

WORD

ISSN: 0043-7956 (Print) 2373-5112 (Online) Journal homepage: www.tandfonline.com/journals/rwrd20

Diglossia

Charles A. Ferguson

To cite this article: Charles A. Ferguson (1959) Diglossia, WORD, 15:2, 325-340, DOI: 10.1080/00437956.1959.11659702

To link to this article: https://doi.org/10.1080/00437956.1959.11659702

Published online: 04 Dec 2015.

Submit your article to this journal

Article views: 83151

View related articles

Citing articles: 474 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rwrd20

CHARLES A. FERGUSON

Diglossia[1]

In many speech communities two or more varieties of the same language[2] are used by some speakers under different conditions. Perhaps the most familiar example is the standard language and regional dialect as used, say, in Italian or Persian, where many speakers speak their local dialect at home or among family or friends of the same dialect area but use the standard language in communicating with speakers of other dialects or on public occasions. There are, however, other, quite different examples of the use of two varieties of a language in the same speech community. In Baghdad the Christian Arabs speak a "Christian Arabic" dialect when talking among themselves but speak the general Baghdad dialect, "Muslim Arabic," when talking in a mixed group. In recent years there has been a renewed interest in studying the development and characteristics of standardized languages,[3] and it is in following this line of interest that the present study seeks to examine carefully one particular kind of standardization where two varieties of a language exist side by side throughout the community, with each having a definite role to play. The term 'diglossia' is introduced here, modeled on the French *diglossie*, which has been applied to this situation,

[1] A preliminary version of this study, with the title "Classical or Colloquial—One Standard or Two," was prepared for presentation at the symposium on Urbanization and Standard Languages—Facts and Attitudes, held at the meeting of the American Anthropological Association in November, 1958, in Washington, D.C. The preliminary version was read by a number of people, and various modifications were made on the basis of comments by H. Blanc, J. J. Gumperz, B. Halpern, M. Perlmann, R. L. Ward, and U. Weinreich.

[2] The terms 'language', 'dialect', and 'variety' are used here without precise definition. It is hoped that they occur sufficiently in accordance with established usage to be unambiguous for the present purpose. The term 'superposed variety' is also used here without definition; it means that the variety in question is not the primary, "native" variety for the speakers in question but may be learned in addition to this. Finally, no attempt is made in this paper 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.

[3] Cf. especially H. Kloss, *Die Entwicklung neuer germanischer Kultursprachen von 1800 bis 1950* (Munich, 1952), with its valuable introduction on standardization in general.

since there seems to be no word in regular use for this in English; other languages of Europe generally use the word for 'bilingualism' in this special sense as well.

It is likely that this particular situation in speech communities is very widespread, although it is rarely mentioned, let alone satisfactorily described. A full explanation of it can be of considerable help in dealing with problems in linguistic description, in historical linguistics, and in language typology. The present study should be regarded as preliminary in that much more assembling of descriptive and historical data is required; its purpose is to characterize diglossia by picking out four speech communities and their languages (hereafter called the defining languages) which clearly belong in this category, and describing features shared by them which seem relevant to the classification. The defining languages selected are Arabic, Modern Greek, Swiss German, Haitian Creole.[4]

Before proceeding to the description it must be pointed out that diglossia is not assumed to be a stage which occurs always and only at a certain

[4] The judgments of this paper are based primarily on the author's personal experience, but documentation for the four defining languages is available, and the following references may be consulted for further details. Most of the studies listed here take a strong stand in favor of greater use of the more colloquial variety since it is generally writers of this opinion who want to describe the facts. This bias can, however, be ignored by the reader who simply wants to discover the basic facts of the situation.

MODERN GREEK: H. Pernot, *Grammaire grecque moderne* (Paris, 1898), Introduction, pp. vii–xxxi; K. Krumbacher, *Das Problem der modernen griechischen Schriftsprache* (Munich, 1902); G. N. Hatzidakis, *Die Sprachfrage in Griechenland* (Athens, 1905); J. Psichari, "Un Pays qui ne veut pas sa langue," *Mercure de France*, October 1, 1928, 63–121; repr. in J. Psichari, *Quelque travaux . . .*, I, pp. 1283–1337 (Paris, 1930); A. Steinmetz, "Schrift und Volksprache in Griechenland," Deutsche Akademie (Munich), *Mitteilungen*, 1936, pp. 370–379.

SWISS GERMAN: O. von Greyerz, *Sprache, Dichtung, Heimat* (Berne, 1933): "Vom Wert und Wesen unserer Mundart," pp. 226–247; A. Senn, "Das Verhältnis von Mundart und Schriftsprache in der deutschen Schweiz," *Journal of English and Germanic Philology* XXXIV (1935), 42–58; K. Schmid, "Für unser Schweizerdeutsch," in *Die Schweiz: ein nationales Jahrbuch 1936* (Basle, 1936), pp. 65–79; H. Kloss, *Entwicklung* (cf. fn. 3), pp. 126–138.

ARABIC: W. Marçais, three articles in *L'Enseignement public* XCVII (1930), 401–409; CV (1931), 20–39, 120–133; J. Lecerf, *Littérature dialectale et renaissance arabe moderne* (Damascus, 1932–33), pp. 1–14; *Majallat al-majmaᶜ al-ᶜilmī al-ᶜarabī* (Dimashq), Vol. 32:1: *ᶜAdad xāṣṣ bilmu'tamar al-'awwal lilmajāmiᶜ al-luġawiyyah al-ᶜilmiyyah al-ᶜarabiyyah* (Damascus, January, 1957); S. J. Al-Toma, "The Teaching of Classical Arabic to Speakers of the Colloquial in Iraq: a Study of the Problem of Linguistic Duality . . ." (Harvard Univ. D.Ed. thesis, 1957); A. Chejne, "The Role of Arabic in Present-Day Arab Society," *The Islamic Literature* X (1958), no. 4 (April), 15–54.

HAITIAN CREOLE: S. Comhaire-Sylvain, *Le Créole haitien* (Wetteren and Port-au-Prince, 1936); R. A. Hall, Jr., *Haitian Creole* (Menasha, Wisc., 1953).

point in some kind of evolution, e.g. in the standardization process. Diglossia may develop from various origins and eventuate in different language situations. Of the four defining languages, Arabic diglossia seems to reach as far back as our knowledge of Arabic goes, and the superposed 'Classical' language has remained relatively stable, while Greek diglossia has roots going back many centuries, but it became fully developed only at the beginning of the nineteenth century with the renaissance of Greek literature and the creation of a literary language based in large part on previous forms of literary Greek. Swiss German diglossia developed as a result of long religious and political isolation from the centers of German linguistic standardization, while Haitian Creole arose from a creolization of a pidgin French, with standard French later coming to play the role of the superposed variety. Some speculation on the possibilities of development will, however, be given at the end of the paper.

For convenience of reference the superposed variety in diglossia will be called the H ('high') variety or simply H, and the regional dialects will be called L ('low') varieties or, collectively, simply L. All the defining languages have names for H and L, and these are listed in the accompanying table.

		H is called	L is called
Arabic	Classical (=H)	*'al-fuṣḥā*	*'al-ᶜāmmiyyah, 'ad-dārij*
	Egyptian (=L)	*'il-faṣīḥ, 'in-naḥawi*	*'il-ᶜammiyya*
Sw. German	Stand. German (=H)	*Schriftsprache*	*Hoochtüütsch*
	Swiss (=L)	*[Schweizer] Dialekt, Schweizerdeutsch*	*Schwyzertüütsch*
H. Creole	French (=H)	*français*	*créole haïtien*
Greek	H and L	*katharévusa*	*dhimotikí*

It is instructive to note the problems involved in citing words of these languages in a consistent and accurate manner. First, should the words be listed in their H form or in their L form, or in both? Second, if words are cited in their L form, what kind of L should be chosen? In Greek and in Haitian Creole, it seems clear that the ordinary conversational language of the educated people of Athens and Port-au-Prince respectively should be selected. For Arabic and for Swiss German the choice must be arbitrary, and the ordinary conversational language of educated people of Cairo and of Zürich city will be used here. Third, what kind of spelling should be used to represent L? Since there is in no case a generally accepted orthography for L, some kind of phonemic or quasi-phonemic

transcription would seem appropriate. The following choices were made. For Haitian Creole, the McConnell-Laubach spelling was selected,[5] since it is approximately phonemic and is typographically simple. For Greek, the transcription was adopted from the manual *Spoken Greek*,[6] since this is intended to be phonemic; a transliteration of the Greek spelling seems less satisfactory not only because the spelling is variable but also because it is highly etymologizing in nature and quite unphonemic. For Swiss German, the spelling backed by Dieth,[7] which, though it fails to indicate all the phonemic contrasts and in some cases may indicate allophones, is fairly consistent and seems to be a sensible systematization, without serious modification, of the spelling conventions most generally used in writing Swiss German dialect material. Arabic, like Greek, uses a non-Roman alphabet, but transliteration is even less feasible than for Greek, partly again because of the variability of the spelling, but even more because in writing Egyptian colloquial Arabic many vowels are not indicated at all and others are often indicated ambiguously; the transcription chosen here sticks closely to the traditional systems of Semitists, being a modification for Egyptian of the scheme used by Al-Toma.[8]

The fourth problem is how to represent H. For Swiss German and Haitian Creole standard German and French orthography respectively can be used even though this hides certain resemblances between the sounds of H and L in both cases. For Greek either the usual spelling in Greek letters could be used or a transliteration, but since a knowledge of Modern Greek pronunciation is less widespread than a knowledge of German and French pronunciation, the masking effect of the orthography is more serious in the Greek case, and we use the phonemic transcription instead. Arabic is the most serious problem. The two most obvious choices are (1) a transliteration of Arabic spelling (with the unwritten vowels supplied by the transcriber) or (2) a phonemic transcription of the Arabic as it would be read by a speaker of Cairo Arabic. Solution (1) has been adopted, again in accordance with Al-Toma's procedure.

1. Function. One of the most important features of diglossia is the specialization of function for H and L. In one set of situations only H is appropriate and in another only L, with the two sets overlapping only very slightly. As an illustration, a sample listing of possible situations is given, with indication of the variety normally used:

[5] Cf. H. O. McConnell and E. Swan, *You Can Learn Creole* (Port-au-Prince, 1945).
[6] H. and R. Kahane and R. L. Ward, *Spoken Greek* (Washington, 1945).
[7] E. Dieth, *Schwyzertütschi Dialäktschrift* (Zürich, 1938).
[8] S. J. Al-Toma, *op. cit.*

	H	L
Sermon in church or mosque	x	
Instructions to servants, waiters, workmen, clerks		x
Personal letter	x	
Speech in parliament, political speech	x	
University lecture	x	
Conversation with family, friends, colleagues		x
News broadcast	x	
Radio "soap opera"		x
Newspaper editorial, news story, caption on picture	x	
Caption on political cartoon		x
Poetry	x	
Folk literature		x

The importance of using the right variety in the right situation can hardly be overestimated. An outsider who learns to speak fluent, accurate L and then uses it in a formal speech is an object of ridicule. A member of the speech community who uses H in a purely conversational situation or in an informal activity like shopping is equally an object of ridicule. In all the defining languages it is typical behavior to have someone read aloud from a newspaper written in H and then proceed to discuss the contents in L. In all the defining languages it is typical behavior to listen to a formal speech in H and then discuss it, often with the speaker himself, in L.[9]

The last two situations on the list call for comment. In all the defining languages some poetry is composed in L, and a small handful of poets compose in both, but the status of the two kinds of poetry is very different, and for the speech community as a whole it is only the poetry in H that is felt to be "real" poetry.[10] On the other hand, in every one of the defining languages certain proverbs, politeness formulas, and the like are in H even when cited in ordinary conversation by illiterates. It has been estimated that as much as one-fifth of the proverbs in the active repertory of Arab villagers are in H.[11]

2. Prestige. In all the defining languages the speakers regard H as superior to L in a number of respects. Sometimes the feeling is so strong

[9] The situation in formal education is often more complicated than is indicated here. In the Arab world, for example, formal university lectures are given in H, but drills, explanation, and section meetings may be in large part conducted in L, especially in the natural sciences as opposed to the humanities. Although the teachers' use of L in secondary schools is forbidden by law in some Arab countries, often a considerable part of the teacher's time is taken up with explaining in L the meaning of material in H which has been presented in books or lectures.

[10] Modern Greek does not quite fit this description. Poetry in L is the major production and H verse is generally felt to be artificial.

[11] *Journal of the American Oriental Society* LXXV (1955), 124f.

that H alone is regarded as real and L is reported "not to exist." Speakers of Arabic, for example, may say (in L) that so-and-so doesn't know Arabic. This normally means he doesn't know H, although he may be a fluent, effective speaker of L. If a non-speaker of Arabic asks an educated Arab for help in learning to speak Arabic the Arab will normally try to teach him H forms, insisting that these are the only ones to use. Very often, educated Arabs will maintain that they never use L at all, in spite of the fact that direct observation shows that they use it constantly in all ordinary conversation. Similarly, educated speakers of Haitian Creole frequently deny its existence, insisting that they always speak French. This attitude cannot be called a deliberate attempt to deceive the questioner, but seems almost a self-deception. When the speaker in question is replying in good faith, it is often possible to break through these attitudes by asking such questions as what kind of language he uses in speaking to his children, to servants, or to his mother. The very revealing reply is usually something like: "Oh, but they wouldn't understand [the H form, whatever it is called]."

Even where the feeling of the reality and superiority of H is not so strong, there is usually a belief that H is somehow more beautiful, more logical, better able to express important thoughts, and the like. And this belief is held also by speakers whose command of H is quite limited. To those Americans who would like to evaluate speech in terms of effectiveness of communication it comes as a shock to discover that many speakers of a language involved in diglossia characteristically prefer to hear a political speech or an expository lecture or a recitation of poetry in H even though it may be less intelligible to them than it would be in L.

In some cases the superiority of H is connected with religion. In Greek the language of the New Testament is felt to be essentially the same as the *katharévusa*, and the appearance of a translation of the New Testament in *dhimotikí* was the occasion for serious rioting in Greece in 1903. Speakers of Haitian Creole are generally accustomed to a French version of the Bible, and even when the Church uses Creole for catechisms, and the like, it resorts to a highly Gallicized spelling. For Arabic, H is the language of the Qur'an and as such is widely believed to constitute the actual words of God and even to be outside the limits of space and time, i.e. to have existed "before" time began with the creation of the world.

3. Literary heritage. In every one of the defining languages there is a sizable body of written literature in H which is held in high esteem by the speech community, and contemporary literary production in H by members of the community is felt to be part of this otherwise existing literature.

The body of literature may either have been produced long ago in the past history of the community or be in continuous production in another speech community in which H serves as the standard variety of the language. When the body of literature represents a long time span (as in Arabic or Greek) contemporary writers—and readers—tend to regard it as a legitimate practice to utilize words, phrases, or constructions which may have been current only at one period of the literary history and are not in widespread use at the present time. Thus it may be good journalistic usage in writing editorials, or good literary taste in composing poetry, to employ a complicated Classical Greek participial construction or a rare twelfth-century Arabic expression which it can be assumed the average educated reader will not understand without research on his part. One effect of such usage is appreciation on the part of some readers: "So-and-so really knows his Greek [or Arabic]," or "So-and-so's editorial today, or latest poem, is very good Greek [or Arabic]."

4. Acquisition. Among speakers of the four defining languages adults invariably use L in speaking to children and children use L in speaking to one another. As a result, L is invariably learned by children in what may be regarded as the "normal" way of learning one's mother tongue. H may be heard by children from time to time, but the actual learning of H is chiefly accomplished by the means of formal education, whether this be traditional Qur'anic schools, modern government schools, or private tutors.

This difference in method of acquisition is very important. The speaker is at home in L to a degree he almost never achieves in H. The grammatical structure of L is learned without explicit discussion of grammatical concepts; the grammar of H is learned in terms of "rules" and norms to be imitated.

It seems unlikely that any change toward full utilization of H could take place without a radical change in this pattern of acquisition. For example, those Arabs who ardently desire to have L replaced by H for all functions can hardly expect this to happen if they are unwilling to speak H to their children.[12]

5. Standardization. In all the defining languages there is a strong tradition of grammatical study of the H form of the language. There are

[12] It has been very plausibly suggested that there are psychological implications following from this linguistic duality. This certainly deserves careful experimental investigation. On this point, see the highly controversial article which seems to me to contain some important kernels of truth along with much which cannot be supported—E. Shouby, "The Influence of the Arabic Language on the Psychology of the Arabs," *Middle East Journal* V (1951), 284–302.

grammars, dictionaries, treatises on pronunciation, style, and so on. There is an established norm for pronunciation, grammar, and vocabulary which allows variation only within certain limits. The orthography is well established and has little variation. By contrast, descriptive and normative studies of the L form are either non-existent or relatively recent and slight in quantity. Often they have been carried out first or chiefly by scholars OUTSIDE the speech community and are written in other languages. There is no settled orthography and there is wide variation in pronunciation, grammar, and vocabulary.

In the case of relatively small speech communities with a single important center of communication (e.g. Greece, Haiti) a kind of standard L may arise which speakers of other dialects imitate and which tends to spread like any standard variety except that it remains limited to the functions for which L is appropriate.

In speech communities which have no single most important center of communication a number of regional L's may arise. In the Arabic speech community, for example, there is no standard L corresponding to educated Athenian *dhimotikí*, but regional standards exist in various areas. The Arabic of Cairo, for example, serves as a standard L for Egypt, and educated individuals from Upper Egypt must learn not only H but also, for conversational purposes, an approximation to Cairo L. In the Swiss German speech community there is no single standard, and even the term 'regional standard' seems inappropriate, but in several cases the L of a city or town has a strong effect on the surrounding rural L.

6. Stability. It might be supposed that diglossia is highly unstable, tending to change into a more stable language situation. This is not so. Diglossia typically persists at least several centuries, and evidence in some cases seems to show that it can last well over a thousand years. The communicative tensions which arise in the diglossia situation may be resolved by the use of relatively uncodified, unstable, intermediate forms of the language (Greek *mikti*, Arabic *al-luġah al-wusṭā*, Haitian *créole de salon*) and repeated borrowing of vocabulary items from H to L.

In Arabic, for example, a kind of spoken Arabic much used in certain semiformal or cross-dialectal situations has a highly classical vocabulary with few or no inflectional endings, with certain features of classical syntax, but with a fundamentally colloquial base in morphology and syntax, and a generous admixture of colloquial vocabulary. In Greek a kind of mixed language has become appropriate for a large part of the press.

The borrowing of lexical items from H to L is clearly analogous (or for the periods when actual diglossia was in effect in these languages, identical)

with the learned borrowings from Latin to Romance languages or the Sanskrit *tatsamas* in Middle and New Indo-Aryan.[13]

7. Grammar. One of the most striking differences between H and L in the defining languages is in the grammatical structure: H has grammatical categories not present in L and has an inflectional system of nouns and verbs which is much reduced or totally absent in L. For example, Classical Arabic has three cases in the noun, marked by endings; colloquial dialects have none. Standard German has four cases in the noun and two non-periphrastic indicative tenses in the verb; Swiss German has three cases in the noun and only one simple indicative tense. *Katharévusa* has four cases, *dhimotikí* three. French has gender and number in the noun, Creole has neither. Also, in every one of the defining languages there seem to be several striking differences of word order as well as a thorough-going set of differences in the use of introductory and connective particles. It is certainly safe to say that in diglossia *there are always extensive differences between the grammatical structures of H and L.* This is true not only for the four defining languages, but also for every other case of diglossia examined by the author.

For the defining languages it may be possible to make a further statement about grammatical differences. It is always risky to hazard generalizations about grammatical complexity,[14] but it may be worthwhile to attempt to formulate a statement applicable to the four defining languages even if it should turn out to be invalid for other instances of diglossia.

There is probably fairly wide agreement among linguists that the grammatical structure of language A is "simpler" than that of B if, other things being equal,

(1) the morphophonemics of A is simpler, i.e. morphemes have fewer alternants, alternation is more regular, automatic (e.g. Turkish *-lar* ~ *-ler* is simpler than the English plural markers);
(2) there are fewer obligatory categories marked by morphemes or concord (e.g. Persian with no gender distinctions in the pronoun is simpler than Egyptian Arabic with masculine–feminine distinction in the second and third persons singular);

[13] The exact nature of this borrowing process deserves careful investigation, especially for the important "filter effect" of the pronunciation and grammar of H occurring in those forms of middle language which often serve as the connecting link by which the loans are introduced into the "pure" L.

[14] Cf. J. H. Greenberg, "A Quantitative Approach to the Morphological Typology of Language," in *Methods and Perspective in Anthropology* (Minneapolis, 1954), pp. 192–220.

(3) paradigms are more symmetrical (e.g. a language with all declensions having the same number of case distinctions is simpler than one in which there is variation);
(4) concord and rection are stricter (e.g. prepositions all take the same case rather than different cases).

If this understanding of grammatical simplicity is accepted, then we may note that in at least three of the defining languages, the grammatical structure of any given L variety is simpler than that of its corresponding H. This seems incontrovertibly true for Arabic, Greek, and Haitian Creole; a full analysis of standard German and Swiss German might show this not to be true in that diglossic situation in view of the extensive morphophonemics of Swiss.

8. Lexicon. Generally speaking, the bulk of the vocabulary of H and L is shared, of course with variations in form and with differences of use and meaning. It is hardly surprising, however, that H should include in its total lexicon technical terms and learned expressions which have no regular L equivalents, since the subjects involved are rarely if ever discussed in pure L. Also, it is not surprising that the L varieties should include in their total lexicons popular expressions and the names of very homely objects or objects of very localized distribution which have no regular H equivalents, since the subjects involved are rarely if ever discussed in pure H. But *a striking feature of diglossia is the existence of many paired items, one H one L, referring to fairly common concepts frequently used in both H and L, where the range of meaning of the two items is roughly the same, and the use of one or the other immediately stamps the utterance or written sequence as H or L.* For example, in Arabic the H word for 'see' is *ra'ā*, the L word is *šāf*. The word *ra'ā* never occurs in ordinary conversation and *šāf* is not used in normal written Arabic. If for some reason a remark in which *šāf* was used is quoted in the press, it is replaced by *ra'ā* in the written quotation. In Greek the H word for 'wine' is *ínos*, the L word is *krasí*. The menu will have *ínos* written on it, but the diner will ask the waiter for *krasí*. The nearest American English parallels are such cases as *illumination* ~ *light*, *purchase* ~ *buy*, or *children* ~ *kids*, but in these cases 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. Also, 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, e.g. *illumination*, *purchase*, and *children* are not fully parallel in their formal–informal range of usage.

A dozen or so examples of lexical doublets from three of the sample languages are given below. For each language two nouns, a verb, and two particles are given.

	H		L
GREEK	*íkos*	house	*spíti*
	ídhor	water	*neró*
	éteke	gave birth	*eyénise*
	alá	but	*má*
ARABIC			
	hiðā'un	shoe	*gazma*
	'anfun	nose	*manaxīr*
	ðahaba	went	*rāḥ*
	mā	what	*'ēh*
	'al'āna	now	*dilwa'ti*
CREOLE			
	homme, gens	person, people	*moun*[15]
	âne	donkey	*bourik*
	donner	give	*bay*
	beaucoup	much, a lot	*âpil*
	maintenant	now	*nô*

It would be possible to present such a list of doublets for Swiss German (e.g. *nachdem* ≅ *no* 'after', *jemand* ≅ *öpper* 'someone', etc.), but this would give a false picture. In Swiss German the phonological differences between H and L are very great and the normal form of lexical pairing is regular cognation (*klein* ≅ *chly* 'small', etc.).

9. Phonology. It may seem difficult to offer any generalization on the relationships between the phonology of H and L in diglossia in view of the diversity of data. H and L phonologies may be quite close, as in Greek; moderately different, as in Arabic or Haitian Creole; or strikingly divergent, as in Swiss German. Closer examination, however, shows two statements to be justified. (Perhaps these will turn out to be unnecessary when the preceding features are stated so precisely that the statements about phonology can be deduced directly from them.)

(a) *The sound systems of H and L constitute a single phonological structure of which the L phonology is the basic system and the divergent features of H phonology are either a subsystem or a parasystem.* Given the mixed forms mentioned above and the corresponding difficulty of identifying a given word in a given utterance as being definitely H or definitely L, it seems necessary to assume that the speaker has a single inventory of distinctive

[15] Not connected with French *monde*.

oppositions for the whole H–L complex and that there is extensive interference in both directions in terms of the distribution of phonemes in specific lexical items.[16]

(b) *If "pure" H items have phonemes not found in "pure" L items, L phonemes frequently substitute for these in oral use of H and regularly replace them in tatsamas.* For example, French has a high front rounded vowel phoneme /ü/; "pure" Haitian Creole has no such phoneme. Educated speakers of Creole use this vowel in *tatsamas* such as *Luk* (/lük/ for the Gospel of St. Luke), while they, like uneducated speakers, may sometimes use /i/ for it when speaking French. On the other hand, /i/ is the regular vowel in such *tatsamas* in Creole as *linèt* 'glasses'.

In cases where H represents in large part an earlier stage of L, it is possible that a three-way correspondence will appear. For example, Syrian and Egyptian Arabic frequently use /s/ for /θ/ in oral use of Classical Arabic, and have /s/ in tatsamas, but have /t/ in words regularly descended from earlier Arabic not borrowed from the Classical.[17]

Now that the characteristic features of diglossia have been outlined it is feasible to attempt a fuller definition. DIGLOSSIA *is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.*

With the characterization of diglossia completed we may turn to a brief consideration of three additional questions: How does diglossia differ from the familiar situation of a standard language with regional dialects? How widespread is the phenomenon of diglossia in space, time, and linguistic families? Under what circumstances does diglossia come into being and into what language situations is it likely to develop?

The precise role of the standard variety (or varieties) of a language vis-à-vis regional or social dialects differs from one speech community to another, and some instances of this relation may be close to diglossia or perhaps even better considered as diglossia. As characterized here, diglossia differs from the more widespread standard-with-dialects in that no segment

[16] For details on certain aspects of this phonological interference in Arabic, cf. C. A. Ferguson, "Two Problems in Arabic Phonology," *Word* XIII (1957), 460–478.

[17] Cf. Ferguson, *op. cit.*

of the speech community in diglossia regularly uses H as a medium of ordinary conversation, and any attempt to do so is felt to be either pedantic and artificial (Arabic, Greek) or else in some sense disloyal to the community (Swiss German, Creole). In the more usual standard-with-dialects situation the standard is often similar to the variety of a certain region or social group (e.g. Tehran Persian, Calcutta Bengali) which is used in ordinary conversation more or less naturally by members of the group and as a superposed variety by others.

Diglossia is apparently not limited to any geographical region or language family.[18] Three examples of diglossia from other times and places may be cited as illustrations of the utility of the concept. First, consider Tamil. As used by the millions of members of the Tamil speech community in India today, it fits the definition exactly. There is a literary Tamil as H used for writing and certain kinds of formal speaking, and a standard colloquial as L (as well as local L dialects) used in ordinary conversation. There is a body of literature in H going back many centuries which is highly regarded by Tamil speakers today. H has prestige, L does not. H is always superposed, L is learned naturally, whether as primary or as a superposed standard colloquial. There are striking grammatical differences and some phonological differences between the two varieties.[19] The situation is only slightly complicated by the presence of Sanskrit and English for certain functions of H; the same kind of complication exists in parts of the Arab world where French, English, or a liturgical language such as Syriac or Coptic has certain H-like functions.

Second, we may mention Latin and the emergent Romance languages during a period of some centuries in various parts of Europe. The vernacular was used in ordinary conversation but Latin for writing or certain kinds of formal speech. Latin was the language of the Church and its literature, Latin had the prestige, there were striking grammatical differences between the two varieties in each area, etc.

Third, Chinese should be cited because it probably represents diglossia

[18] All clearly documented instances 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.

[19] There is apparently no good description available of the precise relations of the two varieties of Tamil; an account of some of the structural differences is given by Shanmugam Pillai, "Literary and Colloquial Tamil," to appear in *Linguistic Diversity in South Asia* (ed. C. A. Ferguson and J. J. Gumperz). Incidentally, it may be noted that Tamil diglossia seems to go back many centuries, since the language of early literature contrasts sharply with the language of early inscriptions, which probably reflect the spoken language of the time.

on the largest scale of any attested instance.[20] The *weu-li* corresponds to H, while Mandarin colloquial is a standard L; there are also regional L varieties so different as to deserve the label 'separate languages' even more than the Arabic dialects, and at least as much as the emergent Romance languages in the Latin example. Chinese, however, like modern Greek, seems to be developing away from diglossia toward a standard-with-dialects in that the standard L or a mixed variety is coming to be used in writing for more and more purposes, i.e. it is becoming a true standard.

Diglossia is likely to come into being when the following three conditions hold in a given speech community: (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, whether as source (e.g. divine revelation) or reinforcement, some of the fundamental values of the community. (2) Literacy in the community is limited to a small elite. (3) A suitable period of time, on the order of several centuries, passes from the establishment of (1) and (2). It can probably be shown that this combination of circumstances has occurred hundreds of times in the past and has generally resulted in diglossia. Dozens of examples exist today, and it is likely that examples will occur in the future.

Diglossia seems to be accepted and not regarded as a "problem" by the community in which it is in force, until certain trends appear in the community. These include trends toward (a) more widespread literacy (whether for economic, ideological or other reasons), (b) broader communication among different regional and social segments of the community (e.g. for economic, administrative, military, or ideological reasons), (c) desire for a full-fledged standard "national" language as an attribute of autonomy or of sovereignty.

When these trends appear, leaders in the community begin to call for unification of the language, and for that matter, actual trends toward unification begin to take place. These individuals tend to support either the adoption of H or of one form of L as the standard, less often the adoption of a modified H or L, a "mixed" variety of some kind. The arguments explicitly advanced seem remarkably the same from one instance of diglossia to another.

The proponents of H argue that H must be adopted because it connects the community with its glorious past or with the world community and because it is a naturally unifying factor as opposed to the divisive nature

[20] An excellent, brief description of the complex Chinese situation is available in the introduction to Y. R. Chao, *Cantonese Primer* (Cambridge, 1947), pp. 1–17.

of the L dialects. In addition to these two fundamentally sound arguments there are usually pleas based on the beliefs of the community in the superiority of H: that it is more beautiful, more expressive, more logical, that it has divine sanction, or whatever their specific beliefs may be. When these latter arguments are examined objectively their validity is often quite limited, but their importance is still very great because they reflect widely held attitudes within the community.

The proponents of L argue that some variety of L must be adopted because it is closer to the real thinking and feeling of the people; it eases the educational problem since people have already acquired a basic knowledge of it in early childhood; and it is a more effective instrument of communication at all levels. In addition to these fundamentally sound arguments there is often great emphasis given to points of lesser importance such as the vividness of metaphor in the colloquial, the fact that other "modern nations" write very much as they speak, and so on.

The proponents of both sides or even of the mixed language seem to show the conviction—although this may not be explicitly stated—that a standard language can simply be legislated into place in a community. Often the trends which will be decisive in the development of a standard language are already at work and have little to do with the argumentation of the spokesmen for the various viewpoints.

A brief and superficial glance at the outcome of diglossia in the past and a consideration of present trends suggests that there are only a few general kinds of development likely to take place. First, we must remind ourselves that the situation may remain stable for long periods of time. But if the trends mentioned above do appear and become strong, change may take place. Second, H can succeed in establishing itself as a 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. Otherwise H fades away and becomes a learned or liturgical language studied only by scholars or specialists and not used actively in the community. Some form of L or a mixed variety becomes standard.

Third, 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. If there are several such centers in different dialect areas with no one center paramount, then it is likely that several L varieties will become standard as separate languages.

A tentative prognosis for the four defining languages over the next two centuries (i.e. to about 2150 A.D.) may be hazarded:

SWISS GERMAN: Relative stability.

ARABIC: Slow development toward several standard languages, each based on an L variety with heavy admixture of H vocabulary. Three seem likely: Maghrebi (based on Rabat or Tunis?), Egyptian (based on Cairo), Eastern (based on Baghdad?); unexpected politico-economic developments might add Syrian (based on Damascus?), Sudanese (based on Omdurman-Khartoum), or others.

HAITIAN CREOLE: Slow development toward unified standard based on L of Port-au-Prince.

GREEK: Full development to unified standard based on L of Athens plus heavy admixture of H vocabulary.

This paper concludes with an appeal for further study of this phenomenon and related ones. Descriptive linguists in their understandable zeal to describe the internal structure of the language they are studying often fail to provide even the most elementary data about the socio-cultural setting in which the language functions. Also, descriptivists usually prefer detailed descriptions of "pure" dialects or standard languages rather than the careful study of the mixed, intermediate forms often in wider use. Study of such matters as diglossia is of clear value in understanding processes of linguistic change and presents interesting challenges to some of the assumptions of synchronic linguistics. Outside linguistics proper it promises material of great interest to social scientists in general, especially if a general frame of reference can be worked out for analysis of the use of one or more varieties of language within a speech community. Perhaps the collection of data and more profound study will drastically modify the impressionistic remarks of this paper, but if this is so the paper will have had the virtue of stimulating investigation and thought.

Center for Middle Eastern Studies
Harvard University
Cambridge 38, Massachusetts