Finding the “Two” in Diglossia

John C. Paolillo

1. Introduction*

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 (1959) classic definition 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, more so than a dialect in relation to its standard language. However there is little agreement on this point, and different researchers give different characterizations of how divergent H and L must be. Four broadly characterizable positions can be cited, as listed below:

\textit{H and L are structurally related but distinct}

1.a H and L 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)

1.b H and L are “optimally” related varieties of the same language. (Britto 1985)

\textit{H and L are distinct and need not be structurally related}

2.a H and L may be two different dialects or languages. Structural relatedness is only important to the extent that there is a distinction between H and L. (Fasold 1984)

2.b H and L are two more or less distinct languages, dialects or styles used for different communicative purposes. (Fishman 1967, 1972)

* An earlier version of this paper was presented at the 22nd annual meeting of the Linguistic Association of the Southwest, University of Texas at Arlington, October 24, 1993. This research was supported in part by NSF grant BNS-9013361.
H and L are structurally related along a continuum

3. A range of mixtures of H and L blurs the distinction between them, creating a continuum with two extreme poles. (De Silva 1974a)

H and L are structurally related but distinct in grammar only

4. H and L are distinct grammatical systems. Lexical code-mixing superficially obscures this distinction. (Gair 1992)

These four characterizations are important because they illustrate the difficulties inherent in trying to distinguish among different types of language situations, of which diglossia is only one. Ferguson’s characterization (1.a) is founded on the assertion that diglossia is distinct from other types of language situations, particularly standard-with-dialects and societal bilingualism. Fishman’s characterization (2.b) is rooted in the observation of similarities among all three situation types. De Silva’s characterization (3) invokes similarities with (post-)creole continua, based on the existence of varieties of language intermediate between H and L, while Gair (4) proposes a weakening of Ferguson’s characterization, in an attempt to maintain the distinctness of a diglossic type.

The theoretical status of intermediate varieties of language in a diglossia is crucial to sorting out what sort of relationship exists between H and L. If an intermediate variety were to constitute a distinct canonical variety, a “third pole” distinct from H and L (Ferguson 1991), then the essential “two-ness” of Ferguson’s (1959, 1991) characterization would break down. Likewise, if H and L were extremes on a uniform continuum of varieties, and were not clearly distinct, the “two-ness” characterization would again break down.

A peculiar fact about research on diglossia is that while its two-ness is a central theoretical claim, and while this claim is not universally accepted, no one appears to have ever developed an empirical procedure for testing this claim. Research on diglossia typically employs the casual observations of the sociolinguist, who impressionistically makes judgments about the relatedness or divergence of linguistic systems. Claims about one, two, three, or uncountable numbers of canonical varieties are simply asserted for others to accept, without further justification. The two-ness of diglossia is therefore something which needs to be empirically established.

In this paper, I examine empirical evidence for two-ness in Sinhala diglossia based on a study of 52 Sinhala text fragments. I conclude that a two-ness does exist in Sinhala diglossia, but that a modified view is necessary. The view I adopt is one where H and L are seen as systems of correspondence among the varieties of language in a diglossia; I adapt Britto’s (1986, 1991) notion of H and L “diasystems” for the purpose of describing these correspondences.
2. Sinhala diglossia

The Sri Lankan Sinhala language situation is one example of diglossia for which there are different assessments of its two-ness. Gair (1968, 1986) describes Sinhala diglossia in terms of Ferguson’s nine defining characteristics (or “rubrics”, Fasold 1984). Each of these rubrics is discussed below.

2.1 Function

*Spoken Sinhala*, the variety used by Sinhala speakers for everyday conversation, is the L variety. *Literary Sinhala*, the variety used for most written communication, is the H variety. The distribution of functions reported in Gair (1968, 1986) closely parallels the distribution predicted by Ferguson (1959), as can be seen in Table 1. The only 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 of these functions, we find what Gair (1968) calls a formal version of Spoken Sinhala, characterized by a mixture of H and L features. The two-ness of the Sinhala situation depends on the nature of this “Formal Spoken” — is it a version of Spoken Sinhala, a version of Literary, or does it have its own unique properties?

Table 1. Distribution of functions in Sinhala diglossia (source: Gair 1986)

<table>
<thead>
<tr>
<th>Function</th>
<th>Sinhala Variety</th>
<th>Predicted Variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions to servants, waiters, workmen, clerks</td>
<td>Spoken</td>
<td>L</td>
</tr>
<tr>
<td>Conversation with family, friends, colleagues</td>
<td>Spoken</td>
<td>L</td>
</tr>
<tr>
<td>Radio “soap opera”</td>
<td>Spoken</td>
<td>L</td>
</tr>
<tr>
<td>Caption on political cartoon</td>
<td>Spoken</td>
<td>L</td>
</tr>
<tr>
<td>Folk literature</td>
<td>Spoken</td>
<td>L</td>
</tr>
<tr>
<td>Sermon in church, mosque, temple</td>
<td>Spoken (formal)</td>
<td>H</td>
</tr>
<tr>
<td>University lecture</td>
<td>Spoken (formal)</td>
<td>H</td>
</tr>
<tr>
<td>Speech in parliament, political speech</td>
<td>Spoken (formal)</td>
<td>H (or Literary)1</td>
</tr>
<tr>
<td>Personal letter</td>
<td>Spoken (formal)</td>
<td>H (or Literary)</td>
</tr>
<tr>
<td>Novels (conversational parts)</td>
<td>Spoken</td>
<td>*</td>
</tr>
<tr>
<td>(non-conversational parts)</td>
<td>Literary</td>
<td>*</td>
</tr>
<tr>
<td>News broadcast</td>
<td>Literary</td>
<td>H</td>
</tr>
<tr>
<td>Newspaper editorial, news story, picture caption</td>
<td>Literary</td>
<td>H</td>
</tr>
<tr>
<td>Poetry</td>
<td>Literary</td>
<td>H</td>
</tr>
<tr>
<td>Government documents, forms</td>
<td>Literary</td>
<td>*</td>
</tr>
<tr>
<td>Airline announcements</td>
<td>Literary</td>
<td>*</td>
</tr>
</tbody>
</table>

* = not in Ferguson (1959)

1 Gair (1986) notes that Literary is used in personal letters and in parliamentary speeches in “special circumstances”. He does not detail what these circumstances are.
2.2 *Prestige, literary heritage, acquisition, and standardization*

The circumstances referred to by these four rubrics are closely related. Literary Sinhala is the prestige variety of Sinhala, and it has a broad body of literature extending back several centuries. In contrast, there is virtually no literature in Spoken Sinhala. Literary is acquired only through formal education, while the only vehicle for learning Spoken Sinhala (within the context of Sinhala society) is naturalistic acquisition. Finally, Literary appears to be more standardized than Spoken.

2.3 *Stability*

Sinhala diglossia shows no signs of weakening or dying out. According to Gair (1986), the stability of Sinhala diglossia can be explained by the relative absence of various “diglossia-threatening factors”. In short, Sinhala exhibits the stability that Ferguson regards as characteristic of diglossia.

2.4 *Phonology*

Ferguson (1959) notes that H and L phonology are not equally divergent in all diglossic situations. The phonological differences between H and L in Sinhala appear to be among the less divergent, in contrast with other diglossic situations of the region (e.g., Tamil, as described in Britto 1986). The primary differences are that Literary has a number of segments borrowed from Sanskrit, including aspirate consonants, which are absent in Spoken, and that certain vowel reduction processes operate in Spoken but not in Literary (Gair 1968).

2.5 *Grammar*

Ferguson’s characterization of diglossia requires H and L to be structurally distinct. Ferguson (1959) 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 structural distinction 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 (...)
- Sinhalese that does not show Literary main verb forms and the attendant subject-verb agreement is *Spoken Sinhalese*. (Gair 1968:10)

---

2 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 (1st–3rd grades).
The main effort of Gair (1968), however, is to confront the fact that Ferguson’s “two-variety scheme requires some modification when the full range of current materials is taken into account”. As already noted, informal writing (e.g., personal letters) and formal speaking are less readily characterized as Literary or Spoken. Gair characterizes these varieties as sub-varieties of Spoken:

Spoken Sinhalese has two main sub-varieties. One, Formal Spoken, is characterized by the employment 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. (Gair 1968:10)

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 from Ferguson’s model, but the two-ness of the model is preserved, as are the observations concerning prestige, literary heritage, stability, etc. Gair’s proposal amounts to a claim about where the two-ness is found in Sinhala: if H and L are delimited according to the finite verb forms, then the two-variety model applies quite neatly. Grammatical features other than agreement are more variable in their distribution over communicative functions; in a given context, they may be found to a greater or lesser degree depending on other factors. In particular, “public presentation tends to bring forth formal characteristics” (1986:329).

Alternative interpretations of the Sinhala situation are also available. One interpretation might propose a three-variety model for Sinhala: Literary, Formal Spoken, and Colloquial. Another interpretation calls into question the discreteness of H and L. De Silva (1979) has characterized Sinhala in exactly the latter way. While De Silva (1967, 1974a, 1976) describes Literary and (Spoken) Colloquial Sinhala very much as Gair does, he differs in seeing 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 [sic], and therefore it would be unreal to speak of any one formal variety that is emerging. (1979:40)

Under De Silva’s view, there are no sharp structural discontinuities between the varieties Literary, Formal Spoken and Colloquial; there are only varying degrees of formality and varying degrees of hybridization between classical Literary and Colloquial.

---

3 In referencing De Silva’s description, I use the terms Literary and Colloquial, rather than 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 (and their shared properties and structure). Thus the terms “Spoken” and “Colloquial” are not equivalent.
2.6 Lexicon

Although Gair’s and De Silva’s views of Sinhala diglossia appear to be incompatible, Gair (1992) notes that they are intended to address two fundamentally different observations. Gair (1968) is primarily concerned with case morphosyntax and subject-verb agreement — the grammatical components of language structure. De Silva’s (1979) arguments are based primarily on the choice among paired H and Spoken lexical items — the lexical component of language structure.

De Silva (1974b, 1979) reports specifically 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. (1974b, cited in De Silva 1979:41)

From this, De Silva concludes that the principal differences between Literary and Colloquial are lexical in nature, and that the grammatical properties of Literary are “redundancies” (1979:40).

These two different lines of argument emphasize two different characteristics of diglossia — two different rubrics, grammar and lexicon, both identified and discussed in Ferguson (1959). But Ferguson’s description implies that the grammatical and lexical properties of a diglossia should co-vary. Perhaps they do not strictly co-vary in Sinhala, creating a problem for the two-variety model.

Another major issue is the status of intermediate varieties in diglossia — an issue only briefly addressed in Ferguson (1959). Ferguson (1991) comments somewhat more specifically:

I recognized the existence of intermediate forms and mentioned them briefly in my article, but I felt then and still feel that in the diglossia case the analyst finds two poles in terms of which the intermediate varieties can be described; there is no third pole. (1991:226)

Thus, in Sinhala we are left with something of a paradox: if we consider the grammatical aspects of language structure to be primary to the description of diglossia, and if we then follow Gair (1968) in treating agreement as the primary structural difference between H and L in Sinhala, Sinhala fits a two-variety model. We are left to explain, however, the more continuous variability of H/L lexical choice, and the relationship of Formal Spoken Sinhala to H and L. Conversely, if we follow De Silva’s lead in accepting the continuous character of lexical variation as characteristic of Sinhala diglossia as a whole, the two-variety model must be abandoned, at least for Sinhala. Both proposals modify
Ferguson’s original model of diglossia, although Gair’s retains the notion of two-ness, while De Silva’s calls it into question.

3. An approach to the problem

What is needed to resolve the problems in the description of Sinhala diglossia is an empirical test for its two-ness. This requires detailed observation of a number of different varieties of language. The varieties selected should represent as accurately as possible the complete spectrum of linguistic possibilities for Sinhala, both in terms of communicative function and in terms of linguistic structure. We therefore need a corpus of texts classified independently according to structural and functional criteria. Evidence of two-ness (or lack thereof) should emerge from comparison of the structural differences among varieties. The clustering of the more and less formal communicative functions that characterize H and L should be readily apparent.

The selection of a representative sample of texts requires a certain amount of attention. In this study, I was limited to a range of texts that I had brought with me from Sri Lanka, and those that were otherwise available in this country. From these materials I selected a number of texts, both oral and written, that would present the most balanced perspective possible. I was careful to select (i) texts representing communicative functions not represented elsewhere in my sample, and (ii) texts which exhibited combinations of grammatical features I had not otherwise encountered. In this manner I assembled a corpus of 29 texts on which to base my analysis. Following selection, each of the texts was “segmented” into units based on the communicative functions represented in it. The principal purpose of this was to separate out conversation from the surrounding text in 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 1). I thus segmented the 29 texts into 52 text segments (see Appendix A), which were then analyzed independently.

Each of the 52 text segments was coded for the occurrence of 16 different grammatical features, listed in table 2. The 16 grammatical features chosen for observation had all been observed to be present predominantly in either Literary or Colloquial Sinhala — that is, they are markers of the diglossic levels of Sinhala. Only presence or absence of each feature was recorded; frequency information was not gathered.

In order to address systematically the role of the lexicon in Sinhala diglossia, it would be necessary to code the distribution of individual lexical items. I have not attempted this in the present preliminary analysis, and therefore the inferences we can draw from this analysis about the role of lexical...

---

4 The corpus is that of Paolillo (1992).
5 Because of the wide range of text segment lengths, the frequency information that could be gathered would not be equally reliable for each of the text segments.
code-mixing in Sinhala are limited. However, since Literary is reported to make extensive use of borrowed Sanskrit words, I coded my corpus for this feature, using “1” when Sanskrit borrowings were present and “0” when they were not.

Table 2. Grammatical features of Sinhala diglossia

<table>
<thead>
<tr>
<th>Category</th>
<th>Literary</th>
<th>Colloquial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite verbal endings</td>
<td>(1) Subject-verb agreement</td>
<td>(2) Non-agreeing finite verb</td>
</tr>
<tr>
<td>Verbal adjectives</td>
<td>(3) Verbal adjective formed on a Classical Sinhala pattern</td>
<td>(4) Verbal adjective formed on past participle stem</td>
</tr>
<tr>
<td>Infinitives</td>
<td>(5) -iimata dat. verbal nouns</td>
<td>(7) Infinitive in -nna</td>
</tr>
<tr>
<td></td>
<td>(6) -nnata dat. verbal nouns</td>
<td></td>
</tr>
<tr>
<td>Manner nominalizations</td>
<td>(8) Manner nominalizations</td>
<td>(9) Manner nominalizations</td>
</tr>
<tr>
<td></td>
<td>a) ayuru</td>
<td>a) widi</td>
</tr>
<tr>
<td></td>
<td>b) lesa</td>
<td>b) heeti</td>
</tr>
<tr>
<td></td>
<td>c) pinisa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d) paridi</td>
<td></td>
</tr>
<tr>
<td>Syntactic operations</td>
<td>(10) Passivization</td>
<td>(11) Presentative postposing after finite verb</td>
</tr>
<tr>
<td>Copula</td>
<td>(12) Lexical copula in</td>
<td>[no copula]</td>
</tr>
<tr>
<td></td>
<td>a) finite clauses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b) non-finite clauses</td>
<td></td>
</tr>
<tr>
<td>Verbal nouns</td>
<td>(13) Verbal noun in -iima</td>
<td></td>
</tr>
<tr>
<td>Case</td>
<td>(14) accusative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(15) locative</td>
<td></td>
</tr>
<tr>
<td>Lexicon</td>
<td>(16) Sanskrit lexis</td>
<td></td>
</tr>
</tbody>
</table>

Each text was coded with a 16-digit binary number, with each digit representing a single feature. The resulting 16-bit numbers were then plotted as points in a two-dimensional projection of a 16-dimensional space, using a HyperCard stack designed for the purpose. Each feature was assigned an independent axis having only two positions, one assigned to the “0” value and the other to the “1” value. The distance between 1 and 0 values on all axes was held to a constant value. This scheme defines a projection of a 16-dimensional space having 65,536 locations. The 2-dimensional location of each data point was calculated by summing the displacements of each of the 16 feature values.

---

6 This results in an isometric projection. All projections, including isometric projections, produce some distortions in the representation of lengths, depending on orientation.

7 Many of these locations are plotted “on top” of other locations, simply because it was not feasible to manipulate the geometry (the lengths and angles of the plotting axes) to prevent this from occurring. The result of plotting all 65536 locations in 2 dimensions would be a figure with a higher density of points in the middle of a roughly circular area.
for that point. Line segments were then added to represent the relationships between each of the points and its nearest neighbors. The resulting plot appears as Figure 1.

Figure 1. Structural proximity relationships among 52 actual text segments

For readability, the plot in Figure 1 was enhanced in a number of ways. The two large circles near the top and the bottom of the diagram represent the Colloquial and Literary “prototype” locations, i.e., the locations consisting of all and only the features purported to be characteristic of Colloquial and Literary, respectively. The points having “1” values among only the 5 “Colloquial” features are plotted as triangles (△, near the top of the diagram), and those having only “Literary” feature values are plotted with small circles (○, mostly along the left edge of the diagram). The points plotted as small diamonds (◇) represent segments having both Literary and Colloquial features. The segments connecting nearest neighbors were also modified to represent the relative distance (in feature values) between the connected points. The dark, wide lines connect points that differ by a single feature value. The gray lines connect points differing by 2 feature values. Points connected with thin black lines differ by 3 or more feature values.

For a diagram generated in this manner, we could have anticipated a configuration of points different from the plot in Figure 1. If there were a more random relationship between the features chosen, the resulting plot might be more like Figure 2, a similar plot of 52 points of randomly-generated data.
A number of contrasts can be seen in comparing these two diagrams. First, neither of the two prototype locations is occupied in the random data of Figure 2, although both are in the actual data in Figure 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. Since this probability is so small, we can conclude that there must indeed be something special about the two prototype clusters of features in Sinhala.

In addition, there is only one data point in Figure 2 that has only Literary features, as opposed to 9 with only Literary features and 6 with only Colloquial features in Figure 1. The probability of obtaining one of the points with only Literary features is about 1 in 32 (since there are only 5 Colloquial features, all of which must be 0), of the points with only Colloquial features about 1 in 2,048. The likelihood of finding one of the Literary-only points in a sample of 52 text segments is thus reasonably good (better than 1), but that of finding 9 such points in the same size sample is vanishingly small.

One can also observe that the tightest concentration of points in Figure 1 — those differing by only single feature values — occurs around the Colloquial prototype in Figure 1, with a smaller cluster around the Literary prototype. The probability of having one such segment in this size data set is about 1 in 3, and the probability of having as many as 6 such segments is about 1 in 43,046,721. The center of the diagram in Figure 1 is also relatively diffuse when compared with the tightly clustered points in Figure 2. All of these observations about the
distribution of the data points merely confirm that the features chosen are not randomly distributed — they have significant patterns of co-variation, as suggested by their classification as features of Literary or Colloquial Sinhala.

The kind of co-variation present in the Sinhala corpus is not the only kind that could exist. Another common relationship among features found in linguistic systems is the implicational hierarchy, as often found in creole studies (e.g., DeCamp 1971; Bickerton 1973). Figure 3 shows another plot, this time of an implicational hierarchy with a scalability factor of 97%, in which the implicational ordering of the features was arbitrarily chosen. The resulting graph has a strikingly different appearance from the other two. First, there is a very visible “spine” representing the different “mesolectal” varieties projected by the hierarchy. Branching off of the spine are a number of “twigs”, points closely clustered around the spine. 

![Figure 3. Structural relationships among 52 points in an arbitrary implicational hierarchy](image)

---

8 In this particular graph, there is a sharp bend in the spine, at the point where the hierarchical order transverses from Colloquial to Literary features. This is an artifact of the order chosen and the plotting angles used. For consistency, the same plotting angles were used in this diagram as for Figures 1 and 2.
There are two general points illustrated by these diagrams. First, the distribution of the points representing the actual Sinhala text samples is clearly non-random. Although it is not immediately obvious at first, there is a pattern to the distribution of the points. Second, the distribution of points does not resemble a linear hierarchy very well. What emerges most clearly is that there is a highly significant clustering of features in two regions of the graph, and a relatively sparsely populated intermediate region — in spite of the fact that sampling practices favored intermediate varieties. This pair of clusters — one with exclusively Colloquial features and the other with exclusively Literary features — is the “two” in Sinhala diglossia.

4. “Diasystems” vs. “varieties”

We have seen that there is a two-ness to Sinhala diglossia, but that this two-ness cannot be characterized in absolute terms. While the most significant concentrations of samples are at the Literary and Colloquial poles, some mixing of features does occur. 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 (1967) notion of “diatypes”.

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)

This innovation explicitly recognizes the non-homogeneous nature of the H and L systems, but without yet defining diasystem precisely. There are two criteria that pertain to the identification of diasystems. The first identifies 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 could not be used to identify diasystems of a diglossia without circularity. The second criterion is Britto’s notion of “optimal distance”. Following Fasold, Britto 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 (Fasold 1984:40). Further,

---

9 A ‘diatype’ or ‘diatypical variety’ is “any variety distinguished according to characteristics of use” (Gregory 1967:194, cited in Britto 1986:305).
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)

Still, the notion of “optimality” described in this way is not very precise. 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”:

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

• 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” conditions, in that they are implicit in Britto’s characterization: the functionally-defined diasystems would not make sense unless the functional characteristics used identified varieties that corresponded in form as well. The clarifying conditions are also clarifying in another way: we can use them to eliminate a potential terminological confusion of Britto’s notion of “diasystem” with Weinreich’s earlier use of the term to mean 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 re-name Britto’s sense “functional diasystem” to reflect the fact that it is characterized in terms of communicative function in addition to shared linguistic structure. 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. This entire linguistic system can be
viewed as a diasystem with special characteristics. Doing so has the advantage that it captures the systemic complementarity and unity of H and L in diglossia. A definition of diglossia along these lines is as follows:

- **DIGLOSSIA** is a linguistic diasystem consisting of two functional diasystems.

Returning to Figures 1-3, we can readily see that neither Figure 2 nor 3 has paired diasystem-like structural relations within it. Real language systems with similar structural relations between varieties would lack the functional two-ness of diglossia as well. Figure 1, on the other hand, shows (as well as we can expect) a clustering of structural similarities around two definite types, representing two diasystems: Literary and Colloquial. Even intermediate varieties are more closely related to one of these two types than to the other, or than to any third such pole. The two-ness of Sinhala diglossia can thus be observed in the plot of Figure 1.

5. Conclusion

We have seen through the analysis of a corpus of Sinhala texts that the 16 grammatical features from Table 2 fall into two predominant clusters. These clusters can be identified with the two-ness of Sinhala diglossia (cf. Gair 1968, 1992). Note however that the analysis presented here is based principally on morphological and syntactic characteristics, and does not address the more difficult problem of the distribution of lexical items. It is yet possible that there is a continuum of lexical substitution, as suggested by De Silva (1979). However, we find no support for such a continuum among the morphological and syntactic features examined.

Thus the characterization of Sinhala diglossia which comports best with the observations made here is one in which there are two largely discrete functional diasystems. This characterization also allows us to account for the observed variation in distribution of H and L features. Finally, we have seen that direct examination of the structural characteristics of language samples can lead to a clear diagnosis of the diasystemic relations among them, providing a more empirical test than impressionistic subjective observation for two-ness in diglossia or other language situations.

*John C. Paolillo is Assistant Professor of Linguistics at the University of Texas at Arlington.*

References

Bickerton, Derek

Britto, Francis

De Silva, M.W.S.
1976 Diglossia and Literacy. Mysore: Central Institute of Indian Languages.
1979 Sinhalese and Other Island Languages in South Asia. Tübingen: Gunther Narr Verlag.

DeCamp, David

Fasold, Ralph

Ferguson, Charles A.

Fishman, Joshua A.

Gair, James W.
“Sinhala diglossia revisited, or diglossia dies hard.” In Bh. Krishnamurti et al. (eds.), 322-336.


Gregory, Michael

Krishnamurti, Bh., Colin P. Masica and Anjamni K. Sinha (eds.)

Paolillo, John C.

Weinreich, Uriel
## Appendix A: 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 monologue from a 2nd 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 6th grade Buddhist moral primer.
5.2 Comprehension exercises for the stories in the above textbook.
5.3 Retellings of stories of the Buddha’s life and Jataka stories.
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.
8.1 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).
10.2 Narrative portions of the above 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.
12.2 Narrative portions of the above book.
12.3 Author’s preface to the above 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 20th century “psychological narrative poem”.
23.2 Introduction to the above poem.
23.3 Narrative and descriptive verses from the same poem.
23.4 The publisher’s preface to the above poem.
25. Publicly posted signs and billboards.
26. Kelaniya University graffiti.
27. Transcript of an oral funerary sermon.
29. An excerpt of the word-by-word gloss of a Pāli religious text.