B, in conformity with Type I classifier with *Numeral* in front, similar to lexical shift in pronouns and numbers described above. As the *Numeral* was moved to the front it took with it the Classifier, and as a result, juxtaposing Classifier and *Noun* together. The resulting structure (Category B) therefore would try to avoid anaphoric confusion by “compulsory” replacement of any Classifier that used to be identical with the *Noun* under Category A (e.g. ‘*ein dta ’ein / rõt sawng rõt*

– as above), by another word normally having similar meaning, but sourced from other constituent languages or dialects, e.g. *hai cơn mưa* (2 rains), *ba chiếc thuyền* (3 boats). This was demonstrated in detail in the foregoing.

It should be noted that there is nothing unusual about Category A counting structure (Noun + *Numeral* + (CL)), as it is in use in many “languages” under format required, or custom, such as in computer programming, inventory (as in contract for the sale of business), restaurant or theatre reservation, or enumeration, such as in the following utterance:

*Trong cái va-li của con má đã để: áo 10 chiếc, quần 5 cái, áo lót 15 chiếc.*

In luggage of son, Mom has put: shirts: 10 (+ CL), trousers: 5 (+ CL), singlets: 15 (+ CL) => In your luggage, Mom has put: 10 shirts, 5 trousers, 15 singlets.

Or:

*Thằng đó vợ 3 bà.*

Fellow that wife 3 lady ~ => That fellow (has) 3 wives.

It should be noted that under this Category, very often when the classifier is omitted, the *Numeral* is changed from a cardinal number to an ordinal one, in meaning. For example: ‘*Vợ 3 Bà*’ means 3 Wives, but ‘*Vợ 3*’ refers to the 3<sup>rd</sup> Wife. It is established therefore that one of the roles of classifiers under Category A, where the numeral comes after the noun, is to mark the numeral as a cardinal number. This type of marker is not dissimilar to interceder “*thật*” (really) being placed between two identical words:

\* *đen thật đen* (black really black) => extremely dark

\* *đói thật đói* (hungry really hungry) => very hungry

which are, respectively, different from:

\* *đen đen* (black black) => rather black, “light” black

\* *đói đói* (hungry hungry) => hungry a little

There are also languages that employ explicitly both Category A and Category B structure in counting, such as in Polynesian languages like Samoan, Fijian, and Tahitian. For example: *fale e lua* OR: *lua fale* (two houses) (Samoan) [15], where it can be seen that the “*lua fale*” structure (Numeral in front) may have resulted from speakers of the language having contact with the West.

## **Towards a Unified Theory of Classifiers**

Consider the following utterance actually addressed to the writer, circa 1978, from a Vietnamese New Caledonian, belonging to the generation of Vietnamese recruited by the French to go to New Caledonia to fill laboring jobs in plantations and mines, around the time of the Second World War.

*Chiều nay, Ông Bà Ông Điều mời cậu tới ăn cơm.*

(This evening, **Mr Mrs Mr** Điều invited you to dinner)

In that utterance, there is a kind of repeater under the word *Ông* being **Mister (Mr)**, and as far as the writer can recall, utterance with repeater construction, using Mister (Mr) in particular, was widely used among the Vietnamese expatriates in New Caledonia of the labor-recruitment generation.

Such anaphoric or repeater construction bears strong resemblance to the Category A structure in counting detailed above, where the counted Noun is followed by Numeral and Classifier, which may be identical, as repeater, to the Noun: rôt *sawng* **rôt** => car-*two-car*: 2 cars.

Consider also the issue of word order and syntactic stress, as suggested by Daladier [48], through the following utterance in the Amwi dialect of Mon-Khmer languages:

*Jao bO tSi ?u*

He eat rice he

He is eating rice (“He” is not in the presence of the speaker)

Relevant of this Amwi utterance to the present discussion is that word for “He” is repeated, though under a different context. Nevertheless, repeater *Ông* (Mr) in “*Mr Mrs Mr Diêu*”, repeater Classifier in both Category A and B, and repeater “He” given in [48], all point to one common “Repeater” characteristic of Vietnamese and its substratum languages, and it is possible that classifier (Type II) was first generated in accordance with “repeater” demand of constituent languages in the basic substratum of Vietnamese.

There could also be another reason for anaphoric construction in Classifiers of Type II, as it can be recalled that in Burmese, and many languages of the hill tribes in Thailand, such as Lahu and Lisu, for numerals ending in a zero, the classifier is not needed [34] [35]. That is, for numerals like: ten, fifty, five hundred, ten thousand, etc. the use of classifiers is not required, as shown in the following example from the Lahu language [35]:

- 555 wild boars:

' <b>heh</b> va^'	<u>nga</u>	<u>ha</u>	<u>nga</u>	<u>chi</u>	<u>nga</u>	<u>hkeh</u>
'wild boar'	5	100	5	10	5	CL

- 550 wild boars:

' <b>heh</b> va^'	<u>nga</u>	<u>ha</u>	<u>nga</u>	<u>chi</u>
'wild boars'	5	100	5	10

- 10 horses:

' <b>i- mvuh</b> ^'	<u>te</u>	<u>chi</u>
'horses'	1	10

- 12 horses:

' <b>i- mvuh</b> ^'	<u>te</u>	<u>chi</u>	<u>nyi</u>	<u>hkeh</u>
'horses'	1	10	2	CL

Consider also the following utterances in Vietnamese:

- a) (thằng đó) vợ hai bà: (fellow that) wife 2 lady - {that fellow (has) 2 wives}  
**Noun** + hai bà (ie. two + CL)
- b) **Lòng** quyết tâm: determination => CL + quyết tâm => **Noun** + quyết tâm
- c) **Niềm** hy vọng: hope => CL + hy vọng => **Noun** + hy vọng
- d) Bốn Con **Cá** (4 fish) => **Cá** bốn con: **Noun** + bốn Con (ie. four + CL)
- e) **Hai đứa** tụi này: two person we this - Two of us => {Hai + CL} + tụi này  
Noun phrase (hai đứa) [49] + tụi này
- f) **Ông Bà Ông Thọ**: Mr Mrs Mr Thọ

Which incorporate all types of classifiers and two structure categories [a) to e)], as mentioned above.

If all the phrases underlined are to act as “descriptive phrases”, and word or words in bold are considered as Nouns, and the morphology of these utterances are examined in conjunction with the examples on the requirement of Lahu classifiers above, it can be seen that a general pattern of construction or morphology is emerging. That is “descriptive phrases” that follow the main noun or noun phrases are normally arranged in pairs of phonemes, or in other words, the number of words in the descriptive phrases should be an even number. This arrangement in pairs of phonemes could well reflect some nostalgic affinity to disyllabic lexical rhythm present in constituent languages of Vietnamese in the long past.

## REFERENCES & NOTES

- [1] Karen Ann Daley (1998). Vietnamese Classifiers in narrative Texts. Summer Institute of Linguistics (University of Texas at Arlington). ISBN-10: 1556710216 & ISBN-13: 978-1556710216
- [2] Ly Toan Thang (1999). Representation of space in Vietnamese classifiers. Mon-Khmer Studies 29: 71-80.
- [3] Kasumi Yamamoto (2005) *The Acquisition of Numeral Classifiers: The Case of Japanese Children*. Mouton de Gruyter, Berlin, New York, 2005. <<http://www.degruyter.de/rs/bookSingle.cfm?id=IS-9783110183672-1&l=E>>
- [4] V.U. Nguyen (Nguyen Nguyen) (2008) Thử đọc lại truyền thuyết Hùng Vương (In search of the origin of the Vietnamese) – (in Vietnamese). (in publication)
- [5] V.U. Nguyen (2008) Vietnamese Personal Pronouns (submitted for publication).
- [6] V.U. Nguyen (2008) Body Parts in Vietnamese (submitted for publication)
- [7] V.U. Nguyen (2008) Basic Colors and Metaphors (submitted for publication).

- [8] Phan Anh Dũng, Nguyễn Thế, Nguyễn Đình Thảng (2007) On-line Viet-Han-Nôm Dictionary. Pub. by Vien Viet Hoc (Institute of Vietnamese Studies): <http://nom.netnam.vn/HanNom/tracuu/GioiThieu.php>
- [9] Alexandre de Rhodes (1651) Dictionarium Annamiticum – Lusitanum – Latinum. Translated by: Thanh Lãng, Hoàng Xuân Việt, Đỗ Quang Chính. Pub. By Vien Khoa Hoc Xa Hoi – HCM 1991.
- [10] James Campbell (2008) Chữ Nôm Characters – (Online Compilation): <http://www.glossika.com/en/dict/viet/nom-v.php>
- [11] Nôm Preservation Foundation (2002) Another Nôm Lookup Tool based on Unicode: [tech@nomfoundation.org](mailto:tech@nomfoundation.org)
- [12] Lê Quý Đôn (1726-1784) Vân Đài Loại Ngữ (1773). Translated by Tạ Quang Phát. Publisher: Văn Hoá Thông Tin (1995). Set of 3 volumes.
- [13] CCDICT v5.1.1: Chinese Character Dictionary by Chineselanguage.org (1994-2008)
- [14] L. V. Hayes (2001) Austric Glossary <http://home.att.net/~lvhayes/Langling/Glossary/Glospag1/glosf019.htm>
- [15] Charles Hamblin (1984) Languages of Asia & The Pacific. Angus & Robertson Publishers. (1988 Paperback Edition).
- [16] đôi tằm = pair of earrings, with “**tằm**” being alphabetically coded from a word similar to [d**toom**-huu] in Thai, noting [dt] is a sound between [d] and [t].
- [17] Chiếu is pronounced similarly in Champa: [ciêw].
- [18] Huỳnh Tịnh Paulus Của (1895) Đại Nam Quốc Âm Tự Vị. Imprimerie Rey, Curiol & Cie. Re-published by Nhà Xuất Bản Trẻ, 1998.
- [19] Gagiang (2008) English-Diojiu Dictionary: <http://www.gagiang.org/content/index.php?p=15>
- [20] Patricia J. Donegan and David Stampe (2004) Munda Lexical Archive. [www.ling.hawaii.edu/faculty/stampe/AA/Munda/ETYM/Pinnow&Munda](http://www.ling.hawaii.edu/faculty/stampe/AA/Munda/ETYM/Pinnow&Munda)  
<http://ling.lll.hawaii.edu/faculty/stampe/AA/Munda/Dictionaries/00README>
- [21] Nguyễn Văn Khang (editor), Bùi Chỉ, Hoàng Văn Hành (2002) Từ Điển Mường-Việt {Mường-Vietnamese Dictionary}. Published by Văn Hoá Dân Tộc – Hà Nội.
- [22] Lá in *lá thư* (letter), or *lá đơn* (application form), took “leaf” as metaphor, directly connected to Thai word [bai] used for leaf [bai], paper [bai], document [bai], application form [bai sa mak] [23], etc.
- [23] On-line English-Thai/ Thai-English Dictionary: [Thai-language.com](http://Thai-language.com) (2007)
- [24] Words for some internal organs in the Hmong language also show little distinction in sounds: [nplooj siab] (liver), [plab] (stomach), [plawv] (heart). It is noted also that these are not totally dissimilar to Mon-Khmer lexicon: [palay] Alak, [palai] Brou, [play] Stieng, [plei] Bahnar, denoting heart or liver.
- [25] The Saturn English-Hmong & Hmong-English Dictionary (St Paul – State of Minnesota): <http://ww2.saturn.stpaul.k12.mn.us/Hmong/dictionary/enghmong/newmenu.html>
- [26] Lê Văn Đức and Lê Ngọc Trụ (1970) Tự Điển Việt Nam (Vietnam Dictionary). Pub. Khai Trí

- [27] Trần Trọng Kim (1971) *Việt Nam Sử Lược* (A Brief History of Vietnam). Pub, Educational Materials Centre – Dept of Education.
- [28] [piLa] in Ho (Munda) is the closest to both *Lá* and *Lách* in *Lá Lách* (spleen). Its equivalent in Kharia [sa?] and Remo [posaigor], both Munda dialects, however appears to be closer to *Sách* in *Lá Sách* for the Omasum (Book).
- [29] Vietnamese *Ô* (northern) and *Dù* (southern) originally do not mean “umbrella” but merely “Rain”. Both was formed under monosyllabic truncation from compound word [Yu San] 雨傘, with [Yu] denoting “Rain” and [San] having pictogram 傘 depicting the tool (umbrella) for rain sheltering. [Yu] gave rise to “*Dù*” [Yù], and equivalent to [hO] in Fujian, which resulted in “*Ô*”, in the northern dialect. Another word showing lexical transfer by “default” or arbitrariness, is word for Glasses (drinking), being “Ly”. Word for “Ly” (Glass) in Chinese is [Bo Li Bei] 玻璃杯 (Pha Lê Bôi), in which only the last character [Bei] originally denotes “Cup” (made of wood in olden times as it has Tree radical [mu] 木 on the left 杯). Under monosyllabic pressure, both Chinese and Vietnamese could use any of the three [Bo] [Li] [Bei], as word to denote Glass, with “Ly” more popular among the southern dialects. Another word for Glass is “*Cốc*”, having connection with: [Gâew] Thai, [Kăc] Bengali, and [Kaca] Munda. [Gâew] in Thai and Khmer is classifier for Glasses, cups, and tumblers.
- [30] Normally “carpet” is known as [di tan] 地毯 in Chinese.
- [31] “Spleen” in Thai is [Maam].
- [32] Stanley R. Witkowski, Cecil H. Brown (1985) Climate, Clothing, and Body-Part Nomenclature. *Ethnology*, Vol. 24, No. 3 (Jul., 1985), pp. 197-214
- [33] Behrens, L. (2003) Classifiers, Metonymies, and Genericity: A Study of Vietnamese. IN: Cornelia Zelinsky-Wibbelt, Cornelia (ed.) (2003) Text, Context, Concepts, pp 65-126, Berlin: Walter de Gruyter (TTCP Series 4). 371pp
- [34] Andrew Simpson (2005) Classifiers and DP structure in Southeast Asia. IN: *The Oxford Handbook of Comparative Syntax*, Guglielmo Cinque and Richard Kayne (eds.), 806-838, Oxford University Press: Oxford.
- [35] David Bradley, Paul Lewis, Nerida Jarkey, Christopher Court (1999) Hill Tribes Phrasebook. Lonely Planet.Publications (Australia)
- [36] English-Thai Dictionary Online: English-Thai-dictionary.com (2007)
- [37] Graham Thurgood (2001) Review of “Vietnamese Classifiers in Narrative Texts” by Karen Ann Daley. *Language*, Vol. 77, No. 1 (Mar., 2001), p. 198.
- [38] Number one (1) in Chaozhou (Teochew) could be [‘it] {一} or [cheç] {蜀}. [‘It] has kinwords as [ichi] in Japanese meaning “1” (one), and [ít] in Vietnamese meaning Some, Few, Little. [Cheç] (meaning “1”), goes with *Chiéc* (singleton), and also with Japanese [chotto], and Vietnamese [Chút] denoting Some, Few, Little, similar to “ít”.
- [39] L. Sagart (2001) Gan, Hakka and the formation of Chinese dialects. Centre de Recherches Linguistiques sur l'Asie Orientale. Ecole des Hautes Etudes en Sciences Sociales (UMR 8563 du CNRS) Paris - France

- [40] Nicola Di Cosmo (2002) *Ancient China and its enemies - The rise of nomadic power in East Asian history*. Cambridge University Press.
- [41] Cognates for *Giáy* (paper) are: [jia] Lao, [jaak] Thai
- [42] Liim, K-H (1997) *Simplest English-Taiwanese-Hakka Dictionary*. EDUTECH, available on-line: <http://www.edutech.org.tw/Index.htm>
- [43] AJ. L. Taberd (1838) *Dictionarium Anamitico Latino* {Annamese Latin Dictionary}. Republished by Văn Học (2004)
- [44] Short name for Fujian is [Min] 闽 which came originally from the state of Min Yue (Mân Việt) prior to the Qin era. Presently Min-Nan is used to denote generally, dialects in the Southern Fujian region.
- [45] V.U. Nguyen (2008) Classifier “Cái” and “Con”. (in preparation).
- [46] V.U. Nguyen (2008) Vietnamese numerical systems (in preparation).
- [47] V.U. Nguyen (2008) Compound words (in preparation).
- [48] Another example showing the Dutch connection in Vietnamese is word for Candle: *đèn cầy* (southern). Northern word is *Nén* coming from Hakka [nen] 且寧 for Wax. Word in the Central is *Sáp* meaning also Wax, with similar Persian pronunciation [sham’], and *đèn cầy* is compound word from Champa [dien], meaning Candle, and from Dutch [kaars] (wax candle), having a Greek connection [κ ε ρ ι].
- [48] Anne Daladier (2002) *Definiteness in Amwi: Grammaticalization and Syntax*. *Recherches linguistiques de Vincennes* 31, pp 61-78.
- [49] *Đũa* can be taken as general classifier for “Person”, it has Thai counterpart [dtuaa] meaning: self, person, body.
- [50] Kinwords over long distance can be seen from some similarities between Vietnamese and Tibetan lexicon:
- Khác (different) ~ [kha] Tibetan
  - Cẳng (leg, foot) ~ [kahng] Tibetan
  - Đụng (hit, collide) ~ [doong] Tibetan
  - Chặt (cut, break) ~ [châg] Tibetan
  - Rách (torn) ~ [rey] Tibetan
  - Chục (10) ~ [chu] Tibetan
- [51] A metaphorical similarity in the connection between “anger” and “liver” can also be seen in the English expression: “to vent your spleen”, where “liver” in Vietnamese, Hakka, Miao, is equivalent to “spleen” in English.
- [52] Shift of Phoneme for Cloud from [un] to [ang] is also evidenced from Thai classifier for small object: [an] which can also be pronounced as [un].