THE STRUCTURE OF NOMINAL CLASSIFIER SYSTEMS[1]

Benjamin K. T'sou
University of California, San Diego

1. *Introduction*

Syntactic manifestation of semantic classificatory systems is a common trait in natural language. Thus, in the familiar languages of the Indo-European family, differences in gender are reflected in concord requirements that exist between ad-nominals and relative pronouns and the nouns. In a three-gender system such as that of Latin and its derivative Romance languages, there is a greater variety of co-occurrence constraints than in a two-gender system such as that of modern Dutch. In the non-Indo-European languages, particularly those in Asia, semantic classification in nominal structures typically takes on a wider domain. The syntactic manifestation may appear either in the nominal structure or in the verbal structure, and the incorrect use of classifiers renders a sentence 'ungrammatical'.

Nominal classifiers are attested among most languages in Asia, including the Sino-Tibetan languages, the Austroasiatic languages, the Malayo-Polynesian languages, even some of the Indo-Aryan languages bordering on these languages,[2] and the Altaic languages. Their use is generally associated with the quantification of objects or nouns and as a rule they usually occur immediately adjacent to the numeral in a measure phrase or in conjunction with demonstratives. Nominal classifiers were also very much

evidence among the native languages in America, rticularly among the Athapaskan languages. In these nguages the semantic classification of nouns is nifested in the verbal structure, perhaps reflect- g the general tendency of verbal incorporation.[3] eneau has posed a very interesting and important oblem concerning the areal spread of classifiers. will be important to consider areal linguistics the light of possible independent phylogenetic velopments as opposed to a possible combination of th independent evolution and structural borrowing cause of prolonged language contact in contiguous ographical areas. A better understanding of the ructure and development of such syntactic traits n aid in tracing genetic relationships, particular- those among Austroasiatic languages, Sino-Tibetan nguages, and Malayo-Polynesian languages. This lls for more than exhaustive taxonomic descriptions classifiers in each language. It will be neces- ry to devise a formal apparatus by which both the nchronic systems as well as diachronic developments nominal classifiers may be compared.

Moreover, the study of nominal classifier sys- ms suggests an important hypothesis that the use nominal classifiers and the use of the plural rpheme are in complementary distribution in natural nguage. More concretely, it suggests that either a natural language has either nominal classifiers plural morphemes, or b) if a natural language has th kinds of morphemes, then their use is in comple- ntary distribution.

2. *The structure of the ad-nominal classificatory system*

Nominal classifiers are the lexical items that usually come between the numeral and the noun in a measure phrase. Under this definition the number of classifiers in a particular language generally ranges from a handful to about two hundred.

The actual range, as we shall show later, is relatively open-ended. I have proposed elsewhere[4] that a four-way distinction in the kinds of classifiers is justified. They may be characterized by two features: [± entity] and [± exactness]. For example, in the case of 'chicken', the Chinese classifiers are (in Mandarin):

(I) $\begin{bmatrix} + \text{ exact} \\ + \text{ entity} \end{bmatrix}$: zhī ('individual', non-human objects)

(II) $\begin{bmatrix} + \text{ exact} \\ - \text{ entity} \end{bmatrix}$: jīn ('cattie', unit of weight)

(III) $\begin{bmatrix} - \text{ exact} \\ + \text{ entity} \end{bmatrix}$: qún ('brood')

(IV) $\begin{bmatrix} - \textit{ exact} \\ - \text{ entity} \end{bmatrix}$: zhǒng ('kind/type')

In (I) the measure refers to an exact quantity and involves discrete physical entities. A parallel case in English would be *sheet* (in *two sheets of paper*), which characterizes certain physical dimensions of 'paper', the mass noun. In (II) the measure is exact but it refers to no discrete physical entity. *Pounds*, *gallons*, and *feet* for example, are commonly known as measure words. Their function is to delimit exact amounts of unstructured and non-entity mass. The measure is applied to the unit of measurement and not to entities of the delimited mass. *Two pounds of chicken* (or *beef*) pre-

nts an exact measure, but it need not be a discrete tity in that more likely than not the 'Shylockian t' could not have been made. In (III) there is a finite sense of a well-defined discrete entity or tities, but the quantity is not exact either by sign or by convention. For example, a *brood of icks (or a plate of chicken)* is not an exact measure t there is a definite sense of physical entity and can be referred to as a unit. This may be con- asted with *two pounds of chicken (legs)* as in *I ught two pounds of chicken (legs) yesterday.* Here ference is made to an exact quantity rather than object and there is the sense of physical entity cking. In (IV), which characterizes mainly stract nouns, the measure is neither exact nor does refer to a discrete physical entity.

Natural language exhibits all four kinds of asure, but the range of each kind of measure may ry in different languages. Mass nouns in English, comparison to those in Chinese, may be good amples of range difference. In the case of attle': *two head(s) of cattle, head* is a $\begin{bmatrix}+\text{ exact}\\+\text{ entity}\end{bmatrix}$ asure; in *two herds of cattle, herd* is a $\begin{bmatrix}\text{exact}\\\text{entity}\end{bmatrix}$ measure; in *twenty thousand pounds of ttle, pound* is a $\begin{bmatrix}+\text{ exact}\\-\text{ entity}\end{bmatrix}$ measure; and in *two nds of cattle, kind* refers to $\begin{bmatrix}-\text{exact}\\-\text{entity}\end{bmatrix}$ measure. unt nouns in English and other European languages ually require no overt markers for $\begin{bmatrix}+\text{ exact}\\+\text{ entity}\end{bmatrix}$ asure, but the contrary is generally true of the nguages in Asia, where (I) embodies a rich and mplex classificatory system. The number of cate- ries is culture-bound and relatively finite for I). It depends on the standard measures of weight, lume, length, temporal extent, etc. (IV) univer-

sally has few categories, but (III) is relatively open-ended, for it includes what are sometimes known as temporary measures,[5] e.g. *two spoonfuls of cough syrup every day, three tables full of paper,* etc. (IV), in English, also involves a complex and idiosyncratic array of classifiers, e.g. *two prides of lions, three schools of fish, a horde of savages, two gaggles of geese, a bevy of ladies,* etc. Many of these terms are derived from a refined tradition of hunting as a sport where colorful inventory taking or enumeration was an inherent part of the sport. They are used only in special contexts. Thus it is correct in colloquial speech to substitute *group* for *pride, school, horde, gaggle, bevy,* etc.

It may be noted that the semantic distinctions of such a system are supported as well by evidence from syntax. Thus, type (I) classifiers in Chinese cannot be used in a modifier construction involving the *de* particle: i.e. sān zhī jī (3-cl-chicken) '3 chickens' has no parallel in *sān zhī*de* jī (3-cl-of-chicken), whereas the *de*-modifier construction may involve classifiers for the other three classes: sān jīn de jī (3-cl-de-chicken) 'three pounds of chicken', 'a chicken of three pounds'; sān qún de jī (3-cl-of-chicken) 'three groups of chickens'; sān zhǒng de jī (3-type-of-chicken) 'three kinds of chicken', 'chicken in three kinds'. On the other hand, ordinal numbers may be used with all but $\begin{bmatrix} + \text{exact} \\ - \text{entity} \end{bmatrix}$ classifiers: e.g. dì-sān-zhī-jī (OM[6]-3-cl-chicken) 'the third chicken'; *dì-sān-jīn-jī (OM-3-cattie-chicken); dì-sān-qún-jī (OM-3-group-chicken) 'the third brood of chickens'; dì-sān-zhǒng-jī (OM-3-kind-chicken) 'the third kind of chicken.' Moreover, fractional numbers cannot be associated with

exact / entity] classifiers. For example, bàn-zhī-jī (half--chicken) 'half a chicken'[7]; bàn-jīn-jī (half-ttie-chicken) 'half a cattie of chicken'; bàn-pán- (half-plate-chicken) 'half a plate (i.e. order) of icken'; *bàn-zhǒng-jī (half-type-chicken). These cts may be summarized by the following table:

Table 1

tity	Exact	De-adjectival	Ordinal number	Fractional number
+	-	✓	✓	✓
+	+	*	✓	✓
-	+	✓	*	✓
-	-	✓	✓	*

nce type (I) classifiers are the most developed in ian languages, we shall focus our interest on them. inese has about seventy-five items belonging to pe (I) classifiers and we know of no language that es much beyond this number.

The underlying classificatory scheme for type) classifiers readily lends itself to feature alysis and the structure of the classificatory stem features is best accounted for by a hierarchy basically binary features. The primary distinc-on between nouns is concrete vs. abstract nouns, d concrete nouns must be further divided into tity and non-entity nouns.

This latter distinction tends to be parallel to e commonly recognized count/mass distinction. ere are inherently [- entity] ([- count]) nouns .g. *water*), which must be measured either through ntainers or through the volume content of standard-ed hypothetical containers. There are also in-

herently [+ entity] ([+ count]) nouns, e.g. *boy*, which must be counted and not measured. Moreover, there are nouns which may be either [+ entity] or [- entity], e.g. *paper* in: *two sheets of recycled paper* and *two pounds of (old) paper(s)*. One way to capture the facts would be to allow for two lexical entries in the grammar: $paper_1$ which has the feature specification $\begin{bmatrix} + \text{ mass} \\ + \text{ count} \end{bmatrix}$ and $paper_2$ which has the feature specification $\begin{bmatrix} + \text{ mass} \\ - \text{ count} \end{bmatrix}$. This implies the same kind of difference that obtains between *cow* and *beef* (or *cattle*), which are distinctly different lexical items. However, this proposal does not account for the fact that in English $paper_1$ and $paper_2$ are realized as a *single* lexical item (as contrasted with *cow* and *beef*). Note also that there is a third manifestation of paper, $paper_3$: *John wrote three papers last year.* $Paper_3$ stands out clearly from $paper_1$ and $paper_2$. $Paper_3$ would be specified as $\begin{bmatrix} - \text{ mass} \\ + \text{ count} \end{bmatrix}$ and its syntactic properties are very different from the hypothetical $paper_1$ and $paper_2$ because, for example, it permits no classifier in the measure phrase.

In parallel examples, *three head of cattle* $\begin{bmatrix} - \text{ mass} \\ + \text{ count} \end{bmatrix}$ and *three herds of cattle* $\begin{bmatrix} + \text{ mass} \\ + \text{ count} \end{bmatrix}$, the syntactic properties as well as the basic feature specification are also parallel. The commonly accepted notion that *three herds of cattle* indicates a collective noun might be formalized into the feature [$\pm$ collective], but this would only be an ad hoc measure. The choice of *head* or *herd* has no bearing on the inherent countability of the noun being classified in the measure phrase. The notion 'collective' goes with the measure phrase and not with the noun *cattle*.

In the system proposed here, the oppositions entity] and [± exact] can readily handle these fficulties. Each lexical entry will include a set hierarchically organized semantic features and ch feature, such as [entity] and [exact] may be rked plus or minus or else left unmarked. Such a nfiguration of feature specifications will define e entry and reflect the classificatory system. us, considering only these two features, abstract uns will be usually marked [- entity, - exact] so that they n only occur with classifiers such as *type* or *nd* (by means of late rewrite rules) and no other nd of classifiers. Since all nouns may be classi-ed with *type* or *kind*, it may be stated as a ta rule that all nouns may be *optionally* specified [- entity, - exact]. Non-abstract nouns which have other tions will need other specifications. For example, mass noun such as *water* will be marked [α entity, -α exact] capture the fact that the values of its feature ecification will be opposite. This will account r *three cans of water* [+ entity, - exact], *three gallons water* [- entity, + exact] and **three pieces of water* entity, exact], as well as *two kinds of water* (soft and rd): [- entity, - exact]. An example such as *boy* will be rked [+ entity, α exact] to provide for *three boys* [+ entity, + exact] d *three groups of boys* [+ entity, - exact] (as well as *two nds of boys* by the optional meta rule). *Cattle* ll be left unmarked and this will provide for *three ad of cattle, three herds of cattle, three thousand unds of cattle,* and *three kinds of cattle.*

Particular languages will require additional atures. Thus, English nouns marked with [+ entity, - exact] ll have other feature specifications that will

determine the use of *horde, herd, bevy, school, pack, pride, brood,* etc. In many Asian and American languages, nouns with specification $\begin{bmatrix}+\text{ entity}\\+\text{ exact}\end{bmatrix}$ will have further feature specifications for the proper use of classifiers pertinent to round objects, long slender objects, flat objects, rigid objects, etc. It has been suggested by numerous writers that dictionary entries should include pertinent classifiers for nouns with the specification $\begin{bmatrix}+\text{ entity}\\+\text{ exact}\end{bmatrix}$. This is excellent for language pedagogy but it misses an important linguistic generalization that is within the competence of the native speakers.

The classificatory system implied by these special classifiers is a highly productive system, governing the perception of the universe of objects and matter. Thus, when confronted with quantification of new objects, native speakers will always know the proper use of classifiers within fairly narrow margins.[8] With allowance for culture-dependen idiosyncrasies, the features of the classificatory systems used in natural language may be generalized. In the case of child acquisition of language, there is clear evidence that processes of generalization develop for classifiers[9] in a way that is parallel to the learning of irregular verbal conjugations and irregular plurals in European languages.

There is also internal evidence that classificatory systems are very much part of the grammar. In most of the languages under discussion, there is a human/non-human (i.e. animal) distinction which is manifested as two distinct classifiers. It is also commonly observed that the non-human (i.e. animal) classifier may be used on purpose with human nouns.[10]

e resulting implication is one of sarcasm or ridi-
le. The resulting sentences are not ungrammatical--
more than such English examples as *intoxicated*
th love and *a grief ago*.[11] These are instances of
tentional violation of selectional restrictions.
ey are very much in evidence and the ability to
oduce them is part of linguistic competence and
st be accounted for in an adequate grammar.

The semantic features for subdividing $\begin{bmatrix} + \text{ entity} \\ + \text{ exact} \end{bmatrix}$
uns in the classificatory systems of Asian lan-
ages may be drawn from the following groups of
tributes:

(1) Geometrical shape/form[12]
  a) spherical or round
    e.g. Burmese loung; Vietnamese gua?
  b) long, slender or cylindrical
    e.g. Burmese chaung; Camb. daəm; Viet. sợi
  c) possessing flat surface, with or without thickness
    e.g. Burmese cha?; Camb. sonlʏk; Viet. bãi, tờ
  d) *horizontal (vs. vertical) orientation*
    e.g. Cantonese[13] po

(2) Natural attributes
  a) fauna
    e.g. Camb. kɔntuy; Chinese zhi
  b) flora
    e.g. Viet. doa; Chinese duo; Malay kuntum
  c) arboreal
    e.g. Camb. tèm; Viet. cây
  d) head
    e.g. Chinese tou

e) tail

e.g. Camb. kɔntùy; Malay ekor; Mandarin wei

f) marine (vs. land-based)

e.g. Chinese shou

(3) Qualitative attributes

a) rigidity

e.g. Chinese tiao (flexible) vs. gen (rigid)

b) size/bulk

e.g. Camb. phaen vs. dom; Chinese li vs. kuai

c) meritorious

e.g. Viet. đức (vs. tính)

d) stationary/mobile

e.g. Chinese jo (site)

(4) Social attributes

a) politeness

e.g. Burmese un vs. yau?; Chinese ge vs. wei

b) social status

e.g. Burmese pa; Camb. 'ɔŋ(k)

c) sex (male/female)

e.g. Viet. chàng vs. nàng

(5) Cultural attributes

a) literary/written records

e.g. Chinese ben; Viet. bản Burmese iŋga

b) culinary

e.g. Burmese na?; Viet. bức

c) clothing

e.g. Camb. sɔmrap; Cantonese tyut

d) instrumental

e.g. Burmese lɛ?; Chinese ba

e) edifice
e.g. Viet. ngoi, gian; Chinese jian

f) mechanical
e.g. Chinese ja

g) vehicular
e.g. Burmese si; Viet. chiec; Chinese liang

h) domesticated
e.g. Chinese zhi (vs. tou)

(6) Congregational attributes

a) duality (pair)
e.g. Burmese yaŋ; Viet. cap; Chinese shuang, tui; Cantonese ma

b) set (non-uniform objects)
e.g. Burmese khain; Chinese tao

c) set (uniform objects)
e.g. Chinese fu

(7) Inherent attributes

a) animate
e.g. Viet. cai (vs. con)

b) human/animal
e.g. Camb. neak; Chinese zhi (non-human)

Most languages have additional idiosyncratic classifiers as well as one or two general classifiers for objects that are not yet completely classified.[14] The semantic space defined by the above attributes may vary in individual languages because the number and type of features may vary. This is to be expected and one finds parallels in distinctive feature systems used in phonological description. For example, certain languages may have only two vowel heights, while others may have four; some may have round and unrounded vowels and others may not;

some may have voicing or tenseness contrast while others may not.

Under close examination, the data usually yield a complex underlying classificatory system with hierarchically nested features.[15] The hierarchy of features can be internally justified.[16] Thus [size] *must be lower in the hierarchy than [shape] or* [edifice] and [shape] must be lower than [mechanical] If [vehicular] is also utilized, the choice as to which of the two features, [vehicular] and [mechanical is the dominant one may vary from language to language. But above all, [duality] and [set] must dominate [mechanical], that is, the use of classifiers characterized by these features (e.g. those meaning 'pair', 'set') invariably preempts any other $\begin{bmatrix}+\text{entity}\\+\text{exact}\end{bmatrix}$ classifiers. For example, a mechanical device may have certain familiar geometrical dimensions, but the classifier *jià* preempts the use of any classifiers relating to shape and if the machines are perceived as pairs, the classifier *jià* is preempted. Furthermore, if a distinction is made between three-dimensional and two-dimensional objects, the three-dimensional features usually dominate.

In our investigations, the hierarchy of features in Chinese is generally parallel to those in Korean[17] and Japanese. This is to be expected because many of the classifiers used in these languages may be traced to a Sino-Korean or a Sino-Japanese origin. There are, in fact, very few classifiers of native origin in Japanese. There are two important basic classifiers [-tsu] for inanimate objects and [-ri] for human beings, and these still coexist with the Sino-Japanese classifiers e.g. otoko yottari ∿ otoko

nin (men four - cl) 'four men'. In the older forms the language only native classifiers were used.[18]

Phylogenetic Developments in Classifiers

The ample data on classifiers from different nguages point to much variation both in the number classifiers and in the classification. The uidity is usually associated with the classifiers nouns in $\begin{bmatrix} + \text{ entity} \\ + \text{ exact} \end{bmatrix}$ measure. There are frequently re changes in the membership of nouns than changes the classificatory system. In the last section we inted out that many languages have one or two neral classifiers for nouns. It is frequently the se that some of the nouns associated with the neral classifier(s) will acquire specific classi-ers in time. This is indicative of the diffusional ture of membership in a classificatory system.[19] us there are two dimensions to developments in minal classifiers: (1) latitudinal development in e features of the classificatory system, including bsequent developments in syntax; and (2) longitu-nal development of the membership of nominal sets fined by features in the classificatory system. om about four thousand years of written records in inese, we can draw a set of comprehensive data on e historical development of classifiers in at least e natural language.

The occurrence of classifiers was already ob-rved in Chinese in the Second Millennium B.C. The llowing is a summary of latitudinal developments:

(I) $\left\{ \begin{matrix} \text{Num N} \\ \text{N Num} \end{matrix} \right\}$ (one-horse; one-man) (horse-one; man-one)

(II) $N_x \text{Num } N_x$ (horse-one-horse; man-one-man)

(III) N Num Cl (horse-one-cl; man-one-person)

(IV) Num Cl N (one-cl-horse; one-person-man

Stage (I) may be characterized by the following rules:

Rule 1 NP → (MP) + N (MP=measure phrase)

Rule 2 MP → Num + (M) (M=measure) (The absence of M is correlated with $\begin{bmatrix}+\text{ entity}\\+\text{ exact}\end{bmatrix}$ measure)

Rule 3 $M \rightarrow \left\{\begin{matrix}\text{'weight units'}\\\text{'volume units'}\\\text{etc.}\end{matrix}\right\}$ (for $\begin{bmatrix}+\text{ entity}\\-\text{ exact}\end{bmatrix}$ measure)

Rule 4 (opt) $[[\text{Num} + (\text{M})]_{\text{MP}} + \text{N}]_{\text{NP}} \Longrightarrow 3 \quad 1 \quad 2$ [20]

Stage (II) represents further development in the measure phrase such that in a $\begin{bmatrix}+\text{ entity}\\+\text{ exact}\end{bmatrix}$ measure phrase, the measure word M is realized as the noun identical to the head of the nominal construction. This may be characterized by modification of Rule 2:

Rule 2a MP → Num + M

Rule 2b $M \rightarrow \left\{\begin{matrix}M_1\\M_2\end{matrix}\right\}$ (M_2 is for $\begin{bmatrix}+\textit{ entity}\\+\text{ exact}\end{bmatrix}$ measure)

and by adding Rule 5a:

Rule 5a $[\text{N} + [\text{Num} + \text{M}_2]_{\text{MP}}]_{\text{NP}} \Longrightarrow 1 \quad 2 \quad 1$

It may be noted that there are *no* instances of:

$$*[\text{Num} + \text{M} + \text{N}]_{\text{NP}} \text{ (where M = N)}$$

Rule 4 feeds Rule 5a and the special new rule may be viewed as an insertion after Rule 4.

In the written records of the 16th Century B.C. ›th (I) and (II) occurred generally in free varia- lon. This indicates that the optionality of Rule 4, ıich provides input for Rule 5a, was maintained at ıe beginning of Stage (II).

By the 3rd Century B.C. (Chin Dynasty) some of ıe derived measure words (i.e. those identical to ıe head nouns) had acquired independent status. ıey became specific classifiers for a small set of ›uns, and the number of classifiers was no longer s large as the number of countable objects (nouns). ıis is basically the inception of generalization in classificatory system that was beginning to develop. ıere were only a handful of such specific classi- iers. Some examples from the period are listed elow:

pǐ	for horses
shèng / liǎng	for vehicles
zhàng	for tents
gē	for bamboo
tóu	for cattle
rén	for human beings

ome of these are transparent classifiers.[21] For xample, rén (person) and liǎng (vehicle) are derived rom the name of the class of nouns, and tóu (head) rom a body part. The selection of noun classes in he membership of the classificatory scheme reflects ne basic aspects of a developed sedentary culture nd this stage may be viewed as the linguistic mani- estation of developments in cultural values.[22] In he matter of linguistic structure, this means that

the rewrite rule for M also underwent changes because certain noun classes were no longer governed by Rule 5a. Moreover, a small set of features had been added to the classificatory system. Since the use of specific classifiers was mostly optional, this situation may be formalized by the optional addition of Rule 5b which may be generalized as follows:

Rule 5b

$$M \begin{bmatrix} + \text{ entity} \\ + \text{ exact} \\ \alpha \text{ feature}_1 \\ \beta \text{ feature}_2 \\ \cdot \\ \cdot \end{bmatrix} \longrightarrow \begin{Bmatrix} a \\ b \\ c \\ \cdot \end{Bmatrix}$$

(where the choice of any of the lexical items a, b, c, ... is dependent on particular combinations of feature specifications).

This primary development in stage (III) is reflected by the emergence of Rule 5b, which replaced Rule 5a. It should be noted that at this stage the following word order was not yet permitted in $\begin{bmatrix} + \text{ entity} \\ + \text{ exact} \end{bmatrix}$ measure:

*Num + M + N (where M ≠ N)

This shows that Rule 5b, like Rule 5a, must be ordered after Rule 4.

However, by the Han Dynasty (2nd Century B.C. to 3rd Century A.D.), stage (IV) had begun and we have:

Num + M + N (where M ≠ N)

One way to account for this would be to suggest that Rule 4 and Rule 5b had reversed order.[23] The

mber of classifiers had also increased greatly by
is period, continuing an enrichment of the classi-
catory system that had begun a short time ago.
is included the development of the hierarchical
ature system implied by additional specific classi-
ers for small round objects, long slender objects
varying sizes and rigidity, flat objects with or
thout significant thickness and variation in size,
d upright objects such as staffs, edifices, trees,
ensils with handles, and flowers. At the same time
ese rapid latitudinal developments (represented by
e displacement of 5a by 5b) had a parallel in
ngitudinal development so that all nouns now parti-
pated in this classificatory system, even though
significant number of nouns were still associated
th a few general classifiers.

Subsequent centuries have seen further enrich-
nts of the classificatory system. Longitudinal
velopments in increasing numbers of nouns which
d been associated with general classifiers are
flected in the reassignment of these nouns to
ecific classifiers. As in most languages, there is
ill a large residue of nouns which are associated
th general classifiers in present-day Chinese.
wever, by as early as the Tang dynasty (7th Century
9th Century A.D.), Chinese had totally left stage
I) so that there were no more instances of classi-
ers identical to the nouns they were associated
th.

Stage (IV), which is modern Chinese, appears to
transitional for it includes stage (III). How-
er, in modern Chinese stage (III) occurs *only*
the special case of enumeration or inventory

taking and even then it occurs infrequently in the colloquial language. There is a touch of bookishness in using the post nominal measure phrase. The special status of enumeration has been very persistent in the development of the measure phrase in Chinese and was probably instrumental in its initial development. In terms of syntactic change, the optionality of rule 4 has been greatly curtailed in modern colloquial Chinese.

4. *Genetic Relationship and Areal Diffusion*

The model of phylogenetic developments in Chinese specific classifiers is paralleled to varying degrees in the other languages of Southeast Asia, as far as it could be deduced from available data. For example, stage (II) is still very much in evidence in Thai, Burmese, Vietnamese, and Malay, while the primary systems in these languages are like stages (III) and (IV). The persistent problem of enumeration and the variable ordering between the head noun and the measure phrase is quite common. In contrast, native Korean, native Japanese, the Altaic languages, and some of the Malayo-Polynesian languages are basically at a stage of development that is similar to the initial phase of stage (III), where a few (general) classifiers are in use. These languages cover an extremely large area and it appears that classificatory systems involving either one or two general classifiers or numerous highly developed specific classifiers are universal to the area.

Jones (1970), following Emeneau's hypothesis on the spread of areal features, suggested that the Tai group of languages was a likely source of influence in the spread of the use of classifiers in Southeast Asia and China. He said (1970:1-2):

'In Cambodian and Malay the incidence of classifiers is relatively high but less so than in Thai or Amoy Chinese, for example; in Javanese the incidence is somewhat lower; in Indonesian they are relatively rare; in the Philippines they are virtually non-existent. Moreover, in the peripheral areas the syntactic structures in which classifiers occur seem less stable and subject to variation.'[24]

The implication of this statement is that ndonesian and Malay are two distinct languages which they are not) and that of the three related anguages, Malay, Javanese, and Indonesian, Malay as the most classifiers, Indonesian the least, and avanese in between. This is incorrect, though, for ndonesian has almost as many classifiers as Malay though they are disappearing more rapidly[25]) and avanese seems to have no specific classifiers. What avanese has for $\begin{bmatrix}+\text{ entity}\\+\text{ exact}\end{bmatrix}$ measure is a general lassifier that parallels the basic general classiiers (-tsu) and (-ari) in the native Japanese ystem. This form is idji, which is related to the alay classifier biji (Indonesian bidji) used for nall globular or roundish objects. The numeral one' in Javanese has incorporated this item: awidji ∿ sidji, and it seems to contain remnants f an older classifier system in Javanese. Lorentz 1972) has surveyed the following widely separated anguages in the Malayo-Polynesian language family nd has found strong evidence of either existing ystems of classifiers or remnants of such systems: tjehnese, Batak, Bolaang Mongondow (on Northern ılawesi), Buginese, Dayak, Makassarese, Niasic, alauan,[26] Tagalog, Rottinese (on Timor), Tontemboan,

Trukese, and Wolio (on Buton).[27]

The assertion that classifiers (either general or specific) are universal to such a wide area immediately invites questions of genetic relationship and possible causes that account for the differential rates of development. Of immediate interest are the different directions of development that are observed for Javanese and Japanese (and Korean). These languages have basically similar systems of simple native classifiers and have been in close proximity to languages that have highly developed specific classifiers. Why is it, then, that Japanese (and Korean) adopted the system from the unrelated Chinese, whereas Javanese apparently remained untouched by the system in its closely related neighbor, Malay/Indonesian? The answer must be sought in the realm of socio-linguistics, for China has dominated Japan (and Korea) culturally, if not politically, until recent times, whereas Javanese society has the dominant role in the Malaysian archipelago. It is therefore possible for Javanese to remain relatively undeveloped in its classifier system while Malay (which in recent times had been dominated by Thai) continues to develop its classificatory system with probable influence from languages of the North.

In considering classificatory systems as evidence for genetic relationship, we have seen that in the case of Javanese and Indonesian there can be divergent developments, even though we can trace similarities in the earlier structure. In considering more remote and unlikely relationship, we can take the example of Chinese and Malay. The general latitudinal spread of classifiers is quite similar. However,

here are deep-seated differences which throw serious oubt on any consideration of genetic relationship. mong the Sinitic languages there are differences in he classifier systems. For example, the features mechanical] and [vehicular] (through the classifiers a and bu respectively) interact differently in andarin and Cantonese, so that while aeroplane is ssociated with ja in both, automobile is associated ith ja in Cantonese, and with bu in Mandarin. These re recently imported cultural items, and one would eadily accept slight variations. On the other hand, here is much overall agreement in the use of classiiers for more traditionally established and fundaental objects. For example, in the area of body arts, both Mandarin and Cantonese (as well as others) ssociate sole objects such as head, nose, throat, orehead, neck, etc. with the general classifier ge nd inherently paired or multiple objects such as yes, ears, hands, fingers, feet, toes, legs, thighs, eeth, etc. with the general classifier for nonuman nouns zhi. There is further agreement on the se of tiao (non rigid, long slender objects) for ail, tongue, and eyebrows.[28]

Now consider the classificatory system as far s it concerns body parts in Malay and Iban (Dayak) Omar 1972:93):

> In choosing classifiers for the various parts of the body, both Malay and Iban divide these parts into two categories: those which occur in pairs and those which do not. The body parts which do not occur in pairs are the fingers, the toes, the chin, the mouth, the neck, etc. These noun-objects do not take any classifier in Malay. In Iban, the toes and the fingers take the classifier lambar, because they number more than one and more than just a pair.

As such, they are treated like other noun-objects (cf. A7), while those body parts which occur as sole objects, like the mouth, the nose, etc., do not require any classifier.

The parts of the body which occur in pairs are the hands, the feet, the legs, the eyes, the ears, and the cheeks. The classifiers for these are Malay: belah; Iban: piak.

Exx. 9. Malay Examples

i. sebelah kaki *a cl foot = a foot*
ii. dua belah kaki *two cl foot = both the feet*
iii. dua belah tangan *two cl hand = both the hands*

Exx. 10. Iban Examples

i. sepiak kaki *a cl foot = a foot*
ii. dua piak kaki *two cl foot = both the feet*
iii. dua piak jari *two cl hand = both the hands*

Supposing these parts of the body, which in ordinary human beings occur only in pairs, happen to consist of more than two in some other being, then no classifier is used.

Exx. 11. Malay Examples

i. lima tangan *five hand = five hands*
ii. tiga telinga *three ear = three ears*

Exx. 12. Iban Examples

i. lima jari *five hand = five hands*
ii. tiga pending *three ear = three ears*

The body parts which form exceptions to the rule given above are the eyes and the head. Besides taking belah (which literally means *half*), the eyes in Malay can also take the classifier biji. Likewise in Iban, the eyes can take the classifier igi or leka as an alternative to piak (cf. A1). Biji in Malay and igi in Iban are used as classifiers for the head, as the head is marked by roundness in shape (cf. A1).

It is quite clear that both latitudinal and longitudinal differences between Chinese and Malay are far greater than those found among closely relate

nembers of the same family. In Chinese, classifiers are used for all body parts and they involve the opposition between singularity and plurality. The fundamentally different system in Malay makes use of the opposition between duality and non-duality and, moreover, sole objects do *not* require any classifier. The Iban case represents a secondary refinement of the Malay system in that the opposition is between paired and unpaired objects, and unpaired objects are separated into singular and plural objects, with inherently single objects taking no classifier. The inter-family difference is in the choice of features and in the difference between having or not having zero-morphs. On the other hand, intra-family difference involves a relatively minor difference in the hierarchical organization of features.

We may consider further the example of units of time, which form another set of fundamental cultural items. It is generally said that nouns representing units of time require *no* classifiers[29] and this fact may be construed as an exception to our hypothesis. This is not the case because units of measure such as those of weight and volume are not used with classifiers, because they themselves are 'classifiers' in the system proposed here. It has also been suggested that syntactic properties of time units tend to be similar to those of units of weight, volume, distance, etc. There is evidence to support this. For example, as mentioned above, both units of this kind and time units are not associated with classifiers. However, time measures are fundamentally different from physical measures. Measurements of time are very old, probably since man became concerned about a calendar. In most of Asia, time has been

measured by means of the lunar calendar whose units are by no means exact. There are, for example, great variations, not only in the form of 'leap years' but also in the form of 'leap months'. Thus a time measure is a $\begin{bmatrix}+\text{ entity}\\-\text{ exact}\end{bmatrix}$ measure. It is different from physical measurement, which is characterized as $\begin{bmatrix}+\text{ entity}\\+\text{ exact}\end{bmatrix}$ measure by the system proposed here. Moreover, there is syntactic evidence to support this. In Chinese, for example, 'pound' cannot co-occur with ordinal numbers, but 'year' can, while both are amenable to fractional numbers and de-adjectival constructions.[30] Thus 'year' has the same kind of syntactic properties as *qun* 'group', which is $\begin{bmatrix}+\text{ entity}\\-\text{ exact}\end{bmatrix}$. *There is a superficial surface difference* between units of time such as 'year' and other $\begin{bmatrix}+\text{ entity}\\-\text{ exact}\end{bmatrix}$ classifiers: the head noun 'time' is usually omitted when it is associated with a measure phrase, whereas the head noun associated with a measure phrase involving 'group', for example, is usually present in the surface structure of languages which do not have too many subdivisions (i.e. specific classifiers) in this category. English has a highly developed infrastructure of subdivisions such as the aforementioned: *pride* (lion), *herd* (cattle), *school* (fish), etc. English also has a general classifier for this category: *group*. Thus, if an Englis speaker uses *herds* by itself, it is clearly understood that he is referring to cattle, and classifiers such as units of time can only refer to time. On the other hand, classifiers such as units of weight and volume can refer to any kind of matter and must therefore include their head nouns when associated with a measure phrase, unless they are clear from specific contexts.

The above discussion suggests that units of time re not relevant to the study of genetic relationship, ut this is not the case. In Modern Chinese, 'month', week', 'hour' and 'second', for example, take classi- iers, whereas 'year', 'day', 'night' (monosyllabic) nd 'minutes', for example, do not take classifiers.[31] his is common to all of the major dialects that we ave studied, and it confirms the genetic relation- hip that is already known.

In Japanese and Korean, these idosyncrasies are ot in evidence, even though their systems of nominal lassifiers are basically of Chinese origin. The omparative evidence therefore underscores the tructural difference underlying the genetic differ- nce. This is further substantiated by the specific nits that were borrowed from Chinese into Japanese nd Korean. Vietnamese, for example, borrowed many nits from Chinese, and Burmese borrowed from anskrit, but the residual original systems in Viet- amese and Burmese are more closely related to each ther than either one is to Chinese. This again nderscores systematic differences between genetic elationship and areal diffusion.

There is no proven discovery procedure for deter- ining genetic relationships, but from about the time f the Neogrammarians, linguists have been continually earching for systematic means to determine genetic elationship. It is the purpose of this paper to how that the comparison of classifier systems may e an important area for historical linguistics, articularly in Asia. Studies of this kind can in ime help eradicate Meillet's (1925:26-27) bias owards the applicability of the comparative method

to linguistic reconstruction in Asian languages:

> '...les langues d'Extrême-Orient qui, comme le chinois ou l'annamite, n'offrent presque pas de particularités morphologiques, n'ont par là même rien où puisse se prendre le linguiste qui essaie de trouver des langues parentes aux parlers chinois ou aux parlers annamites; et la restitution d'une "langue commune" dont le chinois, le tibétain, etc., par exemple, seraient des formes postérieures, se heurte à des obstacles quasi invincibles.'

---

[1]The work reported here was supported in part by the National Science Foundation (GN-244) and by the Rome Air Development Center (F30602-71-C-0116). The ideas in this paper grew out of T'sou 1965 and *a graduate seminar of the same title as this paper* given at UCSD in Spring 1972. The author is grateful to the participating students of the seminar, especially to Alec Bamford, Chanida Chanyapate, Suzette Elgin, Jo-Ann Flora, Masako Inoue, Charles Lee, Ove Lorentz, Alexis Takizala, Truong Ngoc Tinh, and Marina Tsang. He is also grateful to Pete Becker and Don Forman for their enlightening comments to the first draft of this paper.

[2]See Emeneau (1956).

[3]See Haas (1948), Hoijer (1945), and Lander (1964).

[4]T'sou (1965).

[5]See Chao (1968), Nguyen Dinh Hoa (1957).

[6]Ordinal number prefix.

[7]There is a major distinction between fractional numbers and integers. The *de*-modifier construction may involve $\begin{bmatrix} + \text{ exact} \\ - \text{ entity} \end{bmatrix}$ *classifiers if the number in* the measure phrase involves a fraction: bàn-zhī-de-jī, sān zhī-bàn-de-jī.

[8]This phenomenon is similar to the use of gender in European languages as well as in the Bantu languages, which employ much more than a three-way distinction (see Takizala (1972) and Givon (1970)). The classificatory systems do not classify the nouns, but objects and matter in the perceptual universe of the native speakers.

[9]See, for example, Lai-ha Li (1968:44ff).

[10]This is the usual case in Fuzhou, a member of the Min dialect group of China.

[11]Dylan Thomas.

[12]Burmese examples are from Hla Pe (1965); Vietnamese examples from Nguyen Dinh Hoa (1957); and Cambodian examples from Jacob (1965).

[13]Unless otherwise specified, Chinese romanizations are given in the Pinyin system (without tones). Cantonese romanizations are based on T'sou (1972a).

[14]From this we can infer a certain diffusional character in the development of classificatory systems (see Section 3).

[15]See Mohr (1968), Coyaud (1968), and T'sou (forthcoming).

[16]There is evidence in the study of Chinese historical phonology which bears on the question of hierarchical organization in phonological features. For example, it is generally accepted that archaic Chinese (2nd Millennium B.C.) has three kinds of syllabic endings: voiced and voiceless stops and nasals [b, d, g, p, t, k, m, n, ŋ]. In most of the modern dialects in the North (e.g. Mandarin) only two nasals remain: [n, ŋ]. The historical changes that gave rise to this situation generally involved consonantal merger. It is interesting to note that by about 300 A.D. the voiced consonant [b] had merged with [d] and that by 1400 A.D. Early Modern Mandarin had merged [m] with [n]. One good way to account for the reccurrence of labial and dental merger is to postulate a hierarchy of features. Similarly, postulating language specifier feature hierarchy is one possible way to account for why [θ] is commonly manifested as either [t] or [s] in those languages that do not have [θ], but have both [t] and [s].

[17]See Mohr (1968).

[18]See, for example, Samson (1928:82-84).

[19]There can be also decreasing membership in the classificatory system. The case of Indonesian and Malay is a good example (see Omar 1972: 95), as will be discussed later. Such 'receding' developments are usually conditional or accompanied by syntactic changes.

[20]The underlying word order is problematical here. The order N-Num presents an interesting problem as to whether Num is a predicate, because it is in the position of the predicate. If this were the case, then Rule 4 may not be justified because stage (I) may be said to contain two kinds of structures in which numerals may occur. A much more detailed study will be necessary to clear up this matter. The formulation proposed here is tentative and serves as an example for discussion.

[21]See Omar (1972:88).

[22]Some scholars have argued that the rise of specific classifiers were motivated by an attempt at avoiding problems of homophony (see, for example, Samson 1928). For a detailed study of homophony and language change, see T'sou (1971).

[23]For a detailed discussion of reordering in syntactic change, see T'sou (1972b).

[24]Herrfurth (1964: 3) made similar claims about the distribution of classifiers in Asia.

[25]See Omar (1972), Lorentz (1972).

[26]Based on the field notes of Jo-Ann Flora.

[27]On the other hand, it appears that Polynesian languages generally do not make use of classifiers for $\begin{bmatrix} + \text{ entity} \\ + \text{ exact} \end{bmatrix}$ measure.

[28]As may be expected, in reference to the neck of a giraffe, speakers from both languages tend to use tiao, the classifier for flexible long slender objects.

[29]For example: Burmese (Hla Pe 1965:165,181); Nung (Saul 1965:289); Vietnamese (Nguyen Dinh Hoa 1957:131); Malay and Iban (Omar 1972:94); Archaic and Ancient Chinese; Korean; and Japanese.

[30]See Table 1.

[31]The cause for this is related to the names given to the units of the calendar, such as the months in the year, and the ambiguity created by such a system (see T'sou, forthcoming).

REFERENCES

enton, Richard A. 1968. *Substitutes: Classifiers in Trukese.* University of Hawaii M.A. thesis.

erlin, Brent. 1968. Tzeltal Numeral Classifiers: a Study in Ethnographic Semantics. The Hague: Mouton.

erlin, Brent, and A. Kimbal Romney. 1964. Descriptive Semantics of Tzeltal Numeral Classifiers. *American Anthropologist* Special Publications, vol. 66, no. 3, part 2, pp. 79-98.

urling, Robbins. 1965. Burmese Numeral Classifiers. *Context and Meaning in Cultural Anthropology*, ed. by Melford E. Spiro, 243-264.

hao, Yuan Ren. 1968. *A Grammar of Spoken Chinese.* Berkeley.

hou, Fa Kao. 1956. *A History of Ancient Chinese.* [In Chinese.] Academia Sinica Special Publications No. 39.

oyaud, Maurice. 1968. Analyse et classification automatique en Chinois. T.A. Information: Revue International du Traitement Automatique du Langage 1.18-29.

avidson, William, L.W. Elford, and Harry Hoijer. 1963. Athapaskan Classificatory Verbs. *Studies in the Athapaskan Languages*, ed. by Harry Hoijer, 30-41. (University of California Publications in Linguistics No. 29.)

eloustal, R. 1914. Des déterminatifs en annamite. BEFEO, 14:5.

meneau, M. 1951. *Studies in Vietnamese Grammar.* Berkeley.

_____. 1956. India as a Linguistic Area. *Language* 32.

_____. 1962. An Indo-Iranian Areal Isogloss. *Dravidian and Indic Languages.* University of California, Berkeley.

ao, Ming Kai. 1957. *On Chinese Grammar.* [Revised edition; in Chinese.] Shanghai.

ivon, Talmy. 1970. Historical Changes in the Noun-Class System of Bantu, their Possible Causes and Wider Implications. *Working Papers in African Linguistics.*

Haas, Mary. 1942. The Use of Numeral Classifiers in Thai. *Language* 18.

_____. 1948. Classificatory Verbs in Muskogee. *International Journal of American Linguistics* 14. 244-246.

_____. 1951. The Use of Numeral Classifiers in Burmese. University of California Publications in Semitic Philology II. Berkeley.

Haguenhauer, Ch. 1951. *Morphologie du japonais moderne*. Paris: Klincksieck.

Herfurth, H. 1964. Die Klassifikationsaffixe in Chinesischen. *Beiträge zum Probleme des Wortes in Chinesischen*. Vol. II. Berlin: P. Ratchnevsky.

Hla Pe. 1965. A Re-examination of Burmese Classifiers. *Indo-Pacific Linguistic Studies*, ed. by G.B. Milner, and E.J.A. Henderson, part 2.163-85.

Hoijer, Harry. 1945. Classificatory Verb Stems in the Apachean Languages. *International Journal of American Linguistics* 11.13-23.

Huang, Zaijun. 1964. A Study of the Origin and Development of Classifiers in Chinese, beginning with their Use in the Oracle Bone Inscriptions and Jinwen. [In Chinese.] *Zhongguo Yuwen* 6. 432-441.

Jacob, Judith M. 1965. Notes on the Numerals and Numeral Coefficients in Old, Middle, and Modern Khmer. *Indo-Pacific Linguistic Studies*. Amsterdam.

Jones, Robert B. 1970. Classifier Constructions in Southeast Asia. *JAOS* 90.1-12.

Lander, H. 1964. Seven Navaho Verbs of Eating. *International Journal of American Linguistics* 30. 94-96.

Li, Lai-ha. 1968. The Development of the Syntax of a Cantonese Speaking Child from 1 1/2 to 2 1/2 Years of Age. University of Malaya M.A. thesis.

Ling, Quan Fang. 1955. On Unit Classifiers Concerning Man. [In Chinese.] *Yuyan Zhishi*, no. 6.

Liu, Shiru. 1965. A Study of Nominal Classifiers During the Wei Ji Nan Bei Chao. [In Chinese.] *Zhonghua Shuji*.

Lorentz, Ove. 1972. Classifiers, especially Indonesian. University of California, San Diego, unpublished paper.

eillet, A. 1925. *La méthode comparative en linguistique historique*. Instituttet for Sammenlignende Kulturforskning. [Reference to the 1966 re-edition, ed. by Honoré Champion.]

iller, Carolyn P. 1954. The Substantive Phrase in Brou. *Monkhmer Studies I*. Saigon.

ohr, Burckhard. 1968. Numeral Classifiers in Korean. University of California, Berkeley, unpublished paper.

guyen Dinh Hoa. 1957. Classifiers in Vietnamese. *Word* 13.124-152.

kell, John. 1965. Nissaya Burmese: A Case of Systematic Adaption to a Foreign Grammar and Syntax. *Indo-Pacific Linguistic Studies*, ed. by Milner and Henderson, part 2.186-227.

_____. 1969. *A Reference Grammar of Burmese*. Oxford.

mar, Asmah Haji. 1972. Numeral Classifiers in Malay and Iban. *Anthropological Linguistics* 14.87-96.

amson, Geoffrey B. 1928. *A Historical Grammar of Japanese*.

aul, Janice D. 1965. Classifiers in Nung. *Indo-Pacific Linguistic Studies*, part 2.

chafer, Edward H. 1948. Numeral Classifiers in Classical Chinese. *Language* 24.

turtevant, William C. 1964. Studies in Ethnoscience. *Transcultural Studies in Cognition*, ed. by A.K. Romney, and R.G. D'Andrade, 99-131. *American Anthropologist* Special Publications, vol. 66, no. 3 Part 2, pp. 99-131.

akizala, Alexis. 1972. Syntactic Classification, Semantic Classification. University of California, San Diego, unpublished paper.

'sou, Benjamin K. 1965. The Place of Classifiers in a Generative Grammar of Chinese. Quarterly Progress Report, M.I.T.: Research Laboratory of Electronics, vol. 78.221-26.

_____. 1971. Homophony and Internal Change: the Chinese Case. Paper read at the First California May Day Linguistics Conference, Berkeley.

_____. 1972a. *A Basic Cantonese Grammar*. La Jolla.

T'sou, Benjamin K. 1972b. Reordering in Diachronic Syntax. *Papers from the Eighth Meeting of the Chicago Linguistics Society*, ed. by P. Peranteau, 591-606.

_____. (forthcoming). The Structure of Nominal Classifiers in Chinese.

Vesper, Ethel. 1940. Grammatical and Semantic Classifiers as Evidenced in the Possessive Construction in Kusaien. *The Fourth Annual Kansas Linguistics Conference*, ed. by H. Harris. University of Kansas.

Wang, Li. 1958. *A History of Chinese Language*. Vol. II: Syntax. [In Chinese.] Peking.

Waterson, Natalie. 1966. Numeratives in Uzbek: a Study in Colligation of Collocation. *In Memory of J.R. Firth*, ed. by C.Z. Bazell et al, 454-70.