Chapter 2

The genesis of pidgin and creole languages: A State of the Art*

The paper reproduced as the content of this chapter first appeared in a collection put together by Christine Jourdan and Kevin Tuite (2003) (Lefebvre 2003a). The appendix to this chapter, Appendix 1, entitled "A research program on PC genesis for the 21st century" consists of a section of a larger paper, "The field of pidgin and creole linguistics at the turn of the millennium: The problem of the genesis and development of PCs", prepared for the symposium on Pidgin and Creole Linguistics in the 21st Century *held in New York (January 1998) as part of the annual Linguistic Society of America meeting, and published in Gilbert (2002) (Lefebvre 2002).*

The most intriguing question about pidgin and creole languages is no doubt that of how they come about. This question is also a most important one for this question hinges on several disciplines, as will be shown throughout this paper. Due to the circumstances in which they develop, pidgins and creoles constitute an extreme case of languages in contact. Their creation involves second language acquisition, and their development, first language acquisition. The emergence of pidgin and creole languages constitutes a particular case of accelerated linguistic change; it is thus a goldmine for historical linguistics. Two major processes involved in the creation and development of these languages, relexification and reanalysis, are cognitive processes; hence, the study of the emergence of pidgin and creole languages is relevant to cognitive sciences. The other process that plays a major role in the development of pidgins and creoles, levelling, is a social process; hence, the study of the emergence of pidgin and

creole languages is of interest to socio- and ethno-linguistics. This paper is concerned with the various linguistic dimensions of the emergence of pidgin and creole languages.

The problem of the origin of pidgin and creole genesis has been addressed from different points of view over the last century. This paper summarises and evaluates the various theories proposed to account for the genesis of these languages. The paper begins with a discussion of features that any theory aiming at explaining the origin of pidgin and creole languages must be able to account for (section 2.1). The next three sections evaluate competing theories of creole genesis against these features. Section 2.2 reviews theories that share the property of having language varieties as an object of study. Section 2.3 presents the first attempts to shift the focus of pidgin and creole genesis studies from the language varieties to the processes involved in the formation of these languages. Section 2.4 reports on a major long term research project, carried out at the Université du Québec à Montréal, that aimed at providing a unified account of pidgin and creole genesis within the framework of the processes that otherwise play a role in language creation and language change in general. Section 2.5 concludes the paper and provides references identifying avenues for future research on the topic. Appendix 1 provides additional discussion of topics for future research.

2.1. The complex problem of pidgin and creole genesis

The history and structure of PCs are characterised by a number of features.[1] Any theory that seeks to explain the origin of these languages must be able to account for this basic set of features.

First, as was pointed out by Whinnom (1971) these languages are only developed in multilingual communities. Whinnom argues that, in bilingual communities, the speakers of one group will eventually learn the language of the other group.[2]

Second, communities where PCs emerge generally involve several substratum languages whose speakers make up the majority of the population and a superstratum language spoken by a relatively small but economically powerful social group. Crucially, the substratum community does not have one common language. This situation creates the need for a *lingua franca* (see e.g. Foley 1988; Hymes 1971; etc.), not only to permit communication between the speakers of the substratum languages and of the superstratum language, but also to permit the speakers of the substratum languages to communicate among themselves (see e.g. Foley 1988; Singler 1988: 47; Thomason and Kaufman 1991).

Third, in communities where PCs emerge, speakers of the substratum languages generally have very little access to the superstratum language (see Thomason and Kaufman 1991). As Foley (1988: 163) puts it: "the language of the dominant group is not easily made available to the members of the subordinate group(s)." In fact, as has been pointed out on several occasions in the literature, creoles that most resemble their superstratum languages were created in communities where the speakers of the substratum

languages had relatively more access to the superstratum community. Creoles that are more radical (i.e. less like the superstratum language) come from communities where language learners had very little access to the superstratum community (see e.g. Andersen 1983; Baker 1993; Baker and Corne 1982; Bickerton 1977: 55; Thomason and Kaufman 1991; Valdman 1978, 1993). For example, as is argued in Valdman (1993), Louisiana Creole is closer to French than Haitian is because the substratum speakers had more access to French in Louisiana than the African population had in Haiti. Baker and Corne (1982) also discuss this issue on the basis of data from Mauritius and Reunion creoles. On Reunion, French native speakers outnumbered substratum speakers during the formative period of the creole and Reunion Creole grammar displays a significant number of French grammatical categories. By contrast, during the formative period of Mauritius Creole, the proportion of native French speakers was much lower and thus the West African speakers had a much stronger input into the creole. Likewise, during the period where Haitian Creole was formed, the proportion of West African speakers was much higher than that of French speakers (see Singler 1996), such that West African speakers had a very strong input into the creole (see Lefebvre 1998a, and the references cited therein).

A fourth point is that, ordinarily, languages change gradually. Within the span of several generations, speakers of innovative and conservative dialects are able to communicate, even though, over the course of centuries, a new language may evolve (see Lightfoot 1979). By contrast, PCs are created in a relatively short span of time (see e.g. Alleyne 1966; Bickerton 1984; Chaudenson 1977, 1993; Hall 1958; Voorhoeve 1973). This observation

dates back to Van Name (1869–70: 123, cited in Goodman 1964: 135): "Under ordinary conditions these changes proceed at so slow a pace as to be appreciable only at considerable periods of time, but here two or three generations have sufficed for a complete transformation." Hesseling (1933: xi) further reassesses this point in the following terms:

> The genesis of human language is a psychological problem that no single language will ever solve, but from creole one can best learn how a given language emerges from old data and develops, because here something takes shape at a high speed, in a past recognisable to us, something which is the product, in other cases, of many centuries, with a very obscure past in its background.

Thus, in contrast to regular cases of linguistic change, PCs diverge abruptly from their source languages (see Thomason and Kaufman 1991), such that within one or two generations, a different language is created. Hancock (1987: 265) claims that: "most of the principal characteristics that each creole is now associated with were established during the first twenty-five years or so of the settlement of the region in which it came to be spoken." Ferraz (1983), Hymes (1971) and Mintz (1971) suggest that a creole can develop within fifty years or less. Singler (1996) is of the opinion that it takes sixty to eighty years for a creole to form. Whatever the outcome of this issue may be, PCs constitute a unique case of accelerated linguistic change when compared with regular cases of linguistic change.

Fifth, PCs tend to be isolating languages. This observation goes back to Hesseling (1933: xvi) and Schuchardt (1979). It is also found in Hagège (1985: 39). But it was Mufwene (1986, 1990, 1991) who clearly established

this property of PCs and the problem it poses for scholars who work on PC genesis. Indeed, Mufwene has documented the fact that this tendency appears to hold even when the contributing languages are not isolating languages. For example, Mufwene (1986) shows that Kituba, a creole language that has emerged almost exclusively from contact among agglutinative Bantu languages, is an isolating language. "Kituba has selected Kikongo's seemingly marked periphrastic alternative over the more common and apparently unmarked agglutinating system" (Mufwene 1990: 12).

Sixth, it has long been noted in the literature that PCs derive some of their properties from those of the substratum languages and some from those of the superstratum language (see e.g. Alleyne 1966, 1980; Holm 1988; etc.). Moreover, several scholars have noticed that the respective contribution of the substratum and superstratum languages to a creole is not random. For example, Adam (1883: 47) states that:

> J'ose avancer... que les soi-disant patois de la Guyane et de la Trinidad constituent des dialectes négro-aryens. J'entends par là que les nègres guinéens, transportés dans ces colonies, ont pris au français ses mots, mais qu'ayant conservé dans la mesure du possible, leur phonétique et leur grammaire maternelles… Une telle formation est à coup sûr hybride… La grammaire n'est autre que la grammaire générale des langues de la Guinée.
>
> [I go so far as to claim… that the so-called patois of Guyana and Trinidad constitute Negro-Aryan dialects. By that I mean that the Guinean Negroes who were transported to the colonies adopted the words of French but, as much as possible, kept the phonetics and

> grammar of their mother tongues… Such a formation is clearly hybrid… The grammar is no different from the general grammar of the languages of Guinea.]

Speaking of Haitian Creole, Sylvain (1936: 178) observes that: "Nous sommes en présence d'un français coulé dans le moule de la syntaxe africaine, ou (...) d'une langue éwé à vocabulaire français." [We are in the presence of a French that has been cast in the mould of African syntax or … of an Ewe language with a French vocabulary.] Similarly, in his extensive study of French-based creoles, Goodman (1964) observes, over and over again, that particular lexical items in the creoles have a phonological representation similar to a French expression but that these creole lexical items share properties with corresponding lexical items in the African substratum languages. On the basis of data drawn from Djuka, Huttar (1975: 684) also remarks that "the use of morphemes borrowed by a pidgin or a creole language (...) from a European language often diverges from the use of the source morpheme in the source language" and often corresponds to the use of the corresponding word in the substratum languages. Voorhoeve (1973) makes a similar remark on the basis of Sranan and Saramaccan data. Writing about Solomons Pidgin, Keesing (1988: 1–2) remarks:

> I had earlier been struck, when I had learned Solomons Pidgin in the 1960s through the medium of Kwaio, an indigenous language I already spoke fluently, that this learning task mainly required learning Pidgin equivalents of Kwaio morphemes. The syntax of Solomons Pidgin was essentially the same as the syntax of Kwaio, although

> somewhat simpler and lacking some of the surface marking; in most constructions, there was a virtual morpheme-by-morpheme correspondence between Kwaio and Pidgin. (…) Although most of the Pidgin lexical forms were ultimately derived from English, I found this largely irrelevant to my language-learning task. The semantic categories they labeled corresponded to Kwaio ones, not English ones; grammatical morphemes corresponded to Kwaio ones, not English ones. Thus semantically Pidgin *dae* corresponded directly to Kwaio *mae* 'be dead, die, be comatose, be extinguished,' not to English *die*. Pidgin *baebae* corresponded to the Kwaio marker of future/nonaccomplished mode, *ta-*, not to English *by and by*.

These observations suggest that PCs are not formed by an arbitrary mixture of the properties of the languages present at the time they are being created. The general pattern that seems to emerge from the observations reported above is the following: while the forms of the lexical entries of a PC tend to be derived from the superstratum language, the syntactic and semantic properties of these lexical entries tend to follow the pattern of the substratum languages.

Any theory of PC genesis must account for the properties of these languages. Therefore, as has been pointed out in Lefebvre and Lumsden (1989), an optimal theory must account for the fact that PCs emerge in multilingual contexts, where there is a need for a *lingua franca*, and where the speakers of the substratum languages have little access to the superstratum language. It must account for the fact that PCs tend to be isolating languages even when they emerge from contact situations involving

only agglutinative languages. It must also account for the fact that PCs manifest properties of both their superstratum and substratum languages and it must explain why these properties are divided the way they are. Finally, an adequate theory of PC genesis must be stated in terms that are explicit enough so as to be falsifiable.

2.2. Competing theories of pidgin and creole genesis whose object of study consists of language varieties

This section summarises approaches to PC genesis that share the characteristic of having language varieties (as opposed to processes that lead to these varieties) as their object of study. The following proposals will be discussed in turn: The theory according to which PCs constitute reduced linguistic codes; the theory advocating that creole languages are nativised pidgins; the theory according to which creoles are imperfect second language varieties of their lexifier languages; the theories advocating that PCs consist of restructured varieties of their substrate and superstrate languages, respectively; and finally, the theory that creoles consist of language varieties reflecting the properties of Universal Grammar. (This last theory is referred to in the literature as the Language Bioprogram Hypothesis.) Following the methodology in Lefebvre and Lumsden (1989), each of these approaches will be evaluated on the basis of the set of features enumerated in section 2.1.

A methodological proviso is in order at this point. I do believe that a sound theory of PC genesis must be able to account for all the seven features discussed in section 1. The theories that I evaluate on the basis of these features may, however, have not been formulated so as to account for this

particular set of features. My evaluation of each theory is nevertheless based on whether the theory under review, in its current formulation, may account for the particular features in my list, regardless of whether the proposer meant to account for these or not. In adopting this methodology, I do not evaluate each theory for its own sake. However, I do provide the reader with a means of evaluating the various theories against a single set of criteria.

2.2.1. The theory of pidgins and creoles as reduced codes

This theory holds that speakers of the substratum languages were presented with a reduced (baby-talk or foreigner-talk) version of the superstratum language characterised by an absence of functional categories such as gender, number, case, etc. In this view, the plantation owners were voluntarily speaking a reduced version of their own language in order to maximise communication with the slave population. Bloomfield (1933), Göbl-Galdi (1934), Hall (1966), Jespersen (1922: 233) and Schuchardt (1909) (as translated by Goodman 1964: 124) all hold some version of this view. Similar proposals have also been made more recently. For example, Ferguson (1971: 147) advocates the view that "the foreigner talk of a speech community may serve as an incipient pidgin. This view asserts that the initial source of the grammatical structure of a pidgin is the more or less systematic simplification of the lexical source language which occurs in the foreigner talk registers of its speakers, rather than the grammatical structure of the language(s) of the other users of the pidgin." (See Naro 1978, for an extensive discussion of this issue.) Similarly, Foley (1988: 166) writes: " I suggest that a pidgin is a version of a foreigner talk of a superstratum

community that has been conventionalised and accepted, most importantly by speakers of the substrate language(s)."

Does this approach meet the criteria of an adequate theory of PC genesis as outlined in section 2.1? This approach does not explain why PCs only develop in multilingual communities (the first feature). It does not explain the need for the substratum speakers to develop a *lingua franca* (the second feature). It does not account for the fact that these languages are formed quite rapidly (the fourth feature), nor for the type of mix that these languages manifest (the sixth feature). However, this approach may be considered to provide an account of the fact that the substratum population has little access to the superstratum language in situations where PCs emerge (the third feature), and of the lack of inflectional morphology, thus of the isolating character of these languages (the fifth feature). Finally, the theory of reduction is formulated in terms that are precise enough so as to be falsified (the seventh criterion). The next paragraph shows how this approach to PC genesis can be falsified and how it is in fact falsified.

According to some of the proponents of this approach, creole languages would lack the functional category lexical entries of their superstratum language because the speakers of the substratum languages were not presented with these lexical items. This claim can be falsified if it can be shown that the speakers of the substratum languages of a creole were in fact presented with the pertinent data. As has been pointed out in Lefebvre (1998a: 62–65), in a discussion concerning the origin of Haitian Creole, while it could well be the case that French speakers did not use a very elaborate style of French while talking to the African population in Haiti, there is plenty of evidence from the Haitian lexicon that the speakers of the

substratum languages were exposed to the functional category lexical items of French. Indeed, Valdman's *et al*. (1981) dictionary abounds in examples where a Haitian word corresponds to a French expression that includes a French functional item. As is shown in (1) (from Lefebvre 1998a: 64), a simple Haitian lexical entry may contain an agglutinated French determiner (e.g. *la*, *au*), as in (1a), an agglutinated French partitive determiner (e.g. *du*), as in (1b), an agglutinated French complementiser (*que* 'that'), as in (1c), an agglutinated French functional item *à*, as in (1d), or even an agglutinated French conjunction (e.g. *et*, *ou*), as in (1e).

(1)

	HAITIAN LEXICAL ENTRY		FRENCH EXPRESSION	
a.	*larivyè*	'river'	*la rivière*	'the river'
	olye	'instead'	*au lieu*	'instead'
b.	*diri*	'rice'	*du riz*	'rice'
	dife	'fire'	*du feu*	'fire'
c.	*fok*	'complementiser'	(*il*) *faut que*	'there must be'
	tandiske	'while'	*tandis que*	'while'
d.	*afòs*	'by means of'	*à force de*	'by means of'
	apati	'starting from'	*à partir de*	'starting from'
e.	*epi*	'and'	*et puis*	'and then'
	oubyen	'or'	*ou bien*	'or'

These examples, and many more, show that the African population in Haiti must have been exposed to forms containing French functional items. My conclusion, based on Haitian, is in line with Alleyne (1971: 170), who states that the African population of the Caribbean area was exposed to European languages "in their full morphological and syntactic forms." The above data constitute a major drawback to the claim that a PC lacks the functional categories of its superstratum language because the substratum speakers were not exposed to these categories. Instead, they show very clearly that the creators of Haitian, and presumably of other PCs, were exposed to the superstratum functional categories, but that they did not identify them as

such because they did not have enough exposure to the superstratum language.

2.2.2. The theory of creoles as 'nativised pidgins'

The idea that creole languages are nativised pidgins emerged during the late sixties and developed in the seventies. In this approach, a pidgin language is no one's first language. It is a reduced language variety serving as a *lingua franca* in a multilingual community. A pidgin that came to be spoken as the first language of a generation of speakers is said to have undergone nativisation. A nativised pidgin is called a creole. From a linguistic point of view, the nativisation of a pidgin is often seen as being accompanied by expansion or complexification of the source pidgin, the latter being claimed to acquire all the characteristics of a natural language in the process of nativisation (see e.g. Bickerton 1981; Hymes 1971; Labov 1971; Sankoff 1971; Sankoff and Laberge 1973, etc.).[3]

This theory accounts for a number of the characteristics listed in section 2.1. It accounts for the fact that pidgins emerge in multilingual communities (the first feature), for the fact that the members of communities where pidgins emerge are in need of a *lingua franca* (the second feature), and for the fact that creoles emerge rapidly, in this case in one generation (the fourth feature). However, it does not account for the fact that substratum speakers have little access to the superstratum language (the third feature), nor for the isolating character of pidgin and creole languages (the fifth feature), nor for the type of mix that pidgins and creoles manifest with respect to their source languages (the sixth feature). Finally, as has been pointed out in Lefebvre and Lumsden (1989: 253), without linguistic criteria

distinguishing between pidgin and creole languages, the theory is not falsifiable (the seventh criterion).

Precise definitions of pidgin and creole languages are desirable at this point. Pidgins and creoles have long been considered as separate entities on the basis of the following two sets of criteria. While pidgins have been defined as reduced codes, creoles have been defined as expanded versions of these reduced codes (see e.g. Hymes (ed.) 1971, and the references therein). Also, while pidgins have been found to often constitute the second language of the speakers who use them, a creole is considered to be a pidgin that has become the first language of a new generation of speakers, as we saw above (see also Kay and Sankoff 1974; Sankoff and Laberge 1973). In more recent literature, the distinction between pidgins and creoles has been levelled out in view of the fact that there are some pidgins (still used as a second language) that have been shown to have expanded in the same way as languages known as creoles (see e.g. Mühlhäusler 1980, 1986a, for an extensive discussion of this point). Hancock (1980: 64) states: "I prefer not to acknowledge a distinction between *pidgin* and *creole*, and to consider *stabilisation* more significant than *nativisation* in creole language formation." Similarly, Mufwene (1990: 2) uses the term creole to refer "to varieties traditionally called creoles but also to those called pidgins that serve as vernaculars or primary means of communication for at least a portion of their speakers." Moreover, in recent literature in the field, scholars have started referring to pidgins and creoles as PCs, suggesting that they fall into a single category. Furthermore, pidgin and creole languages cannot be distinguished on the basis of the processes that play a role in their formation (see Woolford 1983, for a general discussion of this point). Indeed, the

processes hypothesised to play a role in the formation and development of human languages apply to both pidgins and creoles (see Keesing 1988; Lefebvre 1998a, and the references therein). Since these languages cannot be distinguished on the basis of these processes, no distinction should be made between them (see Lefebvre 1998a: 4). As will be seen further on, this will turn out to be a major drawback to Bickerton's (1981) Language Bioprogram Hypothesis which requires that pidgins and creoles be different and separate entities produced by different processes.

2.2.3. The theory of PCs as crystallised varieties of 'imperfect' second language acquisition

According to the theory of imperfect second language acquisition of PC genesis (see e.g. Alleyne 1971, 1980; Andersen 1980, 1983; Chaudenson 1979, 1993; Mufwene 1990; Schumann 1978; Thomason and Kaufman 1991; Valdman 1980), PCs constitute the crystallisation of an imperfect version of the acquisition of a second language. In this view, the speakers of a hypothesised proto-creole lacking sufficient access to the colonial language data which they were exposed to would have created an approximate simplified system of the type of that found in some cases of second language acquisition.[4]

Does this theory account for the features of PCs enumerated in section 2.1? While this theory of creole genesis accounts in a straightforward way for the fact that speakers of the substratum languages do not have much access to the superstratum language in contexts where pidgin and creoles emerge (the third feature), it fails to account for several of the other characteristics enumerated in section 2.1. It does not account for

the fact that PCs emerge only in multilingual communities (the first feature), nor for the fact that these communities need a *lingua franca* (the second feature), nor for the fact that PCs are created rather rapidly (the fourth feature), nor for the fact that PCs tend to be isolating languages (the fifth feature). More importantly, this theory does not provide an explanation for why PCs have 'crystallised' in the way they have with respect to their source languages (the sixth feature). Finally, this theory does not satisfy the seventh criterion, as it does not appear to be falsifiable. As Lefebvre and Lumsden (1989: 254) point out, this theory "ne définit pas précisément les mécanismes d'acquisition d'une langue seconde, ni en quoi elle diffère de l'acquisition d'une langue maternelle." [does not define in a precise way the mechanisms of second language acquisition, nor its difference with first language acquisition]. Hence, it is not falsifiable (but see section 2.4 for refinements of this approach).

2.2.4. The theory of PCs as restructured varieties

Several proposals may be regrouped under the view that PCs are restructured varieties. The following discussion is organised around three major proposals: PCs as restructured substratum varieties, PCs as restructured superstratum varieties, and PCs as restructured varieties of both of their source languages.

2.2.4.1. PCs as restructured substratum varieties

The substratist theory of creole genesis postulates that Caribbean creole languages have emerged by means of the gradual transformation of

the West African languages (spoken by the slaves) influenced by the European colonial languages (see e.g. Alleyne 1980; Holm 1988).

How does this theory meet the criteria in section 2.1? This theory may account for the fact that creoles only emerge in multilingual communities where there is a need for a *lingua franca*, and where speakers of the substratum languages have little access to the superstratum language (the first three characteristics in section 2.1). However, it does not provide an explanation for why creole languages are created in a relatively short period of time, nor for why they tend to be isolating languages (the fourth and fifth characteristics). While the postulated gradual transformation of the substratum languages influenced by the colonial languages does account for the contribution to the creole of both the substratum and the superstratum languages, it does not predict the principled respective contribution of these languages to the creole (the sixth characteristic). Finally, this theory is not falsifiable (the seventh characteristic), for, as has been pointed out in Lefebvre and Lumsden (1989: 254), it does not account for the facts that distinguish the emergence of PCs from cases of regular change occurring in languages (or in language contact varieties) that are not known as PCs.

2.2.4.2. PCs as restructured superstratum varieties

The superstratist theory of PC genesis holds that PCs constitute restructured dialects of their superstratum language. For example, in this view, French-based creoles would constitute restructured dialects of French, and so on and so forth. The main advocate of this approach is Chaudenson (1973, 1983, 1992).

This approach may be said to account, to a certain point, for the fact that speakers of the substratum languages have little access to the superstratum language (the third feature). It does not account, however, for the fact that creoles only emerge in multilingual communities, where there is a need for a *lingua franca* (the first two features). It does not account for the relatively rapid formation of PCs (the fourth feature) nor for the fact that PCs tend to be isolating languages (the fifth feature). Furthermore, and more importantly, it does not account for the principled division of properties of PC lexicons between their source languages (the sixth property). Is this theory falsifiable? This theory is falsifiable on the following grounds. As is extensively argued in Lefebvre (1998a, 2001d), PCs tend to reproduce the semantic and syntactic features of their substratum languages and hence, from a typological point of view, they pair with their substratum languages rather than with their superstratum languages. On this view, Atlantic creoles tend to reproduce the features of their West African substratum languages (see e.g. Lefebvre 1998a, and the references therein), whereas Pacific creoles tend to reproduce those of their Austronesian substratum languages (see e.g. Keesing 1988; Sankoff 1991). Thus, from a typological point of view, PCs resemble their substratum languages in spite of the fact that the phonological representation of their lexicons are derived from their respective superstratum languages. For example, as has been demonstrated in detail in Lefebvre (1998a), although the bulk of the phonological representations of Haitian words are derived from French, the typological features of Haitian pair with West African languages, not with French which shares features with Romance languages. This situation argues that Haitian Creole cannot be considered to be a dialect of French, for dialects of a given language are

expected to share typological features. In my view, this situation falsifies the theory according to which PCs constitute dialects of their superstratum language. (For further discussion of this issue, see also Mufwene 1996b: 166.)

2.2.4.3. PCs as restructured varieties of both of their source languages

The idea that pidgins and creoles constitute restructured varieties of their substratum or superstratum languages has given rise to a recent collection of papers edited by Neumann–Holzschuh and Schneider (2000) under the title *Degrees of restructuring in creole languages*. This volume contains various papers presenting case studies of pidgins and creoles analysed as restructured varieties. Some papers propose that creoles are restructured varieties of both their substratum and their superstratum sources (see e.g. Alleyne 2000; Chaudenson 2000). According to some authors, creoles may vary with respect to degrees of restructuring (see e.g. Baker 2000; Holm 2000; Winford 2000).

In their introduction to the collection, the editors point out the confusion regarding the definition of 'restructuring' and related concepts. With the exception of the fact that some authors now acknowledge the contribution of both substratum and superstratum sources to the creole, the framework adopted for the papers in the aforementioned collection presents the same problems as the two positions discussed in 2.2.4.1 and 2.2.4.2. with respect to the features that any theory of PC genesis must be able to account for.

2.2.5. The theory that creoles reflect the properties of Universal Grammar

The main proponent of the claim that creoles reflect the properties of Universal Grammar is Bickerton (1981, 1984, 1986, and subsequent work). (Note that, in this view, pidgins and creoles are crucially different entities, see below.) This theory, known as the Language Bioprogram Hypothesis (LBH), rests on the alledged similarity between undoubtedly historically unrelated creoles, such as Haitian, Sranan and Hawaiian, and on the alledged similarity between creoles and child language. According to Bickerton, each person is born with a grammatical model (the Language Bioprogram) enabling him or her to construct a grammar. In contexts where creoles emerge, children are exposed to a pidgin spoken by their parents. According to Bickerton, this pidgin is an impoverished language variety that does not present all the characteristics of a native language. Being faced with this impoverished linguistic model, the children use their hypothesised Language Bioprogram in order to nativise the pidgin. Nativisation of the pidgin is claimed to consist in expanding the pidgin. The language variety so created is claimed to be a creole that reflects both the unmarked grammar that is hypothesised to characterise the language of young children, and the unmarked grammar that is hypothesised to characterise creole languages. Thus, in Bickerton's view, both creole languages and child language are closer to Universal Grammar than other language varieties, for both present the unmarked options of Universal Grammar (see also Seuren and Wekker 1986 for a similar view on this point). Still on this view, in ordinary cases of first language acquisition, children are exposed to linguistic data that are produced by the adults around them. Presumably, in this situation, children have a chance of acquiring the

language specific features of their native language. Bickerton claims that, in the special case of first language acquisition in the context of creole genesis, children are deprived of an adequate adult model, in such a way that the language that they develop has the features of Universal Grammar. Bickerton claims that his theory accounts for both the hypothesised similarity between creole languages and the hypothesised similarity between creoles and child language.[5]

Does this theory account for the seven properties identified in section 2.1? Since Bickerton crucially considers pidgins and creoles as separate entities, I will refer only to creoles in addressing this question. Bickerton's approach does not account for the fact that creole languages emerge only in multilingual communities that are in need of a *lingua franca* and where language learners have little access to the superstratum language (the first three features). Since the creole is nativised in one generation, this theory can be said to account for the rapid development of creole languages (the fourth feature). The theory does not account, however, for the fact that pidgin and creole languages tend to be isolating languages nor for the fact that they manifest the properties of both their substratum and superstratum languages in the way they do (the fifth and sixth features). The theory is formulated in terms that are precise enough so as to be falsifiable. Building on Lefebvre and Lumsden (1989: 255), the next paragraphs discuss several points that falsify the universalist approach to creole genesis.

Crucially, the Language Bioprogram Hypothesis of creole genesis requires that pidgins and creoles be different entities formed by different processes. On the one hand, it has been demonstrated that pidgins and creoles are not qualitatively different from one another (see in particular the

work of Sankoff, e.g. Sankoff and Brown 1980; Sankoff and Laberge 1980). On the other hand, as was mentioned in section 2.2.2. and, as will be shown in section 2.4, pidgins and creoles are not distinguishable on the basis of the processes that are at work in their formation. This constitutes a first major drawback to Bickerton's theory.

Second, the alledged similarity between creole languages falls short in view of detailed comparisons of various creoles. As is shown in Muysken (1988b), while a superficial look at creole languages may yield the conclusion that they are alike, a closer look at the data forces a revision of this conclusion. An example in point is the fact that, while some creoles manifest the serial verb construction (e.g. Saramaccan, Jamaican, Haitian, Papiamento, Tok Pisin, etc.), others do not (e.g. Philippine Creole Spanish, Hawaiian Creole English, Mauritian Creole, Seychellois, Reunionais, etc.). Likewise, while some creoles manifest the predicate cleft construction (e.g. Haitian, Papiamento, etc.), others do not (e.g. Tok Pisin, Solomons Pidgin, Australian creoles, etc.). Furthermore, as is extensively discussed in Lefebvre (1998a, 2001d), creoles tend to reproduce the semantic and syntactic features of their substratum languages. Hence, Atlantic creoles tend to reproduce the features of their West African substratum languages, whereas Pacific creoles tend to reproduce the features of their Austronesian substratum languages. This explains why Atlantic creoles manifest the predicate cleft construction, whereas Pacific ones do not. A comparison of Haitian (and contributing languages) in Lefebvre (1998a) with Solomons Pidgin (and contributing languages) in Keesing (1988) strongly supports this claim. In conclusion, the alleged similarity between creole languages falls short when data from creoles of different geographical areas are considered.

Third, the hypothesised similarity between creole languages and child language receives no support in current literature (see for example the critiques formulated by a number of scholars in a special issue of *Brain and Behavioral Sciences*, 1984).

Fourth, the claim that creole languages reflect the unmarked case is not of much use without a theory of markedness. Indeed, no theory of markedness is formulated in Bickerton's work. Furthermore, based on a theory of markedness, some authors show that creole languages do present marked options of Universal Grammar. Indepth discussions of this point can be found in Koopman (1986), Lefebvre (1998a) and Muysken (1981b).

Fifth, Bickerton's theory looses even more points when historical data are considered. For example, Singler (1996) shows that nativisation of the Caribbean plantation societies was an extremely slow process. First, the slave traders imported twice as many men than women (Curtin 1976); second, the birth rate was very low (Kiple 1984); third, infant mortality was very high (Singler 1993a); fourth, life span of Africans in the Caribbean was short (Singler 1993a). As Singler (1993a: 237–238) comments: "This combination of factors yielded societies unable to reverse the natural population decrease. They were societies marked by both a disproportionately small number of children and an ongoing stream of recently arrived slaves from Africa". Since the bulk of the Caribbean population at the time the creoles were formed was adult, Singler (1996: 199) concludes that the principal agents of creole genesis must have been adults. Moreover, as will be shown in section 2.4, the main process at work in the formation of pidgin and creole languages requires adult language competence.

In conclusion, the Language Bioprogram Hypothesis of creole genesis does not hold in view of all these facts.

2.2.6. Summary

In this section, six approches to creole genesis were reviewed. All share the characteristic that they focus on language varieties rather than on processes that lead to these language varieties. Each of these theories has been evaluated against the seven features that need to be accounted for by any theory that seeks to provided a complete theory of the origin of these languages, as per the criteria established in section 2.1. The results are summarised in Table 2.1, which should be interpreted in light of the methodological proviso at the end of the introduction to section 2.2.

Table 2.1. Summary of the six theories of pidgin and creole genesis with respect to the features that need to be accounted for by any theory that seeks to account for the origin of these languages. (The symbols + and – indicate whether a given theory accounts for a given feature)

Features theories	Multi-lingual commu-nity	Need for a lingua franca	Little access to super-stratum language	Rapid creation	Isolating language	Type of mix	Theory is falsifiable
PCs as reduced codes	–	–	+	–	+	–	+
Creoles as nativised pidgins	+	+	–	+	–	–	–
PCs as cases of imperfect second language acquisition	–	–	+	–	–	–	–
PCs as restructured substratum varieties	+	+	+	–	–	–	–
PCs as restructured superstratum varieties	–	–	+	–	–	–	+
LBH	–	–	–	+	–	–	+

As can be seen from Table 2.1, none of the six theories reviewed in this section can account for all the features that need to be accounted for in an optimal theory of PC genesis. Strikingly enough, while each theory accounts for at least one of the features in the list, none of the theories evaluated thus far can account for the type of mix that is manifested by PCs from among their source languages. As will be seen in section 2.4, the relexification account of PC genesis crucially predicts the principled division of properties of PC lexicons between their source languages.

With the exception of the nativised pidgin theory of creole genesis (section 2.2), all the theories discussed in this section consider PCs as somewhat deprived language varieties (see e.g. the 'baby-talk' or the 'foreigner-talk' approach, the 'imperfect' stage of second language acquisition, the 'restructured' varieties, the alleged similarity between creole languages and child language). For a long time, PCs were considered 'marginal'. To my knowledge, the first state-of-the-art article on PCs was written as late as 1964; it was entitled 'Trade Jargons and Creole Dialects as Marginal Languages', signed by Reineke.[6] In the early seventies, several linguists such as Hall (1966: 121–122), Labov (1971), Whinnom (1971: 109), etc., urged students of PCs to approach these languages in the same way as they do approach other languages. The first substantial collection of papers on pidgins and creoles was edited by Dell Hymes in 1971. This collection is still an extremely valuable source. In the late seventies and early eighties, we begin to see discussions of pidgin and creole genesis within the framework of the processes otherwise known to play a role in language genesis and language change in general.

2.3. Shifting the object of study from language varieties to processes involved in language creation and change

In this section, I undertake the discussion of what I assume to be the second phase of research on pidgin and creole genesis. I estimate this second phase to have begun in the late seventies. It is characterised by the desire to cast the discussion of PC genesis within the framework of the processes otherwise known to be at work in language formation and in language change in general. On the one hand, proposals on the origin of languages referred to as mixed languages[7] are brought into the forum of discussion on the origin of PCs. In this respect, Media Lengua—a Quechua-Spanish mixed language spoken in Ecuador—(see Muysken 1981a, 1988c), Michif—a Cree-French mixed language—spoken by the metis buffalo hunters of Canada and the Northern United States (see Bakker 1989, 1992, 1994; Papen 1988), and Inner Mbugu or Ma'a—a mixed language spoken in Tanzania (see Goodman 1971; Möhlig 1983; Mous 1994, 1995, in press; Thomason and Kaufman 1991), to name but a few, and the theories of their emergence, were brought to the scene of PC studies. On the other hand, scholars started to entertain the idea that the changes observed in pidgin and creole languages are not fundamentally different from regular cases of language change in non-creole languages (see e.g. Adone and Plag (eds) 1994; Baker and Syea (eds) 1996; Hymes (ed.) 1971; Plag 1994a; Sankoff (ed.) 1980; etc.). Finally, sociolinguistic studies seeking to explain the formation of urban varieties out of various rural ones (see e.g. Domingue 1980, 1981; Jourdan 1985; Siegel 1995, 2002; Trudgill 1986) also had their input into PC studies. Three major processes have gradually entered the scene of PC studies. These processes are relexification, shown to play a role

in the formation of mixed languages; reanalysis, a major process in linguistic change; and dialect levelling, a process that takes place when various dialects of the same language come into contact. In this section, I provide definitions for these processes and I show how they began to be applied to PC formation and development. The shift from the study of language varieties to the study of the processes at work in the formation of these varieties constitutes a positive step in the study of PC genesis. As will be seen in the following subsections, however, this shift was not entirely successful on the first round.

2.3.1. Relexification

Muysken (1981a, 1988c) shows that Media Lengua has a lexicon where the phonological forms of major syntactic category lexical items (that is, nouns, verbs and adjectives) are almost entirely derived from Spanish, while the forms of the affixes and of the functional category system are derived from Quechua. The examples in (2) and (3) illustrate this division. (In the Media Lengua examples below, the forms derived from Spanish appear in italics, and those derived from Quechua, in regular characters.)

(2) a. *No sé*. SPANISH
not know.1st
'I do not know.'

b. *Mana yacha-ni-chu*. QUECHUA
No sabi-ni-chu. MEDIA LENGUA
not know.1st.VAL
'I do not know.' (=(3) in Muysken 1981a)

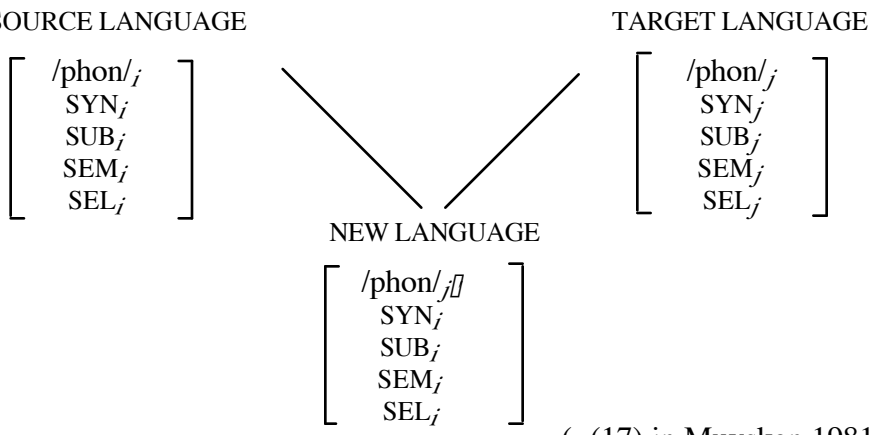
(3) a. *Si llueve demás, no voy a ir*. SPANISH
if rain.3rd too.much, not go.1st to go
'If it rains too much, I will not go.'

b. *Yalli-da tamia-pi-ga, mana ri-sha-chu*. QUECHUA
Dimas-ta *llubi*-pi-ga, *no i*-sha-chu. MEDIA LENGUA
too.much-ACC rain-LO-TO not go-ASP-VAL
'If it rains too much, I will not go.' (=(1) in Muysken 1981a)

Muysken (1981a) demonstrates that although the major category lexical entries of Media Lengua derive their phonological representations from Spanish, their semantic content is derived from Quechua. For example, the phonetic form of the Spanish verb *sentarse* 'sit down' was used to replace the Quechua verb *tiya-ri* 'sit', 'live', 'locative be', 'there is' yielding the Media Lengua verb *sinta-ri* 'sit', 'live', 'locative be', 'there is'. As Muysken (1981a: 56) points out, a single Media Lengua word is substituted for the Quechua word, preserving the various meanings of the latter even when, in Spanish, each of these meanings would be expressed by a separate lexical item: *estar sentado* 'sit', *vivir* 'live', *estar* 'locative be', *hay* 'there is'. Similarly, the phonetic form of the Spanish verb *tener* 'to have, to hold' was used to replace the Quechua verb *chari-* 'to have' yielding the Media Lengua verb *tini-* 'to have' (see Muysken 1988c). The above examples show that, although there is some overlap in the semantics of the verbs that are associated in the replacement, the semantics of the new Media Lengua verbs follows the details of Quechua rather than those of the Spanish lexical entries.

What is the process that produces lexical entries having the division of properties of the type described above? Muysken (1981a: 61) identifies this process as relexification: "Given the concept of lexical entry, relexification can be defined as the process of vocabulary substitution in which the only information adopted from the target language in the lexical entry is the phonological representation." Muysken's representation of the process is reproduced in (4).

(4)



(=(17) in Muysken 1981a)

In this view, relexification is a mental process that builds new lexical entries by copying the lexical entries of an already established lexicon and replacing their phonological representations with representations derived from another language. The nature of this process accounts for the division of properties observed in Media Lengua. The process of relexification has been argued to be an important tool in the creation of other mixed languages (e.g. for Michif, see Bakker 1989, 1992, 1994; for Inner Mbugu or Ma'a, see Mous 1994, 1995, in press).[8]

This process has also been claimed to play a role in PC genesis. For example, Koopman (1986), Lefebvre (1984, 1986), Stewart (1962), Voorhoeve (1973), Whinnom (1977) and others have long claimed that this is so. At one point, Muysken (1981a: 77) also proposed that relexification played a role in the formation of PCs: "If it is the case that the Caribbean creoles show numerous African survivals in their syntax and semantics, then I think we can argue that it is not interference which led to these survivals, but relexification." In a more recent paper (Muysken and Smith 1990: 884), however, Muysken amends his earlier claim, allowing relexification to play a role in language genesis only in bilingual situations, that is, only in the

formation of mixed languages. "We reject the gradual 'relexification' of believers in monogenesis (from a West African Portuguese Pidgin) or Afrogenesis, in situations of communal linguistic confrontation between, e.g. a European planter class and an African slave class. We do accept the possibility of relexification as a mechanism in forming a new language in a bilingual situation." This strong position had the effect of slowing down research on the role of relexification in PC genesis for a while, but, as will be seen in section 2.4, it did not have the effect of killing the idea altogether.

2.3.2. Reanalysis and related phenomena

Reanalysis (and related phenomena, e.g. desemanticisation and grammaticalisation)[9] constitutes a major process of linguistic change (see e.g. Heine and Reh 1984; Hopper and Traugott 1993; Lightfoot 1979, etc.). Reanalysis is a mental process by which a particular form which signals a lexical entry becomes the signal of another lexical entry (see e.g. Lightfoot 1979).[10] A typical example is the reanalysis of verbal expressions as adverbs. For example, the Yoruba verbal expression *sa ere* 'run race' has been reanalysed as an adverb: *sere* 'quickly' (see Bámgbóṣé 1974; Lord 1976).

Since the seventies, several cases of linguistic change that have occurred in PCs have been analysed as cases of reanalysis. Such cases are reported in Baker and Syea (eds) (1996), Bickerton (1988), Foley (1988), Koopman and Lefebvre (1981), Lefebvre (1984), Mühlhäusler (1986a), Muysken (1988b), Plag (1993), Rickford (1987), Romaine (1988), Sankoff (1990, 1991), Sankoff and Laberge (1973), Valdman and Highfield (eds) (1980), Washabaugh (1975), to name but a few.[11]

However, most of these authors were generally looking at data drawn from the pidgin or creole alone, and data from contributing languages to these PCs were generally not considered. As a consequence of this, more cases of reanalysis were postulated to have occurred in PCs than had actually taken place. Indeed, as was subsequently shown by various authors, when the properties of the corresponding lexical entries in the substratum languages are considered as a point of departure for the pidgin or creole lexical entries, there is less need to call upon reanalysis than was first hypothesised to explain the various functions of a particular lexical item in PCs (see e.g. Bruyn 1996; Keesing 1988; Lefebvre 1998a; Traugott 1999). For example, Koopman and Lefebvre (1981, 1982) hypothesised that the complementiser *pou* entered the Haitian lexicon through a process of reanalysis involving both the preposition *pou* 'for' and the irrealis mood marker *pou*. Lefebvre (1998a: 191–193) shows, however, that the corresponding substratum lexical entries cumulate these functions. It thus appears that, in this particular case, there is no need to have recourse to reanalysis to explain the creole data, and that relexification alone turns out to account for the full range of functions of the creole lexical entry.

2.3.3. Dialect levelling

Dialect levelling, as discussed in the literature on dialects in contact (see e.g. Domingue 1980, 1981; Siegel 1985, 1987, 1995, 1997; Trudgill 1986; etc.) refers to the reduction of variation between dialects of the same language in situations where these dialects are brought together. As Siegel (1997: 21) puts it, "dialect differences are reduced as speakers acquire features from other varieties as well as avoid features from their own variety

that are somehow different. This may occur over several generations until a stable compromise dialect develops." Well-documented cases of dialect levelling include Bhojpuri as spoken in Mauritius (see Domingue 1980, 1981) and English as spoken outside of England (see e.g. Siegel 1997; Trudgill 1986).

In the recent literature on pidgins and creoles (mainly in the eighties and nineties), it has been suggested that dialect levelling also plays a role in the further development of these languages (see e.g. Harris 1991: 199; Jourdan 1985; Mufwene 1990: 138–139, 1994c, 1996b: 22; Mühlhäusler 1980: 34; etc.). For example, Siegel (1997: 26) asserts that: "Mixing and levelling may (…) be important in the development of stable pidgin and creole languages... When the various versions of the superstrate are then used as the main means of communication among speakers of different substrate languages (in other words, when vernacularisation occurs), and when these speakers begin to view themselves somehow as a 'community', then levelling begins."

With the exception of Siegel who claims that variation within a given PC comes from the various second language acquisition versions of the lexifier language, at this stage, no precisions are given as to where the variation found in PCs comes from. Documented cases of dialect levelling in PC development are quasi non-existent.

2.3.4. Summary

In the late nineteen seventies and in the eighties there was a shift in the focus of enquiry about PC genesis from the study of language varieties to the study of the processes yielding these varieties. Three major processes

shown to play a role in language genesis and language change in general—relexification, reanalysis and dialect levelling—were hypothesised to also play a role in the creation and development of PCs. As we saw earlier, the use of these processes in the study of PC genesis was not entirely successful on the first round. There are two major reasons for this situation. The first one is due to the lack of systematic and detailed comparative work between PCs and their source languages. For example, the overestimation of the role of reanalysis mentioned in section 2.3.2 is without doubt due to this factor. The second reason is due to the fact that there was no theory on how these three processes apply and interact in PC genesis. Each process was being considered for its own sake. Since none of the three processes taken individually may constitute a theory of PC genesis[12], at the end of this second phase, the field was left, so to speak, in a state of flux. The turning point into the third phase of studies on PC genesis was marked by the debate between substratists and universalists (see Muysken and Smith (eds) 1986) and by Keesing's (1988) publication of the first comparison of a PC with its source languages: the comparison of Solomons Pidgin with English, its lexifier language, and with Kwaio, one of its Austronesian substratum languages. At the same time, another piece of comparative research had been undertaken on an Atlantic creole, Haitian Creole, to which I now turn.

2.4. A unified theory of pidgin and creole genesis

From 1985 to 2000, successive projects on the genesis of PCs were carried out at the Université du Québec à Montréal (hereafter UQAM).[13] These projects were based on the general assumption that it should be

possible to account for the formation of PCs in terms of the processes that are at work in language genesis and language change in general, that is relexification, reanalysis and dialect levelling, and in terms of a sound theory of how these processes interact in PC genesis (see Lefebvre 1986, 1993b, 1998a, and the references therein; Lefebvre and Kaye (eds) 1986; Lefebvre and Lumsden 1989, 1994a, 1994b). Another assumption was that the linguistic account should be compatible with the situation that prevailed at the time the creole was formed. Because of the importance of this research in terms of resources[14] and production[15], this section is dedicated to an overview of the theoretical progress in the account of the genesis of PCs accomplished within the framework of the aforementioned projects. Section 2.4.1. presents an overview of the hypothesis and of the methodology of this research. Section 2.4.2. summarises the contribution of this research to the issue of how the processes involved apply and interact in PC genesis. Section 2.4.3. evaluates the proposed theory on the basis of the characteristics that define an optimal account of creole genesis as provided in section 2.1.

2.4.1. Hypothesis and methodology of the UQAM projects

The basic hypothesis (as formulated in Lefebvre 1986, 1993b; Lefebvre and Kaye (eds) 1986; Lefebvre and Lumsden 1989, 1994a, 1994b; etc.) tested by the research reported on in this section is that the creators of a creole language, adult native speakers of various languages, use the properties of their native lexicons, the parametric values and semantic interpretation rules of their native grammars in creating a PC. The bulk of a PC's lexical entries is created by the process of relexification. Two other

processes, fed by the output of relexification, dialect levelling and reanalysis, also play a role in the development of a PC. As is pointed out in Lefebvre and Lumsden (1994a), this account is a further development of the second language acquisition theory of PC genesis (see section 2.2.3): it is claimed that, in creole genesis involving situations where there is little access to the superstratum language, the process of relexification is used by speakers of the substratum languages as the main tool for acquiring a second language: the superstratum language. The hypothesis was tested on the basis of Haitian Creole. The research program involved two dimensions, a historical dimension and a linguistic one.

The historical research was designed to answer the following questions: when was Haitian Creole formed? What were the salient demographic characteristics of the Haitian population during that period? Who were the people present at the relevant time? What was their linguistic background? (see Lefebvre 1993b). The historical research in the colonial archives of France was carried out by John Singler (see Singler 1993a, 1993b, 1996). In short, Singler establishes the following points. Haitian Creole was formed between 1680 and 1740. As a consequence of a shift from a tobacco and cotton economy to a sugar economy, the number of colonists decreased and the number of slaves exploded; this had the effect of modifying the slave population's exposure to French. The bulk of the Caribbean population at the time Haitian Creole was formed was adult (see details in section 2.2.5). As for the languages that these adults were speaking, Singler (1993b) shows that they were all Niger-Congo languages, more particularly Kwa (Gbe and Akan) and Bantu. During the formative period of Haitian Creole, Gbe speakers made up more than 50% of the

French Caribbean slave-export population. As is pointed out in Lefebvre and Lumsden (1994b), the overall situation found in Haiti between 1680 and 1740 presented all the prerequisites for the emergence of a creole language: there was a multilingual community, in need of a *lingua franca*, and the bulk of the population, the speakers of the substratum languages, had only reduced access to the superstratum language.

The test of the linguistic hypothesis consists in a detailed comparison of the lexicon and grammar of Haitian Creole with those of its contributing languages: French, its superstratum language, and West African languages, its substratum languages. Due to time and resource constraints, we decided to limit the detailed study of the substratum languages of Haitian to one language. Because of the importance of the influence of the Fon culture on that of Haiti (with respect to religion and art, see e.g. Bastide 1967; Herskovits 1975), Fongbe, a language of the Gbe cluster, was chosen as the substratum language to be studied in detail (see Lefebvre 1986, 1993b; Lefebvre and Kaye (eds) 1986).[16] This choice turned out to be a good one, in view of Singler's finding that the Gbe speakers outnumbered speakers of the other West African languages at the time Haitian Creole was formed. As has been pointed out in Lefebvre and Lumsden (1994a), the methodological choices that were made had the effect of making the relexification hypothesis easier to falsify. (For a thorough discussion of the methodology of the research and the validity of the linguistic test, see Lefebvre 1998a: 52–77, and the references cited therein.)

The linguistic test involves a global comparison of the lexicons, parametric values, semantic interpretation rules and concatenation principles of the languages involved. As has been emphasised in Lefebvre and

Lumsden (1994a, 1994b), one or two examples either way are not enough to support or falsify the hypothesis. The test must involve quantity as well as quality.

To the best of my knowledge, this was the first time that sufficient resources have been gathered to make such a detailed and extensive comparison of the grammar and lexicon of a creole language with those of its superstratum and substratum sources. The global results of this threeway comparison can be found in Lefebvre (1998a). Additional results are reported on in Brousseau (in preparation), Lefebvre (1999a, 2001a), and in Lumsden (1999a, 1999b). The bulk of the threeway comparison supports the hypothesis in a way that surpasses my expectations when I started this research.

2.4.2. The interplay of the processes embedded within a scenario of creole genesis

The content of this section summarises the theory developed during the 1989–1994 UQAM project with respect to how the three processes discussed in section 2.3 interact in a scenario of creole genesis. Unless otherwise specified, the theory reported on here is as developed in Lefebvre and Lumsden (1989, 1992, 1994a, 1994b) and in Lumsden and Lefebvre (1994).

Relexification [also referred to as relabelling] applies in creole genesis in the following way. Native speakers of various substratum languages are brought together. Crucially, the speakers of the substratum community do not have a common language, a situation which creates the need for a *lingua franca*. The substratum speakers are exposed to a

superstratum language, the language of the colonists. However, they do not have enough exposure to this language to learn the details of its lexical entries. Due to this situation, speakers of the substratum languages relexify the lexical entries of their respective lexicons on the basis of phonetic strings found in the superstratum language (see the examples in (1)). The relexification of various lexicons on the basis of a single superstratum language provides the speakers of the substratum languages with a common vocabulary. It is the limited direct access to the superstratum language that makes relexification so important in the formation of radical[17] creoles.

It is a well documented fact that in PCs, functional category items as well as major category lexical entries have phonological representations that are similar to some phonetic strings of the superstratum language. These lexical entries, however, do not have the same properties as the corresponding superstratum forms from which they are phonologically derived (see e.g. Carden and Stewart 1988; Lefebvre 1984; Lefebvre and Lumsden 1989, 1992; Mufwene 1991). In the scenario of creole genesis developed during the 1989–1994 UQAM project, it is hypothesised that, because speakers of the substratum languages have very limited access to the superstratum data, they typically fail to identify the functional categories of the superstratum language. These speakers thus try to relexify the functional items of their native languages on the basis of forms found in the superstratum language. It is proposed that the functional category lexical entries of the substratum languages are relexified on the basis of *major* category lexical items (e.g. nouns, adjectives, verbs, adverbs and prepositions) of the superstratum language. For example, the definite determiner of the substratum languages of Haitian Creole is argued to have

been relexified on the basis of the postposed French adverb *là*, yielding Haitian *la* (see Lefebvre 1998a: 79–84). The relexification of functional, as well as major category lexical entries, provides the speakers of the various substratum languages with a common vocabulary in all areas of the lexicon.

The process of relexification is semantically driven in the sense that there must be partial semantic overlap between the source and target lexical entries for it to take place (see Muysken 1981a). Consequently, relexification is constrained by what the superstratum language has to offer in terms of appropriate strings to relexify original lexical entries. This is particularly crucial in the case of functional category items. It is thus possible that some lexical entries cannot be assigned a new phonological form in relexification, either because an original lexical entry has no semantic content (e.g. operators, case markers) or because there is no form available in the superstratum language to provide a new phonological form for an original lexical entry. Due to space limitations, such cases are not discussed here. I refer the reader to Lefebvre (1998a) for extensive discussion of such cases.

The lexicons created by relexification become the basis of a *lingua franca* within the creole community. When the relexified lexicons become the target of the creole community, a new language is born. At this point, the speakers are no longer targeting the superstratum language. They are targeting the common language that they have developed through relexification: the incipient creole. At this stage, two other processes come into play: dialect levelling and reanalysis.

Relexification is a cognitive hence individual process. Situations where creoles emerge involve several substratum languages. Each individual

relexifies his or her own lexicon. Hence, speakers of various substratum languages reproduce the idiosyncratic semantic and syntactic properties of their own lexicons in relexification and thus, the product of relexification is not necessarily uniform across the creole community. The relexification of several lexicons thus creates variation within a creole. This scenario allows for a sound explanation of the facts referred to in the literature on creole studies as the 'cafeteria principle' – a term used first by Dillard (1970) and later by Bickerton. As Bickerton (1981: 49) puts it: "As things stand, we are asked to believe that different African languages contributed different rules and features to particular creoles (…) it is (…) absurd to suppose that a creole could mix fragments of Yoruba, Akan, Igbo, Mandinka, and Wolof (…)." The differences created by the relexification of various lexicons may (but need not) be levelled out with time. The proposal that dialect levelling operates on the output of the various relexified lexicons involved in creole formation provides a principled explanation of the observation that several different substratum languages may contribute features to a given creole. Indeed, in the competition among different creole dialects (created by the relexification of different substratum lexicons), there are winners and losers. As is shown in Lefebvre (1998a), the competition is not always won by speakers of the same relexified lexicon (see also Siegel 1997).

An original lexical entry that was not relexified (either because the creators of the creole did not find an appropriate form in the superstratum language to relexify the copied lexical entry, or because it had no semantic content and thus could not be relexified) may be signalled by a periphrastic expression. For example, a tense or aspect may be signalled by an adverb with a similar meaning. The periphrastic expression may later become the

phonological representation of the functional category in question through the process of reanalysis. The postulated link between relexification and reanalysis accounts in a straightforward way for the paradoxical situation noted in the literature according to which, in the course of their further development, creoles develop lexical entries that manifest the properties of their substratum languages even in situations where the substratum languages have ceased to be spoken (see e.g. Chaudenson 1994; Mufwene 1990; Mühlhäusler 1986a, 1986b; Sankoff 1991: 73). (For an extensive discussion of this point, see Lefebvre 1998a: 108–110, 375–386.)

2.4.3. An optimal account of creole genesis

The theory of creole genesis outlined above provides a straightforward and optimal account of all the properties of creole languages discussed in section 2.1. The following discussion builds on a preliminary one in Lefebvre and Lumsden (1989, 1994a), as updated in Lefebvre (1998a).

The theory accounts for the fact that creole languages emerge in multilingual societies in need a *lingua franca* (the first and second features). It accounts for the fact that the creators of the creole have little access to the superstratum language (the third feature). Only one generation of speakers is required to create a new language by means of relexification (the fourth feature). When it is created, this new language evolves as any other language.

By virtue of the definition of the process, creole lexical entries are predicted to have the same semantic and syntactic properties as the corresponding lexical entries in the substratum languages, but phonological

representations derived from the phonetic strings of the superstratum language. The relexification theory of creole genesis thus accounts for the fact that creoles reflect the properties of both their superstratum and their substratum source languages in the way they do (the sixth feature).

As is observed in Lefebvre and Lumsden (1994a), the fact that creoles are generally isolating languages also follows from the above proposal. Since the functional category lexemes of creole languages derive their phonological forms from major-category lexemes in the superstratum language, or from reanalysis, and since these categories are typically free morphemes, it follows that creoles will tend to be isolating languages (the fifth feature).

Finally, the theory that the bulk of a creole's lexical entries are formed by the process of relexification is falsifiable. As is stated in Lefebvre and Lumsden (1989, 1994a), if the three-way comparison of the lexical properties of a radical creole with the lexical properties of its source languages were to show that the syntax and semantics of the creole are *not* systematically parallel to the syntax and semantics of the substratum languages, then the theory would be falsified. Likewise, the theory that the creators of the creole use the parametric values, the semantic interpretation rules and the principles of concatenation of their own grammars in creating the creole is falsifiable. If a comparison of the grammatical and semantic properties of a creole with those of its source languages were to show that the properties of the creole are *not* systematically parallel to those of the substratum languages, then the hypothesis would be falsified.

Thus, the theory of creole genesis summarised in this section does account for all the features of an optimal theory of creole genesis.

Consequently, if we were to add the above account of PC genesis to the list of theories in Table 2.1, all the features that need to be accounted for would be assigned a positive value.

2.4.4. Summary

The core of the results of the UQAM Haitian projects can be found in Lefebvre (1998a, and the references cited therein). The detailed comparison of Haitian and its source languages overwelmingly supports the relexification account of creole genesis (see also chapter 3). The detailed comparison of Solomons Pidgin with its source languages by Keesing (1988) also supports such an account. The results in Migge (1998b) go in the same direction as well. As more cases are being documented, we deepen our understanding of the process itself and of the constraints upon it.

2.5. Conclusion

Three major phases of investigation on PC genesis have been summarised in this paper. The first phase includes six major approaches. These approaches were shown to have in common the fact that they all focus on linguistic varieties. It was shown that none of them can account for all the features that characterise an optimal theory of PC genesis. The second phase was characterised by an attempt to shift the focus of study from the language varieties to the processes that are at work in creating these language varieties. These processes are relexification, reanalysis and dialect levelling. For various reasons discussed in section 2.3, the first attempt at applying these processes to pidgin and creole genesis and development was not entirely

successful. The third phase of studies in pidgin and creole genesis made a breakthrough in several respects. It provided a unified theory of the origin of PCs cast within the framework of the processes which otherwise play a role in language genesis and language change in general, and it provided a theory of how these processes interact in PC genesis. It also provided detailed and systematic comparisons of pidgins and creoles with their source languages.[18]

What is ahead in the study of pidgin and creole genesis? Surely, the documentation of more cases will provide new questions for the theory (see for example the papers in Siegel 2000). Types of studies that are needed for getting new insights into the theory are discussed in Lefebvre (2002), of which section 3 is here reproduced as Appendix 1. Regardless of what theory of PC genesis will be adopted, researchers must bear in mind the fact that pidgins and creoles constitute a problem for the genealogical reconstruction of language families. Furthermore, if relexification is the main process at work in the creation of these languages, and since relexification is a cognitive process, it cannot be excluded that in the course of history, several PCs were created without our knowing it!

Notes to chapter 2

* This paper has been written as part of the project "Research builds on research" financed by SSHRCC. I want to thank the following colleagues for their most useful comments on an earlier version of this paper: Bernard Comrie, Christine Jourdan, Jeff Siegel and Kevin Tuite. Thanks to Marijo Denis and Virginie Loranger for their help in editing the manuscript, and to Andrée Bélanger for formatting it.

1 The content of this section builds on a preliminary discussion in Lefebvre and Lumsden (1989, 1994a) updated in Lefebvre (1998a: 1–4).

2 There is a general consensus in the literature that multilingualism is a required feature of communities where creoles may emerge. However, this proposal has recently been challenged by Smith, Robertson and Williamson (1987), who claim that Berbice

Dutch emerged out of contact between only two languages: Dutch and Eastern Ijo. Assuming that Berbice Dutch is a true creole, and that Ijo was the sole African language present at the time this creole was formed, this case would constitute the first documented evidence against Whinnom's widely accepted claim. For further discussion of this issue, see also Foley (1988) and Thomason (1997b).

[3] Within this general approach to creole genesis, two views have been advocated: the monogenetic and the polygenetic theory of creole languages. The former approach claims that European-based creoles are derived from a single pidgin, the Portuguese pidgin that emerged during the 15th century on the route of the Portuguese merchants. Among the tenants of this theory, we find Alleyne (1971), Goodman (1964), Hancock (1968), Stewart (1962), Whinnom (1956, 1965, 1971), to name but a few. On this view, the original pidgin would have been diversified with the dispersion of its speakers in the various countries of colonisation. As it became the native language of a first generation of speakers, this original pidgin language would have evolved into mutually unintelligible creoles due to borrowing from different colonial languages (see e.g. Stewart 1967: 47). The polygenetic theory of creole genesis (see e.g. Hall 1966), stipulates that different pidgins gave rise to different creoles. In this view, a French-based pidgin would be the source of French-based creoles, an English-based pidgin would be the source of English-based creoles, and so on and so forth. The debate on this issue has raised enormous methodological problems. Some of these are discussed in Frake (1971), Goodman (1971), Grimshaw (1971), Hymes (1971), Southworth (1971), etc. To the best of my knowledge, no one advocates a monogenesis theory of PCs anymore.

[4] For a discussion of similar and contrastive properties of second language acquisition and creolisation, see e.g. Véronique (1994).

[5] Note that Bickerton's (1981) Language Bioprogram and Chomsky's (1986) Language Acquisition Device are sometimes considered as being equivalent. In my understanding of the two approaches, however, these two devices are quite different. On Bickerton's view, it is the lack of pertinent data that activates the Language Bioprogram, whereas on Chomsky's view, it is the presence of pertinent data that activates the Language Acquisition Device. It thus seems that the two devices are not equivalent with respect to the trigger that activates them. The association of these two devices thus appears to be erroneous.

[6] Thanks to John Reineke for having brought PCs to the attention of linguists.

[7] Languages that are being referred to as mixed languages emerge in contexts where only two languages are spoken in contrast to PCs which emerge in contexts involving more than two languages, as we saw in section 2.1. For an extensive discussion of the differences and similarities between PCs and mixed languages, and the situations in which they emerge, see Lefebvre (1998a: 29–30, and the references therein).

[8] For a summary of the role of relexification in the formation of various mixed languages, see Lefebvre (1998a: 18–29).

[9] There is an ongoing debate on whether cases of grammaticalisation and of reanalysis constitute a single process or two separate ones. For example, Heine and Reh (1984: 97), Hopper and Traugott (1993: 32) and Lefebvre (1998a: 41–45) consider that reanalysis subsumes cases of grammaticalisation. Haspelmath (1998, 1999), however, strongly argues for formal differences between them. Since nothing in the present discussion hinges on possible distinctions between reanalysis and grammaticalisation, I do not pursue the discussion of this point here.

[10] I take the process of reanalysis to apply within a particular language. There are other acceptions of the term, however. For example, some authors consider calques as cases of reanalysis. Such cases are not included in my use of the term reanalysis.

[11] For an extensive discussion of how cases of reanalysis reported to have taken place in PCs are of the same kind as those observed in other languages, see Lefebvre 1998a: 30–33).

[12] For this reason it is not possible to discuss the proposals reviewed in this section against the seven features in section 2.1.

[13] A summary of the history of these projects can be found in the Preface to Lefebvre (1998a).

[14] For a complete statement of the human and financial resources allocated to these projects, see the Preface to Lefebvre (1998a).

[15] See Lefebvre (1998a) and Lefebvre and Brousseau (2002, and the references therein).

[16] In no way does this methodological choice entail that Haitian is Fongbe relexified. For further discussion of this point, see Lefebvre 1998a: 52–77.

[17] Creoles which less resemble their superstratum languages are referred to as radical creoles.

[18] At the symposium "Pidgin and Creole Linguistics in the 21st century" (1998), Glenn Gilbert asked participants the following question: "What place will universalist theories retain in 21st-century pidgin and creole linguistics?". My reply to his question was the following (see Lefebvre 2002: 247–286):

> Universals of language should be assumed to be part of PCs on two grounds: first, by definition, language universals are universal; second, pidgin and creole languages are natural languages and therefore they reflect the universal properties of language just as other natural languages do. The differences between PCs and their contributing languages, on the one hand, and among PCs, on the other hand, are to be found in areas of the grammar that allow for variation between languages. The lexicon is the component *par excellence* where differences between languages exist. This has two major consequences. First, on the relexification account of creole genesis, creole lexicons, above all, should reproduce the idiosyncrasies of their substratum language lexicons; this prediction is borne out by the Solomons Pidgin discussed by Keesing (1988) and by the Haitian data discussed by Lefebvre (1998a). Second, PCs formed on the basis of substrata from different language families are expected to reflect the idiosyncrasies of their respective substratum languages. This is also borne out by data drawn from the two studies mentioned above. While Solomons Pidgin reproduces the specific features of the Austronesian languages, Haitian Creole reproduces those of the West African languages (for a preliminary discussion of this issue, see Lefebvre 1996b, 1998a). By the same token, the approach taken here helps clarify another issue raised by Gilbert (2002): "Will the typology (language grouping) of creoles ultimately link them more strongly to their constituent languages, or will it link them more strongly to each other?" While some authors advocate the latter possibility (for example, McWhorter 1998a), the analysis presented in this paper would tend to favor the former. Indeed, PCs are hybrid languages which derive the semantic and syntactic properties of their lexical entries from their substratum languages and the phonological representations of these lexical entries from their superstratum languages. Given that they may be formed from different substratum and superstratum languages, what unites them is not their actual features, but rather the processes by which they are formed and the fact that they all emerge in language contact situations.

(For further discussion of this issue, see also chapter 8 of this book.)