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, 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, particularly among the Athapaskan languages. In these languages the semantic classification of nouns is manifested in the verbal structure, perhaps reflecting the general tendency of verbal incorporation. Jeanneau has posed a very interesting and important problem concerning the areal spread of classifiers. It will be important to consider areal linguistics in the light of possible independent phylogenetic developments as opposed to a possible combination of both independent evolution and structural borrowing because of prolonged language contact in contiguous geographical areas. A better understanding of the structure and development of such syntactic traits can aid in tracing genetic relationships, particularly those among Austroasiatic languages, Sino-Tibetan languages, and Malayo-Polynesian languages. This calls for more than exhaustive taxonomic descriptions of classifiers in each language. It will be necessary to devise a formal apparatus by which both the synchronic systems as well as diachronic developments of nominal classifiers may be compared.

Moreover, the study of nominal classifier systems suggests an important hypothesis that the use of nominal classifiers and the use of the plural morpheme are in complementary distribution in natural language. More concretely, it suggests that either a natural language has either nominal classifiers or plural morphemes, or b) if a natural language has both kinds of morphemes, then their use is in complementary 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 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) \[\begin{array}{c}
+\text{exact} \\
+\text{entity}
\end{array}\] : zhī ('individual', non-human objects)

(II) \[\begin{array}{c}
+\text{exact} \\
-\text{entity}
\end{array}\] : jīn ('cattie', unit of weight)

(III) \[\begin{array}{c}
-\text{exact} \\
+\text{entity}
\end{array}\] : qún ('brood')

(IV) \[\begin{array}{c}
-\text{exact} \\
-\text{entity}
\end{array}\] : zhòng ('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 entity in that more likely than not the 'Shylockian it' could not have been made. In (III) there is a finite sense of a well-defined discrete entity or entities, but the quantity is not exact either by sign or by convention. For example, a brood of hicks (or a plate of chicken) is not an exact measure but there is a definite sense of physical entity and can be referred to as a unit. This may be contrasted with two pounds of chicken (legs) as in I bought two pounds of chicken (legs) yesterday. Here reference is made to an exact quantity rather than an object and there is the sense of physical entity lacking. In (IV), which characterizes mainly count nouns, the measure is neither exact nor does it refer to a discrete physical entity.

Natural language exhibits all four kinds of measure, but the range of each kind of measure may vary in different languages. Mass nouns in English, in comparison to those in Chinese, may be good examples of range difference. In the case of cattle: two head(s) of cattle, head is a [+ exact] measure; in two herds of cattle, herd is a [+ entity] measure; in twenty thousand pounds of cattle, pound is a [- exact] measure; and in two kinds of cattle, kind refers to [- exact] measure. Count nouns in English and other European languages usually require no overt markers for [+ exact] measure, but the contrary is generally true of the languages in Asia, where (I) embodies a rich and complex classificatory system. The number of categories is culture-bound and relatively finite for (I). It depends on the standard measures of weight, volume, 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, e.g. *two spoonfuls of cough syrup every day, three tables full of paper, etc. (IV), in English, also involves a complex and idio-syncretic 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 [+ exact] classifiers: e.g. dī-sān-zhī-jī (OM3-cl-chicken) 'the third chicken'; *dī-sān-jīn-jī (OM3-cattie-chicken); dī-sān-qún-jī (OM3-group-chicken) 'the third brood of chickens'; dī-sān-zhōng-jī (OM3-kind-chicken) 'the third kind of chicken.' Moreover, fractional numbers cannot be associated with
exact classifiers. For example, bàn-zhǐ-jī (half-chicken) 'half a chicken'; bàn-jīn-jī (half-little-chicken) 'half a cattie of chicken'; bàn-pán-jī (half-plate-chicken) 'half a plate (i.e. order) of chicken'; *bàn-zhǒng-jī (half-type-chicken). These facts may be summarized by the following table:

<table>
<thead>
<tr>
<th>Entity</th>
<th>Exact</th>
<th>De-adjectival</th>
<th>Ordinal number</th>
<th>Fractional number</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
</tr>
</tbody>
</table>

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

The underlying classificatory scheme for type (I) classifiers readily lends itself to feature analysis and the structure of the classificatory system features is best accounted for by a hierarchy of basically binary features. The primary distinction between nouns is concrete vs. abstract nouns, and concrete nouns must be further divided into entity and non-entity nouns.

This latter distinction tends to be parallel to the commonly recognized count/mass distinction. There are inherently [- entity] [- count]] nouns (e.g. water), which must be measured either through containers or through the volume content of standardized 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: paper₁ which has the feature specification [+ mass ] and paper₂ which has the feature specification [+ mass ] . 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 paper₁ and paper₂ are realized as a single lexical item (as contrasted with cow and beef). Note also that there is a third manifestation of paper, paper₃: John wrote three papers last year. Paper₃ stands out clearly from paper₁ and paper₂. Paper₃ would be specified as [- mass ] and its syntactic properties are very different from the hypothetical paper₁ and paper₂ because, for example, it permits no classifier in the measure phrase.

In parallel examples, three head of cattle [- mass ] and three herds of cattle [+ mass ] , 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 [+ 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 \( [\text{entity}] \) and \([+\text{ exact}]\) can readily handle these difficulties. Each lexical entry will include a set hierarchically organized semantic features and each feature, such as \([\text{entity}]\) and \([\text{exact}]\) may be marked plus or minus or else left unmarked. Such a configuration of feature specifications will define the entry and reflect the classificatory system.

Thus, considering only these two features, abstract nouns will be usually marked \([-\text{entity}]\) so that they can only occur with classifiers such as \(\text{type}\) or \(\text{kind}\) (by means of late rewrite rules) and no other kind of classifiers. Since all nouns may be classified with \(\text{type}\) or \(\text{kind}\), it may be stated as a meta rule that all nouns may be optionally specified \([-\text{entity}]\). Non-abstract nouns which have other specifications will need other specifications. For example, mass noun such as \(\text{water}\) will be marked \(\alpha \text{ entity} \)\(\alpha \text{ exact}\) to capture the fact that the values of its feature specification will be opposite. This will account for three cans of water \([+\text{ entity}]\), three gallons \([-\text{entity}]\) and \(*\)three pieces of water \([\alpha \text{ entity}]\) \([\alpha \text{ exact}]\), as well as two kinds of water (soft and hard): \([-\text{entity}]\). An example such as \(\text{boy}\) will be marked \([+\text{ entity}]\) to provide for three boys \([+\text{ entity}]\) and three groups of boys \([+\text{ entity}]\) (as well as two kinds of boys by the optional meta rule). Cattle will be left unmarked and this will provide for three \(\text{ad of cattle}, \text{three herds of cattle}, \text{three thousand unds of cattle}, \text{and three kinds of cattle.}\)

Particular languages will require additional features. Thus, English nouns marked with \([-\text{exact}]\) will 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 $^{+}\text{entity}$ 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 $^{+}\text{exact}$. 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. With allowance for culture-dependent 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 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.
The resulting implication is one of sarcasm or ridicule. The resulting sentences are not ungrammatical—more than such English examples as *intoxicated with love* and *a grief ago*. These are instances of intensional violation of selectional restrictions. They are very much in evidence and the ability to produce them is part of linguistic competence and must be accounted for in an adequate grammar.

The semantic features for subdividing [+entity] + [exact] uns in the classificatory systems of Asian languages may be drawn from the following groups of attributes:

1. Geometrical shape/form
   a) spherical or round
      e.g. Burmese loung; Vietnamese gua?
   b) long, slender or cylindrical
      e.g. Burmese chaung; Camb. daem; Viet. sqú
   c) possessing flat surface, with or without thickness
      e.g. Burmese chaʔ; Camb. sonlvk; Viet. bǎî, tɔ
   d) horizontal (vs. vertical) orientation
      e.g. Cantonese13 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. tɛm; Viet. cây
   d) head
      e.g. Chinese tou
e) tail
e.g. Camb. կոնտույ; Malay ekor; Mandarin wei

f) marine (vs. land-based)
e.g. Chinese shou

(3) Qualitative attributes
a) rigidity
e.g. Chinese tiáo (flexible) vs. gen (rigid)
b) size/bulk
e.g. Camb. phæn vs. dom; Chinese lí vs. kuài
c) meritorious
e.g. Viet. dực (vs. tịnh)
d) stationary/mobile
e.g. Chinese ji (site)

(4) Social attributes
a) politeness
e.g. Burmese un vs. yau?; Chinese ge vs. wei
b) social status
e.g. Burmese pa; Camb. ឈឺ(k)
c) sex (male/female)
e.g. Viet. chàng vs. nàng

(5) Cultural attributes
a) literary/written records
e.g. Chinese běn; Viet. bàn Burmese ႏင္း
b) culinary
e.g. Burmese နာ; Viet. bực
c) clothing
e.g. Camb. sömráp; Cantonese ตยุต
d) instrumental
e.g. Burmese လ၁၇၉၀; 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 yañ; 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. caï (vs. con)
b) human/animal
   e.g. Camb. n̄ḡk; 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. 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. The hierarchy of features can be internally justified. 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 [+entity] classifiers. For example, a mechanical device may have certain familiar geometrical dimensions, but the classifier *jā* preempts the use of any classifiers relating to shape and if the machines are perceived as pairs, the classifier *jā* 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 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 languages point to much variation both in the number classifiers and in the classification. The fluidity is usually associated with the classifiers nouns in [+ entity] measure. There are frequently more changes in the membership of nouns than changes in the classificatory system. In the last section we hinted out that many languages have one or two general classifiers for nouns. It is frequently the case that some of the nouns associated with the general classifier(s) will acquire specific classifiers in time. This is indicative of the diffusional nature of membership in a classificatory system. 19 Thus there are two dimensions to developments in nominal classifiers: (1) latitudinal development in the features of the classificatory system, including subsequent developments in syntax; and (2) longitudinal development of the membership of nominal sets defined by features in the classificatory system.

From about four thousand years of written records in Chinese, we can draw a set of comprehensive data on the historical development of classifiers in at least the natural language.

The occurrence of classifiers was already observed in Chinese in the Second Millennium B.C. The following is a summary of latitudinal developments:

(I) \( \{ \frac{\text{Num}}{\text{N}} \} \) (one-horse; one-man)

(II) \( \times \frac{\text{Num}}{\text{N}} \times \) (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 \[ NP \rightarrow (MP) + N \] \hspace{1cm} (MP=measure phrase)

Rule 2 \[ MP \rightarrow \text{Num} + (M) \] \hspace{1cm} (M=measure)
(The absence of M is correlated with
\hspace{1cm} [+ entity] measure)
\hspace{1cm} [+ exact]

Rule 3 \[ M \rightarrow \begin{cases} \text{'weight units'} & (\text{for } [+ \text{ entity}] \\
\text{'volume units'} & \text{etc.} \end{cases} \] \hspace{1cm} \text{measure}

Rule 4 (opt) \[ ([\text{Num} + (M)]_{MP} + N)_{NP} \rightarrow 1 \hspace{1cm} \frac{20}{2} \hspace{1cm} 3 \]

Stage (II) represents further development in the measure phrase such that in a [+ entity] 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 \[ MP \rightarrow \text{Num} + M \]

Rule 2b \[ M \rightarrow \begin{cases} M_1 \\
M_2 \end{cases} \] \hspace{1cm} (M_2 \text{ is for } [+ \text{ exact}] \text{ measure})

and by adding Rule 5a:

Rule 5a \[ [N + [\text{Num} + M_2]_{MP}]_{NP} \rightarrow 1 \hspace{1cm} 2 \hspace{1cm} 1 \]
\hspace{1cm} \frac{2}{3}

It may be noted that there are no instances of:

\[ *[\text{Num} + M + N]_{NP} \] (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. both (I) and (II) occurred generally in free variation. This indicates that the optionality of Rule 4, which provides input for Rule 5a, was maintained at the beginning of Stage (II).

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

\[
\begin{align*}
pí & \quad \text{for horses} \\
shèng \} & \quad \text{for vehicles} \\
lìăng \} & \quad \text{for tents} \\
gē & \quad \text{for bamboo} \\
tōu & \quad \text{for cattle} \\
rén & \quad \text{for human beings}
\end{align*}
\]

Some of these are transparent classifiers. For example, rén (person) and lìăng (vehicle) are derived from the name of the class of nouns, and tōu (head) from a body part. The selection of noun classes in the membership of the classificatory scheme reflects the basic aspects of a developed sedentary culture and this stage may be viewed as the linguistic manifestation of developments in cultural values. In the 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:

\[
\begin{array}{c}
M \\
+ \text{entity} \\
+ \text{exact} \\
\alpha \text{ feature}_1 \\
\beta \text{ feature}_2 \\
\vdots \\
\vdots
\end{array}
\begin{array}{c}
(a) \\
\{b\} \\
\{c\}
\end{array}
\]

(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 measure:

\*Num + M + N (where M \neq 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:

\text{Num + M + N (where M \neq N)}

One way to account for this would be to suggest that Rule 4 and Rule 5b had reversed order. The
mber of classifiers had also increased greatly by this period, continuing an enrichment of the classificatory system that had begun a short time ago. This included the development of the hierarchical nature system implied by additional specific classifiers for small round objects, long slender objects varying sizes and rigidity, flat objects with or without significant thickness and variation in size, and upright objects such as staffs, edifices, trees, vessels with handles, and flowers. At the same time these rapid latitudinal developments (represented by the displacement of 5a by 5b) had a parallel in longitudinal development so that all nouns now participated in this classificatory system, even though a significant number of nouns were still associated with a few general classifiers.

Subsequent centuries have seen further enrichments of the classificatory system. Longitudinal developments in increasing numbers of nouns which had been associated with general classifiers are reflected in the reassignment of these nouns to specific classifiers. As in most languages, there is still a large residue of nouns which are associated with general classifiers in present-day Chinese.

However, by as early as the Tang dynasty (7th Century to 9th Century A.D.), Chinese had totally left stage (II) so that there were no more instances of classifiers identical to the nouns they were associated with.

Stage (IV), which is modern Chinese, appears to be transitional for it includes stage (III). However, in modern Chinese stage (III) occurs only in 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.'

The implication of this statement is that Indonesian and Malay are two distinct languages (which they are not) and that of the three related languages, Malay, Javanese, and Indonesian, Malay has the most classifiers, Indonesian the least, and Javanese in between. This is incorrect, though, for Indonesian has almost as many classifiers as Malay (though they are disappearing more rapidly) and Javanese seems to have no specific classifiers. What Javanese has for [+ entity, + exact] measure is a general classifier that parallels the basic general classifiers (-tsu) and (-ari) in the native Japanese system. This form is idji, which is related to the Malay classifier biji (Indonesian bidji) used for small globular or roundish objects. The numeral one' in Javanese has incorporated this item: awidji  וה sidji, and it seems to contain remnants of an older classifier system in Javanese. Lorentz (1972) has surveyed the following widely separated languages in the Malayo-Polynesian language family and has found strong evidence of either existing systems of classifiers or remnants of such systems: Jtehnese, Batak, Bolaang Mongondow (on Northern Malawesi), Buginese, Dayak, Makassarese, Niasic, Malauan, 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 doubt on any consideration of genetic relationship. Among the Sinitic languages there are differences in the classifier systems. For example, the features [mechanical] and [vehicular] (through the classifiers ja and bu respectively) interact differently in Mandarin and Cantonese, so that while aeroplane is associated with ja in both, automobile is associated with ja in Cantonese, and with bu in Mandarin. These are recently imported cultural items, and one would readily accept slight variations. On the other hand, there is much overall agreement in the use of classifiers for more traditionally established and fundamental objects. For example, in the area of body parts, both Mandarin and Cantonese (as well as others) associate sole objects such as head, nose, throat, forehead, neck, etc. with the general classifier gen and inherently paired or multiple objects such as yes, ears, hands, fingers, feet, toes, legs, thighs, teeth, etc. with the general classifier for non-human nouns zhi. There is further agreement on the use of tiao (non rigid, long slender objects) for nail, tongue, and eyebrows.28

Now consider the classificatory system as far as 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
members 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 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 \([+\text{entity}][{-}\text{exact}]\) measure. It is different from physical measurement, which is characterized as \([+\text{entity}][{+}\text{exact}]\) 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 \([+\text{entity}][{-}\text{exact}]\). There is a superficial surface difference between units of time such as 'year' and other \([+\text{entity}][{-}\text{exact}]\) 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 English 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 are not relevant to the study of genetic relationship, but this is not the case. In Modern Chinese, 'month', 'week', 'hour' and 'second', for example, take classifiers, whereas 'year', 'day', 'night' (monosyllabic) and 'minutes', for example, do not take classifiers. This is common to all of the major dialects that we have studied, and it confirms the genetic relationship that is already known.

In Japanese and Korean, these idiosyncrasies are not in evidence, even though their systems of nominal classifiers are basically of Chinese origin. The comparative evidence therefore underscores the structural difference underlying the genetic difference. This is further substantiated by the specific units that were borrowed from Chinese into Japanese and Korean. Vietnamese, for example, borrowed many units from Chinese, and Burmese borrowed from Sanskrit, but the residual original systems in Vietnamese and Burmese are more closely related to each other than either one is to Chinese. This again underscores systematic differences between genetic relationship and areal diffusion.

There is no proven discovery procedure for determining genetic relationships, but from about the time of the Neogrammarians, linguists have been continually searching for systematic means to determine genetic relationship. It is the purpose of this paper to show that the comparison of classifier systems may be an important area for historical linguistics, particularly in Asia. Studies of this kind can in time help eradicate Meillet's (1925:26-27) bias towards the applicability of the comparative method.
...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 Chanyapat, 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).


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 [+ exact] classifiers if the number in the measure phrase involves a fraction: bàn-zhǐ-de-jǐ, sān zhǐ-bàn-de-jǐ.

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

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

Dylan Thomas.

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

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

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

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

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 recurrance 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].

See Mohr (1968).

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

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.
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.

See Omar (1972:88).

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).

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

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

See Omar (1972), Lorentz (1972).

Based on the field notes of Jo-Ann Flora.

On the other hand, it appears that Polynesian languages generally do not make use of classifiers for [+ entity] measure.

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.

For example: Burmese (H1a 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.

See Table 1.

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


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


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


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*.

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


'sou. 1972a. A Basic Cantonese Grammar. La Jolla.

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


