

1 Introduction

The term *diglossia* is a fundamental sociolinguistic concept that is used recurrently in the literature with, at times, varying definitions. In most cases, diglossia describes a situation where two linguistic systems coexist in a functional distribution within the same speech community. One system is assigned the status of high variety (H), while the other receives the status of low variety (L). The H variety is used in more formal domains while the L variety is typically limited to oral informal communication. Since the mid-twentieth century, the term, nevertheless, has been applied, at times with major conceptual modification, to a wide array of situations in different parts of the world. Some of these situations included contact between varieties of the same language, languages belonging to the same family, or distant languages. The aim of this book is to refocus the concept of diglossia and situate it within current theories of language contact. My central argument is that diglossia, as a theoretical framework, has not been appropriately discussed in the field of contact linguistics, partly because it has been extended to describe cases that in reality bear very few similarities to classical diglossic cases where genetic relatedness is a key element.

In this introductory chapter, I outline the development stages of the concept of diglossia and lay the theoretical ground for the analyses that follow in the subsequent chapters. I argue that classical diglossia, understood as a situation that involves closely related language varieties, is a more useful concept for studies on language contact than the notion of extended diglossia that describes any pair of languages, or more, which are in complementary distribution regardless of their genetic relatedness.

1.1 Defining diglossia

Taken in its original literal meaning, diglossia describes the coexistence of two words, i.e., vocabulary doublets, in a given language to refer to the same concept or entity. It was first used to refer to the Greek language situation, one of the defining cases mentioned by Ferguson (1959a) in his seminal article discussed in detail below. In Greece, and up to 1976, two varieties of Greek were in use. Demotic was the spoken vernacular that evolved over time from classical Greek.

2 Introduction

It was naturally acquired and widely used by all native speakers of Greek. Katharevousa, on the other hand, was artificially developed in the nineteenth century from classical Greek, without consideration of the vernacular features then present in Demotic. Katharevousa was declared the official language in 1834 and became the language of education and administration until Demotic, which had gone through a process of standardization, was recognized as the new national language, officially ending the diglossic situation in 1976 (Alexiou 1982). Although nowadays, the situation of Arabic is often recognized as the embodiment of diglossia, in fact it was the Greek situation that prompted the adoption of the concept in modern linguistics. Toufexis (2008: 207) indicates that Karl Krumbacher (1902) had already used the term diglossia to refer to this situation around the turn of the twentieth century.

One of the major voices in favor of Demotic Greek was Jean Psichari, who, in his article “Un pays qui ne veut pas de sa langue” (“A country that does not want its language”), also used the term diglossia to describe the debate surrounding the Greek language question at the time:

La diglossie – le fait pour la Grèce d’avoir deux langues – ne consiste pas seulement dans l’usage d’un double vocabulaire, qui veut qu’on appelle le *pain* de deux noms différents: *artos*, quand on est un homme instruit, *psomi*, quand on est peuple; la diglossie porte sur le système grammatical tout entier. Il y a deux façons de décliner, deux façons de conjuguer, deux façons de prononcer; en un mot, il y a deux langues, la langue parlée et la langue écrite, comme qui dirait l’arabe vulgaire et l’arabe littéral. (Psichari 1928: 66)¹

In this early definition, Psichari separates the term from its literal meaning of vocabulary duality and extends it to describe the existence of two linguistic systems that diverge considerably, although they still belong to the same historical language. His inclusion of all levels of structure as part of a diglossic situation and the claim that we are dealing with two nearly separate languages would remain at the center of subsequent studies on diglossia. This first extension of the concept of diglossia would be paralleled later by another extension to practically any situation where two languages are in a functional distribution, regardless of the degree of their genetic connectedness (Fishman 1967, 1980). Initially, however, Psichari, with the Demotic/Katharevousa situation in the foreground, perceived diglossia as the end result of the creation and maintenance of an artificial system, Katharevousa in this case, imposed by the literate elite on the rest of the population. His line of reasoning is that there

¹ “Diglossia – the fact that Greece has two languages – doesn’t consist in the mere use of vocabulary doublets, which means that *bread* is called by two different names: *artos*, when you are an educated man, *psomi* when one belongs in lower class; diglossia concerns the grammatical system as a whole. There are two ways to decline, two ways to conjugate, two ways to pronounce; in a word, there are two languages, the spoken language and the written language, say vulgar Arabic and written Arabic” (my translation).

was a vernacular form in use by all members of the community but the formulation of a norm, which is heavily inspired in the classical language and written sources and which is not the native language of any member of the speech community, created the diglossic situation between the standardized variety and the vernacular. This process is guided by what he calls “contempt for the spoken language,” on the one hand, and “fetishism of the classical language,” on the other (Psichari 1928: 72 and 76). The issue, then, is with writing the vernacular, not necessarily with its oral use in ordinary communication (see Chapter 7 for the role of writing in ending diglossia in the case of the Romance languages). After all, the vernacular is the native language of the entire society and it would be difficult to forcefully prevent its use (Psichari 1928: 118).

This is a key observation since, as I will argue throughout this book, a paradox is observed in the attitude and behavior of speakers in a diglossic situation: they degrade the very same language they faithfully transmit to their children, thus perpetuating the situation of which they consciously disapprove. Both the illiterate speakers, often the ones who are deprived of any possible gains associated with knowing the standard form, and the educated speakers share a negative perception of the vernacular. This is what I propose to call here the *diglossia paradox* and what I believe lies beneath the emergence and perpetuation of diglossic communities in cases where both varieties belong to the same historical language.² This paradox also explains the rapid changes in the vernacular and the static, and at times utterly cumbersome, structure of the standardized form. The resistance to include loanwords, naturally incorporated into the vernacular, is one feature that results from this paradox and may lead, for example, to forced morphological creations in the standard variety that are opaque and, at times, do not follow the morphological rules of the classical or the vernacular forms of the language. As an example, Psichari discussed the case of the Italian word *bomba*, which was adapted as *vomva* in Katharevousa Greek, given that the sound /b/ does not exist in classical Greek, leading to an impossible combination in this language of /mv/.

The application of the concept of diglossia to the Arabic situation, specifically in the context of North Africa, was first attempted by William Marçais (1930) in his article “La diglossie arabe” (“Arabic diglossia”). Marçais, who spent several years in Algeria and Tunisia around the turn of the twentieth century, established that the main characteristic of the language situation in the region was the existence of two varieties of Arabic playing against each other, the written form and the spoken vernacular. While the written form had not changed much since it was codified early on in the Islamic period, the dialects had gone through a considerable number of changes and, at the time of Marçais,

² The situation is different in cases of nongenetic diglossia, given that the indigenous language and the superposed one both have native speakers, even if they do not share the immediate same space.

4 Introduction

they were still not satisfactorily described. As a result, he argued that it was challenging for non-native speakers to learn the dialects, given the great degree of divergence existing between the different North African cities where he had lived. Yet, he knew that the written form was not of much use to the largest sector of the population, who, under the colonial rule, were deprived of access to education altogether.

Although he himself did not express a preference regarding which variety should be taught, Marçais presented diglossia as an irreconcilable situation, comparable to a two-headed monster. He argued that this situation was behind the difficulties faced by the colonial schooling programs with which he was very familiar in his positions as *Inspecteur général de l'enseignement des indigènes* and *Directeur de l'école supérieure de la langue et littérature arabes de Tunis*. He fervently summarized the situation and his position as follows:

Disons deux états d'une même langue, assez différents pour que la connaissance de l'un n'implique pas, absolument pas, la connaissance de l'autre; assez semblable pour que la connaissance de l'un facilite considérablement l'acquisition de l'autre. En tout état, un instrument pour l'expression de la pensée qui choque étrangement les habitudes d'esprit occidentales; une sorte d'animal à deux têtes et quelles têtes! Que les programmes scolaires ne savent trop comment traiter, car ils ne sont pas faits pour héberger les monstres. On conclura peut-être que le premier responsable de la crise de l'arabe, c'est l'arabe (Marçais 1930: 409).³

The anguish that Marçais seems to experience over this situation stems largely from the belief, in his “Western mind,” of what constitutes a favorable situation for education: a monolingual monodialectal population. This is reminiscent of the opinions by early observers of attempts to educate bilingual and bidialectal children in other societies, including native American, African American, and Hispanic students in the United States in the first half of the twentieth century. Whether or not the opinion of directly involved parties, such as Marçais, who for a period oversaw *l'enseignement des indigènes* and was fluent in Arabic, added to the marginalization of Arabic instruction under the French rule, the result was that illiteracy engulfed French North Africa during the colonial rule.

The position that a diglossic situation is unnatural or is not real continued to be at the center of the debate between advocates of a stronger Arabization and advocates of the standardization of the vernacular, although with much less graphic nature. As the French-educated elite accessed the positions of power

³ “Let’s say there are two states of the same language, different enough that knowing one does not, absolutely not, imply knowledge of the other; yet so similar that knowledge of one greatly facilitates the acquisition of the other. In any state, an instrument for the expression of thought that strangely shocks the habits of the Western mind; a kind of a two-headed animal – and what heads! – which school programs do not quite know how to treat because they are not made to accommodate monsters. One might conclude that perhaps the one primarily responsible for the crisis of Arabic is Arabic” (my translation).

immediately after independence of the Maghreb, they felt French was a more appropriate language for education and avoided dealing with the diglossic situation altogether, at least initially. A further discussion of the situation of Arabic and diglossia in the Maghreb is presented in Chapter 3.

A full-fledged theory of diglossia was spelled out thirty years later in the work of Charles Ferguson. In his article, “Diglossia,” Ferguson (1959a: 336), who was particularly familiar with the situation of Arabic and had carried out extensive research on Arabic dialects, formulated the following 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.*⁴

To develop this theory, Ferguson based his observations on four separate situations that became known in the literature as the defining cases of diglossia: Arabic, Modern Greek, Swiss German, and Haitian Creole. After introducing the labels of H variety (the superposed variety) and L variety (the vernacular form), he presented a list of nine features that he argued identify a diglossic situation as such.

First, and foremost, each of the two varieties has a set of functions distributed across the different domains. The H variety is the one used in educational, administrative, and religious discourse, while the L variety is often restricted to personal circles and transactional interactions. Second, the H variety is highly valued by the members of the community, whether they are fully competent in it or not. As a matter of fact, it is considered to be the “real” language they speak or should be speaking while the vernacular is considered a transitory and corrupt version of it and one that came about primarily because of lack of education. Third, the entirety of the literary body that a community records in the written form is in the H variety. Oral poetry and narratives may be produced in the vernacular as long as they are not recorded or published in written form. Fourth, the L variety is the variety that is transmitted naturally and children start to formally learn the H variety only when they start school. The L variety is the native language of all members of the community while nobody has native competence in the H variety. The majority of the speakers never attain higher levels of competence in H variety, especially in societies with high rates of illiteracy. Fifth, the H variety is fully described and dictionaries and grammar books are available for its teaching and learning while the L variety remains under-described. Speakers often have the impression that the grammar of the L

⁴ Italics used in the original text.

6 Introduction

variety is the same as that of the H variety but obscured and corrupted by the lack of knowledge of the “correct” forms. Sixth, diglossia is a stable phenomenon that does not change over a few generations. Both varieties can exist alongside each other for long periods of time. The H variety sees little change from when it was codified based on the written sources while the L variety continues to evolve like any other language that is acquired and transmitted naturally. Seventh, the grammar of the L variety is considerably less complex than the H variety’s grammar, which tends to follow a set of conservative rules that reflect the high attention paid to it through the centuries by generations of prescriptive grammarians. Few people have native-like competence in the H variety and profound familiarity with its grammar is interpreted as an index of high levels of education. Eighth, Ferguson states that the largest part of the vocabulary is shared between both varieties, although the H variety has a more extensive technical and specialized lexicon. In addition, Ferguson mentions the existence of “lexical doublets,” where the two varieties have settled on different terms for the same concept. Finally, at the phonological level, both varieties share “a single inventory,” although the L inventory may show simplification on the one side and interference from other languages on the other. These features will be discussed again as different case studies are introduced in the following chapters.

1.2 Diglossia extended

Ferguson’s definition of diglossia and its features became known in the literature as classical diglossia. Its major hallmark is the fact that, in these situations, the varieties involved are from the same historical language, as in the case of Arabic, or closely related ones as in the case of French and Haitian Creole. Without this premise, several of the characteristics he described for diglossia, especially at the structural and historical levels, as opposed to the functional aspect, will not hold. This was the case when another major amendment to the theory of diglossia was introduced by Joshua Fishman: extended diglossia.

Fishman (1967) argued that functional distribution is the most critical aspect of a diglossic situation and, as such, the term should be extended to include situations with separate languages provided they are in a complementary distribution. He distinguished between bilingualism, as the individual ability to use more than one language, and diglossia as the social reality of the languages in use within the same speech community:

bilingualism is essentially a characterization of individual linguistic behavior whereas diglossia is a characterization of linguistic organization at the socio-cultural level (Fishman 1967: 34).

He proposed that bilingualism can exist with or without diglossia and diglossia can exist with or without bilingualism (Fishman 1967: 30). In cases of

bilingualism without diglossia, speakers have competence in both languages but are able to use either one of them in a wide variety of domains without rigid compartmentalization. In cases of bilingualism with diglossia, speakers have competence in both languages but the usage of each of them is restricted to a set of functions stipulated by the sociocultural powers in place. Fishman discussed the case of Paraguay, where Spanish, an Indo-European language, and Guaraní, a Tupí language, exist in what he described as a situation of bilingualism with diglossia. While Paraguay is one of the countries with the highest rates of societal bilingualism, Spanish is largely considered the language that enjoys higher prestige and whose use is closely associated with the dominant socio-cultural and economic agents.

Since Fishman's proposal, studies on diglossia multiplied (for extensive bibliographical references see Fasold 1984; Fernández 1993; Kaye 2001; Hudson 1991, 1992, 2001, 2002a, 2002b). As the remaining features of classical diglossia, other than functional compartmentalization, seemed to be put aside, a wide range of proposals suggested the possibility of interpreting many cases of bilingualism and multilingualism where a complementary distribution is observed as cases of diglossia, with, at times, some changes in the label. Platt (1977) introduced the term *polyglossia* when he described the situation of the multilingual English-educated Chinese population residing in Malaysia and Singapore. He postulated that, in the case of Singapore, this segment of the population has access to two H varieties (Formal Singapore English and Mandarin), two of what he calls "Dummy H" (DH) varieties, referring to Standard Malay and Tamil which are perceived as prestigious but which are not often used by this group, one Medium variety (M) defined as Colloquial Singapore English, and a number of L varieties including Hokkien, Cantonese, other Chinese dialects and Bazaar Malay. Mkilifi (1978) used the term *double overlapping diglossia* to describe the situation in Tanzania where local vernaculars are used alongside Swahili and English. In these cases, the same variety can play the role of the H and L varieties depending on which other varieties it is compared to. In Tanzania, Swahili is the L variety when it is in distribution with English, but it becomes the H variety when it is in distribution with the local vernacular, all depending on the setting in which the languages are used and the particular status of the speakers. Romaine (1989: 34) used the term *triglossia* to describe the Tunisian situation in particular, referring to the existence of two H varieties (Standard Arabic and French)⁵ and one L variety (Tunisian Arabic):

⁵ Although a difference will be made between Classical Arabic, the language of the religious texts and classical written sources, and Modern Standard Arabic, the modern variety of Arabic used in education and mass media, throughout this volume I will use Standard Arabic to refer to the H variety of Arabic regardless of whether it is Classical Arabic or Modern Standard Arabic.

8 Introduction

There are cases in which societies have two high varieties in conjunction with a single low, or so called triglossia. In Tunisia, for example, Classical and Tunisian Arabic are in diglossic distribution, but French is also used, so the three varieties are in a functional distribution.

The term has to a certain extent become a blanket term for uncounted cases of bilingualism and multilingualism regardless of the nature of the genetic relatedness of the varieties involved. In fact, given the predominance of bilingualism in many parts of the world, with different languages often serving different socio-cultural functions, it might be argued that under extended diglossia, the largest part of the world population lives in some sort of diglossic situation (Winford 1985). It is true, however, that while Fishman calls for a unified theory of societal bilingualism that includes classical diglossia as part of it, he makes it clear that, in his model, “not every instance of societal multilingualism is diglossic. Far from it” (Fishman 2002: 97). Nevertheless Ferguson’s stress that speakers in a classical diglossic situation must perceive both varieties as the same language gives genetic relatedness a defining role, as he reiterated in 1991:

My intention was that the users would always view the two as the same language: I excluded 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).

1.3 Diglossia and language contact

In numerous studies, the objective was to add one more case to the list of diglossic situations without necessarily evaluating the difference in the implications to the languages involved depending on the genetic distance between them. In fact, and interestingly enough, diglossic situations are not always approached as contact situations given that claimed functional compartmentalization allows for the perception that the varieties are indeed separate.

The relevance of diglossia to contact linguistics was explored by Winford, who applied the concept to the Caribbean creole continua (Winford 1985). He shares the observation that the term has been stretched excessively since its original formulation by Ferguson and raises the concern that under extended diglossia “the question of the structural or genetic relationship between the codes becomes secondary” (Winford 1985: 346). This is a concern that has been voiced by several scholars and remains at the heart of the question about what role the degree of genetic connectedness should play in determining if a situation is one of diglossia. Taking the functional distribution of any two linguistic systems as the exclusive yardstick to label a situation as diglossic would limit the usefulness of this concept. As Winford formulates the concern:

Since the functions of any two languages controlled by a single speaker are almost always in a partial or total complementary distribution, clearly then, to equate diglossia with bilingualism is not very useful (Winford 1985: 346).

The fact that we are dealing with two or more distinct languages in cases of societal bilingualism, applying the concept of diglossia is less productive for understanding the mechanisms and outcomes of language contact under both diglossia and bilingualism. One major difference that will be explored throughout the current volume is that, not only do the distribution and the domains of usage of the H and the L varieties differ considerably in diglossia and societal bilingualism, but also the mechanisms of language contact, such as code-switching, are different. The nature and frequency of bilingual code-switching is different from diglossic code-switching, i.e., between varieties of the same language (see Chapter 4). And so, by implication, the type of language change that may happen as a result of contact between the H and L varieties is expected to be different. Finally, what Fishman described as diglossia with bilingualism does not in fact capture a situation such as the one in North Africa, where both classical diglossia and bilingualism exist, and does not offer a more precise theoretical framework that better depicts the situation and its possible outcomes.

The fact that in diglossic cases, the speakers are aware of the relatedness of their L variety to the H variety plays a role in their attitude towards both varieties and how they go about usage and transmission. Bilingual speakers are, on the other hand, fully aware that their languages are separate languages, each with its own history and standing. This is also true in the case of creole languages as in one of the defining cases referred to by Ferguson, French and Haitian Creole. DeJean (1993) argues that the relationship between Haitian Creole and French is categorically different from the other defining cases discussed by Ferguson, mainly because the contact between the two varieties in each of the other cases was never broken. For the majority of Haitians, French remains a foreign language that, unlike other H varieties, they cannot refer to as “theirs.” In fact, recent arguments against the exceptionality of creole languages (Mufwene 2001, 2005, 2008; DeGraff 2003, 2005, 2009) established that creole languages are languages in their own right and not failed attempts at learning the lexifying languages. As such, the Haitian Creole/French situation is not similar to the Greek, Swiss German, or Arabic situations (Coulmas 2002). This awareness of the degree of relatedness of the two codes in diglossic situations, as reiterated in my interviews with Maghrebi speakers who perceive Standard Arabic and their dialects as being practically the same language, highly affects the speakers’ attitude towards interference from third languages in either H or L varieties and, as a result, the direction of contact-induced changes in cases of diglossia and bilingualism.

Eckert (1980) proposed that diglossia could lead to language shift in cases where two different languages are involved, as happened in her case study of Gascon and French in southern France. Her argument is that diglossia is not a stabilizing mechanism for language maintenance but rather that “diglossia actually organized the shift” (Eckert 1980: 157). In her opinion, as one of the languages encroaches upon domains where the other variety is used, the community gradually shifts away from the less prestigious variety. Eckert argues that public education in France is the way French started spreading into the domains of use of the vernacular and eventually led to monolingualism in the majority of the regions in France. While the scenario of Gascon/French is a clear example of how a case of extended diglossia can end up with the installment of the H variety as the language of all domains of communication and the gradual disappearance of the L variety, this does not necessarily apply to cases of classical diglossia. In fact, this argument takes a stab at a central defining feature of classical diglossia: stability. In cases of classical diglossia, the fact that the H variety is one that is not transmitted naturally and whose use is extremely artificial makes it impossible for the H variety to take over the domains of the L variety. Thus, the scenario described by Eckert applies to extended diglossia but not to cases of classical diglossia.

In classical diglossia, the domains shift in an opposite direction than they usually do in societal bilingualism. This allows us to distinguish between two cases of shift: *diglossic shift*, where the vernacular variety of the language sees its use extended to previously off-limit domains, and *language shift*, where the superimposed language gradually displaces the local variety, usually an indigenous or immigrant one. This being so, given the notion of *diglossia paradox* introduced above and the fact that in cases of bilingualism, the H variety is in fact the native language of the socioculturally dominant group. The shift is slower in cases of diglossic shift unless there is an abrupt rupture with the H variety, usually in the form of military intervention such as the Reconquista in Spain or the Norman occupation of Malta, and a stronger unrelated H variety is introduced, Castilian Spanish in Al-Andalus and Sicilian in Malta (see Chapter 5, 6, and 7 for a detailed discussion of these two cases). In cases of language shift, the shift can happen across a generation or two, most often prompted by subtractive bilingualism. The key then for an L variety in a classical diglossic situation to become a recognized language in its own right is its distancing from the H variety to the degree that the speakers do not have access to H variety any longer and stop seeing their native L variety and the historically H variety as “the same language.” In addition to strong nationalistic beliefs and changing sociocultural values increasingly associated with the L variety, the H variety could be displaced by another unrelated H, freeing the way for the L variety to establish itself outside of the previous diglossic situation as was the case for Maltese Arabic.