Language in Society **26**, 269–296. Printed in the United States of America

Sinhala diglossia: Discrete or continuous variation?

JOHN C. PAOLILLO

Linguistics Program, Box 19559
University of Texas
Arlington, TX 76019-0559
johnp@utafll.uta.edu

ABSTRACT

Sociolinguists disagree on how to characterize diglossia with respect to the structural relatedness of the H(igh) and L(ow) varieties: Ferguson 1959, 1991 holds that H and L should be distinct but related varieties of language, while others maintain that a continuum model is more appropriate. Both discrete models (Gair 1968, 1992) and continuum models (De Silva 1974, 1979) have been proposed for Sinhala, as spoken in Sri Lanka. In this article, I employ a computer-generated multidimensional graph of relations between varieties of Sinhala to show that the distribution of H and L grammatical features in a sample of naturally occurring texts supports the discrete H and L model more than the continuum model. A rigorous characterization of diglossia as a distinct type of language situation is proposed, based on the notion "functional diasystem." (Diglossia, Sinhala, Sri Lanka, diasystem, hybridization, continuum, South Asia, standardization)*

Sociolinguists generally agree that a diglossic situation is one in which a single speech community employs two or more varieties of language, a H(igh) variety and a L(ow) variety, for different communicative purposes. Ferguson's classic definition (1959) also includes a structural component: the two forms of language are varieties of the same language, and hence related – but "highly divergent" from one another, moreso than a dialect in relation to its standard language. But on this point there is little agreement, and different researchers give different characterizations of how divergent H and L must be. Four broadly characterizable positions can be cited:

(I) *H and L are distinct, but structurally related.* They are two highly divergent varieties of the same language. Lexicon, grammar, and phonology are all involved in the structural relationship between H and L (Ferguson 1959, Britto 1986).

(II) *H and L may be more or less distinct, and need not be structurally related.* They may be two different dialects or languages. Structural related-

© 1997 Cambridge University Press 0047-4045/97 $7.50 + .10

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

TABLE 1. *Grammar and lexis in four models of diglossia*

Characterization	H/L Grammar	H/L Lexis
I Ferguson	discrete	discrete
II Fishman	discrete or continuous	discrete or continuous
III De Silva	continuous	continuous
IV Gair	discrete	continuous

ness is only important to the extent that there is some difference between H and L (Fishman 1967, 1972, Fasold 1984).

(III) *H and L are not distinct, but are structurally related along a continuum.* A range of mixtures of H and L blurs the distinction between them, creating a continuum with two extreme poles (De Silva 1974).

(IV) *H and L are distinct in grammar only, but structurally related.* While H and L have distinct but related grammars, lexical code-mixing varies continuously, and gives the impression of a continuum (Gair 1992).

These four characterizations illustrate the difficulties inherent in trying to distinguish different types of language situations. Model I asserts that diglossia is distinct from other types of language situations, particularly "standard with dialects" and societal bilingualism. Model II reflects the observation that different language codes tend to specialize for certain domains in all language situations, independent of the degree of structural distinctness of H and L. Models III and IV were advanced principally to account for patterns observed in Sinhala diglossia, but are nonetheless significant for the characterization of diglossia more generally. De Silva's Model III suggests similarities between diglossia and (post-)creole continua, based on the existence of language varieties intermediate between H and L. Diglossia thus exemplifies hybridization, a more general pattern of language change through contact. In contrast, Gair (Model IV) characterizes Sinhala H and L as separate mainly in terms of grammar. The differences among these four models can be summarized as in Table 1.

The pivotal differences among the four models involve the nature of the variation, continuous or discrete, found in both grammar and lexis. There is a similarity between Models I and IV, in that both accord a special status to situations where the variation exhibited is discrete; and between Models II and III, in that both regard diglossia as realizable in situations where there is no sharp discontinuity between H and L. We can thus isolate two competing models of diglossia: one in which a sharp binary distinction is made between H and L (represented for Sinhala by Gair's model) and one in which the differences between H and L are regarded as a continuum (cf. De Silva's model). Which of these two models accounts for the Sinhala situation best?

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

Since the question turns on the discreteness or continuousness of H and L grammar, we need empirical evidence to demonstrate continuousness or discreteness in the grammar of Sinhala.

While the different models of diglossia disagree as to whether they regard H and L as distinct, no one appears to have developed an empirical procedure to test for distinctness in diglossia, or indeed in any other language situation. Instead, diglossia research typically reports impressionistic judgments about the relatedness or divergence of varieties used in a given language situation (e.g. the papers in Hudson 1991, and part III of Krishnamurti et al. 1986). Claims about one, two, three, or uncountable numbers of canonical varieties are simply asserted, with little further justification.

In this article, I have two principal aims: first, to evaluate the specific claims regarding the discreteness or "two-ness" of Sinhala diglossia; and second, to present a methodological approach that could be used to assess discreteness in other language situations, diglossic and otherwise. I do this by examining 52 Sinhala text segments for grammatical relationships among the varieties of language they employ. These relations are then plotted by a computer program in a multi-dimensional graph, from which claims about discreteness and continuousness can be directly interpreted. The results of this investigation reveal that two discrete clusters of varieties exist in Sinhala, but that substantial hybridization also occurs. To account for these results, I propose a model of diglossia in which H and L are systems of correspondence among the varieties of language in a diglossia; I adapt Britto's notion (1986, 1991) of H and L "diasystems" in order to describe these correspondences. The resulting model provides an important refinement of our understanding of the relationships among the varieties present in diglossic and other language situations.

The remainder of this article is organized as follows. The next section provides an overview of Sinhala diglossia, focusing in particular on the continuum and discreteness claims. Following this, I describe the methodology used to identify and graph structural relations among language varieties in Sinhala. This is followed by the results of the analysis; an interpretation of the results, in terms of a new model based on the notion of "diasystem"; and general conclusions.

SINHALA DIGLOSSIA

The Sri Lankan Sinhala language situation has been characterized as diglossia (De Silva 1967, 1974, 1976, 1979, 1986; Gair 1968, 1986, 1992). Gair 1968, 1986 describes Sinhala diglossia in terms of the nine defining characteristics (or "rubrics," Fasold 1984) that were originally identified by Ferguson 1959. In the next two sub-sections, I discuss the six non-structural rubrics: function, prestige, literary heritage, acquisition, standardization, and

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

TABLE 2. *Distribution of functions in Sinhala diglossia*

Function	Sinhala Variety	Predicted Variety
Instructions to servants, waiters, workmen, clerks	Spoken	L
Conversation with family, friends, colleagues	Spoken	L
Radio soap opera	Spoken	L
Caption of political cartoon	Spoken	L
Folk literature	Spoken	L
Sermon in church, mosque, temple	Spoken (formal)	H
University lecture	Spoken (formal)	H
Speech in parliament, political speech	Spoken (formal) (or Literary)[a]	H
Personal letter	Spoken (formal) (or Literary)	H
Novels (conversational parts)	Spoken	L[b]
(non-conversational parts)	Literary	H[b]
News broadcast	Literary	H
Newspaper editorial, news story, picture caption	Literary	H
Poetry	Literary	H
Government documents, forms	Literary	H[b]
Airline announcements	Literary	H[b]

[a]Gair 1986 notes that Literary is used in personal letters and in parliamentary speeches in "special circumstances," but he does not specify these circumstances.
[b]Not in Ferguson 1959.
Source: Gair 1986.

stability. Following this, in separate sub-sections, I discuss each of the three structural rubrics: phonology, lexis, and grammar.

Function

Gair 1968, 1986 gives perhaps the clearest summary of function in Sinhala diglossia. According to him, LITERARY SINHALA, the variety used for most written communication, is the H variety, and SPOKEN SINHALA, the variety used for everyday conversation, is the L variety. The distribution of the functions of Spoken and Literary parallels the distribution predicted by Ferguson, as seen in Table 2.

However, the situation reported in this summary raises certain questions about the appropriateness of a distinctness model for Sinhala. Discrepancies occur in four functions: sermons, university lectures, political speeches, and personal letters – the first three of which can be characterized as formal speaking situations. In all these functions, we find what Gair 1968 calls a formal version of Spoken Sinhala, characterized by a mixture of Literary (H) and Spoken (L) features. Whether a distinctness model works for Sinhala depends on the nature of this "Formal Spoken" variety: is it a version of Spoken Sinhala, or of Literary, or does it have its own unique properties? Is it distinct from Spoken and Literary, or does it blend smoothly into both?

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

Even if distinctness can be established, there are still some difficulties in applying Model I. For Ferguson, it is important that formal speaking contexts use a variety of H (p.c.). Yet Gair's label "Formal Spoken" indicates that a variety of L is used for formal speaking, not H. Formal Spoken varieties of Sinhala therefore offer an important testing ground for the discreteness and continuum models of diglossia.

Prestige, literary heritage, acquisition, standardization, and stability

These five rubrics are discussed as supporting a discreteness model by Gair 1968, 1986, and as supporting a continuum model by De Silva 1974. In the discreteness model, Literary Sinhala is the prestige variety, and it has a broad body of literature extending back several centuries; but there is virtually no literature in Spoken Sinhala.[1] Literary Sinhala is acquired only through formal education, while the only vehicle for learning Spoken (within the context of Sinhala society) is naturalistic acquisition. Literary is also more standardized than Spoken. The entire Sinhala diglossic situation is very stable, which Gair (1986:335) explains through the relative absence of various "diglossia-threatening factors." In short, Literary and Spoken Sinhala exhibit the social distribution regarded as characteristic of diglossia by Ferguson.

In the continuum model of De Silva 1974, a number of different varieties are observed in addition to the Literary and Spoken identified by Gair 1968. The varieties are "hybrids" resulting from a kind of contact between the modern (colloquial) language and the language of an Augustan past, as preserved in classical literature. Creators of new literature are among the most prominent innovators in this process of hybridization. In addition, socio-political forces, such as nationalistic reactions to foreign colonization, promote infusion of classical elements into the spoken language; so no domain is free of hybrid influences. Depending on one's precise political alignment (e.g. with nationalist purists, or with Buddhist monks and Ayurvedic physicians), different elements of the classical grammar may be regarded as prestigious, leading to hybrids competing for standard status in certain functional domains. Mass education and the popularization of literature, rather than reinforcing the distinctness of Literary and Spoken, actively contribute to the process of hybridization, by turning people with imperfect knowledge of the classical language into active users of it. The hybrid varieties coined by these people then become models for subsequent generations of learners, leading to further hybridization. The continuum model thus represents the situation dynamically, with hybridization constantly breaking down any distinction between Literary and Spoken Sinhala.

The two views of the Sinhala situation thus make quite different claims about the five non-structural rubrics. In the discreteness model, prestige, literary heritage, and acquisition have uniform and well-defined roles. The con-

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

sequence of this uniformity is standardization and long-term stability of the Sinhala situation. In the continuum model, the prestige language features are not uniform for all members of the speech community; acquisition promotes hybridization, and hence historical instability and heterogeny, rather than standardization and stability. Thus the two models arrive at very different representations of the situation. Since the continuum account depends on the hybridization of language varieties, it should be possible to confirm or disconfirm this account by looking for such hybrid varieties. This is the approach that is taken in the investigation below.

Phonology

H and L phonology are not sharply divergent in all diglossic situations (Ferguson 1959). The phonological differences between H and L in Sinhala appear to be among the less divergent, in contrast with other diglossic situations of South Asia (e.g. Tamil, described in Britto 1986). The primary phonological differences in Sinhala are (a) that formal writing uses a number of segments borrowed from Sanskrit, including aspirate consonants, which are absent in informal speaking; and (b) that certain vowel reduction processes operate in speech which are not written (De Silva 1974:79–85, Gair 1968, Paolillo 1992:79–83). From Ferguson, it is not clear precisely how phenomena such as fast speech rules are to be treated. One might be tempted to regard them as L, since they do not occur in writing, which is the provenance of H. Yet in the appropriate contexts (e.g. in broadcast news, which in Sinhala is read quite rapidly) one may find them being used with H material as well (Paolillo 1992:81). For these reasons, phonological differences among the different varieties of Sinhala are not be treated here, but left to future study.

Lexis

A prominent argument made by De Silva 1979 in favor of the continuum model concerns choices among more and less formal lexis. De Silva reports specifically (1974, 1979) on a number of experiments he conducted on the effects of lexical selection and comprehensibility of H:

> In some one thousand test cases conducted with a view to studying if illiterate people misunderstood literary texts because of their grammatical peculiarities, I obtained no evidence to support the view that these peculiarities impeded understanding. Any lack of understanding in my test cases was caused by lexical difficulties alone, and as soon as difficult words were replaced by their more commonly known equivalents, the purport of the text passages was understood by everyone tested. (1974:76, cited in 1979:41)

From this, De Silva concludes that the principal differences among different varieties of Sinhala are lexical in nature, and that the grammatical prop-

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

erties of more formal varieties are "redundancies" which are filtered out in comprehension (1974:76, 1979:40). Lexical variation is continuous in nature, since it involves independent choices between H or L forms at any point in a sentence (1974:87).

Gair (1968:6) also observes that some meanings have "ladders" of forms from H to L; such is the case with the set *bævin*, *nisaa*, *hindaa*, *handaa*, all meaning 'because', but expressing different levels of formality from H to L. Sets like this are numerous, and when all the variable choices among the different items are considered, the number of different, finely graded levels of formality that are possible through word choice becomes quite large (Gair 1968). In other words, Gair 1968, 1992 and De Silva 1974, 1979 agree that lexical variation is a resource which permits finely graded expression of formality; for this reason, lexical variation is not a major focus of the present study. However, the characterization of lexical variation as continuous distinguishes both Gair's and De Silva's models from Ferguson's model (Model I), in which lexis is expected to have discrete H and L components. It is therefore desirable to include some features of lexis in this investigation. Two such features are considered: the functor items in manner nominalizations (which form a "ladder" like the 'because' set above), and Sanskritic lexis, a large number of words that have been borrowed into Sinhala from Sanskrit, especially for specialized technical vocabulary (Gair 1968, De Silva 1974).

Grammar

The discreteness model requires H and L to be grammatically distinct. Ferguson (1959:333) summarizes, "It is certainly safe to say that in diglossia *there are always extensive differences between H and L*" (emphasis in original). According to Gair, the principal difference between Literary and Spoken Sinhala is that Literary has subject-verb agreement on finite verbs, while Spoken has none: "Sinhalese showing Literary main verb forms and the attendant subject-verb agreement is *Literary Sinhalese*. This will normally imply the presence of other features of Literary ... [By contrast,] Sinhalese that does not show Literary main verb forms and the attendant subject-verb agreement is *Spoken Sinhalese*" (Gair 1968:10). Gair 1992 argues that this single difference has many consequences throughout the grammar, since a number of other differences between Literary and Spoken Sinhala appear to depend on agreement.

However, a distinctness account must confront the issue that the "two-variety scheme requires some modification when the full range of current materials is taken into account" (Gair 1968:1). As already noted, informal writing (e.g. personal letters) and formal speaking are less readily characterized as either Literary or Spoken. Gair characterizes the language used in such functions as sub-varieties of Spoken (1968:10): "Spoken Sinhalese has two main sub-varieties. One, Formal Spoken, is characterized by the employ-

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

ment of one, and usually more, structural features of the Literary variety with relative consistency. Spoken Sinhalese that does not show such features is (Spoken) Colloquial."

The informal writing and formal speaking varieties all lack subject–verb agreement, and are thus defined by Gair's criteria as varieties of Spoken or L. The communicative functions of H and L, thus defined, come out slightly skewed with respect to Ferguson's model; but the two-ness of the model is preserved. Gair's proposal amounts to a claim about where the two-ness is found in Sinhala: H and L are delimited by the finite verb forms used, and by the other grammatical differences that directly follow from them. Features not dependent on agreement may be more variable in their distribution, depending on other factors. In particular, "public presentation tends to bring forth formal characteristics" (1986:329).

The discreteness view thus formulated is not the only possible interpretation of the situation, however. Alternative perspectives call into question the discreteness of H and L in Sinhala. In Paolillo 1991, I looked at different varieties of Sinhala used in children's literature, and concluded that at least four grammatically distinct varieties are used in that domain alone. De Silva 1974 identifies several other varieties he regards as hybrid. While De Silva 1967, 1974, 1976 describes Literary and (Spoken) Colloquial Sinhala very much as Gair does,[2] he sees Literary and Colloquial as two extreme ends of a continuous range of variation. According to De Silva, "... the various features of divergence between the two varieties are liable to create a number of hybridisms in the process of formalising, and therefore it would be unreal to speak of any one formal variety that is emerging" (1979:40). In this view, there are no sharp structural discontinuities between the varieties Literary, Formal Spoken, or Colloquial; there are only varying degrees of formality and varying degrees of hybridization between the language of the classical literature and the colloquial language.

Thus the discreteness and continuum models of Sinhala differ from one another crucially in whether they regard grammar as continuous or discrete.[3] The present investigation focuses on grammatical features of H and L in an effort to resolve this issue empirically. The results will determine the extent to which a discreteness model of diglossia can be maintained for Sinhala.

AN APPROACH TO THE PROBLEM

To resolve the problems in the description of Sinhala diglossia mentioned above, I designed an empirical procedure to discover whether Sinhala possesses a natural "two-ness," or whether it is more continuum-like. This procedure was carried out in two stages. The first involved detailed analysis of a number of different samples of Sinhala, selected to represent as accurately as possible the complete spectrum of varieties, in terms both of communi-

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

cative function and of linguistic structure. The second stage involved comparing the structural differences among the samples by plotting them in a multi-dimensional graph with the aid of a computer program. The graph was designed such that two-ness, if present, would show up as a pair of distinct clusters of samples; within each cluster, the samples would be more closely related to one another than to those of the other cluster. Types of organization other than "two-ness" could also be observed. For example, a third canonical variety would emerge as a distinct third cluster of texts; a continuum would appear as a uniform sequence of varieties between the two extremes, with no distinct clusters in between.

For the purposes of this analysis, I made use of a corpus I had previously constructed (Paolillo 1992), consisting of 29 texts selected to represent the broadest possible range of communicative functions and the greatest number of combinations of grammatical features. One limitation of this corpus is that the selection criteria caused intermediate varieties to be disproportionately represented, and thus potentially to bias the results of the analysis in favor of a continuum model. However, if "two-ness" were to emerge from this corpus, it would provide even more compelling evidence in support of the discreteness model.

After selecting the texts, I segmented each one into units, based on the communicative functions represented in it. The principal purpose of this was to separate conversation from the surrounding text within narratives, since the distinction between conversation and non-conversation has been observed to influence the choice of H and L (De Silva 1967, Gair 1968; see also Table 2). I thus segmented the 29 texts into 52 text segments (listed in the Appendix), which were then analyzed independently.

Each of the 52 text segments was coded for occurrence of 16 different grammatical features, listed in Table 3. The 16 grammatical features chosen for analysis have all been reported in the literature as present predominantly in either Literary or Colloquial Sinhala – i.e., they are markers of the diglossic levels of Sinhala. Only presence or absence of each feature was recorded; information about frequency within each segment was not gathered.[4] Each feature is described and exemplified in a separate sub-section below; the first line of each example indicates its source, as described in the Appendix. The example numbers of the 14 examples below correspond to the numbering of the features in Table 3 – except for locative case, feature 15, which appears in ex. 14.

Finite verbal endings

Finite verbal endings have been reported to be one of the principal markers of Literary and Colloquial Sinhala (Gair 1968, 1992). Written communications characteristically employ finite verb forms that agree in person, number, and gender with the subject. Examples of these are:[5]

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

TABLE 3. *Grammatical features of Sinhala diglossia*

Category	Literary	Colloquial
Finite verbal endings	(1) Subject-verb agreement	(2) Non-agreeing finite verb
Verbal adjectives	(3) Prenominal past verbal adjective	(4) Past participial verbal adjective
Infinitives	(5) *-iimata* dative verbal nouns	(7) Infinitive in *-nna*
	(6) *-nnata* dative verbal nouns	
Manner nominalizations	(8) Manner nominalizations	(9) Manner nominalizations
	(a) *ayuru*	(a) *widi*
	(b) *lesa*	(b) *hæti*
	(c) *pinisa*	
	(d) *paridi*	
Syntactic operations	(10) Passivization	(11) Presentative postposing after finite verb
Copula	(12) Lexical copula in	[no copula]
	(a) finite clauses	
	(b) non-finite clauses	
Verbal nouns	(13) Verbal noun in *-iima*	[verbal noun in *eka*]
Case	(14) accusative	[syncretized with nominative]
	(15) locative	[syncretized with genitive]
Lexicon	(16) Sanskritic lexis	

(1) Quoted dialog from a sixth-grade Buddhist moral primer (14 1)
budu-hu "mama sitiye-mi to da sita" yi wadaal-oo
Buddha-PL I stand-1SG you also stand IMP QUOT say PAST-3PL
'The Buddha said, "I am standing. You stand too '"

The finite verb form characteristically employed in spoken communication is the non-agreeing finite verb form (Fairbanks et al. 1968:1.6–7). This form takes a subject of any person, number, or gender category:

(2) Conversation in a traditional story told orally by a monk (17.1)
ee-wagee, ee looba bamunaa ee salli tæn-tæn-wala hænguw-aa
that-like, that greedy brahmin that money place-place-PL GEN hide PST-FIN
'And like that, that greedy brahmin hid that money in various places '

Verbal adjectives

Especially in formal writing, a special pre-nominal form of the past tense verbal adjective is often used (Gair & Karunatilaka 1974:109–12). These forms are rarely used in speaking.

(3) Descriptive and narrative portions of the TV news (22 2)
æya palaa giyee wedi hañda æsuu wita pamani
she run PPL GO.PAST FOC bullet sound hear.PAST.ADJ when only.PRED
'It was only when she heard the sound of the bullets that she ran away '

Correspondingly, there is another verbal adjective form, also with past tense reference, which is rarely used in writing; this form is based on the past

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

participle stem, and is called the past participle adjective form (Fairbanks et al. 1968:1.342–44).

(4) Narrative portions of a traditional story told orally by a monk (18.2)
passe karapu dee allala gæṭa gahaa-gana alla-gana genaawaa.
after DO.PPL.ADJ thing grab.PPL knot bind-REFL.PPL grab-REFL.PPL bring.PAST.FIN
'The thing they did after that was grab her and bind her and bring her.'

Infinitives

A number of different alternative infinitive forms are available in Sinhala. Written communication most often makes use of dative verbal nouns ending in either *-iimaṭa* (Gair & Karunatilaka 1974:14) or *-nnaṭa* (1974:90).

(5) Instructional exercises of a second-grade Buddhist moral primer (4.3)
siduhat raja tumaa-t, deewiya-t satuṭu kiriima-ṭa raja gedara
Siddhartha king HON-also queen-also happy DO.NOM-DATking house
hiṭiyoo kawuru da?
be.PAST.3PL who Q
'Who lived in the palace to make King Siddhartha and the queen happy?'

(6) Transcript of an oral funerary sermon (27)
... itin maṭa gam-ee nawati-nna-ṭa awaśya nææ.
so I.DAT village-LOC stay-NOM-DAT necessary NEG
'... so it is not necessary for me to stay in the village.'

The most frequent infinitive form found in spoken communication is the *-nna* infinitive, for which dialect variants exist in *-nḍa* and *-nṭa* (Fairbanks et al. 1968:1.6–7, 2.71).

(7) Narrative portion of a popular children's book (10.2)
... kumaara huňgak praśna ahaṇḍa uṇaa.
Kumaara many question ask.INF become.PAST.FIN
'... Kumara began to ask many questions.'

Manner nominalizations

Sinhala has a number of ways of forming manner nominalizations. Written communication typically uses one of the forms *ayuru*, *lesa*, *piṇisa*, or *paridi* with a modifying verbal adjective (Gair & Karunatilaka 1974:114, 172, 224); but in speaking, typically either *widi* or *hæṭi* is used to form these constructions (Fairbanks et al. 1968:2.58).

(8) Abstract from a journal on cultural and linguistic studies (2)
samaajayehi, deeśapaalana, aarthika, sańskṛtika yanaadii kaṭayutu
society.LOC political economic, cultural etc. duties
sańdahaa bhaaṣaawa yodaa gæniim-ee dii in upari-ma
for language use.PPL take.NOM-GEN while that-INS superior-EMPH
kaaryakṣamataaw-ak iṭu wa-na lesa kaarya
capability-IND fulfillment become.PRES-ADJ manner action
sańwidhaanaya kiriima awaśya weyi.
planning do.NOM necessary be.PRES.3SG
'In order to realize a superior capability of utilizing the language for political, economic, cultural etc. purposes in the society, undertaking "action planning" is necessary.'

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

(9) Narrative portion of a fourth-grade story book (12.1)
ææ æňdum aňdina hæţi ohu balaa hiţiyee ya.
she clothes wear.PRES.ADJ way he-ACC look.PPL be.PRES.3SG PRED
'He was watching the way she got dressed.'

Syntactic operations

There is a passive construction which is used in writing but not typically in speaking: the subject of a transitive verb is marked with the postposition *wisin* 'by', the verb is marked with the auxiliary *labanawaa/læbenawaa* 'receive', and the object is promoted to the status of subject and agrees with the verb. There is no construction with a directly corresponding function in ordinary spoken communication (Gair 1968:5, 1992:192–94).

(10) Conversation from a story in "Grade-School Buddhist Doctrine" (14.1)
maa wisin bohoo sathu dænæ dænæ mara-ṇa lada-ha.
I.ACC by many being.PL know.PPL know.PPL kill-ADJ receive.PAST-3PL
'Many beings have been wittingly killed by me.'

There is also a presentative postposing operation which is more characteristic of informal communication than of formal communication. In this construction, the first mention of a discourse participant appears in the position immediately following the finite verb (usually, though not necessarily intransitive) in the non-agreeing finite form. This construction typically occurs with the important participants in a narrative (Herring & Paolillo 1995).

(11) Narrative portion of a traditional story told orally by a Buddhist monk (17.2)
budu haamuduruwoo sæwæt nuwara wæḍa inna koṭa ehee hiţiyaa
Buddha monk Savathi city go.PPL be.PRES.ADJ when therebe.PAST.FIN
aananda kiyalaa siţuwaray-ek.
Ananda say.PPL millionaire-IND.
'When the Buddha lived in Savathi, there lived a millionaire named Ananda.'

Copula

Sinhala has no lexical copula in informal discourse; equational, adjectival, and other non-verbal predicate sentences are formed verblessly (Gair & Paolillo 1988). But in formal written communication, an overt copula is obligatory in sentences with non-verb predicates (Gair & Karunatilaka 1974:11–12, 27).

(12) Descriptive and narrative portions of the TV news (22.2)
eya iţu kiriima saňdahaa wæya wii æti mudal rupiyal
that fulfillment do.NOM for price become.PPL exist.ADJ money rupees
miliyana hatarak pamana weyi.
million four about be.PRES-3SG
'The money that has been spent for the fulfillment of that is about four million rupees.'

Verbal nouns in -iima

Formal communication tends to use verbal nouns formed in *-iima* quite frequently (Fairbanks et al. 1968:2.52–53); an example can be found in the first line of ex. 12 (*eya iţu kiriima* 'the fulfillment of that'), as well as in ex. 13.

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

(13) Narrative of a book of a famous author's reminiscences (6.1)
mohu hæñdinwiima-ṭa wiśeṣa namak no-mæti wiima
this.3SG designate.NOM-DAT special name.IND NEG-exist be.NOM
kaṇagaaṭuw-ak-i.
unfortunate-IND-PRED
'It is an unfortunate thing that there is no special name to refer to him.'

A corresponding form used informally is made with the verbal adjective (present or past tense) followed by *eka* 'one' (Fairbanks et al. 1968:1.198). The *eka* verbal nouns are rare in my data, and I have not counted them for this study.

Case

Accusative case is reported to be a central characteristic of written Sinhala, although it has also been noted to occur in formal speaking situations (Gair 1968:4, 1992:180–82, Paolillo 1992:211–15). In most spoken communication, the functions of the accusative case are expressed by nominative forms. The locative case is an additional form that occurs in writing and formal speech, but not in informal speech (Gair & Karunatilaka 1974:7–8).

(14) Narrative portions of sixth-grade Buddhist moral primer (5.3)
boosat tumaa kal gewuu-yee owun samaŋga noyek-ut
bodhisatva hon time spend.PAST-FOC they.ACC with various-also
keḷi sellam aadiyehi nitara nitara yediim-en-i.
play game etc.LOC always always engage.NOM-INS-PRED
'It was by always engaging in the playing of games with them etc. that the bodhisatva spent his time.'

Lexis

Written Sinhala often makes extensive use of borrowed Sanskrit words, especially as technical terms (Gair 1968:6). I therefore coded the text segments in my corpus for this feature. Different vocabulary items have been borrowed at different times, and Sanskritic forms differ in the extent to which they show adaptation to Sinhala phonology. As a rule, however, formal writing uses more of the Sanskritic forms without adaptation; therefore I did not count borrowed Sanskrit words that have been phonologically adapted. E.g., I counted a word like *śaalawa* 'hall' as Sanskritic, since it is essentially identical to Sanskrit *śaalaa* 'hall', with the addition of Sinhala morphology, but I did not count a word like *saalaawa* 'hall', because it exhibits adaptation of palatal *ś* to alveolar *s*, even though it too is derived from Sanskrit *śaalaa*.

Coding and displaying the data

Each text was coded with a 16-digit binary number, with each digit representing one of the numbered features in Table 3. For example, text segment 7, a book jacket's coverleaf, has the following features: agreement (1), classical past participial adjectives (3), *-iimaṭa*, *-nnaṭa*, and *-nna* infinitives (5, 6, 7), passive (10), *-iima* verbal nouns (13), accusative case (14), and Sanskritic lexis (16). The presence of a feature was coded with a 1, and its absence with

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

TABLE 4. *Grammatical feature encoding for text segment 7*

Digit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Value	1	0	1	0	1	1	1	0	0	1	0	0	1	1	0	1

a 0, so this text segment was assigned the 16-digit binary number shown in Table 4.

A specially designed Macintosh computer program was then used to plot these 16-digit numbers as points in a diagram in a way that would allow their natural similarities to emerge.[6] Specifically, this was done by treating each of the digits of the 16-bit numbers as a separate dimension with two potential values, 0 being the origin position on that axis, and 1 representing a length along that axis that could be set from within the program. To create the diagram, each of the 16 axes (features) was given a unique angle in two-dimensional space; the 16 angles were arranged in a fan-like sequence in increments of 8 degrees, with the first axis being 8 degrees above the horizontal, the second being 0 degrees, the third being 8 degrees below, etc. The increment of 8 degrees was chosen to make all the angles fall within 180 degrees. This allows a feature's contribution to a data point's location to be inverted by adding 180 degrees to its angle, thereby flipping the relative positions of 0 and 1 on that axis. In this way, features whose presence contributes to formality (i.e. H features) can be separated from those whose presence contributes to informality (L features). Different features were assigned to each of the angles in a number of trial plots, allowing the most informative assignment of features to angles to be chosen.

Feature values such as having agreement, or lacking accusative case, were then treated as vectors from the origin point (whose value may also be set) along the appropriate angle. A feature value of 0 adds no displacement (a vector of magnitude 0) on the associated angle. The distances between 1 and 0 values on all 16 axes were kept the same (30 pixels), so that all features would be represented with vectors of the same magnitude. Each text was then plotted as a vector-sum of the values of all its features. The point 0000000000000000 (representing a text with none of the examined features present) thus falls exactly on the plotting origin; and, using the final plotting angles, a point 1000000000000000 (representing a text having agreement and none of the other features examined) would fall exactly 30 pixels away from the origin, at an angle 8 degrees above horizontal and to the left.

A diagram created in this manner has the virtue that any two texts differing by only one feature will always be plotted the same distance apart, no matter what the feature chosen is. However, the plotting distance between texts differing by more than one feature depends on the features by which they differ, and the plotting angles chosen to represent those features. The

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

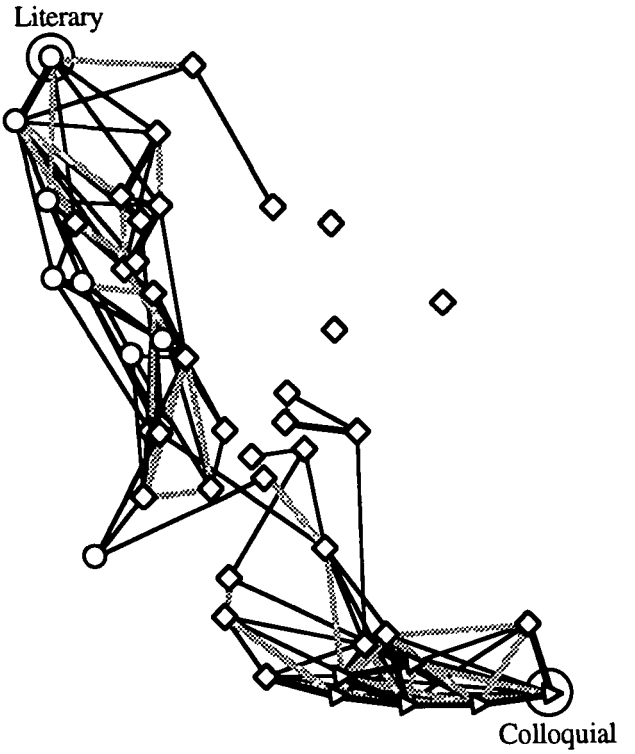


FIGURE 1: Proximity relationships among 52 Sinhala text segments.

result is that many of these locations will be plotted on top of other locations, simply because it is not feasible to manipulate the lengths and angles of the plotting axes to prevent this from occurring. A plot of all 65,536 (2^{16}) locations would be a figure with a higher density of points in the middle of a roughly circular area. To assist in identifying clusters of texts with similar features, another plotting algorithm was used to connect the points corresponding to texts differing from one another by a small number of features. Three different types of line segments were used: dark, wide lines connect points that differ by a single feature value; gray lines connect points differing by two feature values; and thin black lines connect points differing by three feature values. All relations between texts differing by one, two, or three feature values were plotted in this diagram. The resulting plot appears as Figure 1.

To further assist in interpretation, the plot in Fig. 1 was enhanced in two more ways. The two large circles near the top and the bottom of the diagram represent the hypothetical Literary and Colloquial "prototype" locations

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

respectively, i.e. the locations consisting of all and only the features purported to be characteristic of Literary and Colloquial. The points having values of 1 among only the five Colloquial features were plotted as triangles (▷, near the bottom of the diagram); those having only Literary feature values were plotted with small circles (○, mostly toward the left edge of the diagram). The points plotted as small diamonds (◇) represent segments having both Literary and Colloquial features.

RESULTS

A number of striking features are immediately evident from Fig. 1. First, we find that almost every text segment is related to some other text segment by changing no more than three features – a rather small difference, considering the number of features examined and the small number of texts sampled. Only a very small number of texts (three) have no direct relationship to other texts in this way. The result is a relatively smooth transition from the features of one text (or variety) to the features of another. The 52 text segments in the corpus are spread between the two poles, suggesting a continuum of varieties. However, in an area at the bottom of the middle third of the diagram, only two links, of three feature values each, connect the upper and lower parts of the diagram. This suggests a discrete difference between the Literary (upper) and Colloquial (lower) text samples in the corpus.

Plotting a corpus of texts in this manner could easily have resulted in a plot different from that in Fig. 1. If there were a more random relationship between the features chosen, the result might be more like Fig. 2, a plot of 52 points of randomly generated data, using identical angles and lengths.

A number of contrasts can be seen when comparing these two diagrams. First, neither of the two prototype locations is occupied in the random data of Fig. 2, although both are in the actual data in Fig. 1. The probability of having a text segment with either of these two values in a random sample of 52 points is approximately 1 in 1,260. The probability that one would get both in the same sample is much smaller, about 1 in 82,000,000. Thus there is a vanishingly small probability that the plot in Fig. 1, which has both Literary and Colloquial prototype locations occupied, would be the result of random variation among the features observed. In other words, we have empirical confirmation of the two prototype clusters of features, Literary and Colloquial Sinhala.

In addition, only one data point in Fig. 2 has exclusively Literary features, as opposed to nine with only Literary features and six with only Colloquial features in Fig. 1. The probability of randomly obtaining a single text with only Literary features present is about 1 in 32 (since there are only five Colloquial features, all of which must be 0), that of randomly obtaining a text with only Colloquial features present is about 1 in 2,048. The probability of

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

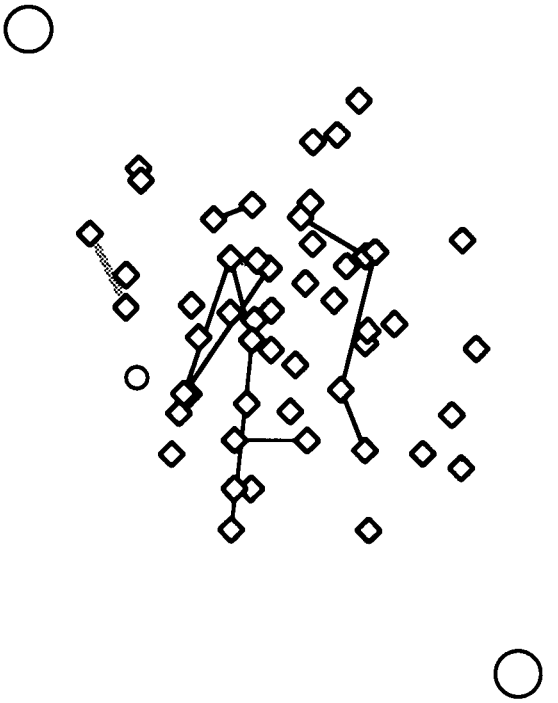


FIGURE 2: Proximity relationships among 52 random points.

randomly obtaining at least one Literary-only text in a sample of 52 texts is thus reasonably good ($1/32 \times 52 = 1.625$), but that of finding nine such points in the same size sample is vanishingly small. This suggests that significant covariation exists among the Literary features: they have a strong tendency to vary together in naturally occurring texts.

One can also observe that the tightest concentration of points in Fig. 1 – those differing by only single feature values – occurs around the Colloquial prototype, with a smaller cluster around the Literary prototype. The probability of having one such segment in a random data set of 52 points is about 1 in 3, and the probability of having as many as six such segments is about 1 in 43,000,000. The center of the diagram in Fig. 1 is also relatively unoccupied when compared with the centrally distributed points in Fig. 2. These observations about Figs. 1–2 merely confirm that the features chosen are not randomly distributed: they have significant patterns of covariation, as suggested by their classification as features of Literary or Colloquial Sinhala.

The kind of covariation present in the Sinhala corpus is not the only kind that could exist. Another common relationship among linguistic features is the implicational hierarchy, as often found in creole studies (cf. DeCamp

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

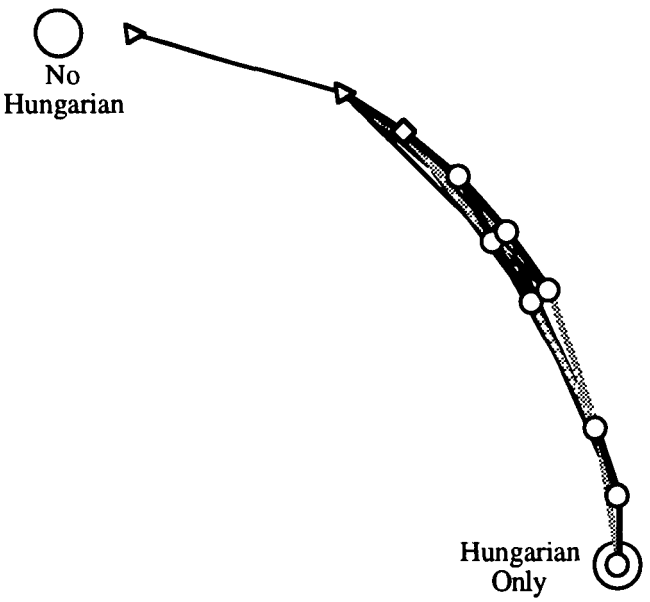


FIGURE 3: Proximity relationships in an implicational hierarchy: Hungarian use among women in Oberwart (Gal 1979).

1971, Bickerton 1973). Figure 3 shows another plot, this time of an implicational hierarchy, based on data from Gal 1979, concerning Hungarian use by women in the Austrian village of Oberwart.[7] In this graph, Hungarian use (or lack of use) with each of 12 interlocutors was assigned to an angle, starting from horizontal and progressing in the order of the implicational scale. The resulting graph has a strikingly different appearance from the other two: there is a very neat arc along which all the data points fall. Moreover, there are no large discontinuities in the data (there are no breaks in the lines forming the arc). While Gal's data are quite different from the Sinhala data examined here, we could nonetheless expect the Sinhala data to produce a similar configuration, if in fact it were implicationally scalable.

A third data set helps make this point. Figure 4 presents a plot of Bickerton's data (1973) on copula use in Guyanese creole.[8] Here we are concerned with a single linguistic feature, the basilectal and acrolectal copula forms. The different dimensions represent the different environments in which a basilectal or acrolectal copula form may be used. These are ordered as presented by Bickerton, in the following implicational scale: continuous verb > locative/existential > present "predicate adjective" > cleft > past "predicate adjective" > NP > impersonal. Again in this diagram, we can see a clean,

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

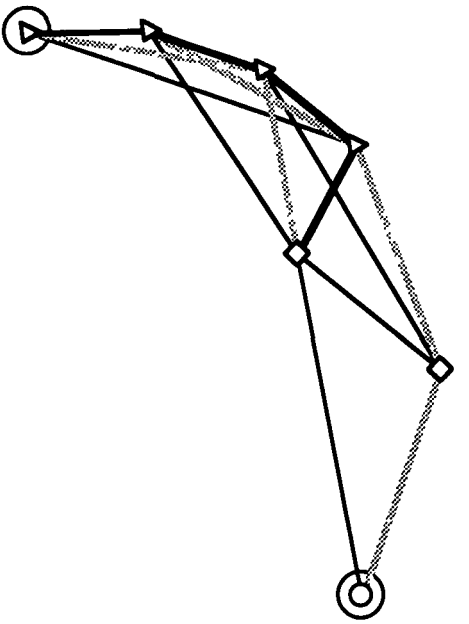


FIGURE 4: Proximity relationships in an implicational hierarchy: Copula use in Guyanese Creole (Bickerton 1973).

unbroken arc, representing the transition from acrolectal to basilectal forms in each of the relevant environments.

When plotting the Sinhala data, I experimented with several different orderings of the plotting angles, in an attempt to reveal any implicational scales that might be present. As it turns out, the data are not very easily scaled; only among the Colloquial cluster of texts did anything like a scale emerge. This scale shows up on the bottom edge of the lower cluster of Fig. 1. The relevant features, and the order of their scaling, are:

(15) *Implicational scaling of features among Colloquial text segments* (numbers correspond to numbers in Table 3)
(9) + *widi* and *hæṭi* manner nominals
(4) + participial adjectival form
(7) + *-nna* infinitives
(11) + postposing
(6) − *-nnaṭa* dative verbal noun
(9) > (4) > (7) > (11) > ~(6)

No matter which feature is chosen as the next one, however, there is a sudden discontinuity. From this point upward, the Literary (upper) cluster of the diagram could not be coerced to fall along an arc in the manner of the Colloquial (lower) portion of the diagram. Rotations other than the one cho-

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

sen here tended to straighten out the upper portion of the diagram, and sometimes interfered with the scaling of the lower portion. Thus, of the 16 chosen features, only 5 appear to be potentially involved in an implicational scale. The evidence in favor of the continuum model for Sinhala diglossia is thus rather weak. Instead, there seems to be a broader tendency toward covariation of H and L features, with a moderate degree of hybridization being tolerated.

To sum up, two general points are illustrated by Figs. 1–4. First, the distribution of the points representing the actual Sinhala text samples is nonrandom in a way that supports the two-ness of Ferguson's and Gair's models of diglossia. Second, the distribution of points does not resemble an implicational hierarchy very closely. There is a highly significant clustering of features in two regions of the graph, and a sharp discontinuity between these two regions – in spite of the fact that the corpus employed favored intermediate varieties, and therefore also continuity.

While the results do not show a uniform continuum, they do show evidence of the hybridization observed by De Silva 1974: points within both the Colloquial and Literary regions of Fig. 1 often have features of the other variety, as indicated by the large number of diamond-shaped points. Thus an adequate model of Sinhala diglossia needs to address the following two observations: there is a pair of clusters of varieties, oriented toward Literary (H) and Colloquial (L), and there is also substantial variation involving hybridization of Literary and Colloquial features within these clusters.

DIASYSTEMS VS. VARIETIES

We have seen that a two-ness exists in Sinhala diglossia, but that it cannot be characterized in terms of absolute covariation of H and L features. While the most significant concentrations of samples are at the Literary and Colloquial poles, a good deal of mixing of features occurs. How then are we to characterize this two-ness in a way that explicitly acknowledges the observed variability? One approach to the problem of variation in diglossic systems is that of Britto 1986, 1991, who proposes that the notion of H and L "varieties" be cast instead in terms of "diasystems," modeled after Gregory's notion (1967) of "diatypes".[9]

> Ferguson uses *variety* in two senses. In one sense it refers to the two major components, H and L, of diglossia, and in another sense it refers to each of the many subsystems that constitute a language. The word *diasystem* here is used to refer to the first sense, so that one may speak of a diasystem as composed of different varieties. (Britto 1991:60)

In terms of the data presented here, we can think of each individual text segment as representing a (potentially) distinct variety. These different varieties

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

then may each belong to one or the other diasystem, H or L, represented by the clusters of texts in Fig. 1.

Britto's innovation explicitly recognizes the non-homogeneous nature of the H and L systems. In defining the notion of diasystem more precisely, he identifies two principal criteria. The first characterizes them in terms of shared communicative functions:

> ... a diasystem is a collective denomination, or an abstract label, to refer to all varieties sharing certain common features (e.g. the fact that they are not superposed, the fact that they are used for conversation, etc.) ... (Britto 1986:14)

However, such a criterion alone is insufficient to identify diasystems of a diglossia without circularity. The second criterion is what Britto terms "optimal distance." He characterizes the structural relatedness between H and L in diglossia as intermediate between that of two different languages, on the one hand, and two different dialects or registers of the same language, on the other (following Fasold 1984:40):

> For convenience of reference, this intermediate relationship that is presumed to exist between H and L may be said to be optimal. H and L may then be termed optimally distant varieties or optimal varieties, so that they are distinguished from "languages" (which would be super-optimal varieties) and styles and accents, etc. (which would be considered sub-optimal varieties). (Britto 1986:10)

In other words, an intermediate degree of structural discontinuity between H and L is found in diglossia. Still, the notion of optimality described in this way does not greatly clarify the difference between diglossic and bilingual situations. Different languages can be as different as Spanish and Guaraní, or as similar as Danish and Norwegian. Optimality needs to be refined further by reference to the relations among the members of a particular diasystem, rather than just the relations between diasystems in the larger situation. We can thus add a pair of clarifying conditions to Britto's notion of optimality:

(a) Members of a particular diasystem must be more closely related in structure to the other members of that diasystem than to members of the other diasystem(s).

(b) Similarity of form among members of a diasystem corresponds to similarity of communicative function. Likewise, divergence in form corresponds to divergence in communicative function.

These conditions are clarifying, since they are implicit in Britto's characterization: functionally defined diasystems would not make useful descriptions unless the functional characteristics employed also identified varieties that corresponded in form. They are also clarifying in another way: we can

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

use them to eliminate a potential terminological confusion of Britto's notion of diasystem with Weinreich's earlier use of the term (1968). Weinreich was referring to a system of correspondence between linguistic systems with partial similarities, whether dialects, languages, or other language varieties:

> A "diasystem" can be constructed by the linguistic analyst out of any two systems which have partial similarities (it is these similarities which make it something different from the mere sum of the two systems). But that does not mean that it is always a scientist's construction only: a "diasystem" is experienced in a very real way by bilingual (including "bidialectal") speakers and corresponds to what students of language contact have called a "merged system." (Weinreich 1968, cited in Britto 1986:305)

Adopting Weinreich's definition as the basic sense of diasystem, we can rename Britto's sense "functional diasystem," to reflect the fact that it is characterized in terms of communicative function in addition to shared linguistic structure.[10] In other words, what is special about the optimal diasystems of a diglossia is that they consist of functionally coherent, structurally similar varieties.

A diglossic situation consists of a number of varieties having shared linguistic structure, and this relation "is experienced in a very real way" by the speakers of the diglossic community. Viewing this entire linguistic system as a diasystem, with special characteristics, has the advantage that it captures the systemic complementarity and unity of H and L in diglossia. A definition along these lines is as follows: DIGLOSSIA is a linguistic diasystem consisting of two functional diasystems.[11]

The two clear clusters in Fig. 1 correspond to the two functional diasystems in this model of diglossia. The variation within the clusters corresponds to the variation within functional diasystems, where each text is taken to represent a variety. Fig. 2 clearly lacks both these attributes; if a language situation were to plot out like Fig. 2, it would be characterized by a single diasystem (e.g. a monolingual, non-diglossic speech community), rather than by two functional diasystems. In contrast, Figs. 3 and 4 represent unbroken continua. If a language situation were to plot out like Figs. 3 or 4, it would not consist of paired diasystems, but of a single diasystem which stretched between two poles. Conversely, if a language situation were to produce a plot in which there was a broken continuum like Fig. 1, but with more than one break, then we would have three or more basic diasystems. Only Fig. 1 presents a situation in which there is a relatively discrete choice among the members of two distinct diasystems, and these results support the discreteness of Gair's model of diglossia. However, within the diasystems of Fig. 1, there is considerable variation in which features of the other diasystem are employed; this is the "hybridization" of De Silva's model. While we can confirm the existence of hybridization from the diamond points in Fig. 1, it is

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

important also to note that hybridization does not break down the distinctness of the two diasystems. Intermediate varieties appear more closely related to one of these two types than to the other, or than to any third such pole.

CONCLUSION

Through the analysis of a corpus of Sinhala texts, we have seen that the grammatical features of Table 2 are distributed in two predominant clusters, Literary and Colloquial. These clusters provide support for the claim that a discrete "two-ness" exists in Sinhala diglossia (cf. Gair 1968, 1986, 1992). Hybridization of Literary and Colloquial is also observed in Fig. 1, but Literary and Colloquial are quite distinct, contrary to De Silva's claim (1979) that there are no sharp structural discontinuities between them. The model of Sinhala diglossia which comports best with the results presented here is one in which there are two largely discrete functional diasystems. This model preserves many of the features of Ferguson's model, as well as important aspects of Gair's model, while placing both on a more sound empirical footing. The diasystem model also allows us to account for the variation in distribution of H and L features observed by De Silva and others, since variation is explicitly acknowledged in the notion of diasystem.

Of course, the analysis presented here is based principally on morphological and syntactic characteristics, and does not address the distribution of lexis. If lexical substitution is more continuous, as suggested by De Silva 1979 and Gair 1968, we would then have evidence that H and L grammar and lexis in a diglossia may show very different kinds of variation – a situation not expected under Ferguson's model (Model I), in which grammar and lexis strictly covary. In fact, it seems that different types of variation may be simultaneously possible under all three structural rubrics: phonology, lexis, and grammar. Investigation of this hitherto unexplored structural dimension of diglossic typology might help us better understand different processes of linguistic change.

What remains to be seen is whether diglossic situations like that of Sinhala are substantially different from other situations, such as the "standard with dialects" type or the creole type. For purposes of comparison, analyses similar to this one need to be carried out on other situations. A good test case would be a classical creole continuum, such as Guyanese Creole (Bickerton 1973, Rickford 1981, 1987). It has been well established that continuous, implicationally scalable variation holds on the level of individual features, such as basilectal and acrolectal copula forms. However, it is not clear from such results whether the same pattern would be found on the global level of analysis: if we examined patterns of covariation among a number of features in a creole situation, would they cluster, as they do in Sinhala diglossia, or

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

not? If they do, then a single, more general theory could account for language variation in both diglossic and creole language situations; if they do not, then different theories of variation are needed for each. Another related case is Haitian Creole French, employed by Ferguson as one of his four defining situations for diglossia. However, since it is a creole, one might expect it to exhibit properties of a continuum model. Indeed, there is still some controversy over whether Haitian Creole is diglossic in Ferguson's sense (Valdman 1988), and an empirical approach to the situation could be revealing.

Parallel investigations could also be undertaken among standard-with-dialects situations. We might anticipate, for example, that Standard and Nonstandard varieties of American English would reveal greater continuity than the diglossic Sinhala situation. If so, different theories are needed to account for the variation found in each. Conversely, if it were to turn out that diglossia and the standard-with-dialects situation yield the same type of plot, then we should expect the same patterns of variation in both (e.g. style, class, and gender stratification); and the linguistic and social processes which affect situations of both types should be the same. Finally, analyses of bilingual situations (e.g. Spanish and Guaraní, Rubin 1968) would presumably reveal greater discontinuity in structure between the H and L extremes. Each of these hypotheses is empirically testable using a methodology similar to that of this article. Examining situations like these would allow us to establish what degree of difference constitutes "optimal distance" in Britto's sense (1986) – the structural difference between diasystems necessary and/or sufficient for a situation to be diglossic.

The approach to linguistic variation piloted here is one which focuses on the details of linguistic structure, as does the (micro-)variationist approach of Sankoff 1988, Labov 1994 (and references therein), and others working in that tradition. However, it differs from this latter approach in that it looks at patterns of covariation among a collection of variables simultaneously, rather than the distribution of individual variables separately. The questions which can be asked in this approach are correspondingly different. Rather than seeking to identify the distribution of specfic markers, this macro-variationist approach addresses a typological question about the overall patterns of language variation in a given situation. The answers to such questions have much to contribute to understanding linguistic variation more generally. Until we understand how overall patterns of linguistic variation differ in diglossic, creole, and standard language situations, we will not know whether studies of micro-variation from such different situations can be interpreted comparably. There is thus a need to initiate a new program of research into what Ferguson 1991 has called "sociolinguistic typology." The present approach is intended to make a contribution to that effort.

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

NOTES

* An earlier version of this research was presented at the 22nd annual meeting of the Linguistic Association of the Southwest, University of Texas at Arlington, October 24, 1993, and published in the UTA Working Papers in Linguistics (Paolillo 1994). This research was supported in part by NSF grant BNS-9013361. I thank Susan Herring, James Gair, and Bh. Krishnamurti for comments on earlier drafts of this article.

[1] What literature does exist in Spoken Sinhala is either not taken seriously by the public (De Silva 1974), or is used at the lowest educational levels (first–third grades).

[2] In referencing De Silva's description, I use his terms *Literary* and *Colloquial*, rather than Gair's *Literary* and *Spoken Sinhala*. For both Gair and De Silva, *Colloquial* is a particular variety used in the vernacular communicative functions; for Gair, *Spoken Sinhala* is a cover term for both Colloquial and Formal Spoken (with their shared properties and structure). Thus the terms *Spoken* and *Colloquial* are not equivalent.

[3] Gair (1992:180) interprets the differences between his model and De Silva's as differences in focus: Gair 1968 treats the grammar rubric as crucial, while De Silva 1974 treats the lexicon rubric as crucial. However, De Silva specifically addresses hybridization and continuousness in grammar, as well as lexicon (1974:85–89).

[4] Because of the great variation in length of text segments, the frequency information that could be gathered would not be equally reliable for each.

[5] Examples which come from written sources are transcribed here according to the conventions of the Sinhala writing system. Examples from spoken sources were first transcribed in close phonetic transcription and then normalized. Some phonemic differences, notably final short *a* vs. final schwa, are therefore not represented in this transcription system. The abbreviations used in glossing the examples are as follows:

1SG	first person singular	INS	instrumental
3PL	third person plural	LOC	locative
3SG	third person singular	NEG	negative
ACC	accusative	NOM	nominal(izer)
ADJ	adjectival	PL	plural
DAT	dative	PPL	past participle
FIN	finite	PRED	predicate marker
FOC	focus	PRES	present tesne
GEN	genitive	PAST	past tense
IMP	imperative (familiar)	Q	question marker
IND	indefinite	QUOT	quotation marker
INF	infinitive	REFL	reflexive

[6] The program used to draw the figures presented here, called DrawMD, was written in the Prograph Classic visual language. It may be obtained by anonymous ftp from ling.uta.edu. Source code is available from the author upon request.

[7] The present version of the program only handles binary, not ternary, oppositions; hence German use was excluded from the data used for this diagram. This exaggerates the scalability of Gal's data somewhat.

[8] The data here are confined to representing the presence or absence of acrolectal copula forms. This again somewhat exaggerates the scalability of Bickerton's data. My purpose is to illustrate how an idealized implicational scale would be drawn using the current method, so the idealizations made in representing Bickerton's and Gal's data are appropriate for this purpose.

[9] A "diatype" or "diatypical variety" is "any variety distinguished according to characteristics of use" (Gregory 1967:194, cited in Britto 1986:305).

[10] While Weinreich's notion of diasystem has a psychological component, this is not crucial to the use I make of it. I seek to explain the relationships among systems revealed by an analysis of structure. Whether such systems are cognitively manifested among speakers' individual competences is a quite different question that requires separate consideration.

[11] This formulation leaves open the question of whether the two diasystems in a diglossia

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

need be "the same language" in some sense. According to Model II, this should not be necessary for a language situation to exhibit diglossic variation; however, Model I specifically excludes this possibility. This is an empirical issue which could be addressed by examining other language situations, using a methodology similar to that employed here.

REFERENCES

Bickerton, Derek (1973). The structure of polylectal grammars. In Roger Shuy (ed.), *Sociolinguistics: Current trends and prospects* (Georgetown University Roundtable on Languages and Linguistics 1972), 17–42. Washington, DC: Georgetown University Press.

Britto, Francis (1986). *Diglossia: A study of the theory, with application to Tamil*. Washington, DC: Georgetown University Press.

_____ (1988). A diglossic typology. *Sophia Linguistica: Working Papers in Linguistics* 23/24:253–62. Tokyo: Sophia University.

_____ (1991). Tamil diglossia. An interpretation. In Hudson 1991: 60–84.

DeCamp, David (1971). Toward a generative analysis of a post-creole speech continuum. In Dell Hymes (ed.), *Pidginization and creolization of languages*, 349–70. Cambridge & New York: Cambridge University Press.

De Silva, M. W. S. (1967). Effects of purism on the evolution of written language. *Linguistics* 36:5–17.

_____ (1974). Convergence in diglossia: The Sinhalese situation. *International Journal of Dravidian Linguistics* 3:60–91.

_____ (1976). Diglossia and literacy. Mysore, India: Central Institute of Indian Languages.

_____ (1979). *Sinhalese and other island languages in South Asia*. Tubingen: Narr.

_____ (1986). Typology of diglossia and its implications for literacy. In Krishnamurti et al. 1986: 304–11.

Fairbanks, Gordon; Gair, James W., & De Silva, M. W. S. (1968). *Colloquial Sinhalese (Sinhala)*, vols. 1–2. Ithaca, NY: South Asia Program, Cornell University.

Fasold, Ralph (1984). *The sociolinguistics of society*. Oxford: Blackwell.

Ferguson, Charles (1959). Diglossia. *Word* 15:325–40. [Reprinted in Pier Paolo Giglioli (ed.), *Language in social context*, 232–51. Harmondsworth, UK: Penguin, 1972.]

_____ (1991). Epilogue: Diglossia revisited. In Hudson 1991: 214–34.

Fishman, Joshua (1967). Bilingualism with and without diglossia, diglossia with and without bilingualism. *Journal of Social Issues* 23:29–38.

_____ (1968), ed. *Readings in the sociology of language*. The Hague: Mouton.

_____ (1972). *Language in sociocultural change*. Stanford, CA: Stanford University Press.

Gair, James W. (1968). Sinhalese diglossia. *Anthropological Linguistics* 10:1–15.

_____ (1986). Sinhala diglossia revisited, or diglossia dies hard. In Krishnamurti et al., 1986: 322–36.

_____ (1992). AGR, INFL, case and Sinhala diglossia, or: Can linguistic theory find a home in variety? In Braj Kachru et al. (eds.), *Dimensions of South Asia as a sociolinguistic area: Papers in memory of Gerald B. Kelley*, 179–98. New Delhi: Oxford & IBH.

_____, & Karunatilaka, W. S. (1974). *Literary Sinhala*. Ithaca, NY: South Asia Program, Cornell University.

_____, & Paolillo, John C. (1988). Sinhala non-verbal sentences and argument structure. *Cornell Working Papers in Linguistics* 8:39–78.

Gal, Susan (1979). *Language shift: Social determinants of linguistic change in bilingual Austria*. New York: Academic Press.

Gregory, Michael (1967). Aspects of varieties of differentiation. *Journal of Linguistics* 3:177–97.

Herring, Susan, & Paolillo, John C. (1995). Focus position in SOV languages. In Pamela Downing & Michael Noonan (eds.), *Word order in discourse*, 163–98. Amsterdam: Benjamins.

Hudson, Alan (1991), ed. *Studies in diglossia*. (*Southwest Journal of Linguistics*, 10:1.) Denton, TX: Linguistic Association of the Southwest.

Krishnamurti, Bh.; Masica, Colin P.; & Sinha, Anjani K. (1986), eds. *South Asian languages. Structure, convergence, and diglossia*. Delhi: Motilal Banarsidass.

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

Labov, William (1994). *Principles of linguistic change, I: Internal factors.* Oxford: Blackwell.
Paolillo, John C. (1991). Sinhala diglossia and the theory of Government and Binding. In Hudson 1991:41–59.
_____ (1992). *Functional articulation in diglossia: A case study of grammatical and social correspondences in Sinhala.* Ann Arbor: University Microfilms.
_____ (1994). Finding the "two" in diglossia. *University of Texas at Arlington Working Papers in Linguistics* 1:15–31.
Rickford, John (1981). A variable rule for a creole continuum. In David Sankoff & Henrietta Cedergren (eds.), *Variation omnibus*, 201–08. Edmonton, Canada: Linguistic Research.
_____ (1987). *Dimensions of a creole continuum: History, tests and linguistic analysis of Guyanese Creole.* Stanford, CA: Stanford University Press.
Rubin, Joan (1968). Bilingual usage in Paraguay. In Fishman 1968:512–30.
Sankoff, David (1988). Sociolinguistics and syntactic variation. In Frederick J. Newmeyer (ed.), *Linguistics, the Cambridge Survey, IV: Language, the socio-cultural context*, 140–61. Cambridge & New York: Cambridge University Press.
Valdman, Albert (1988). Diglossia and language conflict in Haiti. *International Journal of the Sociology of Language* 71:67–80.
Weinreich, Uriel (1968). Is structural dialectology possible? *Word* 14:388–400. Reprinted in Fishman 1968:305–20.

APPENDIX: TEXT SEGMENTS INCLUDED IN THE CORPUS

1 An article from a journal on cultural and linguistic studies.
2 An article summary abstract from the same journal as above.
3.1 Radio advertisements dramatizing real-life situations.
3.2 Radio advertisements with non-dramatic content.
4.1 Quoted monolog from a second-grade Buddhist moral primer.
4.2 Quoted conversational portions of stories in the above textbook.
4.3 Instructional exercises of the above textbook.
4.4 Narrative portions of the above textbook.
5.1 Quoted dialogue from a sixth-grade Buddhist moral primer.
5.2 Comprehension exercises for the stories in the above textbook.
5.3 Narrative portions of the above textbook.
6.1 Author's foreword in a book of reminiscences.
6.2 Publisher's foreword to the above book.
6.3 The narrative portion of the above book.
7 Coverleaf from the above book on the history of Buddhism.
8 Monthly publication on (Buddhist) doctrine.
9.1 Introduction to "The Prosperous Path of Divinity."
9.2 Foreword to "The Prosperous Path of Divinity."
9.3 "The Prosperous Path of Divinity" (scholarly pamphlet on morality).
10.1 Conversational exchanges in a popular children's book.
10.2 Narrative portions of the above book.
11.1 Conversational exchanges in a first-grade children's book.
11.2 A quoted inscription in the above book.
11.3 Narrative portions of the above book.
12.1 Conversational exchanges in a fourth-grade story book.

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press

12.2 Narrative portions of the above book.
12.3 Author's preface to the above book.
13.1 Conversational exchanges in a sixth-grade story book.
13.2 Narrative portions of the above book.
14.1 Conversation from a story in "Grade-School Buddhist Doctrine."
14.2 Narrative portions of the above text.
15 A monthly statement form from the People's Bank.
16 A subscription form for an academic journal.
17.1 Conversation in a traditional story told orally by a monk.
17.2 Narrative portions of the above story.
18.1 Conversation in a traditional story told orally by another monk.
18.2 Narrative portions of the above story.
19 A personal letter by a young Sinhala woman.
20 A personal letter by another young Sinhala woman.
21 The introductory portion of a university lecture on historical linguistics.
22.1 Quoted speech of various participants in TV news stories.
22.2 Descriptive and narrative portions of the TV news.
23.1 Conversational exchanges in a "psychological narrative poem."
23.2 Introduction to the above poem.
23.3 Narrative and descriptive verses from the above poem.
23.4 The publisher's preface to the above poem.
24 A transcript of a public lecture on Marxism and socialism.
25 Publicly posted signs and billboards.
26 Political graffiti on the Kelaniya University campus.
27 Transcript of an oral funerary sermon.
28 Excerpt of a speech by Sri Lanka President J. R. Jayawardhana.
29 An excerpt of the word-by-word gloss of a Pāli religious text.

https://doi.org/10.1017/S0047404500020935 Published online by Cambridge University Press