Numeral Classifier Systems:
A Southeast Asian Cross-Linguistic Analysis
By Donald R. Goral
University of California, Berkeley

I. INTRODUCTION AND GENERAL REMARKS

Numeral classifier systems occur in languages throughout Southeast Asia and are attested in such non-Asian languages as Tzeltal, a Mayan language (11) and Tarascan, an Amerindian language (29). Classifiers are usually introduced in the context of numeral phrases, corresponding to English constructions such as 'two sheets of paper' or twenty head of cattle', where the classifiers are underlined. However, as Adams et al. (1, p. 2) point out: "Besides their function in numeral noun phrases classifiers in various languages function as nominal substitutes, nominalizers of words in other form classes, markers of definiteness, relativizers, markers of possession, and as vocatives; serve to disambiguate sentences, establish coherence in discourse and regularly mark registers and styles within a language." They also raise the question, without providing an answer, of whether the diverse phenomena listed above can be considered as a single Southeast Asian areal grammatical feature. Beyond this dilemma, there is often the problem of determining when a particular occurrence of a morpheme can be designated as the occurrence of a classifier.

In the literature, as will be discussed further below, classifiers are usually not clearly defined. Instead, a few examples are given, the grammatical category is assumed to exist, and various lexical forms are included or excluded from the set of classifiers of a given language seemingly by the unsubstantiated whim of the investigator. One of the objectives of this study will be to provide rigorous characterizations of classifier systems in various Southeast Asian languages.

The development of the concept of the classifier will begin with the presentation of Vietnamese examples. Then, the general conclusions of various investigators will be discussed, postponing final evaluations until detailed analyses are made of data from individual languages. Guided by the general problems and theories proposed and exposed in the literature, the syntax and semantics of classifiers (hereafter abbreviated CL) will be investigated in Vietnamese, Burmese, Cambodian, Lahu, Chinese, Thai, Lao, and Indonesian. The genetic affiliations of these languages are illustrated in Appendix 1, as explained by J. Matisoff in his course: Linguistics of Southeast Asia, Spring, 1976, U.C. Berkeley. The Austro-Thai hypothesis, developed by P. Benedict, is still controversial.

In order to provide a typological standard by which the various Southeast Asian languages can be compared, the structure of each language was first examined in terms of the universals proposed by Greenberg (31). Vietnamese has the basic order subject-verb-object (SVO). It is prepositional rather than postpositional, though what are translated as prepositions in English can also be interpreted as verbs (4, p. 201). The dominant position of qualifying adjectives is noun + adjective, although items borrowed from Chinese maintain the order adjective + noun. Again, the Vietnamese equivalent of English adjectives are usually interpreted as verbal elements. The genitive appears after the governing noun.
The basic order typology for Lahu was obtained from Matisoff (45), for Indonesian from Kwee (40), for Lao from Roffe (57), and for the remainder from Greenberg (31, pp. 87-88). The results are shown in Appendix 1.

Finally, the comparisons and conclusions not made in the course of the language-specific studies will be explicitly presented and explained in XIII.

II. VIETNAMESE EXAMPLES OF CL'S

The following examples are taken from Nguyen Dinh Hoa (48, 49) and from Nguyen Phu Phong (50). The Vietnamese data are written in the standard Vietnamese orthography, and the CL's are underlined. Since phonetic detail is not the primary concern of this study, the data from each language is written in the transcriptions which are used in the sources.

1) ba cái bát (3 + non-living thing + eating bowl) 'three eating bowls'
2) bốn con mèo (four + living thing + cat) 'four cats'
3) hai tờ báo (two + sheet + newspaper) 'two newspapers'
4) năm can thit (five + pound + meat) 'five pounds of meat'
5) cây dạo (tree, plant + peach) 'peach tree'
6) cánh dạo (branch + peach) 'branch of peach tree'
7) hoa dạo (flower + peach) 'peach blossom'
8) quả dạo (fruit + peach) 'peach (the fruit)'
9) một con gà (one + animate + chicken) 'one chicken'
10) hai cân gà (two + unit of measure + chicken) 'two kilograms of chicken'

The above limited corpus illustrates several problems and properties of Vietnamese CL's. #1-2 show the contrast inanimate object/animate object. Here, the CL's are evidently required in the context of a numeral phrase. #3 shows the semantic feature sheet of paper. #5-8 show how the choice of CL can change the meaning of a word or phrase. #9-10 show the contrast quality/quantity.

From #1-4 and #9-10, a first attempt at characterizing Vietnamese CL's can be made on syntactic grounds.

Def. 1: A CL is any lexical item which can appear in the frame

Num + ___ + N, where Num = numeral, N = noun.

What about the CL's in #5-8? These were given by Nguyen Dinh Hoa (49, p. 127) without any numerals, though presumably each of these examples could be preceded by a numeral. Therefore, Def. 1 still holds, as long as it is realized that the CL may also occur outside of the numeral phrase environment.

Are there any syntactic differences between CL's given by Def. 1? It will be shown that CL's can not all occur in the same non-numerative environments. What are the semantic functions of CL's? Frequently, linguists prefer to categorize "quantifiers", such as the CL's in #4, 10, separately from "qualifiers", such as the other CL's. Can this distinction be made rigorous? What are the syntactic and/or semantic grounds for associating nouns and CL's? These questions can be extended to CL's in other languages, though the answers will sometimes vary.

III. THEORETICAL APPROACHES TO THE STUDY OF CLASSIFIERS

III A. Semantic Basis of CL Systems
Adams & Conklin (2, p. 1), after surveying 37 Asian languages, claimed that: "Numeral classification is based primarily on the parameters of animateness, shape, or function which are attributed to the head noun." They considered CL's to be the means "by which nouns are classed for counting." The main problem that they are concerned with is determining the basis, cultural or cognitive, of the noun classes associated with the various CL's. Adams, Becker, & Conklin (1) expanded the earlier treatment of CL's to include non-numeral phrase occurrences, but nowhere is there an explicit statement of how CL's were identified. Because CL's are used (1, p. 4) to make nouns into countable units, it was claimed that there is a tendency for "naturally countable" items such as time periods and body parts not to take CL's, presumably because they are already countable. Similarly, abstract quantities are supposed to be unclassified in most languages, presumably because they are inherently uncountable, although Burmese was admitted to be a counterexample to this generalization. The general pattern discovered by Adams & Conklin (2) was that animate nouns formed their own class or classes; inanimate nouns were categorized by shape; the primary dimensions of shape were length, flatness, and roundness, or in other terms, 1-dimensionality, 2-dimensionality, and 3-dimensionality. These dimensions were also expressed by the plant metaphor: stem, leaf, and fruit. A major problem with all of the above generalizations is that exceptions are admitted and no quantification is attempted. Thus, the claims are not in forms suitable for testing.

III B. Burmese CL System: A Spatial Metaphor

Becker (5) interpreted the Burmese CL system as a "spatial metaphor." Animates are located in various "orbits" from the "center", which is identified with Buddha. Inanimate objects are located in various orbits about the self and are also associated with the head or body. Concepts are located in orbits about sacred (Buddhist) concepts. Becker also emphasized (5, p. 111) that "the numerative classifier system, then is not a folk taxonomy, in which items are classified on the basis of objective features, but rather a system much more like a paradigm, in which items are located relationally." He stated (personal communication, 1975) that he did not know if such an analysis could be applied to Vietnamese. This analysis is clearly culture-specific but is not inconsistent with the Adams & Conklin generalizations. Even though all the CL's are related to each other, in Becker's system, through variations along a deictic dimension, the animate/inanimate distinctions are still maintained, and the 1-dim, 2-dim, 3-dim trichotomy of shape can still be imposed on inanimate objects. The implication of Becker's work is that a given CL system may have more than one level of structure. One level, including such features as the shape trichotomy, may be based on "universal" or "cognitive" principles, while other levels may be culture- or language-specific.

III C. Psychological Basis of CL's

The psychological approach was adopted by Clark (19). After reviewing Adams & Conklin (2), Clark (19, p. 7) decided that "... much basic categorization depends on the visual form of the entities classified. Secondary classifications usually involve additional physical, perceptible properties. In addition, some classifications are made on the basis of .
function, but these tend to be language- or culture-specific, unlike the classifications based on shape." Clark then claimed to see the same patterns of classification in the over-extensions made by children and concluded (19, p. 17) that from visual perception can be derived a "common cognitive basis that allows (and even encourages) the formation of natural, universal categories of the sort considered in children's over-extensions and in classifier systems." An examination of Vietnamese data, described below, is sufficient to prove that Clark's views of CL systems are oversimplified at best. Again, the vagueness of the claims concerning "primary" and "secondary" classifications preclude any possible test.

III D. Importance of Shape

Friedrich (29) studied numeral CL's, classificatory verbs, and locative suffixes in Tarascan, and then surveyed studies of other languages. He saw shape as an important semantic component in many languages. More specifically, he claimed that shape is primarily analyzed into the complex concept of orifice/(curved) edge; the trichotomy of long, flat, round (1-dim, 2-dim, 3-dim); and the overall shape of the zoological body. Once again, it is difficult to prove or disprove a general tendency, but a quantitative analysis of Vietnamese data proves that the primacy of shape is not universal.

III E. T'sou's Syntactic Analysis of CL Systems

On a more abstract level, T'sou (61) proposed a four-way distinction for all CL's, based on ± values of the two features entity and exactness. He illustrated this feature system with English (61, p. 4): two head (+exact, +entity) of cattle, two herd (-exact, +entity) of cattle, twenty thousand pounds (+exact, -entity) of cattle, and two kinds (-exact, -entity) of cattle. The (+exact, +entity) CL's are claimed to comprise most of the rich, complex CL systems in most Asian languages. The (+exact, -entity) words, such as gallons and kilograms, are commonly known as measure words, are culture-bound, and are relatively small in number. The (-exact, +entity) words are the "temporary measure" words, such as in "three tables full of papers". The (-exact, -entity) CL's have "universally few categories" and are most common for abstract nouns. T'sou constructed a hierarchy of semantic features to subclassify the (+exact, +entity) CL's (61, pp. 10-13; see Appendix 2).

T'sou claimed (61, p. 14) that the above features are organized in a nested hierarchy, with the implication that a lexical form based on a dominant feature in the hierarchy will tend to replace a form based on a less dominant feature. This argument is not presented clearly. Examples of this dominance are: size is dominated by shape and edifice, shape is dominated by mechanical, mechanical is dominated by duality and set, mechanical and vehicular vary in relative dominance between languages. No more of the hierarchical structure was given.

Finally, T'sou stated (61, p. 3): "The hypothesis advanced in this paper is 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) a natural language has either nominal classifiers or plural morphemes, or b) if a natural language