Introduction

The 3,000 or so entries in Mauro Fernández’s bibliography on diglossia (1993) might well lead one to the conclusion that everything that needed to be said on the subject had already been said. Yet, forty years after the publication of the late Charles Ferguson’s historic paper on diglossia (Ferguson 1959), a coherent and generally accepted theory of diglossia remains to be formulated. In his last major statement on the subject, Ferguson himself was moved to comment on how research over the course of the preceding three decades had failed to adopt a comparative or typological approach to the study of diglossia, having been pre-occupied instead with individual case studies focused largely upon the issue of whether a particular language situation was or was not an instance of diglossia (1991: 219). This assessment of the situation may perhaps be seen as a particular example of Chambers’s criticism of sociolinguistic research more generally, to the effect that “the social science content of sociolinguistics has overshadowed its theoretical implications” (1995: 11).

It is abundantly clear that Ferguson’s original contribution was intended to reach beyond the purely descriptive and classificatory levels of analysis to the theoretical. The abstract notion of diglossia is derived from the specifics of four prototypical cases, and its synchronic, typological aspect is intimately related to its diachronic, evolutionary one. In Ferguson’s own words, “what I tried to do was to characterize this situation in the hopes that here we would have discovered one possible element in a general typology of socio-linguistic situations” (1963a: 163). This typology was to have led to a “set of principles or frame of reference,” in terms of which patterns of language use and the evolution of verbal repertoires might be conceptualized (1991: 215). Again in Ferguson’s words, “My goals, in ascending order were: clear case, taxonomy, principles, theory” (1991: 215).
Consideration of a score or more of sociolinguistic situations that bear some significant resemblance to those discussed by Ferguson, and in particular to the Arabic, Greek, and Swiss German cases, suggests that these situations may form a coherent, bounded, and theoretically motivated subset of all possible sociolinguistic situations, the shared social and linguistic characteristics of which may be causally attributable to common underlying sociohistorical circumstances. In terms of their synchronic characteristics as well as what may be common patterns in the natural histories of these situations, diglossic situations may usefully be compared to and contrasted with other instances of standards-dialects or with instances of societal bilingualism in the most generally understood use of these terms. The resulting classification is not as categorical as we might like; there are examples of societal bilingualism that bear some resemblance to diglossic situations, just as there are, or have been, examples of diglossia that, in certain aspects of their social evolution, resemble societal bilingualism. Finally, and as Ferguson himself has noted (1991: 219), it becomes clear upon a close examination of a number of case studies of diglossia that this subset itself is not completely homogeneous, and that even the most unequivocal cases of diglossia, if such there be, present contrasts in detail that are at times considerable, yet always instructive for a theory of diglossia and for sociolinguistic theory more generally.

These caveats aside, one of the purposes of this study, in furtherance of Ferguson’s interest in “the sources and outcomes of different language situations” (1991: 223), is to attempt to draw a workable distinction between the phenomenon of diglossia in the strict sense of the term and that of diglossia in some broader sense, and to motivate this distinction theoretically in terms of the sociogenesis of each, the course of development of each, and the long-term consequences of each for the verbal repertoires that they constitute, in whole or in part. I will argue, in the end, for restricting the scope of the term “diglossia” essentially to that first envisioned by Charles Ferguson in 1959, on the grounds that diglossia and societal bilingualism, two major types of sociolinguistic arrangement often regarded as surface variants of the same underlying phenomenon, are, in fact, fundamentally different in their social origins, evolutionary courses of development, and resolutions over the long term, and, furthermore, that inclusion of these two phenomena under a single rubric obscures rather than clarifies sociolinguistic theory.

There are, in fact, two principal reasons for attempting to forge the distinction between the type of sociolinguistic situation described by Ferguson and its counterpart in societal bilingualism. The first of these concerns the potential relationship between the nature of the
complementary functional distribution of linguistic varieties, on the one hand, and either stability or direction of displacement of codes in the event of shift, on the other hand. The second concerns the relationship between linguistic form, or the formal structure of linguistic repertoires, on the one hand, and language function, as well as change in language function, on the other hand.

**Functional complementarity of codes, acquisition, and stability**

The Arabic, Greek, and Swiss German situations are clearly distinct from typical cases of standard-with-dialects and from societal bilingualism in that the former are instances of unalloyed register variation, which is to say, linguistic variation that is stratified by context of use only and not by the social identity of the user (Halliday 1968). What may have been left implicit in Ferguson’s first formulation of diglossia is stated with absolute clarity in his retrospect on the subject: “If we assume that there are two basic dimensions of variation in language, dialect variation correlating with the place of the speaker in the community and register variation correlating with occasions of use, then the H and L varieties of diglossias are register variants, not dialect variants” (Ferguson 1991: 222). Such variation has also been characterized as diatypic, rather than dialectal, or as use-oriented, versus user-oriented (Britto 1986: 37–39). In other words, stratification of variation in diglossia should then show sensitivity to differences in situational context without much, or indeed any, sensitivity to differences in social class (see Bell 1984: 153, figure 3c; Walters 1996: 175, figure 2d).

The fact that in Ferguson’s canonical cases no section of the community uses the H variety for ordinary conversation is arguably “the most important factor in a diglossic situation and one that makes for relative stability” (Keller 1982: 90). Taking German-speaking Switzerland as prototype, no native Schwyzertüütsch speaker of any social background would ever use Standard German for personal, informal interaction with another. Students of the Swiss case have insisted, for example, “dass das Verhältnis von Hochdeutsch und Mundarten nicht ... ein soziolektales Verhältnis ist” (Häcki Buhofer and Burger 1993: 21), “dass in der deutschen Schweiz Dialektsprechen gerade nicht schichtspezifisch [ist]” (Werlen 1993: 9), and, finally, “[dass] alle Einheimischen aller sozialen Gruppen Dialekt sprechen, der Arbeiter wie der Hochschulprofessor, der Bauer wie der Bundesrat” (Trüb 1992: 21). As a result, there exists no reference group within the speech community that uses H for conversational purposes, and therefore no social motivation
whatsoever for others to imitate this practice in their everyday conversation. Parents never speak H to each other (Keller 1982: 91), children have no opportunity to acquire H as their native variety (Keller 1982: 91), and, as Ferguson himself has pointed out (1959: 331), no one is ever comfortable in H to the extent that they are in L. As the younger generation reaches child-bearing age, the cycle repeats itself. L maintains itself in this case, due to the fact that the only native speakers of H are members of another speech community (or tourists or immigrants therefrom) that do not serve as social role models for the diglossic speech communities of German-speaking Switzerland. In the case of Arabic, too, most, if not all, speakers of H and L select between the elevated and vernacular speech styles based upon situational context rather than social status of the speaker (Kaye 1994: 58), and “even Classical Arabic literature and grammar professors … go home and speak their colloquial dialects with their children, families, and friends” (Kaye 1994: 60). As Keller (1982) correctly notes, this is a critical factor in the definition of diglossia and in differentiating it, therefore, from other types of sociolinguistic situations such as the better-known cases of societal bilingualism. The general principle for sociolinguistic theory that emerges from contrasting the two types of situation seems to be that the social motivation that drives shift from a lower-status language or variety to a higher one derives from the prestige accorded those speakers who use the higher variety as a vernacular, and not from the prestige of the social contexts in which H is employed, provided the same norms of functional allocation obtain throughout the entire speech community.

It should also be recorded that, just as the realization of H versus L does not vary as a function of the social identity of the speaker, H as opposed to L may not be exchanged asymmetrically between interlocutors in order to signal social distance or inequality of social standing in a given interactive event, as may, for instance, traditional Javanese address styles. In particular, the latter have been differentiated quite explicitly from diglossia by Errington, who notes that “Javanese address styles can be exchanged symmetrically like L and H varieties [in diglossia], but, unlike L and H varieties, they can also be exchanged asymmetrically” (1991: 200), and that this distinction reflects a “fundamental interactional difference” between honorific language use and the use of H and L in diglossia (1991: 203). Others would seem to take issue with this position, as, for instance does Schiffman (1997: 213), when he claims that “the use of L may be an expression of solidarity and may not be offered to speakers whose social position is superior or distant” or that “H may be the only variety appropriate in a given situation because the use of L would imply a solidarity that is reserved only for members of a particular in-group.”
Ferguson himself, indeed, has envisaged the possibility that in certain “idealized” speech communities, where everyone knows both varieties, “choice of the H or L variety can be used in the same way that the choice of pronouns of address is used in other communities … [and that] there could be times when it would be appropriate for one person to use the H variety and for the addressee to use the L variety to mark power relationships” (Ferguson 1991: 228). Switzerland seems close to just such an idealized situation, yet nonreciprocal use of H would be considered nothing short of condescending, if not absurd. In this connection, it has been observed that “it is psychologically impossible for any two Swiss of any class or occupation ever to address each other privately in anything but the ‘Low’ variety” (Keller 1982: 91). Exceptions to the reciprocity principle may be the use by Sanskrit dramatists of H as the language of kings and priests, and L as the language of the common people, although in real life “everyone spoke in one or another L variety and used H only for religious, literary, and official documents and certain public occasions” (Ferguson 1991: 228; see also Lee 1986: 151). Similarly, “in the early twentieth century an Arab playwright (Mikhail Nu’aimeh) … had the educated people speak H and the less educated people speak L, although in real life everyone spoke L in the situations presented in his dramas” (Ferguson 1991: 228). It seems clear enough, however, that in both of these examples, the nonreciprocal use of H by upper-class characters, and of L by lower-class characters, is a dramatic device that does not accord with contemporary social reality, and that it is, in fact, a type of metaphorical switching that transposes the horizontal, situational stratification of H and L into a vertical one intended to highlight class differentiation.

Social-class stratification of the use of H and L (or its absence) must be distinguished from social stratification of the opportunity to acquire H, and from social stratification of access to those social contexts in which H is appropriate — a distinction not explicitly recognized by Ferguson in his original contribution (but see Ferguson 1991: 227). The typical case, indeed, seems to be that “la diglossie ne touche pas toujours les mêmes couches de la population au même moment” (Mackey 1989: 16), and that only “some speakers possess the requisite linguistic versatility to be able to use H under one set of conditions and L under another” (Ferguson 1959: 325). As Schiffman has noted, “in some linguistic cultures, all speakers exhibit diglossic behavior (i.e., use both H and L varieties in complementary distribution), while in others, only some members of the society do” (1997: 212), and “in many diglossic situations, only a minority or elite control the H domain successfully” (1997: 206). It goes without saying that whereas social access to the
informal situations in which L is appropriate is universal, access to those more formal situations in which H is appropriate is asymmetrically distributed in favor of those educationally privileged, literate, or otherwise specialized classes in society most likely to have had the opportunity to acquire H formally. It is only in this sense that Ferguson’s position can be reconciled with that of Henry and Renée Kahane, who view diglossic systems as “the linguistic representation of a class system,” where “H, the prestige language, is used by, and therefore becomes the mark of, a sector of society which excels through power, education, manners, and/or heritage,” and where “L, the everyday idiom, is the language of the others, and is often used by H speakers in their non-H roles” (Kahane and Kahane 1979: 183). Britto has aptly formulated this distinction as follows: “Total superposition does not imply that every member of a diglossic community knows H and uses H, but merely that there is no portion of the community which actually knows H and uses H without also knowing L and using L” (1991: 61).

In sum, diglossia, in its ideal form, may be conceived of as the quintessential example of linguistic variation where linguistic realization as opposed to language acquisition — here, grossly oversimplified, the use of H or L — is a function solely of social context, and not of social identity of the speaker. In diglossia, it is context, not class, or other group membership, that controls use. Sociolinguistic situations therefore may be compared to each other in terms of the degree to which the variation between two or more alternants in their respective code matrices is determined by social context, as opposed to the social identity of the speaker, and in situations such as those described for Switzerland, Greece, most Arabic-speaking countries, and numerous other cases, the bulk, if not all, of the variance in the use of H and L appears to be explained by situational context.

Communicative systems may also be compared with regard to the degree of functional compartmentalization of their linguistic alternants, and it is clear that Ferguson’s view of diglossia is one where the elevated and vernacular codes are in relatively sharp complementary distribution (Ferguson 1959: 328). Others have proposed that Ferguson’s version of diglossia is simply the extreme end of a continuum ranging from “rigid diglossia,” where “there is minimal functional overlapping between the codes,” and “fluid diglossia,” where “several functions are less rigidly attached to a particular code” (Pauwels 1986: 15). It is a commonplace that in every speech community, some degree of differential functional allocation of linguistic varieties is to be expected, whether categorical, as in the prototypical diglossic case, or statistical, and viewed from a surface synchronic perspective alone, diglossia may not appear to differ, except
in degree, from other instances of situational alternation either between two languages or between two varieties of the same language.

What distinguishes diglossia from other instances of interlingual or intralingual situational alternation — even more than the sharp complementary distribution of linguistic varieties per se — is that the functional distribution of codes in the diglossic case is one that specifically protects the role of the L variety as a natively learned variety. Like standard varieties generally, H “is not ‘native’ to anyone, being a higher cultural endowment with functions that cannot be mastered until after the period of normal first-language acquisition” (Joseph 1987: 17). Emphatically, then, diglossia is a special case of sharp functional differentiation of registers in which the H variety, or set of varieties, is nobody’s mother tongue. According to Ure, diglossia is a special case of separate sets of registers “in which the marked set ... is not the mother tongue of any members of the community” (1982: 16). Coulmas too has noted that diglossia generally “is characterized by a long-term coexistence of native spoken varieties of a language, on the one hand, and a written variety which is the mother tongue of nobody, on the other” (1987: 117). Stated differently, since adults “invariably use L in speaking to children and children use L in speaking to one another, ... L is invariably learned by children in what may be regarded as the ‘normal’ way of learning one’s mother tongue,” and the learning of H “is chiefly accomplished by the means of formal education” (Ferguson 1959: 331). Finally, as Schiffman observes, “In diglossia no one speaks the H variety as a mother tongue, only the L variety. In the standard with dialects situation, some speakers speak H as a mother tongue, while others speak L varieties as a mother tongue and acquire H as a second system” (1997: 207). Given the express, widely held view that only L is acquired as the natural mother tongue in a diglossic speech community, it is remarkable that time after time in the sociolinguistic literature this critical feature of diglossia is disregarded, as, for instance, in the case of Paraguay, where Spanish and Guarani are in fact the mother tongues of two distinct segments of the community.

It is scarcely imaginable, as Ferguson himself has noted, that any change toward full utilization of H could take place without the willingness on the part of adults to speak H for conversational purposes, especially with their children (Ferguson 1959: 331). True, in other instances of language or dialect contact, the vernacular role of the L variety is also relatively protected in the initial stages, but what further distinguishes diglossia from these latter is the absence of any prestige group that employs H as its vernacular and could, therefore, provide the social impulse for shift away from L as the vernacular. Language shift,
like internal language change, requires social motivation, but, as noted above, this motivation derives from the adoption as role models of high-prestige individuals who themselves already employ H as a native, vernacular variety, and not from its association with prestigious, high-culture domains of interaction. In the absence of such a reference group, and given the functional protection afforded L as a natively acquired variety, stability in that role is to be expected. Short of stability, then shift from H to L in formal domains of interaction, or shift from H to a new vernacular-based standard, is to be anticipated, as opposed to shift from L to H in informal domains.

It was Ferguson himself who pointed out that the strictly complementary functional allocation of the elevated and vernacular varieties in the Arabic, Greek, Haitian, and Swiss German cases had existed for periods of no less than several centuries in each case. From this observation, it has been tempting to conclude that situations like these are inherently stable, and therefore that the attainment and maintenance of sharp functional compartmentalization of codes may be the key to minority-language maintenance and even reversal of language shift within multilingual or multilectal speech communities (see Fishman 1991; Paulston 1994). To argue this is in essence to beg the question, however, since the means whereby such functional compartmentalization — entailing the eradication of the “H” variety from all domains of personal, informal interaction and therefore from all first-language-acquisition contexts — might be motivated, attained, and maintained are not suggested by the model. Moreover, stability is a relative matter, and the fact remains that Colloquial Arabic, Dhimotiki, and Schwyzertütsch have all gained ground in the twentieth century, to one degree or another, at the expense of their coterritorial elevated counterparts. It is clear, then, that functional compartmentalization at a given point in time is in and of itself no warranty against eventual shift. Whether the long-term maintenance of a vernacular variety necessarily follows from functional compartmentalization of codes, or whether its displacement by a more prestigious rival follows from the absence of such functional compartmentalization, is far from obvious. Nonetheless, the fact that in the cases just mentioned, as in numerous others, it is precisely the elevated or culturally prestigious variety, and not the vernacular, that has been displaced to some extent is reason enough to distinguish these from other cases where it is typically the vernacular that has been displaced, often to the point of extinction, by a prestigious competitor.

Although most often described, and problematically so, in terms of synchronic functional distribution of linguistic varieties, it is the diachronic aspect of diglossia — its sociogenesis, course of development
across time, and ultimate resolution — that is of greatest significance where sociolinguistic theory is concerned. The variables discussed by Ferguson may in fact be viewed as contextual, linguistic, or temporal in nature. Function, prestige, acquisition, literary heritage, and standardization are all in one way or another contextual aspects of diglossia. Grammar, lexicon, and phonology are linguistic aspects. Stability, clearly, refers to the temporal axis along which the various contextual and linguistic variables may vary. In the main, scholars to date have treated Ferguson’s nine characteristics of diglossia as a potentially open-ended checklist of features or parameters of a sociolinguistic typology or have tended to focus their attentions on one feature only, or a smaller number of such features, most notably the functional compartmentalization of codes and the structural distance between codes. The position advanced here, however, is (1) that Ferguson’s features are not an open-ended list, but rather a closed, tightly interdependent set of features, (2) that functional compartmentalization, acquisition, and stability of L as vernacular are causally related, and (3) that the abstract social factor underlying these latter three features is in turn related in a nonrandom way to the structural relationships between the diglossic codes.

Structural relationships between codes in diglossia

The four cases described by Ferguson are remarkable for the fact that the elevated and vernacular varieties in each bear a close structural resemblance to one another — not so close as to be readily mutually intelligible, but also not so distant as to be unhesitatingly regarded as separate languages. While acknowledging the difficulty, on the strength of linguistic criteria alone, of identifying different language varieties as varieties of the same or of different languages, and setting aside too the vexed question of the nature of the creole–standard relationship in the Haitian situation, it is hardly contestable that the linguistic affinities between Classical or Modern Standard Arabic and Colloquial Arabic, between Standard German and Schwyzertüütsch, or between Katharevousa and Dhimotíκι, are of an order quite apart from that of the affinities between, say, Spanish and Guarani in Paraguay, Spanish and Nahuatl in Mexico, or even Spanish and English in the American Southwest. Other pairs, such as Dutch and Frisian in the Netherlands, French and Provençal in France, and Catalán and Castillian in Spain, may, admittedly, be more problematic to place linguistically between the extremes of linguistic affinity, but the question that nonetheless presents itself is whether proximity of linguistic relationship between varieties,
however measured, is causally associated with the potential for
displacement of one by the other.

Ferguson sets his initial discussion of diglossia solidly within the
context of language situations where “two or more varieties of the same
language are used by some speakers under different conditions” (1959: 325
[emphasis added]) and further asserts that “no attempt is made … to
examine the analogous situation where two distinct (related or unrelated)
languages are used side by side throughout a speech community, each
with a clearly defined role” (1959: 325, note 2). He later recognizes
explicitly the difficulty of measuring linguistic distance and of identifying
two varieties as members of the same or different languages (Ferguson
1991: 220). Elsewhere, too, he adverts to the possibility that some future
sociolinguistic typology might extend the term diglossia to include certain
situations where the codes in question might be structurally unrelated
(Ferguson 1963b: 174), and, more particularly, to the possibility that the
relationship between Spanish and Guaraní in Paraguay might also
reasonably be described as diglossic (Ferguson 1973: 41–42). Never-
theless, his intention, by his own admission, had been “that the users
would always view the two [varieties] as the same language” and that
“diglossia” should not extend to cases “where superposed on an ordinary
conversational language is a totally unrelated language used for formal
purposes, as in the often-cited case of Spanish and Guaraní in Paraguay”
(Ferguson 1991: 223).

Over time, opinion with regard to the theoretical significance of the
relatedness between codes in diglossia has varied in the extreme. Some
scholars, following Ferguson, have maintained the distinction between
related and unrelated language varieties. In proposing the term
“diglossia” for situations where two relatively closely related varieties
are in complementary functional distribution, and the term “biglossia”
for circumstances where the varieties in question are varieties of different
languages, Fellman has implied a fundamental distinction between the two
(1975a: 39). More recently, Daltas too has proposed that “Fergusonian
diglossia should be preserved as a useful notion to refer to a specific type
of variation involving an H and an L variety of (what is societally seen as)
the same language” (1993: 348).

Others have taken a view of the relationship between codes in
diglossia as being intermediate between that involved in multilingual
situations, on the one hand, and stylistic shifting on the other (Daltas
1993: 341). An important issue here is whether structural intermediacy
in this sense is directly or indirectly asserted to connote sociological
intermediacy as well. Wexler, for instance, seems not to be concerned
with the latter, but rather, notwithstanding the methodological dilemmas
involved, with establishing the minimum distance between codes in
diglossia, so that the latter may be distinguished from more common
cases of standards-with-dialects by the existence of “a broad structural
gap between the standardized written norm and the unstandardized (as
a rule) spoken dialects” (1971: 336). Coulmas, likewise, assumes “that a
large structural gap between the spoken norm and the written norm of
a speech community is the most salient feature of diglossia,” although
“it is far from obvious how the distance between spoken and written norm
can be measured, and how the gap in one language can be compared with
that in another” (1991: 126).

For other scholars, too, the difference between diglossia and societal
bilingualism appears largely to reduce to the degree of linguistic difference
between varieties in the code repertoire. Fasold, for instance, views the
diglossia of the Arabic-speaking world, Greece, and German-speaking
Switzerland as “a convenient midpoint in the possible range of relatedness
to be found in broad diglossia” (1984: 54), thus distinguishing the former
from so-called superposed bilingualism, where the prestige and vernacular
varieties are more distantly related to each other, and from style-shifting,
where the two varieties are more related structurally. Such a categori-
zation, however, is at one time overinclusive and underinclusive. It is
overinclusive in that it admits to the category of diglossia sociolinguistic
situations such as Dutch–Frisian in the Netherlands and French–
Provençal in France, neither of which, in all probability, properly
belongs. Likewise, it excludes situations such as Yiddish–Hebrew in
Jewish communities in Europe prior to World War II and among certain
groups of Ultra-Orthodox Jews in Israel today, all of which deserve
serious consideration as candidates for the designation of “diglossia.” The
difficulty with such approaches to the identification of diglossia is that
they focus almost exclusively upon structural relationships between codes
in the community verbal repertoire at the expense of social and functional
distribution.

Along much the same lines, Britto regards the intermediate distance
presumed to exist between H and L in diglossia as “optimal,” so that these
linguistic varieties may be distinguished “from ‘languages’ (which would
be Super-Optimal varieties), and “styles,” “accents,” etc. (which would be
Sub-Optimal varieties)” (1986: 10). Recognizing that “it is not always
easy to judge when a pair of varieties are optimally distant” (Britto 1986:
19) and that “there are at present no adequate techniques to measure it
[optimal distance] empirically” (1986: 22), Britto proposes nevertheless
that “if such clear norms are established to identify the Optimal vari-
eties from other varieties which are too closely or too distantly related,
then it may be possible to apply these norms to a given situation and
evaluate whether the situation is diglossic or not diglossic” (1986: 20). This reasoning borders on tautology, in that if there are no empirical measures by which the structural relationship between two codes may be identified as “optimal” other than that the sociolinguistic situation in which two codes coexist is already known to be diglossic, then it is clear that optimal distance between codes cannot be used as a diagnostic in sociolinguistic situations, the status of which is not known by independent criteria. The best that can be said of this approach is that it identifies the linguistic distance between codes in the Arabic, Greek, Swiss German, and Tamil cases as prototypical for diglossia, assuming them to be commensurate in the first instance. In the end, establishing a certain distance between codes as criterial for diglossia proves no less of a conundrum, and perhaps no more theoretically relevant, than ascertaining whether two linguistic varieties should be identified as the “same” language. Perhaps all that can be said, then, along with Ferguson, is that “in diglossia there are always extensive differences between the grammatical structures of H and L” (Ferguson 1959: 333 [emphasis in original]).

Other attempts to formulate the relevant linguistic questions expose a lingering preoccupation with structural distance as a defining characteristic of diglossia. Winford, for example, has asked, “What are the outer limits of the structural gap, beyond which a situation becomes bilingual rather than diglossic in Ferguson’s sense” and “what is the minimal structural gap between varieties that will qualify a situation for diglossic status?” (1985: 351). Ironically, even Ferguson himself regarded “the failure to make clear how far apart (or how close together) the high and low varieties have to be for a language situation to be characterized as diglossia” as a shortcoming of his original article (1991: 223). Paolillo, more recently, has proposed that examination of a variety of different sociolinguistic situations “would allow us to establish what degree of difference constitutes ... the structural difference between diasystems necessary and/or sufficient for a situation to be diglossic” (1997: 292). Upon reflection, these turn out to be curious, linguistically deterministic ways of framing the problem, implying either that diglossia is arbitrarily defined by what appears as theoretically unmotivated reference to surface grammatical characteristics of the diglossic codes, or, alternatively, that structural discrepancies of a given magnitude between the elevated and vernacular varieties in diglossia could be held to account theoretically for the remaining features of diglossia, particularly those concerning functional complementarity of codes, the protection of L as the sole variety acquired as a native variety, and the stability of diglossic situations or anticipated direction of shift.
upon their dissolution. This last proposition seems on the face of it implausible, casting linguistic structure as the principal determinant of diglossia.

At the other extreme, Joshua Fishman has implicitly dismissed the degree of structural proximity between codes as irrelevant to the definition of diglossia, asserting that "diglossia exists not only in multilingual societies which officially recognize several ‘languages’ but, also, in societies which are multilingual in the sense that they employ separate dialects, registers or functionally differentiated language varieties of whatever kind" (1967: 30), and provided too that "without schooling the written/formal-spoken [variety] cannot be understood" by speakers of the vernacular (1980: 4). This view is also championed by Pauwels, who employs the term diglossia for functional differentiation of codes within speech repertoires generally but distinguishes between interlingual diglossia, where the codes are not varieties of the same language, and intralingual diglossia, where they are (1986: 15). This is as much as to suggest, for example, that the asymmetric functional allocation of Nahuatl and Spanish in Mexico is to be distinguished from that of Katharevousa and Dhimotiki in Greece primarily on the basis of the structural discrepancies between the codes in each case. Clearly the two situations are worlds apart, however, not only in their surface linguistic dimensions, but, more significantly, in their sociohistorical origins, evolutionary courses of development, and ultimate resolutions.

Fishman was among the first to propose that the role of linguistic relatedness between codes in diglossia be minimized, on the grounds that it is social consensus rather than structural affinity that confers the status of distinct languages on two or more linguistic varieties (1967: 33, note 4). A similar theme has been sounded more recently by Berger, who adds that "individual languages which are clearly distinguishable from a structural point of view are not necessarily distinguishable for the speaker," and, conversely, that "speakers may consider their idiom as a separate language although this is not tenable from a linguistic point of view" (1990: 290). Mackey is almost certainly correct in his suggestion that the theoretical difficulty inherent in establishing language identity has been responsible for the extension of the notion of diglossia to include any set of linguistic varieties, related or not, and that this extension in turn has resulted in the massive multiplication of the number of situations throughout the world that might be called diglossic (1986: 239).

Fishman’s original extension of the scope of diglossia may also be considered in the light of an early thrust in sociolinguistic thought toward the theoretical integration of dialect variation, diglossia, and societal bilingualism as surface variants of the same underlying phenomenon.
John Gumperz had argued that the distinction between bilingualism and bidialectalism was not always a significant one in the social characterization of speech communities (1968 [1962]: 463), and that linguistic difference need not necessarily correspond to difference in social function (1968 [1962]: 464). Fishman himself felt that a “single theory … enabling us to understand, predict and interrelate both of these phenomena [diglossia and bilingualism] is an instance of enviable parsimony in the behavioral sciences” (1967: 32–33), and, in the same volume, Dell Hymes took the position that “cases of bilingualism par excellence (as for example French and English in Canada, Welsh and English in Wales, Russian and French among pre-revolutionary Russian nobility) are salient, special cases of the general phenomenon of variety in code repertoire and switching among codes” (1967: 9). Later still, Sankoff states that the shifting or switching among various codes typical of individuals in multilingual societies “does not differ qualitatively from the behaviour of monolinguals (shifting of style or level)” (1972: 33). Upon reflection, however, it is not at all clear that the French–English situation in Québec and the Welsh–English situation in Wales are of a piece with the Russian–French situation in prerevolutionary Russia, nor that the Nahuatl–Spanish situation in Mexico might not have a great deal more in common with the Occitan–French situation in southern France than either has with the Schwyzertüütsch–German situation in Switzerland. The emphasis on asymmetric functional differentiation of codes as the common denominator of all these situations may, rather than disclosing them to be surface variants of the same social phenomenon, instead have resulted in the surface merger of fundamentally distinct sociolinguistic situations. What is actually called for in the further study of diglossia is the establishment of a balance, and an interaction, between two complementary approaches: one, a universalist approach such as that discussed above, which emphasizes the similarities between diglossia and other sociolinguistic situations, and another, particularistic approach, which focuses upon the differences.

In fact, too much has been made of the degree of structural proximity between constituent codes in a verbal repertoire as a diagnostic of diglossia. Defining the codes in diglossia a priori as varieties of the same language or otherwise is an arbitrary gesture and in itself contributes nothing of value to sociolinguistic theory. As others have noted, the issue is an empirical one, amenable to investigation via comparison of a variety of linguistic situations (Paolillo 1997: 294, note 11). The position taken here is that comparative analysis of a sufficiently broad spectrum of linguistic situations is likely to reveal that the communicative arrangement characterized by Ferguson for Greece, Switzerland, and the Arab
world is a social arrangement more fundamentally than it is a linguistic one, and that there is, therefore, no absolute causal connection between linguistic relatedness of the codes and the direction of eventual or potential shift. This point of view situates diglossia within the more general framework of prestige languages, as this term has been employed by Henry and Renée Kahane, and shares the view expressed by these scholars with regard to structural relatedness between codes: “Genetically ... H may be a variety of L, either similar to it (say, early Medieval Latin vs. early Romance in the Carolingian age); or it may be dissimilar (late Medieval Latin vs. French in the 15th c.); or it may be a truly foreign language to speakers of L (Norman French vs. the English of the Conquest period)” (Kahane and Kahane 1979: 183).

This much said, however, if the structural difference between codes in diglossia is viewed as an outcome of the social circumstances giving rise to diglossia in the first place, rather than as a defining feature of diglossia, there is ample reason to suppose that language varieties in diglossia will in fact show a strong statistical tendency to be varieties of the same language, or, given the difficulties inherent in this notion, to evince structural relationships on the order of those obtaining in the Arabic, Greek, and Swiss-German cases. While it is possible to recall instances where comparable social relationships exist between decidedly distinct languages, such as Yiddish and Hebrew in certain Ultra-Orthodox Jewish communities, this appears to be more the exception than the rule. Rather, it is in the very nature of a diglossic accommodation between codes that the linguistic varieties involved tend more often than not to bear a relatively close linguistic relationship to one another, although the precise nature of this relationship and the specifics of its history may vary from case to case, as will be demonstrated below. Should this be the case, definitions of diglossia that are predicated upon linguistic distance between the constituent codes will prove to be essentially beside the point. The formulation of the problem adopted here is, given a particular pattern of functional complementarity of codes, a particular pattern of first-language (or dialect) acquisition, and a particular diachronic pattern whereby H tends either to converge upon or to be displaced by L, what structural characteristics are to be expected in such a code matrix and what structural characteristics are actually found? The position taken here is that the linguistic arrangement characterized by Ferguson for Greece, Switzerland, and the Arab world is fundamentally a sociological rather than a grammatical one, and that there is, therefore, no direct causal connection between linguistic distance between the component codes in such a code matrix and direction of eventual or potential shift.
While degree of linguistic affinity between codes may not be criterial to the definition of diglossia, it is highly relevant, as Walters (1996: 158) has remarked with respect to Arabic, to inquire as to the nature of the variation found in the preponderance of cases where \( H \) and \( L \) are indeed relatively closely related varieties. Ervin-Tripp has suggested that code use in diglossia may be accompanied “by more cooccurrence restriction than is style shifting where the common features of the styles may outweigh their differences” (1971: 44–45). In cases similar to that of traditional Javanese linguistic etiquette, cooccurrence restrictions appear to be absolute: selection of a given item from one particular lexical set absolutely controls selection of the appropriate item from a second, independent, lexical set (Errington 1991; Geertz 1968), and these patterns of selection, or “linked conjugates” or “stylemes,” as Geertz refers to them (1968: 287), clearly identify the social level of the language variety in question. Ferguson’s claim that the use of the \( H \) or the \( L \) member of a lexical doublet immediately stamps a text as \( H \) or \( L \) suggests that the lexical cooccurrence restrictions in diglossia are more akin in their stringency to those operative in the situation described by Geertz (although the social dynamics involved are very different) than they are, say, to the rather looser cooccurrence restrictions obtaining in English between *illumination*, *purchase*, and *children*, on the one hand, and *light*, *buy*, and *kids*, respectively, on the other (Ferguson 1959: 334). In the case of the English doublets, “both words may be written and both may be used in ordinary conversation: the gap is not so great as for the corresponding doublets in diglossia” (Ferguson 1959: 334). Further, “the formal–informal dimension in languages like English is a continuum in which the boundary between the two items in different pairs may not come at the same point” (Ferguson 1959: 334). Biber (1988) and others have developed a factor-analytic technique for the cross-linguistic study of cooccurrence restrictions in written and spoken language, which, applied to diglossic code matrices, might help to provide a more clinical assessment of the extent and stringency of the cooccurrence restrictions governing the use of \( H \) and \( L \) in various diglossic contexts. To date, the only study of a diglossic code matrix to approximate this approach, though without employing factor analysis as a statistical procedure, has been Paolillo’s (1997) study of continuity versus discreteness in Sinhala diglossia, to be discussed further below.

Where the realization of one variable does not completely control the realization of another, other kinds of cooccurrence constrains are possible. The most obvious alternative to strict cooccurrence restrictions are the implicational scalings of the type proposed by DeCamp for Jamaican Creole English (DeCamp 1971) or by Bickerton for Guyanese Creole
English (Bickerton 1971). It should be noted, however, that Paolillo’s attempt to scale H and L variables in Sinhala has met with only limited success, compared to those reported for various creole situations: out of sixteen Sinhala variables investigated, only in the case of five, all of these within the colloquial cluster of texts, “did anything like a scale emerge” (Paolillo 1997: 287 [no coefficient of scalability given]). On the strength of this, Paolillo concludes that there seems “to be a broader tendency toward covariation of H and L features, with a moderate degree of hybridization being tolerated” (Paolillo 1997: 288). Whether Paolillo’s findings in the case of Sinhala will be confirmed in other cases of diglossia remains to be empirically tested. In any case, it is clearly possible to examine the cooccurrence restrictions between linguistic variants in any given code matrix and to establish whether such restrictions are categorical, implicational, or correlational in nature, and further to explore whether the variation documented in cases of diglossia spans the full range of such possibilities or only a significantly restricted segment of it.

To the extent that more formal and less formal varieties in a code matrix may be clearly distinguished, the general question has been raised whether the more elevated varieties are less tolerant of variation than are the vernacular varieties. In that the H variety or varieties in diglossia are standardized, according to Ferguson, “there is an established norm for pronunciation, grammar, and vocabulary which allows variation only within certain limits” (1959: 332). By contrast, where the L variety or varieties are concerned, “there is no settled orthography and there is wide variation in pronunciation, grammar, and vocabulary” (1959: 332). Substitution or switching has long been described as typical in casual utterance generally (Voegelin 1960: 62), and, conversely, code consistency has been proposed as a universal characteristic of more formal communicative events (Irvine 1984: 215). It may be, therefore, that the H variety or range of varieties in diglossia presents nothing more than an instance of a general principle of relative invariance in formal codes. However, in postdiglossic speech communities, as the standard gradually accommodates to the influence of the vernacular in writing, and where the standard is no one’s native variety, it may be that the relatively “ill-defined” standard evinces a greater degree of variation than the relatively “well-defined” vernacular (Kaye 1970, 1972).

Related to the question of cooccurrence restrictions in diglossia is the question of discreteness as opposed to continuity in linguistic variation. Stated simply, do the individual styles found in a diglossic code matrix, like those found in certain postcreole continua, arrange themselves in a unidimensional series of minimally different varieties, reaching from an extreme basilectal vernacular variety at one pole to an extreme acrolectal
standard at the other? Or is this sequence of minimally different lects punctuated at one or more points along its progression? One view, at least, is that diglossia is “the most striking example” of a divided register range, or of multiple sets of registers within the same language (Ure 1982: 16). Ferguson himself, while conceding that a continuum of variation can be documented in every one of his canonical cases, nonetheless believes 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). Paolillo’s Sinhala data in fact suggest “a broader tendency toward covariation of H and L features” (1997: 288).

Although there is no absolute separation of texts into those characterized exclusively by H features and those characterized exclusively by L features, some 52 text segments group themselves empirically into two broad clusters, such that the texts within each cluster are more closely related to each other than are any two texts from different clusters. These two clusters appear to represent texts in which Literary Sinhala features are relatively prominent, on the one hand, as opposed to those in which Colloquial Sinhala features are relatively prominent, on the other (Paolillo 1997: 288).

Yet another question concerning the grammatical relationships between codes in diglossia is, to the extent that the codes are indeed varieties of the same language in some real or putative sense, does the nature of the grammatical relationships between H and L tend to be consistent across cases? For example, is the morphology of H consistently more complex than L in the operational sense employed by Ferguson, to wit, that morphemes in L have fewer alternants and morphemic alternation is more regular, fewer obligatory categories are morphologically realized, paradigms are more symmetrical, and/or there is stricter concord and rection (1959: 333–334)? The Arabic, Greek, and, arguably, Swiss-German cases seem to be consistent with this hypothesis (Ferguson 1959: 334), but to the extent that this may be generally true of diglossia, how can these relationships be explained? Are they, for example, the normally expected consequence of discourse-pragmatic constraints upon the structure of formal registers (Givón 1979), are they the consequence of universal principles governing the cultural, but pragmatically arbitrary, expression of formality in communicative events generally (Irvine 1984), or, last, are they due to differing rates of change in formal and informal linguistic structures, such that H invariably reflects an earlier form of L, and L, therefore, a later stage upon an expected, perhaps universal, path of grammatical development (Bybee et al. 1994)? Literary Kannada, for example, is viewed by one scholar as “an archaism, a stage which the language reached some centuries ago, when it became
‘frozen’ by social convention’ (Bright 1976: 66). The colloquial dialects, on the other hand, continued to change, with the result that “modern literary Kannada represents, to a large extent, an earlier historical stage of the modern colloquial dialects” (Bright 1976: 66).

**Social origins of diglossia**

Beyond the synchronic description of the social and linguistic dimensions of diglossia, of greater significance to sociolinguistic theory is an understanding of the evolutionary relationship between linguistic form and social function. In diglossia research, if not in sociolinguistic research more generally, it is remarkable how little attention has been paid to the role of macrosocietal and broader cultural processes in the differentiation of various types of verbal repertoires. The position of diglossia within an evolutionary taxonomy of speech repertoires may, perhaps, be gauged with reference to Gumperz’s typology of linguistic communities (Gumperz 1968 [1962]). While even small bands of hunters and gatherers reveal differences between “casual every-day speech and non-casual styles used in singing, recitation, myth-telling and similar ritually defined situations” (Gumperz 1968 [1962]: 466), the opposition, central to diglossia, between a vernacular variety learned in the home and a superposed variety learned after childhood becomes salient in intermediate societies constituted of peasant, herder, or even tribal populations integrated to varying degrees into the dominant society and exhibiting a high degree of social stratification and occupational specialization (Gumperz 1968 [1962]: 467–468). Such societies are likely to develop special sacred and administrative codes, in addition to other special parlances, which are “characterized by extreme codification,” and which require, therefore, “a large investment of time in the study of grammar and rhetoric,” as well as schools, with their complements of scholars, for the pursuit of such study (Gumperz 1968 [1962]: 469). These codes “serve as the language of special administrative and priestly classes” and function, at least in part, “to maintain group exclusiveness” (Gumperz 1968 [1962]: 469). When government “remains in the hands of a small ruling group,” it is possible, even in the face of increasing mobilization of the population, to maintain considerable language distance between the administrative and sacred H-codes, on the one hand, and the rest of the code matrix on the other (Gumperz 1968 [1962]: 469). It may be reasonably hypothesized, therefore, that intermediate societies as described by Gumperz have offered singularly fertile ground for the emergence of classical diglossia.
Gumperz’s intermediate societies resemble quite closely Sjoberg’s “preindustrialized civilized societies,” in which the bulk of the written tradition consists mainly of the society’s sacred writings, and where writing is restricted to, and is perpetuated by, a small, educated, priestly group (Sjoberg 1964: 892). It is this educated elite that, “in formal situations of various kinds, speaks in the traditionally high-status manner,” while “in other situations, especially to communicate effectively with less educated or uneducated persons, they must use a more informal speech style, one that is more akin to the colloquials of the area” (Sjoberg 1964: 893). As a result, “the upper status, educated group typically employs at least two speech styles, in some cases more,” all of which “differ from the speech of the common man — in the lexicon and often the phonology and grammar” (Sjoberg 1964: 893). Furthermore, the formal speech style “tends to be perpetuated over centuries with relatively little change, a phenomenon that results from the high prestige accorded it and its close tie with the written language” (Sjoberg 1964: 893). While neither Gumperz nor Sjoberg explicitly mentions diglossia as such, their social descriptions of “intermediate” and “preindustrialized civilized” societies, and their corresponding verbal repertoires, bear more than a passing resemblance to many cases of diglossia up to recent times.

A distinction of potential significance for evolutionary studies of speech repertoires is that drawn by Besnier between determinants of linguistic form that have their basis in the “physical and psychological characteristics surrounding the situation of use,” on the one hand, and those that have their basis in “the norms of communication at play in each context,” or “the cultural ‘value’ of a communicative context” (1988: 731), on the other hand. It has been asserted that “functional” variation, where the variation is intrinsically related to the means or the purpose of communication, as opposed to the cultural value of a communicative context, is more characteristic of modern societies, while “nonfunctional” variation, in which the choice of variable lexical and syntactic features is not determined by strictly pragmatic demands, is more characteristic of premodern societies (Neustupny 1974: 39–40). In the process of modernization, the nonfunctional opposition between classical and colloquial varieties in premodern diglossic speech communities is replaced by the functional variation involved in an ever more differentiated repertoire of scientific and technical varieties (Neustupny 1974: 40). With particular reference to diglossia, “most of the variation between a Classical and a Modern standard in the case of a premodern diglossia is non-functional in the described sense,” and it is, therefore, “not a matter of chance that the diglossia, a typical case of non-functional
variation, is most often removed at an early stage of modernization” (Neustupny 1974: 40).

Ferguson himself assumes a polygenetic stance when it comes to the emergence and development of diglossia: “Diglossia is not assumed to be a stage which occurs always and only at a certain point in some kind of evolution, ... [but] may develop from various origins and eventuate in different language situations” (Ferguson 1959: 326–327). If, as is argued elsewhere in this paper, the critical distinction between the canonical cases of diglossia and the more typical cases of societal bilingualism indeed resides in the presence or absence of a prestige group of native H-speakers, then it makes sense to look for the social origins of diglossia in circumstances where cultural–linguistic traditions develop or acquire new registers without also acquiring native speakers of these new varieties who might become prestige role models in their host speech communities and thereby provide the social motivation for language or dialect shift. At least three sets of circumstances suggest themselves as having this potential. In the first, the H-variety in question historically has never been used as a vernacular by any native speakers. In the second, the H-variety may have been used as a native vernacular at one time, but the population of speakers so using it has become extinct, whether through the process of intergenerational language shift or internal language change. Third, the H-variety still enjoys a vernacular speaker base in a separate speech community, the members of which do not regard themselves, and are not regarded as, members of the community that employs H for nonvernacular purposes only.

Examples of these three sets of circumstances are not as difficult to come by as might be supposed. Thus, although there is some controversy surrounding the matter, some have contended that Classical Chinese was an artificial language that was never spoken by anybody (Li and Thompson 1982: 84; Rosemont 1974). Likewise, in Japan, kanbun was nobody’s mother tongue, regardless of whether one’s native language was Chinese or Japanese (Coulmas 1991: 130). Modern Standard Arabic has also been characterized as a “somewhat artificial medium” (Kaye 1994: 49), which never had any native speakers (Kaye 1994: 51, 59). The same might also be said of Old Church Slavic, which was never a native vernacular but rather the written version of a Bulgaro-Macedonian dialect of Slavic developed in the ninth century by the Christian missionaries Constantine and Methodius (Comrie 1991: 160). Finally, Katharevousa, the official standard of Greece prior to 1976, “was never a spoken language of any historical period,” but was instead “an artificial compromise between archaism and colloquialism, created in the early nineteenth century and imposed as the official language of the Greek State
in the first Greek constitution of 1834” (Alexiou 1982: 158). While every standard language is of necessity to some extent artificial, the more typical case, unlike diglossia, is that the written and spoken standards are more firmly anchored in actual spoken varieties, often the careful spoken varieties of the educated urban classes.

The second scenario alluded to above, the extinction of the population of speakers upon whose vernacular the contemporary written standard is historically based, in its most familiar guise, is the process whereby change in written language proceeds at a significantly slower pace than change in spoken language. The process in its most general form has been described as follows: “The two varieties of language involved become increasingly dissimilar, the variety of the original body of texts remaining largely unchanged (although likely codified) while the spoken variety, which does not undergo processes of linguistic standardization, continues to change in ‘natural’ ways” (Walters 1996: 161). With particular reference to Egyptian, Pulgram observes that “while the spoken language of the masses goes along its undisturbed linear development, each classical language, from its inception, is held to a level standard, without major changes, as long as the society which employs it remains stable” (1950: 461). Historically, the widening discrepancy between Sanskrit and the Prakrit dialects has been attributed to the fact that, “while Prakrits went on changing, Sanskrit remained unchanging” (Deshpande 1991: 38). Literary Kannada, too, has been described as a frozen, archaic stage of the language reached some centuries ago, while “the colloquial dialects, which are spoken as everyone’s first language, continued to change’ (Bright 1976: 66). In Persia, by the end of the Sassanid Empire in the seventh century, literary Middle Persian “had become markedly distinct from contemporary spoken Persian (dari) in Western Iran” (Jeremiás 1984: 273). In the case of Modern Standard Arabic, “classicisms of all shapes have worked to combine with the forces of linguistic evolution to ... slow down the rate of the linguistic change of this somewhat artificial medium” (Kaye 1994: 49).

A general model of the acceleration of the vernacular away from the standard, punctuated at socially significant intervals by the correction of the latter back toward the former, has been proposed for the historical relationship between spoken and classical Egyptian (Sethe 1925: 316), for written and spoken Latin (Pulgram 1950: 462, and figure 2), and for spoken and various varieties of literary Hebrew (Rendsburg 1990: 31, 174–175). Finally, whereas in the cases just discussed the vernacular might be said to have been changed over time via replacement by its own linguistic descendant, the case of Ge’ez and Amharic in Ethiopia suggests that a vernacular may also be replaced by other than its own descendant
and still result in a case of diglossia. Although Ge’ez was replaced as a vernacular by Amharic and other languages between the ninth and the twelfth centuries (Cooper 1978: 460), it survived as a literary language into the nineteenth century, and, vestigially, as the liturgical language of the Ethiopian Copts up to the present time (Cooper 1978: 461). While Ge’ez and Amharic are related languages within the Ethio-Semitic group, Ge’ez is not a direct linguistic ancestor of Amharic.

A third scenario for the emergence of diglossia is the acquisition of an H-variety from an external speech community, unaccompanied by any significant in-migration of native speakers of H. One of the most interesting of these cases is the adoption of the Chinese language as the means of writing in the earliest documents in Japan, dating from the early seventh century (Coulmas 1991: 129). Although this written variety was subsequently nativized and was ultimately to evolve into the literary Japanese of the nineteenth and early twentieth centuries (Coulmas 1991: 128), it does not seem amiss to regard the differential functional allocation of Chinese to the written sphere and of Japanese to the spoken in the earliest stages of this development as an instance of diglossia involving two unrelated languages.

In many of their particulars, the individual cases of diglossia referred to in the course of this paper differ strikingly from one another. On the surface of things, it might justifiably be asked whether, for example, the opposition between Classical and vernacular Japanese up until the beginning of the twentieth century should be considered of a piece with the Swiss-German diglossia of today, or whether the traditional opposition between literary and liturgical Ge’ez, on the one hand, and vernacular Amharic, on the other, should be considered socio-linguistically akin to that between Katharevousa and Dhimotiki in Greece prior to 1976, or to the centuries-old opposition between Hebrew and Yiddish among Ashkenazic Jews. As Ferguson himself noted, “The four cases I described are not identical; each one is quite different in some respects from the other three, though they have many features in common” (Ferguson 1991: 219). The realization, however, that the essential characteristic of diglossia is the coexistence between an elevated code, which has no native speakers within the speech community in question, and an everyday vernacular introduces a higher order of homogeneity into what on the surface may appear as heterogeneity. This point of view liberates the notion of diglossia from the demand that the two codes in question be varieties of the same language, although for reasons discussed elsewhere, this is to be expected as the usual state of affairs. It further removes the requirement that diglossia be limited to literate speech communities, although, as discussed in the following
paragraphs, the development of writing, when combined with long-term restricted literacy, is particularly conducive to the emergence of diglossia.

The development of writing produces an unprecedented alteration in the communicative ecology of any speech community, and although it clearly does not in every instance lead to the emergence of diglossia, it creates a particularly hospitable environment for it, when combined with other supporting factors. As I have noted elsewhere (Hudson 1994: 308), four principal mechanisms account for the manner in which the development of writing may contribute to the linguistic diversity of a speech community, and therefore to the potential emergence of diglossia.

In the first instance, the sociocultural norms operative in contexts where writing is appropriate commonly dictate that the grammatical structure of written text be less casual and in some sense more elevated than the grammatical structure of spoken utterances. Second, the different pragmatic constraints imposed upon the realizations of written and spoken text by the real-time limitations on production and processing and by the physical, psychological, and social immediacy of the audience determine to some degree the relative frequencies of occurrence of certain grammatical forms and structures in the two modalities. Third, to the extent that the opportunity for the acquisition and use of literacy skills is asymmetrically allocated across the various segments of a speech community for any significant period of time, the grammatical features of written or literate text may come to signal membership in those, typically more privileged, social groups within the community that have access to the occasions on which written or literate text is appropriate. Finally, writing as a medium acts as a fixative agent for linguistic structure. Except for memorized oral text transmitted intact from generation to generation, the structure of oral language is in general much more subject to diachronic change, due to phonetic erosion, than is the structure of written language. To the extent, therefore, that written language in general, or a particular genre of written language, is not directly influenced by developments in the structure of oral language, the potential arises for major disparity between the two and, therefore, for the emergence of diglossia.

Coulmas has pointed to the need for “a sociolinguistic theory of writing and written language which accounts for the nexus between literacy, writing system, and diglossia” (1987: 122). While allowing that writing may be neither a necessary nor a sufficient condition for the emergence of diglossia, he maintains, along with Ferguson, that “writing introduces the possibility of a permanent rift between characteristically different varieties into every speech community” (Coulmas 1987: 122), thereby leading to “potential diglossia in every speech community that becomes
literate” (Coulmas 1987: 114). As Ferguson, too, has noted, graphization, or the reduction of a language to writing, “adds another variety of language to the community’s repertory” (1968: 29), and “communities, as they begin the regular use of writing, generally do not feel that ordinary, everyday speech is appropriate for written use” (1968: 29–30). In nineteenth-century Japan, for example, this view was succinctly expressed as follows by one of the leading intellectuals of the time: “In our letters at present it is inappropriate to write as we speak as well as inappropriate to speak as we write, because the grammars of speech and writing in our tongue are different” (Nishi 1874, quoted in Coulmas 1991: 128). So pervasive was this view that Japanese scholars and government bureaucrats travelling to the Netherlands at the time were surprised to find that the language of writing there was very much closer to the language of everyday conversation than was the case with Japanese, “because the idea of committing anything of serious import to writing in a colloquial variety was alien to them” (Coulmas 1991: 134).

According to Ferguson, diglossia is likely to emerge as a consequence of the long-term monopoly of a small elite on literacy and, therefore, on direct access to the literary heritage of a speech community (1959: 338). In particular, diglossia is likely to come into existence where (1) “there is a sizable body of literature in a language closely related to (or even identical with) the natural language of the community, and this literature embodies ... some of the fundamental values of the community,” (2) “literacy in the community is limited to a small elite,” and (3) “a suitable period of time, on the order of several centuries, passes from the establishment of (1) and (2)” (Ferguson 1959: 338). In Coulmas’s version of this theory, alluded to above, the degree of difference between the written and spoken norms, and therefore the prospect for the emergence of diglossia, is a function of (1) the degree of association between the literary tradition and other great cultural achievements of a religious or artistic nature, (2) the period of time over which the literary tradition flourishes, (3) the extent to which the written language is cultivated by a small caste of scribes and is prevented from adjusting to changes in the spoken language, (4) the literacy rate of the speech community, (5) the acceptance of the written language by the illiterate mass of the population as the only valid manifestation of their language, and (6) the relative fit of writing system and orthography for the language in question (1987: 121–122, 1989: 13). Others, too, have made the connection between the extent of popular literacy and the likely distance between the written and the spoken codes: Parker supposes that “the varying and uncertain distance between spoken language and written language in any time and place ... may be inversely associated with the
level and extent of literacy” (1983: 334–335). As explained by Walters, “the existence of diglossia hinges on a tradition of restricted literacy involving the written variety of a language that becomes increasingly distant (and therefore distinct) from the native variety of language spoken in a speech community that is overwhelmingly illiterate” (Walters 1996: 161–162). While Walters’s account may give the impression that this process is the only one that gives rise to diglossia, Ferguson, as noted elsewhere, claims that “diglossia may develop from various origins” (1959: 327). Of these various origins, however, it is likely that the processes described by Coulmas and Walters are indeed those that most typically give rise to diglossia.

The association between diglossia and literacy, in the sense of both the incidence of individual literacy skills and the existence of a literary tradition, is not an accidental one, since social stratification of literacy may give rise to the independent development of two or more increasingly divergent language varieties. As the linguistic varieties become more divergent, more extensive training is required for the mastery of the literary variety, a development that in turn confers additional prestige upon it: “With the development of writing and a complex and introspective literature, the language variety so employed will often be accorded such high value because of the recorded nature of the medium and the need to be trained to read and write it” (Abrahams 1972: 15). The creation of literary varieties may in turn serve to reinforce the exclusiveness of a class structure based on literacy or, more precisely, on the control of knowledge via literacy. In preindustrial civilized societies, traditional writings “have provided the basis for standardizing thought and action for the literati — who are the leaders in the society — over time and space” (Sjoberg 1964: 893). Thus, “a substantial body of writing furnishes a language with a depersonalized standard that can be fixed and, to some extent, manipulated, that is, cultivated in a conscious way” (Coulmas 1987: 121).

A handful of cases serves to illustrate the intimate relationship between literacy, literary tradition, restricted literacy, and diglossia. Arabic diglossia in large measure has derived its stability from the fact that the H-variety is associated with pre-Islamic poetry, the sacred texts of Islam, later commentary on both of these, and the works of medieval Arabic philologists, among them Sibawaihi of Basra, who based their linguistic analyses on these earlier texts (Rabin 1955: 20; Walters 1996: 161). In the case of Chinese, the linguistic differences between the literati and the general population were probably due to the lengthy accumulation of literary tradition among the former, to the early fixation of the form of writing by the printing press in China (printing from wood
blocks since at least the beginning of the eighth century and from movable type since the eleventh (Morton 1982: 86), and extremely low literacy (DeFrancis 1972: 8). Vernacular Hebrew is generally not attested in the Biblical literature, particularly of the First Temple period, “most likely due to the influence of the prophetic schools ... which acted as a check or brake on the written language” (Fellman 1977: 108). The maintenance of Latin as a prestige language vis-à-vis the emergent Romance vernaculars has been attributed to a “virtual monopoly of knowledge” on the part of the Church, acquired as a result of the emphasis on clerical literacy, monastic control of manuscript production and reproduction, and the decline in literacy among the laity between the fourth and the tenth centuries (Parker 1983: 336–337). In the case of Sanskrit, it was the traditional grammarians’ concern with correct ritual speech that led to the development of grammars such as that written by Panini (Hock and Pandharipande 1978: 21–22), and, in general, over the course of the linguistic history of India, Sanskrit has retained its high status “primarily because literacy was highly restricted to the Brahmanical classes” (Deshpande 1991: 38). Finally, the separation of court and literary Turkish from the vernacular, which began during the sixteenth century, was accelerated by the social cleavage between the vast ruling order composed of bureaucrats, poets, the religious establishment, merchants, and other community leaders on the one hand, and the masses on the other, as well as by the lack of an educational system, which “prevented the dissemination of high class Turkish among the masses” (Karpat 1984: 189).

While it is clear that the development of writing is not in itself a sufficient condition for the emergence of diglossia, the question arises as to whether it is a necessary one. Ferguson himself has suggested the possibility that it is not: “All clearly documented instances [of diglossia] known to me are in literate communities, but it seems at least possible that a somewhat similar situation could exist in a non-literate community where a body of oral literature could play the same role as the body of written literature in the examples cited” (Ferguson 1959: 337, note 18). The distinction between written and spoken language in literate traditions has been recognized as resembling the distinction between the high and the low varieties in diglossia (see Tannen 1982: 15, note 3), and the point has frequently been made also that ritual or artful genres of expression in oral traditions differ in linguistic form and social function from everyday talk in much the same way that written and other noncasual types of discourse differ from colloquial speech in literate cultures (Akinnaso 1982: 8; Chafe 1982: 49–50; Feldman 1991: 47–48). Thus, Akinnaso claims that “in nonliterate (traditional) societies, ritual communication
(for example, ritual chants and divination verses) is different from everyday talk in much the same way that written language differs from ordinary conversational language in literate societies” (1982: 8), or, more explicitly, that “there are no formal, functional, or structural differences between written language in literate societies and oral ritual communication in nonliterate societies” (1982: 29, note 3). Others, such as Chafe (1982: 49–50), have demonstrated that the same features that differentiate spoken and written language may also differentiate colloquial and ritual speech. Also, no less than written codes, “any speaking code used ceremoniously will accumulate the sense of power inherent in the occasions of its use” (Abrahams 1972: 15), and, in terms of functional allocation, opportunity for acquisition, and potential for linguistic divergence from the vernacular, will also offer the prospect of the emergence of diglossia. More than the development of writing as such, then, it seems that an important precursor to diglossia is the existence of a body of literature in the sense of “that body of discourses or texts which, within any society, is considered worthy of dissemination, transmission, and preservation in essentially constant form” (Bright 1982: 272; see also Martin Joos, cited in Voegelin 1960: 60, note 4). Such a body of literature, though typically associated with the written medium, “may also be composed orally, and regularly performed in that same medium” (Bright 1982: 171). The oral literature of ancient India offers a remarkable case in point. The need to preserve the religious effectiveness of the orally composed Vedic hymns by transmission in their exact original form, and the desire to preserve the special status and distinctness of the language of the educated classes, culminated in the Sanskrit grammar of Panini (c. 500 BC), itself in all likelihood an oral composition also (Bright 1982: 273; Hock and Pandharipande 1978: 21–22).

Rise and decline of diglossia

Diglossia has been described by Ferguson as being very stable: “Diglossia typically persists at least several centuries, and evidence in some cases seems to show that it can last well over a thousand years” (Ferguson 1959: 332). Coulmas, too, sees stability as “one of the most remarkable characteristics of diglossia” (1987: 117), particularly where “linguistic differences are not aggravated by political or religious differences” (1987: 118). Others, however, have questioned whether “the stability factor is so crucial to diglossia … since societal changes such as modernisation, urbanisation, the breakdown of rigid class barriers, etc., have made diglossic situations as described by Ferguson (1959) rather
rare” (Pauwels 1986: 16). Finally, as if to proclaim the entire discussion moot, Mackey has declared, “contrairement aux modèles de Ferguson et de Gumperz, les situations diglossiques ne sont pas stables; elles ont chacune leur dynamique” (1989: 16). This small sampling of opinion as to the stability of diglossic situations points as clearly as any discussion could to the general failure of sociolinguistic theory in dealing with diglossia as a distinct communicative arrangement. For further discussion of this and related points, see Hudson-Edwards (1984: 6).

Scholars clearly reach differing conclusions on this subject for a variety of reasons: their characterizations of the facts with which they are dealing differ; they treat different sets of facts as though they were the same; they differ as to their interpretations and definitions of stability; or, finally, they differ as to their understandings of the universe of sociolinguistic situations that might be regarded as diglossic. It is most likely this last, combined perhaps with inevitable differences of opinion as to what in fact constitutes stability in the first place, that accounts for the varying perspectives on the subject of the long-term survivability of diglossic arrangements. This particular conundrum corresponds very closely, in fact, to what Labov refers to as a “paradox of evidence” (1994: 14). Such dilemmas may perhaps be resolved by reexamination of the internal evidence for the validity of the conflicting viewpoints, or by bringing evidence to bear “from neighboring fields of expertise … which may show that the contradictions were really the result of a lack of homogeneity in the object being studied” (Labov 1994: 14). It is past time, indeed, for such a reconciliation of perspectives in order to advance understanding of the essence of diglossia and of its natural course of evolution, and thereby to advance a significant part of sociolinguistic theory more generally.

In the end, however, the notion of stability is at worst ill-conceived and at best relative, and whether diglossic situations are in any sense more stable than other types of sociolinguistic situations, particularly those involving two or more distinct languages, remains to be determined. Some cases of societal bilingualism will almost certainly be found to have endured for longer periods of time than some cases of so-called diglossia, and, on the other hand, comparing instances of long-term diglossia with instances of short-term bilingualism makes no case for the relative stability of diglossia. In the end, too, even the most persistent cases of diglossia eventually show change in the functional distribution of codes, as well as in the linguistic substance of the codes themselves, thus evolving into new and different types of speech economies. Whereas the sociolinguistic situation in German-speaking Switzerland has been described as being “relatively stable” (Keller 1982: 75), at least with regard to the functional status of the two varieties involved if not with regard to their
substance (Keller 1982: 76–77), Arabic diglossia has been described as
“not a relatively stable situation” (Kaye 1970: 391), and, in the case of
Greek, as of the late 1970s at least, “functional and situational differences
between High and Low are far less stable and clear-cut than in Swiss
German or Arabic” (Alexiou 1982: 165).

What seems closer to the theoretical mark than stability, in fact, is
the likely direction of change once the inevitable realignment of codes and
their social functions finally occurs. In cases of societal bilingualism, the
general tendency appears to be for the higher-prestige language eventually
to invade the domain of the home, ultimately displacing the language of
lesser prestige as a first language in the community. This diachronic
process has been documented in strikingly similar terms for Nahuatl–
Spanish language contact by Hill (1983) and for French–Occitan language
contact by Eckert (1980), but also in innumerable other sources for many
other cases throughout the sociolinguistic literature. It is also the situation
described by Fishman under the designation of “bilingualism without
diglossia,” the outcome of which he describes as follows: “Without
separate though complementary norms and values to establish and
maintain functional separation of the speech varieties, that language or
variety which is fortunate enough to be associated with the predominant
drift of social forces tends to displace the other(s)” (1967: 36, see also
1980: 9). In diglossic contexts, on the other hand, it is H that tends to be
displaced by L through a process of structural convergence resulting in
the emergence of a new standard more closely related to certain educated
varieties of the vernacular. On the very face of it, this outcome, which
appears to be nonrandomly, though not categorically, associated with
diglossia in Ferguson’s narrow sense, reveals the fundamental socio-
linguistic distinction between true diglossia situations and situations of
societal bilingualism, including those cases of pseudodiglossia, such as
Catalán–Castillian, Frisian–Dutch, Galician–Portuguese, and Occitan–
French, where the appearance is presented of relatively close structural
affinity between the codes in question.

The potential decline of diglossia may be anticipated by the admission
of the vernacular into domains formerly reserved exclusively for the high
variety, and it is reasonable to ask, therefore, whether the course of
functional expansion of L follows a fixed pattern once instability in the
functional assignments of H and L has set in. Repeatedly, where popular
command of H has declined, this has provided the impetus for the
adoption of L in religious or secular texts, whether written or oral, aimed
primarily at the largely illiterate populace. In India, by the second century
BC, “the Prakrit languages were widely used for secular literature, royal
inscriptions, and especially for the propagation of ... Buddhism and
Jainism” (Deshpande 1991: 24). In China, too, literature intended to entertain or edify the uneducated masses had been written in *gudai baihua*, “a vernacular Chinese close to the spoken language,” since the Tang dynasty (618–907) (Peyraube 1991: 110; see Barnes 1982: 262), while in modern times, a purportedly declassicized Chinese had been mandated for use in political propaganda and other literature intended for popular consumption in the People’s Republic of China (Li and Thompson 1982: 85). Hall reports that the authorities of the Western Church had become so concerned over the failure of the common people to understand the Latin sermons delivered at Mass that the Council of Tours in 813 decreed “that priests should thenceforth translate their sermons in *rusticam romanam linguam aut theotiscam* ‘into the rustic Roman or German tongue’” (1974: 105; see also Parker 1983: 336). As Pulgram aptly states, “what lies behind the reform is a resigned recognition that you cannot talk to people in a language they have long since ceased to understand, nor thereby save their souls” (1950: 461). In thirteenth-century Japan, the use of *hentai kanbun* for official documents during the Kamakura shogunate signalled a declining command of *jun-kanbun* among the military elite (Coulmas 1991: 129–130). In the sixteenth century, European Jesuit missionaries in Ethiopia used Amharic rather than Ge’ez for written purposes and thus were able to reach a great proportion of the Ethiopian populace “for whom Ge’ez remained a sealed priestly mystery” (Fellman 1975b: 180).

Other factors, too, motivate the spread of L. In Tamil, the realistic portrayal of natural conversation between the characters in novels, plays, and films introduced the use of Tamil L into a domain that had formerly been the exclusive preserve of Tamil H (Britto 1991: 69–70), and, as a result of this, Tamil L came to be regarded as appropriate for all monologic discourse, such as sermons, political speeches, and lectures, where the speaker wished to create the impression of “actually conversing with the audience” (Britto 1991: 70). Medieval and modern Romance standard languages, as opposed to Latin, “were typically first used, in the Middle Ages, for compositions which did not threaten established intellectual or professional interests” as, for example, the recitation of epics, the singing of troubadours, and romances, while “the use of the vernacular in fields such as law, medicine or religion... was strongly resisted” (Hall 1978: 111). The spread of vernacular Greek from poetry to nonfictional prose and thence to creative literature more broadly, from the latter to serious critical writing, and thence to scholarly, technical, scientific, and official writing has also been well documented (Alexiou 1982: 159; Mirambel 1964 [cited in Alexiou 1982: 174]).
According to Ferguson, certain social developments betoken the eventual decline of diglossia. These developments may be broadly characterized as tendencies toward mobilization and democratization and are described by Ferguson as follows: (1) more widespread literacy, (2) broader communication among different regional and social segments of the community, and (3) the desire for a standard national language as an attribute of autonomy or sovereignty (1959: 338). Enlarging somewhat upon this list, the decline of diglossia is repeatedly (though not inevitably, as the case of Switzerland clearly indicates) associated with the processes of modernization, urbanization, mercantilism, and industrialization, and the demands that these create for a literate labor force. Related to these forces also are the disestablishment of small ruling groups, the breakdown of rigid class barriers and increased fluidity of role relationships, and the democratization of education, literacy, and knowledge that tend to accompany these. Finally, nationalism, in the form of the desire for autonomy, sovereignty, and political unity, often engenders the desire for, and loyalty to, a standard national language. Such trends help forge “firmer bonds between the spoken and written forms of the language,” so that it becomes “increasingly difficult to separate educated from non-educated speech” (Sjoberg 1964: 894).

Modernization and its various components, culminating in mass education and literacy, have been repeatedly associated with the decline of diglossia. It has been argued, for instance, that the elevation of the status of the vernacular, and the resulting adjustment of the writing system to this, is one of the main types of linguistic reform in modernizing societies (Sjoberg 1964: 897). It has also been suggested, as noted previously, that diglossia “is most often removed at an early stage of modernization” (Neustupny 1974: 40). Further, the urbanization of intermediate societies is reported to result in structural convergence between the subcodes in their speech repertoires, to the point where, in some highly urbanized communities, the discrepancy between standard and local dialects diminishes virtually to the point of extinction (Gumperz 1968 [1962]: 469).

Examples abound. The perceptible growth in urban commercial economic activity by the thirteenth century was accompanied by “a significant increase in the demand for literate individuals to meet the growing requirements of administration, law, and commerce” (Parker 1983: 338), and, in turn, by the emergence of lay schools in many of which “at least some of the instruction was in the vernacular” (Parker 1983: 338–339). In the eighteenth century, partly motivated by the modernizing efforts of Peter the Great (reigned 1682–1725), there emerged a growing desire in intellectual circles to bring the linguistic
situation in Russia more into conformity with that in France, particularly with regard to the perceived uniformity of the written and spoken language, thereby jeopardizing the status of Old Church Slavic (Comrie 1991: 168). In the course of the transformation of Japanese society from a feudal into a capitalist society, the need to translate Western books, the desire to disseminate knowledge broadly throughout the population, and the recognition of the importance to industrial and military development of a unified written standard language that was close to the spoken language all led to a restructuring of the relationship between the written and the spoken language such that diglossia between Classical and colloquial Japanese was essentially eliminated by the end of World War II (Coulmas 1991: 135, 140–141; Neustupny 1974: 36). With the entry of new administrative, legal, educational, and military terms into the Ottoman world during the nineteenth century, “the Ottomans began to realize that their language [Osmanlica] was insufficient to the task of keeping up with European advances” (Gallagher 1971: 161), and that “Ottoman linguistic obscurantism was a barrier to the political reform and increased freedom they were seeking” (Gallagher 1971: 162). Further, a major stimulus for the development of Modern Standard Arabic was the discussion of modern philosophical, literary, and historical topics that took place in various newspapers and journals in Cairo, Beirut, and Baghdad, as Western-influenced Arab intellectuals found Classical Arabic unable to cope with such topics as rationalism, liberalism, socialism, democracy, and nationalism, thus creating the problem of “developing within a few decades … a new and refined Arabic capable of expressing a material and intellectual civilization that had evolved over centuries in Europe” (Abdulaziz 1986: 16).

In conclusion, however, it should be noted that these and/or similar processes, while seemingly contributing to the decline of diglossia, do not always and without exception favor the elevation of the low variety to high status. Thus, the emergence of Standard German at the expense of the local dialects in Germany over a period of more than 500 years may be attributed to the invention of the printing press, Luther’s translation of the Bible into Eastern Middle German, the gradual acceptance of Luther’s German in the Catholic south, political unification in 1871, and universal education in the twentieth century (Mackenzie 1994: 249).

The complex web of processes entailed in the establishment of a new social order inevitably poses a threat to the maintenance of diglossia. It has been claimed that verbal repertoires become more homogeneous and that old administrative codes tend to be replaced by more vernacular varieties as local populations are absorbed into dominant groups and as increasing proportions of the population are drawn into national
life (Gumperz 1968 [1962]: 469). Like prestige languages generally, the H-variety in diglossia “is seriously weakened by a new class structure which gives increased power to groups previously at the margins of, or below the range of, elite society” (Kahane and Kahane 1979: 190). Also like prestige languages, the H-variety “comes in with status and elitism … [and] goes out under the pressures of popular developments and movements which we may call nativist rebellions” (Kahane 1986: 498). In fact, the disestablishment of ruling elites has been portrayed as a necessary condition for the elimination of diglossia: “Unless the ruling class is replaced by another, there is no loosening of the diglossic control” (Sotiropoulos 1982: 19). As class barriers crumble and formerly well-defined social strata are less sharply differentiated, it becomes less clear “who sets the standards for either the written or the spoken forms of the language” (Sjoberg 1964: 894). The general principle appears to be that the code matrix becomes less and less diverse as role distinctness decreases (Gumperz 1968 [1962]: 469), and “that shallow linguistic contrast in styles is a direct correlate of the fluidity of roles symbolized by the distinction between caste and class” (Gumperz 1968 [1962]: 470).

Decline of a classical variety is often accompanied by catastrophic political events involving the breakdown of classical society itself: “A breakdown of this society involves the breakdown of its classical language; and the new socio-historical structure creates a new literary language out of the spoken language then current” (Pulgram 1950: 461–462). In India, between 500 BC and the turn of the millennium, the rise of anti-Vedic religions such as Buddhism and Jainism, both founded by warrior-princes who defied the traditional religious authority of the Brahmins, resulted in the elevation of local vernacular Prakrit languages to the status of languages worthy for religious instruction (Deshpande 1991: 25–26). In the case of Hebrew, it was with the success of the popular-based Maccabaeian revolution in 167 BCE that the vernacular began to rise in importance, later to come into its own with the compilation and codification of the Oral Law in the Mishna and related works, roughly between 100 BCE and 200 AD (Fellman 1977: 108). As described by Rabin, the Maccabaeian revolution “presents all those conditions in which we might expect a change-over of linguistic habits: a protracted period of cultural anarchy and dislocation of education on the one hand, and the rise of a new intellectual elite to take the place of a thoroughly discredited upper class on the other” (Rabin 1958: 156–157 [quoted in Fellman 1977: 108]). In Ethiopia, it was the influence of the clergy that sustained the position of Ge’ez vis-à-vis Amharic; when the clergy lost its political sway, as during the decline of the Axumite
Empire in the tenth century, and again following the political and economic reforms of Menelik II in the late nineteenth and twentieth centuries, Ge’ez lost its superior position relative to Amharic (Fellman 1975b).

In Europe, the decline in the relative importance of the clerical orders in secular administration contributed significantly to the weakening of the Church’s monopoly of knowledge by the fourteenth century, and probably “reduced barriers to, and ultimately increased the demand for, subsequent introduction of the vernaculars in certain spheres of official language use, and thence generally” (Parker 1983: 339). Conversely, among other considerations, the growth of an educated European public, particularly in the fifteenth and sixteenth centuries, “replaced the medieval class system of the learned elite and the non-educated masses, symbolized by the dichotomy of Latin vs. the vernacular, with a system of the new, educated (upper-?) middle class and the non-educated masses, symbolized by the dichotomy of the new standard language vs. the vernacular” (Kahane and Kahane 1979: 188). In Japan, the collapse of the shogunate in the 1860s and the subsequent opening of the Japanese ports led to the influx of cultural influence from the West and thus provided the initial stimulus for the ultimate unification of the written and spoken languages (Coulmas 1991: 134–135). In the case of Chinese language reform, it was only when the old social order of the Qing Dynasty began to crumble that the Chinese script was subjected to critical scrutiny (DeFrancis 1972: 10), and the eventual victory of the vernacular baihua movement appears to have been associated with the replacement of the traditional scholar-bureaucracy by a republican form of government in 1911 (Barnes 1982: 261). The decline of Katharevousa in Greece, foreshadowed in the rise of “a new and pragmatic middle class” toward the end of the nineteenth century, was to reach its climax, after more than a century of political vicissitudes, with the ouster of the military junta and the restoration of official democracy in 1974, and with the reestablishment of Dhimotiki as the national language two years later (Kahane and Kahane 1979: 190). The emergence of Educated Standard Arabic has been attributed to the rise of a new elite with different communicative needs from those of the 1950s and 1960s, who rejected Modern Standard Arabic “either because they did not master it, or because it was not modernised enough to adequately express the technological and educational advances that … [had] shaped their lives” (Mahmoud 1986: 245).

The various processes of modernization as well as the replacement of older elites or traditional social structures by newer ones often converge, at least in modern times, upon the phenomenon of democratization, in its narrower political sense as well as in the sense of the democratization of
education, literacy, and knowledge. Thus, the language unification, or *genbun itchi*, movement in Japan was not only stimulated by the collapse of the shogunate and by the modernization of Japanese society, as noted above, but, as an aspect of the latter, coincided with the people’s rights movement (*minken jiyū undō*) and the movement for a popularly elected assembly (*minsen gin undō*). As a result, “the new colloquial standard which the proponents of *genbun itchi* created and which the society as a whole accepted, ... is to be understood as the linguistic counterpart of the societal changes heralded by these movements” (Coulmas 1991: 139). Within the realms of education and literacy more specifically, the higher-than-average rates of both in Japan have “turned the written norm into a variety that also serves as a supradialectal speech form” (Coulmas 1991: 126, also 126–127).

The decline of Latin at the end of the medieval period may be attributed to “the broadened participation of the middle classes in a general, vernacular-based, education” (Kahane 1986: 498–499). Associated with the official vernacularization of the Chinese literary language were the initiation of public education in 1903 and the abolition of the Confucian civil service examination system in 1905 (Barnes 1982: 262). Likewise, across the Arabic-speaking world, the massive growth in educational participation by people from all walks of life has brought about the end of a tradition of restricted literacy, more widespread access to Classical Arabic or Modern Standard Arabic on the part of the general population, and a measure of stylistic differentiation of Arabic appropriate to modern communicative needs (Abdulaziz 1986: 13–14; Ibrahim and Jernudd 1986: 6; Walters 1996: 166). Finally, the diglossic situation in Greece, which persisted until the 1976 reform, originated in the last two decades of the nineteenth century and the first two of the twentieth, during which period important social and economic changes marked the economy’s passage into capitalism and the definitive taking of power by the bourgeois forces (Frangoudaki 1992: 367–368). “The demoticists represented at the time economic and social progress, industrialization, a free parliament, and aggressive nationalism,” and demoticist leaders fought for compulsory education, technical and vocational education, and the linguistic integration of minorities (Frangoudaki 1992: 368). Katharevousa, on the other hand, was defended by the precapitalist social forces of the time as the high variety, the language of the educated minority, used in all formal contexts (Frangoudaki 1992: 368). The use of demotic in formal contexts carried strong connotations of revolt and social unrest (Frangoudaki 1992: 368).

Closely related to the effect of mass literacy is that of the mass media. Throughout the Arabic-speaking lands, radio, television, and cinema in
the wealthier states, and even newspapers and magazines, which are often transmitted orally to the illiterate, “have greatly helped to spread the knowledge of MSA and the urban forms of spoken Arabic to such an extent that it is claimed that even the peasants in places like Egypt, Syria, and Iraq can ‘comprehend’ news in MSA” (Abdulaziz 1986: 15–16). In Japan, the mass media, as well as universal education, have worked to ensure that “the entire population has access to the standard” (Coulmas 1991: 126–127). According to one source, the spread of literacy, the emergence of the mass media, and the growth of modern literature in the vernacular variety have slowly but steadily eroded the social prestige and functions of the elevated variety in Telugu, resulting in the gradual elimination of diglossia (Radhakrishna 1980: 238).

The relationship between national identity, language loyalty, and diglossia is hinted at in Gumperz’s typology of linguistic communities. Whereas in intermediate societies — those most likely to develop some form of diglossia — language loyalty may attach to “codes which may be quite distinct from the vernacular” (Gumperz 1968 [1962]: 469), in highly urbanized communities “language loyalty … is bestowed on the standard, which now closely reflects the majority speech” (Gumperz 1968 [1962]: 470). The Kahanes formulate the relationship as follows: “Where the dichotomy of prestige language and vernacular exists, national pride is symbolized by the latter if the former is not deeply rooted in a prestigious tradition” (Kahane and Kahane 1979: 192). Thus, the decline of Medieval Latin was linked to the rise of the modern European nation-states, as loyalty to these was symbolized by consciousness of and loyalty to their respective evolving standard languages (Kahane 1986: 499; Kahane and Kahane 1979: 193). The elevation of vernacular Armenian to the status of a new standard literary language in place of the earlier standard, which was based on a religious argot and was incomprehensible to the largely illiterate peasant population, was accomplished during the latter half of the nineteenth century in the face of resistance from the Armenian Apostolic church and its conservative supporters (Nercissians 1987: 635–636). While national conservatives warned against undermining the role of the church in national affairs, liberal and progressive thinkers of the day recognized that “in the absence of the other factors which unify a nation, language must replace religious affiliation as the main national identity marker” (Nercissians 1987: 636). The ultimate success of the vernacular language movement in China “was made possible by its identification with the cause of Chinese nationalism” (DeFrancis 1972: 11), or, more specifically, by the replacement of the premodern Sinocentric view of Chinese domestic institutions and international relations with a national consciousness “committed to the
eradication of political and territorial encroachments and to the organization of a modern nation-state capable of guaranteeing China's sovereignty” (Barnes 1982: 261). In Ethiopia, at the turn of the twentieth century, Menelik II succeeded in making Amharic “the linguistic symbol of national unity and sovereignty,” relegating Ge’ez to liturgical and ceremonial functions only (Fellman 1975b: 181–182). Across the Arab world, in recent decades, “governments and pan-national organizations like ALESCO have sought … to promote the use of CA/MSA [Classical Arabic/Modern Standard Arabic] as a unifying tool and symbol across the Arab world” (Walters 1996: 162–163).

According to Ferguson, H is likely to remain as the standard “only if it is already serving as a standard language in some other community and the diglossia community, for reasons linguistic and non-linguistic, tends to merge with the other community” (1959: 339). As noted previously, however, change toward full utilization of H cannot be expected to take place without a radical change in the diglossic pattern of first-language acquisition, which is to say that shift to H is unlikely unless parents evince a willingness to speak H to their children (Ferguson 1959: 331). Absent these conditions, “H fades away and becomes a learned or liturgical language studied only by scholars or specialists and not used actively in the community” (Ferguson 1959: 339). Such a description of things seems to be incontrovertibly the case for Classical Chinese, Egyptian, Ge’ez, Classical Irish, Japanese, Latin, Old Church Slavic, Sanskrit, and Classical Welsh. In other cases, however, it may not be appropriate to speak of the complete displacement of the H variety by L, but rather of the emergence of a third norm, “which represents a merger of the original two norms” (Wexler 1971: 345–346, note 22). In Ferguson’s terms, “if there is a single communication center in the whole speech community, or if there are several such centers all in one dialect area, the L variety of the center(s) will be the basis of the new standard, whether relatively pure L or considerably mixed with H” (1959: 339). This scenario seems closer to the facts in the cases of Arabic, Greek, Sinhala, and Turkish, and perhaps also in those of Tamil and the other Dravidian languages affected to any degree by diglossia. According to Britto, “most Tamilists say that there exists, or is evolving, a Standard Colloquial, which is based on the speech of the educated, non-Brahmin, middle class, and which has currency all over Tamil Nadu” (Britto 1991: 65). In the case of Turkish, “Ottoman mandarin style” was completely moribund by the end of the war in 1918, and, in its place, “there was coming into being a flexible, living language arising in good part from the spoken language of the educated classes of Istanbul and the larger cities” (Gallagher 1971: 163).
Whether H is completely replaced or merges with the vernacular to produce a new standard, the lexicon, like those of prestige languages generally, survives in the new standard in the form of a large-scale transfer of terminology characteristic of upper-class civilization, abstractions, and professional technologies (Kahane and Kahane 1979: 194). The process has been similarly formulated by other students of diglossia: “The phonological, morphological, and syntactic structures of the new language prevail. However, one element of the H form that survives in ... the new standard, is the lexicon, especially words for abstract notions and professional and scientific terminologies” (Sotiropoulos 1982: 19). This is evident in the pervasive influence of Russian Church Slavic on Modern Standard Russian, where “Slavonisms tend to concentrate in religious vocabulary and in more learned and abstract spheres, while vernacular elements tend to concentrate in more concrete and everyday vocabulary” (Comrie 1991: 169; see pp. 169–171 for examples). It is evident too in the imprint of Katharevousa upon the emerging Demotic standard in Greece: “Puristic is receding, but its lexicon of intellectualism, science, and technology has been channeled into the evolving standard language of the educated” (Kahane and Kahane 1979: 195). In each of the Arabic-speaking countries, “a spoken variety of growing prestige that takes the dialect as the matrix or basis and borrows lexicon, set expressions, and discourse markers ... from CA/MSA [Classical Arabic/Modern Standard Arabic] exists, and its use continues to expand even to the extent that it serves as a spoken standard of some sort” (Walters 1996: 169). Last, in the case of Bengali in India, although the apparently diglossic relationship that has existed between sadhu bhasa (H) and colit bhasa (L) is now virtually dissolved, the lexicon of L “has undergone a degree of convergence with the norms of the H, as more and more lexemes formerly confined to H are incorporated into everyday usage in L” (Klaiman 1993: 162).

In a final scenario, Ferguson proposes that “if there are several [communication] centers in different dialect areas with no one center paramount, then it is likely that several L varieties will become standard as separate languages” (1959: 339). The rise of the vernacular-based Romance standards in Europe and of the Apabhramsa languages in northern India vis-à-vis their classical ancestors, Latin and Sanskrit, respectively, are probably the most striking examples available of this process at work.

Conclusions

Far too much has been made of complementary distribution of codes and of the degree of structural affinity existing between codes as criterial
attributes of diglossia. The situations described by Ferguson (excluding, perhaps, that of Haitian Creole), and the many other potentially comparable situations alluded to in the course of this paper, do not constitute a meaningful conceptual or theoretical class of cases based solely upon asymmetric functional distribution of codes and/or upon the linguistic distance between them. Their conceptual unity inheres in a quite specific set of relationships between functional compartmentalization of codes, the lack of opportunity for the acquisition of H as a native variety, the resulting absence of native speakers of H, and the stability in the use of L for vernacular purposes. In principle at least, the codes involved in this configuration might be varieties of totally unrelated languages as readily as they might be minimally distinct isolects of the same language. In practice, however, it is no accident that these codes tend to be closely related structurally and generally to be regarded as varieties of the same language, albeit significantly different varieties.

In cases where, over time, a single cultural-linguistic tradition differentiates its endogenous linguistic resources according to function, it is obvious that new forms, and new varieties of language, develop (or may be borrowed) to accommodate new (or imported) social functions. Insofar as these functions generally pertain to the more formal and ontogenetically later aspects of the culture, their associated varieties tend to serve precisely those functions that are not the main channels for first-language acquisition (see Joseph 1987: 17, quoted above). Accordingly, there typically does not develop a prestige community of native speakers of H that might serve as a reference group for native speakers of L, thereby providing the social impulse for shift from L to H as a native variety. Exogenous language contact, on the other hand, is an entirely different matter. Here, two separate cultural-linguistic traditions come into contact and therefore into competition with each other for at least some of the same socioecological niches. The linguistic varieties involved may be relatively closely related, as in the cases of Frisian and Dutch, Catalan and Castillian, or Occitan and French, or they may be essentially unrelated languages such as Nahuatl and Spanish, English and Navajo, or Arabic and French. In either case, however, these varieties must be viewed socially, politically, historically, and phenomenologically as different languages, representing different group identities, if not different cultures. In contrast with diglossia, it is axiomatic to say that both cultural-linguistic traditions in the case of exogenous language contact already have their own established functions as vernaculars, although they may differ to varying degrees as to the number and nature of high-culture functions for which appropriate registers have been
developed and deployed. Under such circumstances, then, there is always the potential for one community of speakers — typically, if not by definition, the speakers of H — to come to serve as a linguistic reference group for the speakers of L, and thus to furnish the social motivation necessary for the eventual shift from L to H as the native language. It is therefore the particular social etiology of diglossia that provides the explanation for the security of L as the vernacular over the long term, as well as for the generally expected degree of relatedness between the constituent codes. For this reason, it may not be realistic to think of diglossia, in any narrow sense of the term, as providing an appropriate model for the stabilization of endangered languages or for their eventual revitalization as fully functioning native languages.

In a recent essay on diglossia, Schiffman comments as follows: “It remains to be seen whether the same kind of imbalance of power exhibited in nongenetic diglossia can be said to exist with regard to classical or genetic diglossia” (1997: 206). This is surely one of the most fundamental questions that can, and should, be asked about any sociolinguistic condition vis-à-vis another. The argument that has been advanced in this paper is that, allowing for some exceptions in both directions, there is a general tendency for the imbalance of power found in cases of so-called “nongenetic diglossia” not to be found, or to be found with the opposite polarity, in cases of so-called “genetic diglossia,” and that this very fact is reason enough on its face to distinguish between the two types of verbal repertoire at a deeper level than that of genetic relatedness between the codes in the respective code matrices. It is, in other words, the most fundamental reason for distinguishing between diglossia and other sociolinguistic situations often referred to as societal bilingualism. Surface phenomena, in this case the differential functional allocation of codes within a speech economy, may be interesting in and of themselves, but they attain far greater theoretical significance when understood in relation to the underlying social processes of which they are a product, all the more so, perhaps, when apparently similar surface patterns reflect radically different underlying processes.

Anyone who is familiar at first hand with the situation in German-speaking Switzerland, for example, knows full well that this situation differs radically from the type of federal multilingual relationship existing between French, German, Italian, and Romansch at the national level in that country, and from the sociolinguistic relationship obtaining between Romansch and German in Graubünden, where Romansch, despite its recently attained constitutional status as the fourth official language of Switzerland, is demographically very much on the defensive against Swiss German in the vernacular domains and Standard German.
in the high-culture domains. Those familiar with other situations resembling the Swiss-German one will presumably also be compelled by their experience to differentiate those situations from others mentioned in the course of this paper, such as Frisian and Dutch in The Netherlands, French and Occitan in France, Irish and English in Ireland, Spanish and Nahuatl in Mexico, Navajo and English in the United States, and, not improbably, even Spanish and Guarani in Paraguay. All of these situations also may be characterized by a greater or lesser degree of asymmetric functional distribution of codes among their respective communities of bilingual speakers, and not a few of them, if not perhaps all, have been treated at one time or another as instances of diglossia in the sociolinguistic literature. Sociolinguistic theory is not well served by the treatment of all of these situations as though they were simply surface manifestations of the same underlying phenomenon.

Some time ago, a Swiss-German–speaking acquaintance of mine in Basel, failing to comprehend my interest in the local language situation, remarked to me that diglossia was such a quotidian affair as to seem undeserving of the attention that it received from sociolinguists such as myself. Indeed, in the German-speaking cantons of Switzerland, diglossia is as ubiquitous, and therefore as unremarkable, as the air people breathe. In actuality, however, the type of sociolinguistic situation found in German-speaking Switzerland, in the Arabic-speaking world, and at least until recently in Greece is anything but commonplace today. On the contrary, it is worth remembering, as noted above, that “societal changes such as modernisation, urbanisation, [and] the breakdown of rigid class barriers ... have made diglossic situations as described by Ferguson (1959) rather rare” (Pauwels 1986: 16; note also Baker and Prys Jones 1998: 121; Scotton 1986: 408–409). Indeed, it is the rarity, not the ubiquity, of such situations that confers upon them their considerable theoretical interest for the sociologist of language. The remarkable, yet still not widely understood, facts of the sociolinguistic situations in German-speaking Switzerland, in the Arabic-speaking world, in Greece, in the Tamil-speaking regions of India and perhaps other areas of South Asia, in medieval Ireland and Wales, in Jewish Palestine during the late Second Temple Period, in China from Confucian times up to the present century, in Japan from the seventh century on, and in medieval Latin- and Romance-speaking Europe present a wealth of data from which potentially to construct a concept of diglossia and a theory of its functioning, and it is worth examining whether these situations, and others besides, may be sufficiently alike to warrant their inclusion under a single designation distinct from those of other sociolinguistic situations.

In the final analysis, what the term “diglossia” should be understood
as designating is a matter of linguistic convention within the discipline of sociolinguistics. The meaning is not given in the term itself, nor is the meaning intended by Ferguson in 1959 necessarily immutable for all time. Fishman's (1967) argument that the term “bilingualism,” referring to individual linguistic versatility, should be distinguished from the term “diglossia,” referring to societally held norms governing differential functional allocation of codes, is a crucial one, although, as I have remarked elsewhere, it may be unfortunate that the term “diglossia” should have been coopted to refer to patterned situational variation more generally. At the least, this paper has argued for a uniform use, whatever that may be, of the term “diglossia” in sociolinguistic research and theory construction. Equally important, it also argues that the similarities that exist between the sociolinguistic situations described by Ferguson are sufficiently different from other types of sociolinguistic situation to warrant their recognition as a category distinct from these other types, such as standard-with-dialects and societal bilingualism.

University of New Mexico

References


Nishi, A. (1874). *Y ji o motte kokugo o shosuru no ron* [A plea for writing the national language with Western letters]. *Meiroku zasshi* 1.


