BETWEEN LEXICAL AND LEXICO-GRAMMATICAL CLASSIFICATION:
NOMINAL CLASSIFICATION IN SINHALA

MARA HENDERSON

University of California, Santa Barbara

1. INTRODUCTION. This paper investigates a small set of specific-general noun sequences (SGNs) and their role as a system of nominal classification in Sinhala. Systems of nominal classification have typically been described in terms of three sub-types: 1) lexical systems (class terms and measure terms), 2) lexico-grammatical systems (classifiers), and 3) grammatical systems (noun class markers and gender) (Grinevald and Seifart 2004:261). These three subtypes can be seen as occupying positions on a typological continuum as well as reflecting a diachronic pattern of language change from class terms or measure terms to noun class markers of agreement or gender as illustrated by Figure 1, below.

Lexical   Lexico-grammatical   Grammatical
(class/measure terms)   (classifiers)   (noun class markers/gender)
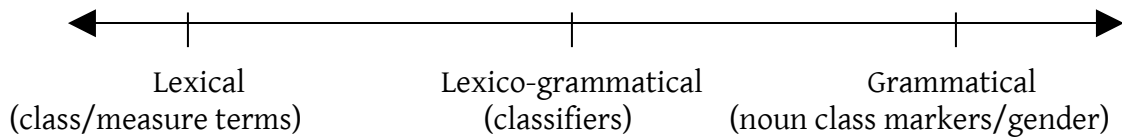
FIGURE 1. Systems of nominal classification (Grinevald and Seifart 2004:261)

Each of these systems and the patterns by which they may be distinguished from each other are elucidated in section 2. For the purposes of this paper, I focus on differentiating semantic and morphosyntactic patters of lexical and lexico-grammatical systems of nominal classification in an attempt to explicate the role of the general noun of SGNs in Sinhala.

Sinhala is an Indo-Aryan language spoken primarily in Sri Lanka (Ethnologue 2004). According to statistics from 1993, approximately 72 percent of the population of Sri Lanka are native Sinhala speakers (Ethnologue 2004). The Sri Lankan language community is primarily made up of Sinhala, English, and Tamil speakers. Sinhala functions as the language of most domains (i.e. government, marketplace, Buddhist temples), while English functions as the lingua franca in private business and education. Sinhala has been described as possessing one system of nominal classification; namely a system of animacy and honorific marking on numerals (Gair and Paolillo 1997:22). Gair and Paolillo (1997:22) describe this as a gender system--that is, it is a system of agreement between the noun and the numeral which quantifies it. A partial reconstruction of that system is as follows:

		Inanimate		Animate	
	Stem	Definite	Indefinite	Definite	Indefinite
one	ek	ekə	ekak	ekenaa	ek**kenek**/**kenek**
two	de	deka	dekak	de**nna**	de**nnek**
three	tun	tunə	tunak	tun**denaa**	tun**denek**
four	hatərə	hatərə	hatərak	hatərə**denaa**	hatərə**denek**

TABLE 1. Numeral gender system in Sinhala (Gair and Paolillo 1997:22)

In this system numerals are obligatorily marked for animacy and definiteness. Here animacy is more aptly described as humanness as the animate forms only appear with humans and not other animate beings like animals. The forms for 'one' have complex patterns of use in Sinhala, which I do not discuss further other than to mention that they are likely a result of a contrast between 'one' and general singular indefiniteness. Other systems of nominal classification, specifically lexical or lexico-grammatical systems have not been described for Sinhala. From a typological perspective, although Sri Lanka is not a linguistic area identified as possessing nominal classification systems, Aikhenvald (2003:77-78, 121-122) does include languages spoken in southern India in her typological study of noun class markers and numeral classifiers. Furthermore, Emeneau (1956:10) identifies Indo-Aryan as the historical source of nominal classification (specifically, noun class markers, measure terms, and numeral classifiers) in Dravidian and Munda languages of India. The presence of lexical systems of nominal classification is perhaps not all that surprising, as they are common crosslinguistically. However, lexico-grammatical systems are far more restricted--that is, they are typically described as an areal phenomenon with a high concentration in Southeast Asia. In fact, Emeneau (1956:16) notes the possibility of Southeast Asia as the source of classifiers in Indo-Aryan. It is, therefore, not too surprising to find a lexico-grammatical system of nominal classification emerge from a study of Sinhala. The evidence provided in this paper suggests that a system of nominal classification not unlike a classifier system indeed exists in a limited semantic domain of language use in Sinhala.

    2. Nominal Classification Systems. The terminology used to discuss nominal classification systems typologically is not consistent in the literature. In particular, researchers tend to use the term classifier to describe both lexico-grammatical and grammatical systems of classification. For the purposes of this paper, I use the term nominal classification as a broad cover term to mean a system through which language or language users mark nouns based on categories, which would include class terms, measure terms, classifiers, and noun class markers. Class terms occur as part of endocentric nominal compounds in which the class term is taken from a higher position in the taxonomy than the other element in the compound, which specifies the type (DeLancey 1986:440). In English, for example, *snake* functions as a class term in compounds like *rattlesnake, king snake,* and *grass snake* where *snake* denotes the basic category and *rattle, king*, and *grass* denote the type of snake (DeLancey 1986:440). Measure terms are terms that denote a quantity of the entity they modify. In English, for example, *pound* functions as a measure term in phrases like *a pound of butter, a pound of sugar,* and *a pound of oranges* (Grinevald and Seifart 2004:261). Classifiers are defined broadly as "morphemes which occur 'in surface structures under specifiable conditions', denote 'some salient perceived or imputed characteristics of the entity to which an associated noun refers' (Allan 1977:285), and are restricted to particular constructions types known as 'classifier constructions'" (Aikhenvald 2003:13). Aikhenvald defines classifiers constructions as "morphosyntactic units...which require the presence of a particular kind of a morpheme, the choice of which is dictated by the semantic characteristics of the referent of the head of a noun phrase" (2003:13). This definition of classifiers is decidedly broad to include a full range of classifier types, however, only two classifier types (numeral and noun) are of particular salience for this discussion of Sinhala nominal classification. Noun classifiers are morphemes in classifier constructions that appear in the noun phrase, typically next to the noun, and typically denote generic semantic characteristics of the noun they categorize, such as, men, women, plants and

animals (Grinevald and Siefart 2004:262-263). Dixon (1982:1992 ff. in Aikhenvald 2003:2) provides an example of a noun classifier in Yidiny:

(1)   bama   waguja
            CL:PERSON   man
      'a man'

NUMERAL CLASSIFIERS too are morphemes in classifier constructions that appear in the noun phrase, however these classifiers occur in numeral phrases and typically denote characteristics of the physical shape of the entity they categorize, such as, 1D long-rigid, 2D flat-flexible, 3D round. Rehg (1981:130 in Grinevald and Seifart 2004:262) provides the following examples of a numeral classifiers in Ponapean:

(2)   tuhke rioapwoat
          tree   two.CL:LONG
          'two trees'
 (3)   pwihk riemen
          pig   two.CL:ANIMATE
          'two pigs'

These examples illustrate the diversity in semantic denotation of numeral classifiers. Although prototypically numeral classifiers denote physical properties such as shape, they have also been found to denote animacy. NOUN CLASS MARKERS (aka, NOUN CLASSES, CLASS MARKERS, GENDER, CONCORDIAL CLASSIFIERS) are "an obligatory grammatical system where each noun chooses one from a small number of possibilities" (Dixon 1986:105). Aikhenvald further states that they are "grammatical agreement classes, based on such core semantic characteristics as animacy, sex, or humanness" (2003:1). An example of noun class markers is found in Portuguese (Aikhenvald 2003:2):

(4)   o   menin-o   bonit-o
            ART.MASC.SG   child-MASC.SG   beautiful-MASC.SG
      'the beautiful boy'

This example clearly illustrates the agreement function of noun class markers through the masculine singular form that appears on the article, noun, and the adjective in the noun phrase.

   Many of these nominal classification systems share semantic and morphosyntactic characteristics which makes distinguishing them from each other rather difficult. The following table is an attempt to clarify the characteristics of each system that may distinguish them from each other.

	Class terms (CT)	Measure terms (MT)	Classifiers (CL)	Noun class markers (CM)
Size	*open; restricted*	*open; restricted*	open, varies; large	small finite set
Distribution	?	?	1N:>1CL some Ns may not take a CL	1N:1CM all
Boundedness	lexically bound; may be but, often not independent nouns	*varies; free form, affix*	free form; occurs in the same NP as the N it qualifies; not independent noun/independent noun	closed grammatical system; affixes, GW, clitics
Scope	*noun*	*noun*	never any reference outside the NP	marking is never entirely within the noun word
Semantics	consistent, circumscribed, hyponym	*provide the measure for a specified quantity*	animacy, shape, functional, generic	animacy, sex, humanness
Inter-speaker Variation	?	?	use varies across registers or styles	little variation between speakers

TABLE 2. Distinguishing nominal classification systems[1]

The following elucidates Table 2.

(a)   Size refers to whether or not it is an open or closed class and in general the number of terms typically found in these kinds of systems cross-linguistically.   An explicit discussion of size for class terms and measure terms was not found, however, I believe that it is safe to state that these are open classes, but typically restricted.   Dixon (1986:106) describes classifiers as typically quite large crosslinguistically (50-400), although there are languages such as Indonesian which have very small sets of classifiers (7)[2].   This is additionally, highlighted by Aikhenvald (2003:81) in her discussion of noun classifiers, she states that the size of the inventory may vary crosslinguistically from a small closed set to a large open set.

(b)   Distribution refers to which nouns in the language take the classifying morpheme. Information was not found regarding class terms and measure terms, however, Dixon (1986) and Aikhenvald (2003) provide some typical characteristics of classifiers and noun class markers.   Dixon (1986:106) provides that typically in languages with classifiers systems not all nouns take classifiers.   Nouns that do not take classifiers are typically mass

---

[1]The information compiled for this table was taken from Dixon (1986:105-108) and Aikhenvald (2003), which explicates the distinctions between noun class markers and classifiers, and DeLancey (1986:440-444), which discusses some differences between class markers and classifiers. The information in italics is my own educated guess and question marks indicate that this information was not found and is therefore unknown.

[2] Thank you to Robert Englebretson for pointing out this particular example.

nouns, time units, and some of the most frequent nouns (Dixon 1986:106). Additionally, a single noun typically is able to take more than one classifier with a resulting change in meaning (Aikhenvald 2003:81, 98, Dixon 1986:106, Greenberg 1972:8). Noun class markers, on the other hand, classify ALL the nouns in a language--that is, there is a 1 to 1 ration of nouns to noun class markers and their distribution is fixed (Aikhenvald 2003: 21, Dixon 1986:106).

(c)   Boundedness refers to the classifying morpheme's realization as bound or free. By definition class terms are lexically bound and may function as independent nouns in other contexts, though they often do not (DeLancey 1986:439). However, these properties are again best described as tendencies since class terms may occur as independent nouns and are not obligatory in all cases (DeLancey 1986:439). The realization of measure terms crosslinguistically is bound or free. Dixon claims that "noun classifiers are always separate lexemes, which may be included with a noun in certain syntactic environments" (1986:105). However, according to Aikhenvald noun classifiers may appear as clitics or nominal affixes via grammaticization or phonological reduction processes (2003:91, 101). Noun class markers typically emerge as affixes, grammatical words, or clitics (Dixon 1986:106). They are often fused with other grammatical morphemes such as definiteness, case, or number (Dixon 1986:106).   In Delancey's work on Tai class terms and classifiers, he alludes to a prototypical property of classifiers, which states that classifiers would not function as independent nouns or part of compounds (1986:439). However, Greenberg states "in the majority of instances, the classifier is itself a noun with its own lexical meaning and may, in fact, have its own classifier when it functions as the head of a noun phrase" (1972:7). Conflicting findings on the function of class terms and classifiers as independent nouns within classifying languages provides evidence that a crosslinguistic explanation of the terms' ability to operate as independent nouns is not a defining feature and therefore should not be heavily weighted in distinguishing it from other nominal classification systems.

(d)   Scope refers to the classifying morpheme's domain--that is, the noun itself, the noun phrase, or outside of the noun phrase. I believe that it is safe to state that the scope of class terms and measure terms is the noun. Classifiers are specifically distinguished from noun class markers based on scope. According to Dixon (1986:106-107) noun class marking is "never entirely within the noun word" rather other elements in the sentence are obligatorily marked with the same marker, while classifiers are never referenced outside of the noun phrase (Dixon 1986:107, Aikhenvald 2003:81, 98).

(e)   Semantically, there is some degree of overlap between the categories. Class terms are described generally as semantically consistent and circumscribed (DeLancey 1986:441). Furthermore their relationship with the entity they categorize tends to be taxonomic. Measure terms can be simply described as denoting the measure for a specified quantity. Classifiers range semantically from animacy, shape, generic, or function (i.e. clothing, transport, food), often depending on the type of classifier (Aikhenvald 2003:1-2, Grinevald and Seifart 2004:263-264).   Noun class markers tend to denote "such core semantic characteristics as animacy, sex, and humanness" (Aikhenvald 2003:1).

(f)   Finally, Dixon (1986:107) also notes a tendency for inter-speaker variation in the use of classifiers based on register or style shifts, while little variation between speakers is found in systems of categorization described as noun class marker systems.

It is clear from the details discussed above that SGNs in Sinhala are not noun class markers.

The general noun of SGNs in Sinhala do not distribute obligatorily in all cases or with all nouns, nor do they reference any other element of the sentence beyond the noun phrase. In these ways, they are clearly not operating as agreement systems and therefore not noun class markers. However, they do display characteristics typical of class terms, measure terms, noun classifiers and numeral classifiers. Because SGNs exhibit properties of lexical and lexico-grammatical systems of classification it is important to this analysis to focus on the properties that distinguish these system types, namely their realization and semantic relationship to the entity they categorize. Specifically, class terms are lexically bound and measure terms and classifiers vary in boundedness crosslinguistically. Therefore, if it can be established that the general noun of SGNs in Sinhala are not bound, then they can be distinguished from class terms. To determine this, I focus on properties of wordhood, obligatoriness, and anaphoric reference of the general term of SGNs in Sinhala (§4). Although there is some overlap, as discussed above, the semantic relationships between the classifying term and the entity they categorize differ for prototypical class terms, measure terms and classifiers.   These prototypical patterns are discussed in relation to the semantic patterns of SGNs in Sinhala to aid in the analytical distinction between them (§4).

While wordhood, obligatoriness, and anaphoric reference provide important features that could distinguish lexical from lexico-grammatical systems, lexico-grammatical systems themselves can be further subdivided into types of classifiers. Most important to the analysis here is the distinction between noun classifiers and numeral classifiers. Therefore, some discussion of the properties which distinguish the two subtypes of relevant classifiers is necessary before moving on to the specifics of SGNs in Sinhala.

Noun classifiers and numeral classifiers share several of the same properties: a) they appear in the noun phrase, b) their selection is based on semantic properties of the entity they categorize, c) their level of grammaticization varies, d) they are characterized as open lexical classes, e) there is evidence of inter-speaker variation, f) some nouns do not take the classifier, while others may vary the classifier with a resulting change in meaning, g) they are typically realized as free forms, and h) they may be used for anaphoric reference (Aikhenvald 2003:81, 98, Greenberg 1972:6). The distinguishing property is their specific location within the noun phrase and their tendency toward types of semantic categorization. As previously mentioned, noun classifiers typically denote generic semantic categories, while numeral classifiers typically denote animacy or physical properties (i.e. size, shape, structure) (Aikhenvald 2003:98, Grinevald and Seifart 2004:262-263). Numeral classifiers more specifically occur in quantifying expressions and numeral noun phrases (Aikhenvald 2003:98). Greenberg (1972) further points to the individuating function of numeral classifiers, a function that has not be claimed for noun classifiers. He states

> [I]n the usual classifier language...classifiable nouns in their isolated form, that is when not accompanied by a classifier or a plural marker, are like collectives in their semantic non-specification of number and in their avoidance of a direct number construction. The classifier is an INDIVIDUALIZER which performs the same function as a singulative derivational affix in languages with the collective/singulative opposition (Greenberg 1972:26, emphasis added)

This individuating function along with the tendencies of a semantic denotation of shape and the syntactic distribution in numeral phrases distinguishes numeral classifiers from the more generic non-individuating noun classifiers.

The implications for an analysis of SGNs as members of a nominal classification system based on semantic properties and relationships between elements of SGNs are discussed further after a brief description of the overall patterns of SGNs in Sinhala in section 3, below.

3. Description of SGNs in Sinhala. A set of nine terms that appear to fulfill a classificatory function have been identified through the examination of elicited sentences from one Sinhala speaker. The nine terms are as follows in Table 3.

geḍi	fruit-like.thing.PL	*palu*	section.PL
mal	flower.PL	*æṭə*	seed.PL
karal	pod-like.thing.PL	*kææli*	piece.PL
peṭi	flat.thing.PL	*kæṭə*	block.PL
alə	potato.PL		

Table 3. Identified classificatory terms in Sinhala

These terms appear in noun phrases following a more specific noun which they classify as in the following example.

(1) **hatu**　　　**mal**　　　narakwelaa
　　　mushroom　　**flower.PL**　　rotten
　　　'The mushrooms are rotten.'

3.1. Semantic Properties.

Semantic Domains. SGNs have so far appeared largely in the domain of food, but they also occur with plants, medicine, and other small objects such as dice and beads. Some examples of their semantic distribution is as follows:

dehi geḍi	'limes'	*boonci karal*	'green beans'
vambotu geḍi	'eggplants'	*behet karal*	'capsules'
keek geḍi	'cakes' (whole)		
paaŋ geḍi	'loaf of bread'	*dehi æṭə*	'lime seeds'
		wii æṭə	'rice grains'
mannel mal	'blue lotuses'	*pabəlu æṭə*	'beads'
hatu mal	'mushrooms'		
kehel mal	'banana stalks'	*pipiŋña kææli*	'chopped cucumbers'
		bætəri kææli	'batteries'
daadu kæṭə	'dice'		
ais kæṭə	'ice cubes'		

dodaŋ palu	'orange segments'	*roosə peti*	'rose petals'
suduluunu palu	'garlic cloves'	*behet peti*	'tablets'
		maalu peti	'fish fillets'

The SGNs above show that *geḍi* may occur with fruit, vegetables, and whole breads of a certain shape. Although *geḍi* is often thought of as meaning fruit, evidence shows that this term may be used with other items that possess properties often thought of as pertaining to fruit, but not necessarily only fruit. According to the consultant, a more precise semantic description of *geḍi* would be a small, hard, fruit-like things. The terms *mal* and *æṭə* appear with all types of things they denote (i.e. flowers and seeds), but also appears with things that are judged to come in the form of flowers (e.g. stalks of bananas and mushrooms) or seeds (e.g. beads). The terms *kæṭə*, *palu*, and *karal* have so far been found with only a few terms. It is unclear what their full semantic distribution is, however, they appear to denote the shape of the items they categorize. The last two terms *peti* and *kææli* have a much wider semantic distribution. In many cases, *kææli* denotes a changed state as in the example above (*pipiŋña kææli* 'chopped cucumbers') or *darə kææli* 'chopped wood'. It may even appear with mass nouns as in *harakmas kææli* 'beef pieces'. However, it also appears with small items that come in sets, such as batteries (*bætəri kææli*) or magnets (*kandaŋ kææli*), in which case the term is used to individuate a single or number of items from the set. A similar case arises from the patterns of *peti*, which may denote a change of state to a mass noun as in *harakmas peti* 'slices of beef' or *maalu peti* 'fillets of fish'. Like *kææli*, it also appears with items that do not undergo a change in state, but rather occur in groups, such as medicine tablets, *behet peti*, and flower petals, *mal peti.*

Semantic Relationships between Elements. The SGNs in this study fall into three types of semantic relationships: 1) hypernym-hyponym taxonomic relationships, 2) noun + shape/physical property denotational relationships, and 3) noun + quanification relationships. The following examples illustrate these findings.

Hypernym-hyponym taxonomic relationships
(2)	*arəliyə mal*	'frangipani flowers'
(3)	*vattakka æṭə*	'pumpkin seeds'
(4)	*batələ alə*	'yam potatoes'

Noun + shape/physical property denotational relationships
(5)	*goowə geḍi*	'cabbages' (lit. cabbage small.hard.fruit-like.things)
(6)	*boonci karal*	'long beans' (lit. bean long.pod-like.things)
(7)	*daadu kæṭə*	'dice' (lit. die blocks)

Noun + quantification relationships
(8)	*dehi palu*	'lime sections'
(9)	*kukulmas kææli*	'pieces of chicken'
(10)	*maalu peti*	'fillet of fish'

It should be noted here that these examples are representative of the semantic tendency of the relationship between the elements for each of the nine classifying terms; however, a few instances have been found that pattern outside of a single term's general tendency (e.g. *kehel*

mal 'banana stalks', *hatu mal* 'mushrooms', *pabəlu æṭə* 'beads' *gammiris æṭə* 'peppercorns' *wii æṭə* 'rice grains' *kaju æṭə* 'cashews'). Because in an overwhelming majority of cases, *æṭə* and *mal* appeared in SGNs that could be categorized as having a taxonomic relationship with the specific noun they classified, I attribute the occurrence of these few terms to semantic extension.

INDIVIDUATION VS. NON-INDIVIDUATION. These classificatory terms appear with some but not all count nouns in the semantic domains previously mentioned. The most striking semantic motivation for the presence or absence of the general noun is the degree of individuation of the referent.

(11)   **sudu-luunu**   suddə-kara-nnə   amaarui
       white-onion   peel-do-INF   difficult
       'Garlic cloves are hard to peel.'

(12)   mee   **sudu-luunu   (palu)**   suddə-kara-nnə   amaarui
       these **white-onion section.PL** peel-do-INF   difficult
       'These garlic cloves are hard to peel.'

(13)   maŋ   laŋgə   **sudu-luunu   paluwak**   tiye-n´wa
       1SG   near   **white-onion   section.SG.IND**   exist-IMPF
       'I have a clove of garlic.'

(14)   **kærət**   tiye-nəwa=də
       carrot exist-IMPF=Q
       'Do you have carrots?'

(15)   maŋ   laŋgə   **kærət   (alə)**   dahayak   tiye-nəwa
       1SG   near   **carrot potato.PL** ten.IND   exist-IMPF
       'I have ten carrots.'

In examples 11 and 14 the referent is non-individuated. In these cases, the general term is dispreferred. However, when specifying a set or a number of the referent, the general term may appear as in examples 12, 13 and 15. Examples 12 and 15 show that the general term in these cases are not entirely required. Further, although the general term in example 15 is judged to be optional, my consultant states that he prefers that the general term appear in constructions such as this one.

VARIATIONS. The general nouns of the SGNs may vary with one specific noun with a resulting change in meaning as in the following example.

(16)   kehel   geḍi   kehel   mal
       banana   fruit.PL   banana   flower.PL
       'bananas'   'banana stalks'

However, there does appear to be a default general term for each specific noun. This is evidenced by the patterns of obligatoriness--that is, while the default term may be optional in some cases, if the specific noun occurs with a general noun other than the default term it becomes obligatory. For example, in example 16 above, *geḍi* is the default term for bananas. It is therefore optional in some cases. However, *mal* is not the default term for bananas; it is

therefore, required in all cases--that is, *kehel* alone may not mean banana stalks, it may only mean bananas.

In sum, SGNs in Sinhala occur in a small semantic domain, primarily with count nouns and only with mass nouns when the mass noun undergoes a change of state that produces countable pieces. Three types of semantic relationships between elements emerge: 1) taxonomic, 2) the general term denotes the shape or physical property of the referent, or 3) the general term quantifies the referent. Further, the general terms serve to individuate a set or number of referents. The general nouns may alternate with a single noun with a change in meaning and one general term operates as the default term for a specific noun. Before moving on to the implications of these patterns for the analysis of SGNs as a system of nominal classification in Sinhala, I discuss the morphosyntactic patterns of SGNs in Sinhala.

3.2. Morphosyntactic Patterns

Morphological Marking. In most cases, nominal morphology, such as case, definiteness, number or question marking, may only appear on the general noun, which in these cases are obligatorily present in the noun phrase as in the following examples.

(17) mæssa   **miris**   **karələkə**   wæhuw-a
      fly.SG   chili.pepper   pod.like.thing.SG.IND.LOC   land-PST
      'The fly landed on a chili pepper.'

(18) *mæssa   **miris-yəkə**   wæhuw-a
      fly.SG   chili.pepper-SG.IND.LOC   land-PST
      'The fly landed on a chili pepper.'

(19) mona   **paaŋ**   **gediyə=də**   narakwelaa   tiye-nne
      which   bread   fruit.like.thing.SG.DEF=Q   rotten   exist-FOC.NPST
      'Which loaf of bread is rotten?'

(20) *mona   **paaŋə=də**   narakwelaa   tiye-nne
      which   bread.SG.DEF=Q   rotten   exist-FOC.NPST
      'Which loaf of bread is rotten?'

However, some specific lexical items have been found to carry the nominal morphology, such as pineapples or mangos. In these cases, the classifying term may be omitted.

(21)   mee   **annaasi**   **gedi-yə**   pæniraha-i
       1PROX   pineapple   fruit.like.thing-SG.DEF   sweet-PRED
       'This pineapple is sweet.'

(22)   mee   **annaasi-yə**   pæniraha-i
       1PROX   pineapple-SG.DEF   sweet-PRED
       'This pineapple is sweet.'

Most of the examples elicited involving nominal morphology required the presence of the general noun. Examples such as 22 above were rare.

Constituency. SGNs display two patterns of constituency--that is, they are cohesive and they move as a unit. As of yet, SGNs always appear together without any intervening lexical or

grammatical formatives. Furthermore, SGNs move as a unit as illustrated in the following examples.

(23) maŋ   laŋgə   **dehi**   **geḍi**   tiye-nəwa
      1SG   near   **lime**   **fruit.PL**   exist-IMPF
      'I have limes.'

(24) **dehi** **geḍi**   maŋ   laŋgə   tiye-nəwa
      lime **fruit.PL** 1SG   near   exist-IMPF
      'I have limes.'

The specific and general nouns may not be separated throughout the phrase and no lexical or grammatical formatives have been found to intervene between them. These patterns of cohesiveness and movement provide evidence for the analysis of SGNs as a lexical constituent.

Related to constituency is the general terms' ability to operate as independent nouns. Although a couple of these nouns were judged to not operate as independent nouns (*palu* and *geḍi*), they have been found outside of SGN contexts in the following examples.

(25) atə   tiyenne   janeele   dakunu   **paluwe**
      hand.SG.DEF   exist-FOC.NPST window.SG.DEF   right   section.LOC
      'The hand is on the right section of the window.'

(26) **alə**   wələ   **æṭə**   næǣ
      potato.PL   seed.PL   NEG
      'Potatoes don't have seeds.'

(27) wiiduruwə   **kæǣli**   wələṭə   kædun-a
      glass.SG.DEF piece.PL   break-PST
      'The glass broke into pieces.'

(28) **mal**   lasənai
      flower.PL   beautiful-PRED
      'The flowers are beautiful.'

(29) annaasi   kiyanne   **geḍiyak**
      pineapple   known.as   fruit-like.thing.IND
      'The thing known as a pineapple is a small, hard fruit like thing.'

According to my consultant, the nouns *alə*, *æṭə*, *kæǣli*, and *mal* are easily identifiable as independent nouns. However, according to my consultant, *palu* and *geḍi* are not typically thought of as independent nouns although, as example 25 and 29 show *palu* may be used in a possessive construction that individuates the thing possessed (the window's section) and *geḍi* may be found independently when talking specifically about the category. The other three terms *peti*, *kæṭə*, and *karal* are judged as unable to operate as independent nouns and no data has been found to the contrary.

ANAPHORIC REFERENCE. SGNs also display patterns of anaphoric reference. SGNs may be anaphorically referenced by inanimate pronouns, such as *eekə* or *eeva*, or by the general noun in individuating contexts as illustrated by the following examples.

(30)   **peera**   **(geḍi)**   narakwelaa
          guava   (fruit-like.things)   rotten
          'The guavas are rotten.'

(31)   **eeva**   narakwelaa
          3p.INAN   rotten
          'They're rotten.'

(32)   **mee**   narakwelaa
          these   fruit-like.things   rotten
          'These are rotten'

(33)   ?***geḍi**   narakwelaa
          3p.INAN   rotten
          'They're rotten.'

Example 31 illustrates the pronominalization of guavas from example 30 using the third person inanimate pronoun, *eeva*. Although, this is the most preferred form, the consultant also provided the example given in 32 as an alternative employing the general term of the SGN in 30 to refer to a specific set of guavas. Here the context is more individuated and therefore, the anaphoric function, illustrated in the gloss, of the general term is judged acceptable. Finally, example 33 illustrates a case in which the consultant judged the use of the classifying term without an individuating deictic as highly dispreferred. However, a couple of examples were obtained in which the general term of an SGN could be function anaphorically without an individuating deictic, as illustrated below.

(34)   waṇḍura   **kehel**   **malə**   uḍə-ṭa   pænn-a
          monkey   **banana**   **flower.SG.DEF**   top-DAT   jump-PST
          'The monkey jumped on the stalk of bananas.'

(35)   **eeka**   uḍə-ṭa   waṇḍura   pænn-a
          3S.INAN   top-DAT   monkey   jump-PST
          'The monkey jumped on it.'

(36)   **malə**   uḍə-ṭa   waṇḍura   pænn-a
          3S.INAN   top-DAT   monkey   jump-PST
          'The monkey jumped on it.'

Although anaphoric reference itself is not a test for constituency, example 36 displays a pattern similar to the pronominalization in 35, a classic test for constituency. Here the general term of the SGN functions as a pronominal, as indicated in the gloss. This pattern provides further evidence for analyzing SGNs as lexical units. Though it must be noted that this is pattern emerged only with a select few examples and is not representative of the patterns of SGNs more generally as shown in examples 31-33.

   The morphosyntactic patterns discussed above demonstrate that 1) in the vast majority of cases nominal morphology may not appear on the specific noun, rather the general term is required in cases where nominal morphology must occur on the noun or noun phrase, 2) SGNs operate as lexical constituents based on patterns of cohesiveness and movement. Additionally, the general term was shown to function anaphorically for the SGN in individuating contexts.

4. Sɪɴʜᴀʟᴀ's Sʏsᴛᴇᴍ ᴏғ Nᴏᴍɪɴᴀʟ Cʟᴀssɪғɪᴄᴀᴛɪᴏɴ. Many of these properties of SGNs in Sinhala are properties of nominal classification systems as discussed in section 2 above. In this section, I aim to describe how Sinhala's nominal classification system fits into the larger framework of nominal classification systems described in the literature. I begin by discussing how the patterns of SGNs implicate an analysis of them as lexical or lexico-grammatical systems. I conclude this section with a discussion of the semantic properties of SGNs that implicate their placement among subtypes of lexico-grammatical classification systems.

4.1. Dɪsᴛɪɴɢᴜɪsʜɪɴɢ Lᴇxɪᴄᴀʟ ғʀᴏᴍ Lᴇxɪᴄᴏ-Gʀᴀᴍᴍᴀᴛɪᴄᴀʟ Sʏsᴛᴇᴍs ɪɴ Sɪɴʜᴀʟᴀ. Class terms and measure terms, though both lexical systems, possess quite different defining characteristics. As a result, I discuss these separately in relation to classifiers. Differentiating class terms from classifiers is accomplished by comparing SGNs and compounds in Sinhala based on three morphosyntactic patterns: 1) wordhood, 2) obligatoriness, and 3) anaphoric reference. However, measure terms are best differentiated from classifiers based on the semantic function of individuation.

Cʟᴀss Tᴇʀᴍs ᴠs. Cʟᴀssɪғɪᴇʀs. As discussed in section 2, class terms are by definition part of compounds, and classifiers are prototypically separate lexemes (Dixon 1986:105). With this in mind, the first step for distinguishing class terms from classifiers is to determine SGN's status as a word. The fact that SGNs operate as a single unit and that nothing has been found to intervene between the specific and general nouns, complicates their differentiation from compounds. However, by examining language internal patterns of wordhood, obligatoriness of elements, and patterns of anaphoric reference, we may contrast SGNs with compounds in Sinhala.

For this part of the analysis, I focus on the contrast between a few general terms of SGNs (*æʈə*, *karal*, *mal*, *alə*) and a few compounds that use the same general terms as the second element in the compound. These compounds are: *muŋæʈə* 'mung beans', *mæækaral* 'Chinese long beans', *innələ* 'potato' (particular kind), *muhudumal* 'coral'.

Wᴏʀᴅʜᴏᴏᴅ. According to Dixon and Aikhenvald, a compound is a single grammatical word made up of one or more phonological words (Dixon and Aikhenvald 2002:19). It therefore may be useful to investigate phonological and grammatical criteria for wordhood for both SGNs and compounds in Sinhala. If SGNs are shown to operate as more than one grammatical word, then we can confidently say that the general noun is not part of a compound and therefore not a class term. However, if SGNs are found to operate as a single grammatical word then other methods of distinguishing them from compounds must be explored. The general nouns would not automatically be discounted from being classifiers since, as Aikhenvald points out, classifiers may emerge in various stages of grammaticization and therefore do not always appear as separate words. In this case, more evidence would be needed to assert their status as classifiers.

Dixon and Aikhenvald provide a set of crosslinguistic criteria for determining grammatical and phonological wordhood (2002:19-21). Grammatical words are identified as having the following universal criteria: a) cohesiveness (the elements always occur together, b) a fixed order, and c) a conventionalized and coherent meaning (2002:19). Universal criteria for phonological words are not as easily explicated. However, Dixon and Aikhenvald point to stress, phonotactics, and phonology the primary areas where distinct patterns may be found

for language internal criteria for phonological wordhood. In the case of Sinhala, although there are some interesting isolated cases where phonological criteria points to one analysis over the other, there is no overwhelming evidence that phonological wordhood is crucial to the distinction between compounds and SGNs.

While the criteria for grammatical wordhood could potentially provide key evidence for the classification of the general nouns of Sinhala's SGNs as class terms or classifiers, the morphosyntactic patterns of SGNs and compounds are remarkably similar. As I have shown above, SGNs occur in a fixed order (specific noun followed by general noun), when both elements are required they always occur together as evidenced by their operation as a syntactic unit that may not be separated by other lexical or grammatical formatives, and they have conventionalized and coherent meanings as a unit as evidence by change in meaning accompanying variations in general terms with specific nouns. Therefore, according to Dixon and Aikhenvald's universal criteria, SGNs qualify as a single grammatical word. Furthermore, as expected the lexical items identified as compounds in Sinhala for this study also conform to this set of criteria. The following examples illustrate the similarity in grammatical wordhood status.

(37)　**muhudumal**　　lasənai
　　　 coral.PL　　　 beautiful
　　　 'The coral is beautiful.'
(38)　**hatu**　　　　**mal**　　　 narakwelaa
　　　 mushroom　 **flower.PL**　 rotten
　　　 'The mushrooms are rotten.'

In example 37, the order of the elements in the compound *muhudumal* may not be reversed, nor may they be separated by other grammatical or lexical formatives. Furthermore, the elements as a unit have a coherent and conventional meaning. In these ways, the SGN, *hatu mal* in example 38 is similar to the compound in example 37. The similarity in the patterns of wordhood between SGNs and compounds leads to the investigation of other patterns which may distinguish SGNs from compounds.

OBLIGATORINESS. The second morphosyntactic pattern that may provide evidence for the analysis of SGNs in Sinhala is patterns of obligatoriness of the elements of SGNs and compounds. Although the obligatoriness of the elements of compounds varies cross-linguistically, we may expect that the conditions for the omission of elements to be fairly restricted since the elements are by definition lexically bound. Classifiers, on the other hand, have been identified as being optional in many languages (Greenberg 1972:6). Therefore, if the general nouns of SGNs are classifiers rather than class terms, we would expect patterns of obligatoriness to be more restrictive for compounds.

The general nouns of SGNs are optional in all cases in which the specific noun is not being individuated or is otherwise morphologically unmarked. However, the elements of compounds are for the most part not optional in Sinhala. The following examples illustrate the distinct patterns of SGNs and compounds.

(39) oyaa   mæækaral   wikunə-nəwa=də
      2sɢ   Chinese.long.bean.ᴘʟ   sell-ɪᴍᴘꜰ=ǫ
      'Do you sell Chinese long beans?'

(40) oyaa   miris   wikunə-nəwa=də
      2sɢ   chili.pepper   sell-ɪᴍᴘꜰ=ǫ
      'Do you sell chili peppers?'

(41) oyaa   muŋæʈə   wikunə-nəwa=də
      2sɢ   mung.seed.ᴘʟ   sell-ɪᴍᴘꜰ=ǫ
      'Do you sell mung beans?'

(42) oyaa   pabəlu   (æʈə)   wikunə-nəwa=də
      2sɢ   bead   seed.ᴘʟ   sell-ɪᴍᴘꜰ=ǫ
      'Do you sell beads?'

(43) oyaa   innələ   wikunə-nəwa=də
      2sɢ   potato.ᴘʟ   sell-ɪᴍᴘꜰ=ǫ
      'Do you sell potatoes?' (a particular type)

(44) oyaa   kærət   wikunə-nəwa=də
      2sɢ   carrot   sell-ɪᴍᴘꜰ=ǫ
      'Do you sell beads?'

Examples 39, 41, and 43 demonstrate that the general term in these compounds is obligatory even when the referent is non-individuated. As examples 40, 42, and 44 illustrate, the general term as a part of SGN constructions is either dispreferred or optional in cases when the referent is non-individuated.

Aɴᴀᴘʜᴏʀɪᴄ Rᴇꜰᴇʀᴇɴᴄᴇ. Patterns of anaphoric reference may point to an analysis of the general terms as classifiers or class terms. While classifiers have been described typologically as having an anaphoric function, class terms have not. Since I have already shown that the general term of SGNs functions anaphorically in context, I investigate the patterns of compounds in this regard. Both compounds and SGNs may be anaphorically referenced by the standard inanimate pronouns. However, unlike the second/general element of compounds, the general nouns of SGNs may be used anaphorically in context. The following examples illustrate the distinct patterns of compounds in Sinhala.

(45)   eeva   narakwelaa
        3p.ɪɴᴀɴ   rotten
        'They're rotten.' (Chinese long beans)

(46)   *karal   narakwelaa
        3p.ɪɴᴀɴ   rotten
        'They're rotten.' (Chinese long beans)

(47)   eekə   lasənai
        3s.INAN   beautiful-PRED
        'It's beautiful.' (coral)

(48)   mal   lasənai
        3s.INAN   beautiful-PRED
        *'It's beautiful' (coral)
        'The flowers are beautiful.'

Examples 46-48 demonstrate the finding that the second element of these compounds may not be used anaphorically to refer to the entity denoted by the compound.[3] Example 48 further illustrates that while the sentence may be grammatical, the omission of an element of a compound may simply alter the meaning of the sentence. So, that even in context, sentence 48 would not make sense in reference to the term *muhudəmal* 'coral' as it would mean 'The flowers are beautiful' not 'It's beautiful' (coral).

The evidence presented suggests that SGNs do not operate as compounds in Sinhala even though they function as a single grammatical word. The patterns of obligatoriness and anaphoric reference are clearly different in the examples found. While the general term of SGNs is optional or dispreferred in non-individuating contexts and optional when morphologically unmarked, both elements of compounds are required in the same contexts. Further, while the general term of SGNs may be used anaphorically in individuating contexts, neither element of the compound may be employed in the same fashion. Since class terms are by definition compounds, these patterns clearly distinguish SGNs from class terms.

Measure Terms vs. Classifiers. Measure terms and classifiers, particularly numeral classifiers, can be difficult to distinguish from each other. As the precursors of numeral classifiers, measure terms perform similar functions and often occur in the same syntactic position. One key difference is that measure terms typically occur with mass nouns. Additionally, Greenberg (1972:9) describes the case for Khmer in which classifiers are optional as a general rule except in instances in which the classifier is functioning as a measure term with mass nouns. Furthermore, as mentioned in section 2, numeral classifiers have been identified as having an individuating function. Both of these characteristics are relevant to the patterns of SGNs in Sinhala.

Most SGNs collected may not occur with mass nouns. However, two (*kææli, peti*) SGNs have been found to occur with mass nouns. This suggests that these two terms are best categorized as measure terms. However, characteristics of at least a couple of the examples of *kææli* displayed characteristics more suggestive of numeral classifiers--that is, it was used to individuate items from a set. The following examples illustrate their patterns as measure terms and as numeral classifiers.

(49)   darə   kææli   dahayak
        firewood   piece.PL   ten.IND
        'ten pieces of firewood'
(50)   maalu   peti   dahayak
        fish   flat.thing.PL   ten.IND

---

[3] However, it must be noted that very few noun-noun compounds were found and their patterns more generally in this regard were not found in the existing literature.

'ten fillets of fish'

(51) bætəri   (kææli)   dahayak
      battery   piece.PL   ten.IND
      'ten batteries'

In examples 49 and 50, *kææli* and *peti* function to provide the unit by which the mass nouns *darə* and *maalu* may be quantified. It should be further noted that *kææli* and *peti* are not optional in these examples. In example 50, however, *kææli* is optional and functions to individuate a number of batteries from the set that batteries usually come in. Only a very small number of items were found with the term *peti* all of which pattern more like example 50 than 51. Additionally, the term *palu* has been found to occur with only a couple of items that are best described as items whose parts constitute a countable whole (oranges, garlic). In these cases, *palu* also patterns like examples 49 and 50, above, and therefore, it is best characterized tentatively as a measure term.[4] The issue of individuation is revisited in the following section as a characteristic which aids in locating the position of Sinhala nominal classification within the lexico-grammatical system.

4.2. DISTINGUISHING BETWEEN SUBTYPES OF LEXICO-GRAMMATICAL SYSTEMS IN SINHALA. As previously discussed the two primary differences between noun classifiers and numeral classifiers is their location within the noun phrase and individuation. While noun classifiers are found next to the noun in noun phrases and typically do not function to individuate the referent, numeral classifiers are typically found in numeral or quantifying phrases and are used in individuating contexts. While the distinction between the classifier's location in the noun phrase sounds clearly distinguishable, it is not so clear cut. In fact, as Greenberg points out, "in many languages the classifiers are not compulsory even for the restricted set of nouns that have them" (1972:6). An, in fact this is the case for SGNs in Sinhala. While SGNs are sometimes preferred in numeral phrases, they are typically not mandatory unless otherwise morphologically marked. Therefore, at first glance, they may appear to behave more like noun classifiers, however, their semantic properties are more suggestive of numeral classifiers. That is, they are used for individuation and for the remaining six terms (*geḍi, karal, kæṭə, æṭə, mal, alə*) the semantic relationships between elements are not clearly taxonomic, but rather the general terms carry some information about the shape or form of the specific nouns they accompany. Furthermore, their semantic consistency comes more from the properties of shape or form than their taxonomic relationship. The emergence of shape as a device for categorization is a typical semantic feature of numeral classifiers. However, in the case of Sinhala, the it is clear that it is not *only* shape that categorization relies upon. In this way, the general nouns of SGNs do not emerge as semantically prototypical numeral classifiers. However, along with the individuating function these terms emerge as more numeral than noun classifier-like.

5. CONCLUSION. The patterns described for SGNs in Sinhala suggest the presence of both lexical and lexico-grammatical systems of nominal classification. Three of the nine general terms of SGNs (*kææli* 'pieces', *peti* 'flat.things', and *palu* 'sections') investigated in this paper displayed characteristics more typical of measure terms (quantification in non-individuated contexts, use with mass nouns) while also showing signs of lexico-grammatical systems (individuation, denotation of shape). Of the remaining six classifying terms two (*æṭə* 'seed', *mal* 'flowers') are

---

[4]More evidence on the distribution of this particular term is needed.

more characteristic of class terms semantically since in most cases the semantic relationship between the elements is taxonomic; however, they displayed morphosyntactic patterns more characteristic of classifiers than of compounds based on language internal patterns of endocentric nominal compounding (obligatoriness of elements, anaphoric reference). Two of the remaining four classifying terms (*karal* 'pod-like.things', *alə* 'root.vegetables') also proved to pattern more like classifiers based on obligatoriness of elements and anaphoric reference. Additionally, these two terms exhibit semantic properties typical of classifiers (denotation of shape/physical properties, individuation). The remaining two classifying terms (*geḍi* 'small.hard.fruit-like.things', *kæṭə* 'block-like.things') display properties characteristic of classifiers, both semantically (denotation of shape, individuation) and morphosyntactically (anaphoric reference in individuated contexts, obligatoriness of elements). These classificatory terms in Sinhala may best be described as residing synchronically on the continuum of noun classification systems between class and measure terms and classifiers. The following figure attempts to demonstrate how we may locate Sinhala's SGNs among the systems of nominal classification.
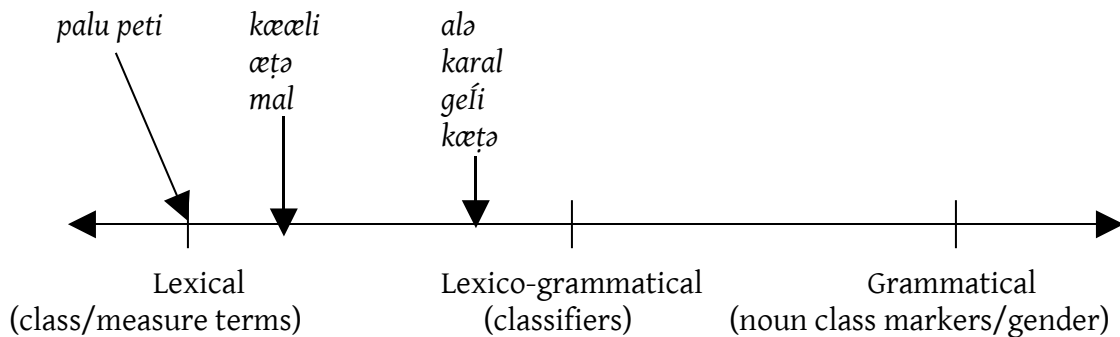

FIGURE 2. Sinhala's system of nominal classification

Although, several of Sinhala's SGNs pattern like lexico-grammatical systems, they are not the best exemplars of classifiers crosslinguistically. Furthermore, the semantic and morphosyntactic evidence suggests that the distinction between classifier subtypes too, may not be so clear cut.

As we have seen even among those whose characteristics are most suggestive of classifiers, the patterns are not prototypical of noun or numeral classifiers. The patterns described here suggest that those terms most like classifiers conform closest to the semantic properties of numeral classifiers, while less representative of numeral classifiers morphosyntactically. The following figure attempts to clarify the classification of these terms in Sinhala.
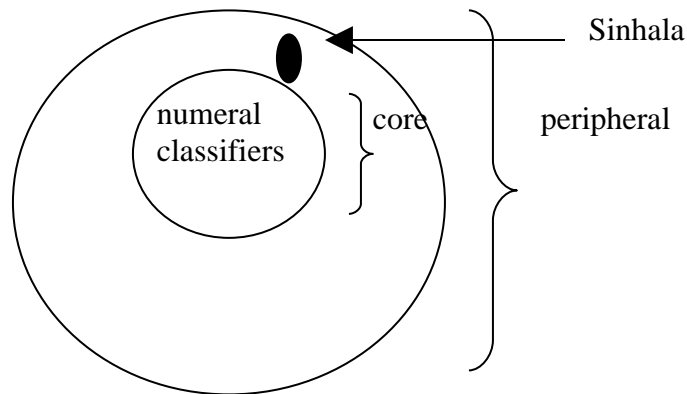
FIGURE 3. Classifier system in Sinhala

This figure attempts to illustrate the location of some SGNs in Sinhala as peripheral members of the category numeral classifiers. Here the inner circle represents the class of numeral classifiers that display the most prototypical characteristics of numeral classifiers. While the outer circle, within which I have placed Sinhala, represents the class of classifiers that do not possess clearly core characteristics or do not pattern systematically in the way those in the core class do, throughout the language.

This paper contributes to the growing body of literature on spoken Sinhala by exploring a small and until now underdescribed aspect of the grammatical system of Sinhala. It further contributes to typological work on Indo-Aryan languages by providing evidence for another system of nominal classification not yet described. There is much left to explore in the semantic and morphosyntactic patterns of nominal classification in Sinhala. This research would be much enhanced by an investigation in to naturally occurring discourse patterns among speakers of Sinhala in Sri Lanka.

REFERENCES

AIKHENVALD, ALEXANDRA Y. 2003. *Classifiers: A typology of noun categorization devices*, ch. 3, 81-97. Oxford: Oxford University Press.

ALLAN, K. 1977. Classifiers. *Language* 53.284-310.

DELANCEY, SCOTT. 1986. Toward a History of Tai Classifiers Systems. *Typological Studies in Language: Noun Classes and Categorization*, vol. 7, ed. by Colette Craig. 437-452. Amsterdam: John Benjamins Publishing Company.

DIXON, R.M.W. 1986. Noun Classes and Noun Classification in Typological Perspective. *Typological Studies in Language: Noun Classes and Categorization*, vol. 7, ed. by Colette Craig. 105-112. Amsterdam: John Benjamins Publishing Company.

DIXON, R.M.W. and ALEXANDRA Y. AIKHENVALD. 2002. Word: A typological framework. *Word: A cross-linguistic typology*, ed. by R. M. W. Dixon and Alexandra Y. Aikhenvald, 1-41. Cambridge: Cambridge University Press.

EMENEAU, M. B. 1956. India as a Linguistic Area. *Language* 32.1.3-16.

GAIR, JAMES W. and JOHN C. PAOLILLO. 1997. *Sinhala.* München: Lincom Europa.

GREENBERG, JOSEPH H. 1972. Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type. *Working Papers on Language Universals* 9.1-39.

GRINEVALD, COLETTE, and FRANK SEIFART. 2004. Noun Classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology* 8.243-285.

University of California, Santa Barbara
Department of Linguistics
3607 South Hall
Santa Barbara, CA 93106
mhenderson@umail.ucsb.edu