

ama

ata  
ana

vater

*father*

padre  
madre

mutter

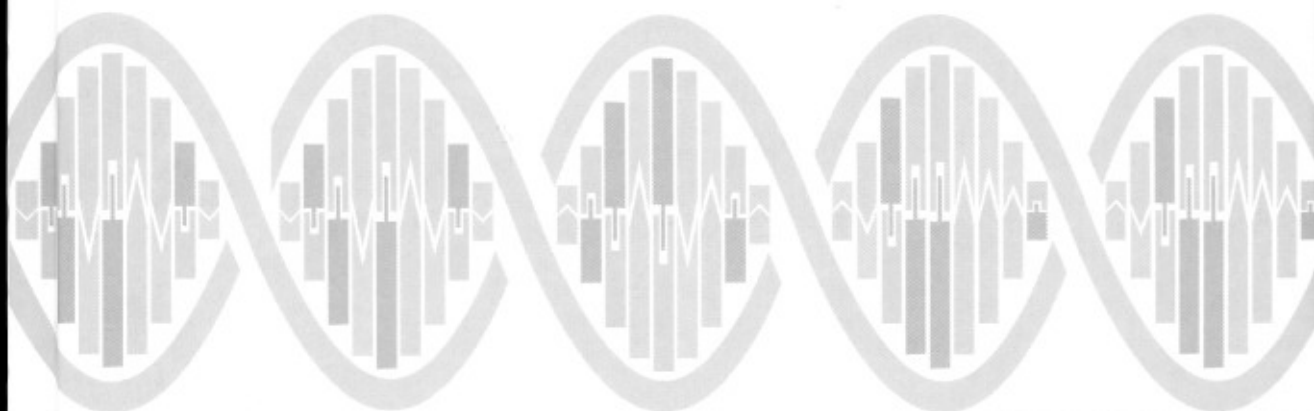
*mother*

padre  
madre

Günter Hauska (Hrsg.)

# Gene, Sprachen

und ihre Evolution



Salikoko S. Mufwene

# Language Evolution: The Population Genetics Way\*

## 1. Introduction

I focus in this essay on only some aspects of language evolution, viz., language speciation and language death, although I also discuss some structural changes, more specifically phonological and syntactic ones. I argue, contrary to the dominant practice in the 20<sup>th</sup> century, that a biological approach to evolution is applicable to languages, although I must clarify at the outset that languages should be analogized to species rather than to organisms. Moreover, the application works best when linguistic species are analogized with viral rather than with animal species. I argue eventually that the approach is analogical only to the extent that it is inspired by scholarship on biological evolution. Otherwise, linguistic species can be posited in their own right (Mufwene 2001). They share properties with other species, biological and otherwise, while they also differ from the latter in interesting ways that are specific to their ontogenetic, architectural peculiarities. Thus genetic linguistics can contribute to theories of evolution, adding hypotheses that are specific to, and/or inspired by, languages as species.

The invocation of population genetics in the title of this essay underscores two working assumptions of mine since Mufwene (2001), viz., 1) the agency of the evolution of a language lies in the individual communicative acts of its speakers, which is similar to the fact that a biological population or species is cumulatively affected by the experiences or activities of the individuals it consists of;<sup>1</sup> 2) the communicative activities that produce language evolution are largely determined by the socio-economic ecologies in which speakers evolve, which is similar to saying that the ecology rolls the dice in evolution. However, I should now explain why I return to an approach that has been unpopular until recently and justify some of the terminology that I now prefer to use in my work, such as *evolution*, *competition*, and *selection*.

\* I am grateful to David Hull for feedback on an earlier version of this essay. We disagree on whether viruses can be grouped into species, on a par with animal species. I saw no reason to change my position, as I also assume that the evolutionary peculiarities of various species are largely determined by their respective ontogenetic properties and their particular modes of gene or feature transmission. I am solely responsible for the shortcomings of the positions submitted here.

<sup>1</sup> Paul (1891, chapters 1 and 2) says something that amounts to this position.

## 2. Why this return to a biological approach?

By the time Charles Darwin published *On the origin of species* (1859), genetic linguists had already been addressing some of the questions that I address here. According to Maher (1983:xxv), August Schleicher felt so vindicated by this book that he responded to it with an elaborate discussion of analogies that he perceived between linguistic and biological evolution (Schleicher 1863, 1869). He covers notions such as ‘variation’, ‘selection’, ‘life of languages’ (as new ones emerge and some of the current ones die), ‘gradual development’ from the most “primitive” to more advanced and complex stages, ‘genus’/‘family’, and ‘species’ (not in the way I use it in the present discussion). He also notes that confusion on whether a language variety should be identified as a language or as a dialect is parallel to what Darwin reports about identifying a class of biological organisms as a species or a subspecies. Overall, he concludes, in the spirit of his self-professed uniformitarianism: “The rules now, which Darwin lays down with regard to the species of animals and [of] plants, are equally applicable to the organisms of languages, that is to say, as far as the main features are concerned” (cited from Koerner, ed. 1983:30).

Schleicher saw in Darwin’s principle of natural selection an explanation for the changes that produce the language speciation patterns represented by the *Stammbaum* ‘genetic tree’ that he was developing, which apparently inspired the biologists’ cladogram. He also assumed, like Darwin (Radick 2002), that different human populations had not reached the same stages of evolution – some being rather primitive and others particularly advanced – and that their languages reflected this putative variation. He thus posited an evolutionary trajectory of languages of the world from the isolating morphosyntactic type to the fusional morphology type. The former type, which he considered as the most primitive, includes, for example, Chinese, Thai, and the Kwa languages of West Africa, whereas the second type, which he claimed to have reached the most advanced level of evolution, includes Western European languages (Schleicher 1863, 1869).

This position, which was apparently widely accepted by the late 19<sup>th</sup> century – even by Charles Darwin (Hull 2002:13; Radick 2002:7, 13)), led French philologists such as Adam (1882, 1883) and Vinson (1882, 1888) to treat creoles and pidgins, which tend to have a predominantly isolating morphosyntax, as languages in their pristine, natural, or primitive state, i.e., as degenerations from the “refined” structures of the European languages they have evolved from (see DeGraff 2003 for an elaborate discussion). Although creoles and pidgins have continued to be associated with non-ordinary, contact-based developments and as special cases that putatively do not fit naturally in the domain of genetic linguistics (however, cf. Mufwene 2003a), the Schleicherian evolutionary model has generally been either discredited or ignored over the past century. Except for Bickerton (1984, 1990), who, through his language bioprogram hypothesis, has suggested that creoles and pidgins give us an idea of the protolanguage in human phylogeny, no language typologist has ever suggested any particular evolutionary ranking of structural types.<sup>2</sup>

<sup>2</sup> Assuming mistakenly that creoles have evolved from pidgins, Bickerton has argued since his *Roots of language* (1981), that pidgins have no syntax or grammar, having been made by adults, who presumably no longer had (full) access to Universal Grammar (UG) or the biological endowment for language. Accordingly, children, guided

It is not clear why the whole biological approach was abandoned in the 20<sup>th</sup> century, except that the comparison of languages with organisms was not particularly insightful, for reasons which I articulate below<sup>3</sup>. Hoenigswald (1990:11) also notes that a “language is no corporeal something with physical boundaries, or molecular permanence, or genome.” According to him, this ontological difference from a biological organism or species would make it difficult to account for language evolution on the biological model. He adds, “Try as we may, we can only think of it as the totality of the very traits that are subject to change” (11). Hull (2002:18) is right on the mark in observing that “most of [the] effort in [genetic] linguistics was expended to working out the *histories* of various languages. Much less attention was paid to the *processes* by which languages change.”

Indeed, the comparative method, on which genetic linguistics has critically relied, is used to determine the extent to which languages share linguistic materials (typically, words, sounds, and morphemes) and can be claimed to be genetically related. However, it cannot be used alone to determine whether the shared forms and structures have been inherited from a common ancestor or borrowed from a common, influential non-ancestral language they all came in contact with. Nor can it be used alone to exclude the possibility of mutual influence among the relevant languages if they have been spoken in the same geographical area or in adjacent ones. In other words, used alone, the comparative method cannot distinguish between “genetic relatedness” (based on materials inherited from a common ancestor), “areal diffusion” (due to borrowings from the same language), and “convergence” or *Sprachbund* (due to mutual influence in a contact area).

Areal linguistics (see, e.g., Heine & Kuteva 2005) underscores the importance of factoring the history of population contacts in genetic linguistics. Moreover, as shown in Mufwene (2003a), there has been very little cross-pollination between genetic linguistics and genetic creolistics (which focuses on the development of creoles, under the conditions of population contacts in plantation settlement colonies). It is not clear why linguists, including some of the most influential creolists, have ignored the fact that genetic creolistics has dealt with ecological details of recent cases of language speciation. Yet, these bear on genetic linguistics, especially in regard to the seemingly gratuitous assumption of asexual

predominantly by this language organ, would have imposed a UG-based grammar onto their parents' structureless pidgins, which would account for world-wide structural similarities among them. The same UG must account for the emergence of protolanguage, for which he has produced no structural evidence, among our hominid ancestors. He assumes that its protosyntax must have been similar to that of child language, which shares features with that of creoles. (Anybody who pays attention to cross-creole structural variation would speak of creole grammars, in the plural, instead!) Every one of these claims is highly disputable, supported by tenuous evidence, but it would be digressive to discuss this issue here. See, e.g., Mufwene (1991, 1992a), Lieberman (1991, 2002), McNeill (1992), and Pinker (1992) for informative discussions of Bickerton's thesis and Mufwene (2001) and Chaudenson (2001) for alternative interpretations of the facts that constitute the sociohistorical ecologies of the evolution of creoles. DeGraff (1999a, 1999b, 2005) is equally informative, from a language acquisition perspective. Suffice it to note here that Bickerton (1984: 141, 157–158) also suggested that creoles instantiate languages in a primitive developmental stage.

<sup>3</sup> Linguists such as Franz Bopp and Karl Ferdinand Becker seem, however, to have used the term *organism* with the meaning of *species*, a term that also occurs in the work of Schleicher. They speak of interindividual, now known also as interdialectal, variation within a language, to which Darwin's evolutionary principle of natural selection could also be invoked to apply (Schleicher 1863). One important 19<sup>th</sup>-century linguist whose views foreshadowed some of those expressed in this article is Paul (1891).

parentage (without mating/contact), i.e., the iterated splitting of languages into new varieties without external causation (see below).

On the other hand, genetic linguistics has stuck loyally to Schleicher's *Stammbaum*, despite its significant share of shortcomings. For instance, the approach has focused more on the outcomes of changes than on the mechanisms of changes themselves. Among other things, it does not account for "substrate" influence, from the other languages that the relevant language has been in contact with. Thus the role of substrate influence in the speciation of Latin into the Romance languages has received only lip service. The role of Celtic languages in the evolution of British English dialects other than Irish (either on the Germanic languages that the Jutes, Angles, and Saxons brought from continental Europe or since the development of Old English) is only now starting to receive some attention (e.g., Tristram 1999, 2000; Vennemann 2001; Filppula et al. 2002).

The monoparental, *Stammbaum* approach also makes it impossible to account for hybridization in language evolution as a natural process. I argue below that the approach has been a major obstacle to learning from the development of creoles lessons that should apply to all cases of language speciation such as the development of the Romance languages or, on a larger scale, the diversification of Indo-European and of Bantu languages.

In Mufwene (2001), I submit the following reasons to explain why communal languages are more like species than like organisms, which I claim to be the counterparts of idiolects. Languages are indeed populations of idiolects<sup>4</sup> and are as abstract as species, which are mere categories (Hey 2001) projected from structural or ontogenetic properties shared by classes of individuals or organisms.<sup>5</sup> As the universe of our experience and knowledge is continuous, the boundaries of both species and languages are naturally fuzzy and operationally arbitrary, imposed by particular ideologies or other practical or theoretical considerations, especially within the same genetic family. It must be difficult to draw the boundary between adjacent vernacular varieties of Italian and French or between those of Dutch and German, a problem aggravated by the arbitrariness of political boundaries.

Unlike organisms, languages qua species are heterogeneous beyond the non-monolithic architecture of idiolects, a property that follows from the fact that idiolects vary among themselves and are alike by the family resemblance principle.<sup>6</sup> They are thus like biological populations in which each organism preserves its genotypic and phenotypic individuality, despite the many features it shares with its conspecifics. Languages evolve

<sup>4</sup> For non-linguists, idiolects are the distinct ways in which individuals, as opposed to groups, speak their language, each one with his/her own peculiarities.

<sup>5</sup> The categories can of course also be posited a priori, based, for instance, on assumed common ancestry (O'Hara 1994), with structural or ontogenetic features being invoked only to justify the grouping or to clarify their internal division into sub-categories.

<sup>6</sup> Even Chomsky (2000) now adheres to this particular view of a language, explaining that lumping idiolects into the same language is a matter of likeness, not of sameness or identity. On the non-monolithic nature of language architecture – i.e., how the different structures coexist, with the functions of some of them overlapping or sometimes conflicting with each other – see Mufwene (1992b). As explained in Mufwene (2001), the piecemeal way in which a language is "acquired," with features selected incrementally from diverse inputs/idiolects, accounts for this state of affairs.

in non-uniform ways, with some idiolects, sociolects (social varieties), or dialects (regional varieties) being more engaged in some changes than others. They are thus like biological species, whose members are differentially affected by, or react in different ways to, changes in their ecologies. As noted in the Introduction, changes affecting languages are a cumulation of changes affecting individual speakers' idiolects, which raises the following interesting question: Under what particular conditions do ontogenetic features of individuals or organisms spread to become those of (part of) a population? We should note, along with McMahon (1994:248), that there are many idiosyncratic innovations that are short-lived and do not so spread.<sup>7</sup>

Languages are like populations/species in the ways they die and are born, viz., in a protracted way (Mufwene 2004a). A language is dead when it has no more speakers left, just like a species is dead when no more specimen is left that can instantiate it.<sup>8</sup> Cases of genocides precluded, languages do not typically suffer sudden deaths, because their speakers die or suffer linguistic dysfunction at different times – which entails gradualism. Moreover, they can die in one geographical or social setting but thrive in another, as is evidenced by Old World languages that have died in different places of the New World but continue to be spoken in their homelands or other parts of the New World. Like species, languages are not born in the same way as organisms. They have no moment of conception; nor do they have a gestation or incubation period. They are identified as new ones *post factum*, after a particular variety (of idiolects or of organisms) has been identified that is considered (significantly) different from an earlier population.

Other characteristics can be invoked, but the present list should suffice to support the position that as communal extrapolations, languages should not be thought of as organisms. This position will enable us to capitalize on the interaction of both variation and ecology to account for evolution. I assume that a communal language exists only as a social construct, suggested by the ability of speakers in a particular setting to communicate successfully when they use similar sequences of spoken or signed gestures.

The bottom line is that every individual speaks or signs in a way that is internally systematic. Communication with other individuals triggers mutual accommodations, which make the individual systems similar both in the ways individual sounds or signs are produced and in how they combine together into more complex interpretable units. What is really required in such settings is ability on the part of each participant to understand and be understood by others. Nobody is required to speak or sign in exactly the same way, and sameness hardly happens among idiolects. However, doing things in similar ways must facilitate mutual intelligibility. Languages as communal systems are construed on the basis of similarities in the ways people speak or sign in order to facilitate mutual intelligibility.

<sup>7</sup> Similar ideas can be seen in Paul (1891, chapters 1 and 2).

<sup>8</sup> As Thomason (2001) points out, the common characterization of language death that I just repeated may be oversimplified, because the lone speaker would not have anybody else to speak it with and his/her knowledge would probably have fallen into attrition by then. One sense in which the characterization may be correct is when a language is thought of as knowledge rather than as practice. Then arises the question of the death of a language by transformation. What proportion of its original features must it preserve in the last speaker or in the last group of speakers in order to be considered the same and alive?

Perhaps it is no accident that folk definitions of languages do not include notions such as systems (consisting of units and principles) but rather boil down to ‘the way a particular group of people speaks’. Thus *German* means the way the people called *Germans* speak and *(Ki)Swahili* means the way the people identified as *(Wa)Swahili* speak. The fact that world languages such as English and French are spoken in different parts of our planet now does not contradict my observation, because these languages can also be defined etymologically as, respectively, ‘(originally) language of the Angles or English people’ and ‘(originally) language of the Franks or French people’.

All these considerations explain why it is necessary to focus on the agency of individuals and dynamics of their interactions. Accommodations made during individual communicative acts not only bear on the development of their communal properties but also drive evolution within their populations. Interesting questions include the following: Does a population interact collectively with its ecology? Or, instead, is what is associated with a population a cumulation of the effects of individual interactions of its members with the relevant ecology? What kind of equation can be posited to account for such a cumulation of effects? How artificial is it to assume that the ecology acts on a population? If the ecology operates on individuals, what precisely is the ecology of the individual? If it is normal to assume an ecology for a population too, is the ecology of a population the same as the ecology of an individual? What’s the significance of these epistemological questions? I must first explain some concepts that are fundamental in this intellectual exercise.

### 3. Competition, selection, ecology, evolution, and other related notions

I should clarify at this point that I discuss ‘language evolution’, not ‘the evolution of language’. The latter deals only with the higher abstraction of what various languages share, with the ability to encode/store and communicate information through a spoken or signed system, and with how this capacity evolved in mankind. *Language evolution* can apply to specific languages in a way similar to *language change*, to which it is closely related semantically. Since the latter term is more established in historical linguistics, the following question arises: why don’t I stick to and use it?

The term *evolution* covers more than the traditional term *change*. In addition to traditional concerns with structural and pragmatic changes, it also covers language speciation and language birth and death, processes to which the term *change* has not applied in linguistics. Also, although language speciation has been a concern of genetic linguistics, a branch of historical linguistics, seldom has the topic been related to those of language birth and death, which are so germane to it. Thus, the development of creoles and pidgins has typically been treated as anomalous or unusual (see, e.g., Hock and Joseph 1996), if not as an aberration, despite the fact that the topic is, at least from a phylogenesis perspective, closely related to historical dialectology.

The traditional practice is made more bizarre by invocations of substrate influence in both genetic creolistics and genetic linguistics, especially in Romance linguistics, where the term *substratum* originated. It is as necessary to invoke substrate influence from the Celtic

languages to account for the speciation of Latin into the Romance languages as it is to invoke African substrate influence to account for the evolution of French and Portuguese into various creoles, the role of other ecological factors discussed in Chaudenson (2001, 2003) and Mufwene (2001) notwithstanding. Both cases are clear instantiations of Pyrrhic victory – where the prevailing language is so clearly affected by the displaced ones – which makes the distinction between creole and non-creole languages more sociological than linguistic (Mufwene 2001, DeGraff 2003). While the term *change* has been used in cases of substrate influence, it has not been applied to those of language speciation as consequences of divergent structural changes. The term *evolution* offers the advantage both of its applicability to this case and of aligning research on language evolution with that of species evolution, which covers a similar range of developments.<sup>9</sup> At least in the way I approach the subject matter, looking for causation in the communicative activities of individual speakers (see below), the term *evolution* brings along a refreshing perspective, viz., the alternative of interpreting the relevant evolutionary processes as adaptive responses to changing communicative ecologies, both external (including other speakers) and internal.

Contact plays an important role in my approach to language evolution, the most significant part being contact between individuals rather than between populations (Mufwene 2001, chapter 2). In language evolution, the coexistence of two populations in the same geographical area is not a sufficient condition for language contact. They must interact with each other. At the level of linguistic communication, this is made possible by interacting individuals, who can spread features from the other language among monolingual members of their respective languages. Moreover, as pointed out by Weinreich (1953), the real locus of language contact is the minds of individual speakers.

Given the piecemeal way in which language acquisition proceeds (see below), speakers' minds are the arenas where selection resolves the competition that takes place among features received from various speakers and, in the case of multilingual speakers, also between features of the target system and linguistic systems that they have already worked out. The phenomenon known in contact linguistics as *interference* represents those cases when features of, say, a speaker's mother tongue are substituted for features of the target language – for instance when the English *th* in *think* is pronounced as *s* or *f* (thus *sink* or *fink*) by a non-native speaker. From the point of view of the development of linguistic or communicative competence, the total amount of linguistic knowledge speakers have, including the variants that compete for the same structural or communicative functions, is comparable to a 'gene pool' in biology, both at the individual and at the population levels.<sup>10</sup> In the case

<sup>9</sup> Those who are not impressed by the distinction I make between *evolution of language* and *language evolution* should note that in biology the term *evolution* is also ambiguous between the two senses that I highlight here. The advantages that the term offers over the traditional term *change* in broadening the scope of genetic linguistics with topics which are germane to language speciation outweigh the inconvenience of an ambiguity that is easily resolved by the context of the discussion.

<sup>10</sup> One clarification is needed here. Although the number of variants at the population level is a set-theory union of variants at the level of individuals, the strengths of the variants at the population level is a function of different kinds of dynamics of interaction and influences among individual speakers. This is also the level where it becomes more obvious that selection does not necessarily eliminate disadvantageous variants. As observed by Kretzschmar & Tamasi (2003) in American dialectology, populations have long memories of variation. Even the most marginal features can be resilient.



of language, the term *feature pool* (Mufwene 2001) seems to be an apt analog. All of these observations make more sense once we re-examine the concepts of “language acquisition” and “transmission,” as I do below.

*Language transmission* is a convenient misnomer in linguistics, because no speaker ever transmits a ready-made system to any learner, although caretakers simplify the learning task to the child by seriously limiting the number of topics for communication and the complexity of utterances. Both the range of topics and structural complexity (in terms of utterance length and embedding) increase as the child matures in age and in cognitive capacity, including processing skills. Otherwise, the burden of developing communicative competence rests on the learner, who, undoubtedly aided by Universal Grammar, must discover the units and principles of the targeted language or dialect and must (re)construct (an approximation of) its system.<sup>11</sup>

Language acquisition is thus a (re)construction process comparable to gene recombination in biology (Mufwene 2001, chapter 2). It can be called a *restructuring* process, though the term has mistakenly been associated with structural divergences that make creoles different from the European languages they have developed from. The main differences from gene recombination in biology lie in the fact that, in the best known cases, gene recombination takes place at the conception of an organism, while its genotype is being formed, with no agency on the part of the gene carrier. In those species of organisms such as viruses whose genotypes can change during their life time, one can argue that gene recombination takes place more than one time.

Languages are species whose phenotypes – the linguistic features on which typological classifications are based – correspond to no genotypes.<sup>12</sup> This peculiarity, which may appear strange to a biologist, is a consequence of the piecemeal way in which speakers develop competence in a language. From the point of view of population genetics, interesting questions arise regarding, for instance, why idiolects are not more different from each other than they are; why they wind up with similar, though not identical, combinations of features; and how some, but not all of, the individual changes they undergo can amount to communal changes. These questions are similar to those regarding how, with genotypes that are different from each other’s, organisms of a biological species do not wind up being more different internally than they actually are and why their phenotypes are so similar.

<sup>11</sup> This otherwise useful statement is problematic. The contents of UG – that part of human predisposition to cognition that putatively specializes in the development of linguistic and communicative competence – has not been explicitly articulated, at least not in terms of how what it is claimed to do cannot be accomplished by a general learning module of the mind. It is not evident either that, while he/she develops competence in a particular language, the learner really purports to develop a system, as systematic as he/she wants to be. One can also argue that a system simply emerges to the analyst – or is projected by him/her – out of the routines that an individual develops while purporting to communicate with particular sequences of spoken or signed gestures. Fortunately, these issues do not bear (significantly) on the main theses of this essay.

<sup>12</sup> Clearly languages and idiolects cannot have genotypes, because they are not biological systems. What is meant here is that they have nothing similar to a genotype in a biological organism or species. I also deliberately ignore here the fact that there is no isomorphism between phenotypes and genotypes, as well as the fact that ecology plays an important role in determining the phenotypes of a population. It is nonetheless true that only some genotypes can be associated with particular phenotypes, for instance the kinds of noses, hairs, lips, ears, limb morphology, and complexion that are found only among humans but not among other mammals.

Needless to say biology will not always inspire linguists who address such issues, nor will insights have to flow only one way in such comparisons between biological and language evolutions. Linguistics can contribute research avenues for biology too, if ontogenetic properties of languages and their peculiar mode of transmission can provide alternative and independent ways of addressing them. One way of accounting for the piecemeal way in which a language is “acquired” is the fact that its features are copied (typically with modification) in ways that are closer to horizontal transmission in biology, as in epidemiology, than to vertical, generational transmission.<sup>13</sup>

Although interactions between caretakers and children suggest that language is “vertically transmitted,” there is much more peer influence from the time the child interacts with other children, despite the fact that every child takes something from the older people they interact with. The importance of social interaction in language development should actually question the wisdom of capitalizing on age differences, rather than on experience, in the way language is “acquired”/“transmitted.” More remains to be thought out on this topic.

Another important feature of language development is also polyploidy, which makes it possible for the learner to incorporate in their emergent idiolects features originating in diverse idiolectal sources, including variants. These include alternative terms for the same concept (e.g., *pail* vs. *bucket*), alternative pronunciations for the same word ([fɪ'næns] vs. [ˈfaynæns] *finance*), or alternative grammatical strategies (e.g., *there's*/*there are* many things to do vs. *it's* many things to do).<sup>14</sup> Identifying and articulating the principles which regulate learners' selection in the development of their own idiolects (i.e., determination of preferred ways of saying things) is a challenge that linguists must face. Individual learners make their selections here in ways that do not seem so similar to the selection that takes place during the formation of a genotype, for instance, what particular genes the process of recombination makes dominant (determining the carrier's phenotypes) or recessive and under what particular conditions. To begin with, no agency of the gene inheritor is involved in the biology case. If ecology plays a role at the level of the formation of genotype, it is not in the same obvious way it seems to do during the formation of idiolects. In other words, it is not clear what factors or what particular selection algorithm a biologist would invoke to account for why particular genes become dominant or recessive in the genotype of a particular organism or for why a particular individual winds up

<sup>13</sup> I speak of “copying” tongue in cheek here. As Sperber (1996:141) observes, “1) De façon générale, les représentations ne sont pas copiées, elles sont transformées dans le processus de la transmission. 2) Les représentations se transforment par l'effet d'un processus cognitif constructif.” That is, 1) Generally speaking, representations are not copied, they are transformed during the process of transmission. 2) The representations are transformed by a constructive cognitive process. David Hull (personal communication, 2004) remarks that selection as an evolutionary mechanism is restricted to vertical, not to horizontal, transmission. I wonder whether the difference does not lie in the complexity of the process, rather than in whether or not selection applies. After all, it is the outcome that shows whether a gene or feature has been preserved in a population, not how it has been preserved or eliminated.

<sup>14</sup> The process is very similar, if not identical with, what Sperber (1996:147) identifies as “synthesis.” The bearer of, for instance, an idea or a story in the case of culture, or of a feature or even an idiolect in the case of language, receives input from several individuals, some of whom influence him/her more than others. The inputs contribute selectively and differentially to the formation of the new version of the original idea or story, or to the formation of the new speaker's feature or idiolect.

with a particular combination of phenotypical traits.<sup>15</sup> Linguists can invoke factors such as the statistical frequency of a feature, semantic transparency, regularity, salience, and social status of the model speakers.

For instance, a learner of English who is focused on regularities could easily and incorrectly substitute *falled* for *fell* as the past tense of the verb *fall*. However, although this kind of error often occurs, especially in child language, it usually does not crystallize into an idiolectal idiosyncrasy. The reason is simply that it cannot compete with the high statistical frequency of *fell*, which makes it less competitive. Interestingly, this is contrary to those cases where some speakers produce *dove* instead of *dived* as the past tense of *dive*, although the basic principle at work is the same. They do this by analogy to *drove* as the past tense of *drive*, because the latter also has a high statistical frequency.

Likewise, because the comparative construction with *more* is more common, many speakers will say *more fit* instead of *fitter*, which should be expected to be more typical of monosyllabic adjectives such as *short* > *shorter*. On the other hand, one also hears *funner* instead of *more fun*, simply because it is more consistent with *fitter* and *shorter*, despite the fact that its status as an adjective is less clearcut. There are thus various factors that affect what particular linguistic habits (features) a speaker internalizes in his/her idiolect. This is how selection works, especially at the population level, often not resolving the competition in exclusive terms, only in terms of dominance and differentially from one speaker to another.

Before proceeding, it may not be redundant to clarify that, like in biology, the terms *competition* and *selection* entail no agency on the part of genes or features. The term *competition* refers to the condition of inequality that obtains among variants in a feature pool, with some factors of their internal or external ecologies (dis)favoring some of them for dominance. Contrary to what is suggested by Dawkins' (1967) memetics, it is speakers who make the selections. They are in fact the ones who impose the inequality condition, based on factors, sometimes non-structural, that guide their preferences.

However, the way in which linguistic competence develops also suggests that in the linguistic species selection applies at the level of features (units or combinatoric principles). Although some ethnographic considerations suggest that selection also applies at the level of languages, when speakers target primarily features of a particular language over those of others, what we know about language mixing and the development of creoles suggests otherwise. Languages are selected indirectly through the fact that their features (sounds, words, combinatoric rules, and particular ways of packaging meanings) wind up

<sup>15</sup> We should remember here that selection does not necessarily operate in terms of total exclusion; it operates in terms of dominance. As explained by Sober (1984), it works more like in a golf tournament, in which scores are added up for every player. While the winner is the one with the highest score, the other players are still part of the overall competition and each occupies the position determined by their respective scores. The competition of linguistic features proceeds more or less the same way, making allowance for those features that are not dominant to coexist with the dominant ones in particular idiolects and to be used every now and then. Because non-dominant features in some speakers' idiolects can be dominant in other idiolects, languages can have longer memories than an oversimplified approach to language evolution, especially one conceiving of languages as organisms, may lead us to expect. The lives of linguistic features are thus to some extent comparable to those of genes, subject to mating/interaction practices.

constituting the majority of those selected from the combined feature pool of the language varieties in contact. Although clearly favored, the indirectly selected language (variety) also bears the influence of (some of) the disfavored varieties and is therefore modified into a new variety. This is what I identified above as Pyrrhic victory. It appears to explain how, under the influence of the Celtic languages over which it prevailed in especially southwestern Europe, Latin evolved into the Romance languages. The same evolution by selection, under the influence of substrate languages, is also true of creoles in former European plantation settlement colonies of the New World and the Indian Ocean. As the relevant socioeconomic history of the relevant plantation settlement colonies shows, a European vernacular was typically selected over African and other European vernaculars, but the emergent colonial variety also bears the influence of these other languages.

Thus, in partial support of Dawkins' (1976) hypothesis that it is genes rather than species which are involved in the selection process, it is indeed still features<sup>16</sup> which are the units of selection, although being associated with a particular language becomes one of the factors that favor those features.<sup>17</sup> In contact settings that produced creoles, only one target language, often misnamed the *superstratum* or *lexifier*, lends most of its vocabulary and grammar to the emergent vernacular.<sup>18</sup> Since the target has typically been structurally heterogeneous, availing competing variants, the selection of particular options, rather than others, into the emergent creoles shows again that selection applies to smaller units and combinatoric principles that collectively make up a language; it applies only implicitly or indirectly to languages themselves.

<sup>16</sup> Following Dawkins (1976), Hull (1988) opts for the term *replicator* to identify units of selection, with basically the same meaning as what I express here with linguistic *feature*, viz., units and principles that the learner of a language, in the present case, would endeavor to replicate. Croft (2000) follows both, alternating between *replicator* and *lingueme* 'linguistic meme'. I am sticking to the traditional term in sociolinguistics, because the others represent no improvements over it. They merely identify features as the elements that can be "replicated," albeit imperfectly – which is just a difference of perspective. Actually, the term *lingueme* conveys as much vagueness and inaccuracy as *meme*, especially because a language is not a body of utterances or texts (*pace* Croft) but rather the "system" that produces them.

<sup>17</sup> Although Lewontin (1970:14) makes allowance for populations to be selected, his primary position is that "many [population] adaptations turn out to be explicable by simple selection at the individual level" (13). This takes us back to idiolects and the piecemeal pattern of their emergence, through the cumulative selection of individual features and their (re)integration into a new "system." This process justifies the position I defend here. Another way of explaining it is that particular units, such as words (which also implies sounds and morphemes), and particular combinatoric rules (such as in syntax) are associated with specific languages. In practice, however, it is individual units and combinatoric rules which are perceived and can be selected in or out of the emergent idiolects. In multilingual or multidialectal contexts, units and rules from different systems (languages or dialects) are often mixed and can lead to the emergence of significantly different language varieties, such as the Romance languages (in relation to Latin and the Celtic substrate languages) or creoles (in relation to western European and African languages in European plantation settlement colonies of the New World and Indian Ocean). As should also be obvious from the literature on code-switching or mixing, what language or dialect a speaker claims to be speaking is sometimes determined more by the speaker's intention than by the actual text of his/her discourse.

<sup>18</sup> The terms are convenient misnomers because, in the case of *lexifier*, it is inconceivable that one would naturally target only the vocabulary of a language and ignore the grammar associated with the words, thus hoping to get it from some other source (Chaudenson 2001, 2003). The terms *superstratum* and *substratum* are also inaccurate in the case of creoles because they are based on the social stratification of the populations in contact but not on the time their languages arrived in the settings where creoles developed (Goodman 1993). Technically, in parts of the New World and the Indian Ocean where creoles developed, the relevant African languages arrived later than the European languages that the slaves appropriated as vernaculars and modified.

Selection is constrained by the specific external and internal ecologies of linguistic interaction. The external, socio-economic ecology imposes a particular ranking of variants (dialects and structural features) to which the selections made by speakers are sensitive. It is not enough of course to know whether a feature is standard or nonstandard. Another important factor is whether a particular feature will promote or discourage the acceptance of a learner/speaker by the particular socio-economic group that he/she wishes to fit in. The literature on covert prestige shows that in informal settings, where vernacular varieties are spoken, nonstandard features are normally preferred to their standard counterparts. This behavior accounts very well for the resilience of nonstandard vernacular features, as stigmatized as they are by the elites of various populations. The personality of the speaker/learner is also another important factor, which partly determines which group he/she wants to be associated with, if given a choice. Similarity to epidemiology is not difficult to establish here, as social practices bear on how viruses spread in a population, although different species are subject to different constraints specific to them in the selection of their features.

The internal ecology, which is no less important, consists of all the other variants that a particular one coexists with. The factors that contrast them with each other determine, relative to the linguistic background of the learner (part of the external ecology), which of them is likely to become dominant. For instance, in the case of creoles, the periphrastic comparative (with *more* in English) has been favored not only because the variant exists in the European language but also, and quite significantly, because the substrate languages typically have only a similar periphrastic option. This partial congruence of structural features favored the generalization of the periphrastic comparison in the new, plantation colony variety of the European vernacular. It is also important to remember in this case that the population of linguistic variants consists of those attested in the 17<sup>th</sup> and 18<sup>th</sup>-century non-standard varieties of European colonial languages, excluding some of the variants that we may be more familiar with today. These considerations prepare us not to be too easily impressed by the fact that structures of creoles are quite different from those of the standard varieties of the same languages today. Thus, we should not claim uncritically and too hastily that the origins of their present nonstandard peculiarities must be African. Rather, they should prompt us to compare creoles with non-standard varieties of the same languages, which should give us a closer sense of the extent to which they have diverged.

Then we must also recall that the target for those who made the creoles consisted of several nonstandard varieties competing with each other, which should make it normal to notice that features of creoles have diverse “super- and substrate” origins and therefore can differ maximally from non-creole varieties that did not develop under similar ecological conditions.<sup>19</sup> Even the European target itself was structurally heterogeneous, includ-

<sup>19</sup> This phenomenon is a consequence of both polyploidy and the fact that populations meet not like armies or sport teams engaging in contact at the same time and in an organized way, but rather as individuals engaging in particularistic contacts at different times and in different settings, all of them having the potential to influence the development of each relevant speaker's idiolect in some respect. How competition and selection of features apply within a population, under specific constraints determined by the relevant ecologies, accounts for the specific

ing not only features of diverse dialects but also xenolectal features from especially the European indentured servants who did not speak it natively. In the case of non-creole colonial vernaculars (such as White American English varieties), whose beginnings have been associated with koinéization (Montgomery 1995, Mufwene 2003b), the situation is thus comparable to one where members of different subspecies can mate with each other, across the subspecies boundaries. In biological terms, it is like having a colony where different subspecies of dogs can interbreed across their normal social boundaries and produce new, hybrid canine sub-species. In the case of creoles, the situation would amount to one where the dogs can also interbreed with members of related species such as foxes, jackals, and wolves, always bearing in mind that the modes of “transmission” in animal biology and in language are different, especially the fact that features of one particular language are likely to be heavily favored by the particular socio-economic setting of the contact.<sup>20</sup>

However indirect the process may be, community-wide targeting or selection of a particular language or dialect over (an)other alternative(s) can have particular ethnographic consequences. Languages that are less often selected for communication are endangered and may eventually die, as explained above. There are various reasons why speakers in a multilingual or multilectal community would prefer a particular language or dialect, or would often/usually not choose to speak another language or dialect. They are all practical, associated with the *hic et nunc* conditions of interaction, and need not be discussed here. They can also vary from one speaker to another in the same population. In the vast majority of cases around the world, the decisions are individual and taken at different times (however repeatedly) and in different tokens or kinds of settings, though members of the population do influence each other.

Similarities between language and biology arise here too in how competition and selection operating on features/genes and on organisms can lead to evolution at the species level. Future research should shed more light on details of similarities, how far they go and where they end. Note, for instance, that when such selections eventually lead cumulatively to situations where members of the population can no longer speak their language or dialect (fluently), loss of competence does not proceed uniformly within the overall population, confirming Mayr’s (2001:86) observation that “Since all changes take place in populations of genetically unique individuals evolution is by necessity a gradual and continuous process.”<sup>21</sup>

kinds of mix and heterogeneity that distinguish one language or dialect from another. Although the transmission mechanisms by which a viral species can display such genetic mixing and heterogeneity are not the same (see Part 4), the similarities in outcomes are too close to escape notice.

<sup>20</sup> I assume successful communication to be the closest counterpart of biological interbreeding. Differences between these notions are quite consistent with the fact that idiolects have no genotypes and develop gradually. An individual can learn only a limited number of linguistic features from a particular communicative event. As utterances share many features, there are many communicative events from which little or nothing new is learned, even during the early stages of the development of idiolects traditionally identified as “language acquisition.”

<sup>21</sup> This is of course in the context of species, with the process of evolution premised on variation, inheritance, and differential reproduction, although layman’s language allows individuals to evolve. As explained in Mufwene (2001, chapter 1), inheritance need not be interpreted in too strictly biological terms of gene transmission. It can be extended to “information copying,” as in culture and language, *pace* Fracchia & Lewontin (1999).

#### 4. Ecology and language evolution

An important dimension that for too long has received little attention in historical linguistics is the motivation for change, i.e., the (chain of) factors that trigger the process. Weinreich et al. (1968:102) formulate it under the name “*actuation problem*”: “What factors can account for the actuation of change? Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?”

McMahon (1994:248) articulates it in the following fairly complementary way: “The real actuation question is why some of these innovations [by individual speakers] die out and others catch on, spreading through the community, or why certain instances of variation become changes while others don’t.”<sup>22</sup>

Having focused on the subject matter for over twenty years, Labov (2001:466) reformulates the “problem” as follows: “Why here and now? The beginnings of change [applied to a population] are as mysterious as ever. Why *not* here and *not* now? Endings are equally difficult to understand. The obverse of the actuation problem is *continuation*. If change has already begun and is not coming to an end, it is continuing. What was the force that was missing a hundred years ago, that fuels the engine of the Northern Cities Shift<sup>23</sup> today and keeps it moving?” He thus widens the scope of the question, showing in the rest of his book that a variety of social ecological factors account especially for the ways changes are driven in particular language communities, often affecting only segments of the relevant populations but not others.

Recasting the actuation question/problem the population genetics way, I submit that the interaction of the external ecology of a language with its internal ecology should shed light not only on causes of linguistic changes and how these phenomena spread. Here too, there are similarities between, more specifically, linguistic and viral species in the ways changes occur, as they are affected by the social practices of their hosts. Critical to this approach is recognizing that changes often consist of only modifications of patterns of variation within a language, not necessarily of the introduction of new variants (McMahon 1994:248) or loss of some others. They can consist, at the population level, of shifts in the statistical frequencies of the variants, with some becoming dominant that used to be recessive. My basic assumption is that variation within a population is likely to remain stable unless something happens in its external ecology that disturbs the “balance of power” between the competing variants.

In the case of linguistic and viral species, factors such as migration to a new habitat, contact with another population, and other changes in the composition of the population bear on actuation. Migration also implies split of the proto-population, which, if proceeding randomly, can affect variation both in the migrating, colonial sub-population and in the metropolitan one (the individuals left behind in the motherland). Depending on

<sup>22</sup> It would be misguided to assume that the actuation question does not apply to language or dialect choice, as indirect as this process is. It is important to examine what particular changes in the ethnographic settings of communication would have speakers prefer to use one but not (an)other language(s) or dialect(s).

<sup>23</sup> The term applies to chain-style changes undergone by vowels in urban American English dialects.

how it was (self-)selected, the colonial sub-population may carry less than the totality of variants that obtained in the proto-population. The “balance of power” between the variants can shift in the colony, so that (some of the) features that continue to be marginal in the metropolitan population can become dominant in the colony. In the case of language, such a shift in the constituency of variants can eventually produce a new language variety, or it can simply cause the colonial and metropolitan sub-populations to evolve in divergent directions. Although this is not the full explanation, it contributes to accounting for why, for instance, English on the Falkland Islands is not the exact match of any dialect in England, despite the fact the Britons settled permanently in this archipelago in the late 19<sup>th</sup> century, much later than the English colonized North America, in the early 17<sup>th</sup> century. Other reasons for the linguistic divergence emerge from the discussion that follows.

The above observations should remind us that even if the English had not come in contact with other European populations in North America and in Australia, their language would have speciated anyway. Shifts in the constituency of the variants or in their statistical frequencies would have changed the “balance of power” and (dis)avored different variants. Traditional accounts that have simply invoked geographical separation by rivers, mountains, and oceans to account for language speciation had barely addressed the question, especially if they assumed a common ancestor for all the new varieties.<sup>24</sup> Likewise, changes in the patterns and dynamics of social interaction – which migrations help produce – affect the vitality of a particular (strain of a) virus in two or more allopatric populations that have the same ancestor. Socio-economic speciation in language and epidemiology is in some ways analogous to geographic diversification.

Language contact can also disrupt the extant pattern of variation, provided the host and immigrant populations interact regularly with each other, at least through some of their members. We are quite aware of the most drastic effects of recent European migrations to settlement colonies of the Americas and Australia on indigenous languages. Owing to specific population structures<sup>25</sup> – for instance who holds political and economic power and who makes concessions to whom – significant proportions of indigenous languages have been driven to extinction. An important reason lies in the changing socio-economic ecologies which have prompted the indigenous populations to assimilate the European populations’ practices, including the adoption of the colonists’ languages as vernaculars. In other cases where European vernaculars were appropriated by

<sup>24</sup> Paul (1891) often invokes shift in the frequency of variants to account for language change. It is curious why this factor has figured so little in 20<sup>th</sup>-century qualitative historical linguistics.

<sup>25</sup> I borrow this term from Goodnight & Wade (1999) and Wimsatt (1999), from whom the meaning is more implicit than explicitly articulated. It is clearly used in reference to the internal social organization of a population, as suggested by another term used by Goodnight & Wade: *population subdivision*. I use it below for a variety of social factors, including how a population is variously sub-categorized (often in overlapping terms, such as by gender, age, race, education, profession, and socio-economic class), how it is stratified, and who interacts or socializes with whom. My working assumption is that linguistic features spread along speakers’ patterns of socialization. Needless to say that I am rearticulating, from a population perspective, parameters that sociolinguists have traditionally invoked to account for language variation. My primary intention is to highlight not only ecological similarities between the linguistic and viral species but also methodological ones between linguistics and biology (too often ignored in traditional, 20<sup>th</sup>-century genetic linguistics) that must be used to advantage in order to better understand certain general mechanisms of evolution.



subordinated non-indigenous populations that have not been integrated by the Europeans, typically Africans in plantation settlement colonies of the New World and Indian Ocean, new language varieties have emerged that have been substantially influenced by African languages, which were also driven to extinction.<sup>26</sup> These are the new, colonial vernaculars that have been disfranchised from the Indo-European language family by the name *creoles*.

One somewhat oversimplified but still plausible account of such developments is that the other languages that the European languages came in contact with affected the patterns of variation within the European languages, owing to various cases of (partial) structural congruence between them and the other features that the non-European languages sometimes introduced into the systems (see Mufwene 2001 for a more accurate and nuanced account). The same processes operated in less drastic ways to produce new colonial varieties of European languages in settings where only these came in contact with each other under varying population structure conditions. White American English varieties are generally byproducts of contacts among various metropolitan English dialects and other European languages.<sup>27</sup> *Mutatis mutandis*, the same can be said of colonial varieties of other European languages. The whole evolution is similar to bringing populations infected by various strains of a particular virus in contact with each other, creating ecologies in which new strains can emerge, with each ecology producing its own dominant kind. In both cases, population structure, in terms of who interacts with whom, is an important ecological factor, to which I return below.

Both contact and migration influence the ways competition and selection affect variants differentially in the colonial and the metropolitan (sub-)populations. Eventually they produce new and divergent language varieties. Population structure – interpreted now as who lives where and under what conditions he/she can interact with members of another group

<sup>26</sup> The reasons for the extinction of the African languages in the plantation colonies are not identical to those for the extinction of the indigenous languages, although the institution of the new socio-economic world order by European settlers was a major stochastic event in both cases. As explained in Mufwene (2004a, 2004b), the particular way in which the plantation societies developed played a central role in an evolution that could otherwise be described as “chaotic” (in the context of chaos theory). They started from homesteads on which the Africans were minorities, integrated, and their children acquired colonial varieties of European languages as their primary, if not exclusive, vernaculars. These creole, locally-born slaves would become cultural, and especially linguistic, models for the bozal, incoming African-born slaves of the larger plantation societies, in which African languages were underrated. Societal multilingualism among the slaves just compounded the problem, leading to the abandonment of African languages in the same way that many of them are now losing ground to urban vernaculars in African cities. On the other hand, Native American languages would be endangered much later by European languages, when the new American societies would assimilate them and, in the case of North America, lured more and more of them out of the reservations, where living conditions worsened and the preservation of their traditions became more difficult.

<sup>27</sup> In the case of the United States, segregation by nationality among the European colonists during the colonial period must have reduced, by retardation, influence from continental European languages. Gradual language shifts from continental European languages must have taken place after the American English koinés had “crystallized” primarily from the contact of metropolitan English dialects, in the same way as on the Falkland Islands. Continental European linguistic influence must have been kept in check in the same way that influence from today’s waves of immigrants’ xenolectal speech is. By the Founder Principle, children acquire the extant variety natively and their parents die with their xenolectal features identified as such and largely not accepted by the host populations.

or population – has a great deal to do with how evolution starts and proceeds. Changes in the population structure also affect whether or not a particular change ends. In the way that Labov (2001) discusses actuation, changes occur even when the structure of a population does not seem to have changed. Assuming several changes to have started from below, in the way societies are economically or ethnically stratified, he considers social mobility and non-conformity as critical factors in the causation of changes.

Social mobility is actually another way the structure of a population changes, when individuals leave one socio-economic class for another and thus join new networks of interaction. “Non-conformity,” which means refusal to abide by all the norms of the new class, suggests the expectation on the part of the current members that one’s language remains the same all the time, thus that the current pattern of variation is there to stay.<sup>28</sup> Other factors of course determine whether the non-conformists will have followers and which of their “innovations” will spread and be recognized communally as change. In population genetics, they are more or less like individuals who introduce new viruses in a population and who can spread them largely thanks to how the host population interacts with them.

All the above observations make more sense once one factors in the following: 1) populations are constructed from individuals behaving singly most of the time, 2) they are not homogeneous, and 3) their members typically interact within small, overlapping networks. At least, from a geographical perspective, they are more like what ecologists call *metapopulations*, i.e., habitat patches interconnected by dispersing individuals (Hanski 1996). They are basically convenient groupings of individuals who interact with each other, within and across accepted social boundaries (racial, ethnic, gender, age, professional, economic class, etc.), as individuals, not as team members.<sup>29</sup>

An important question is: How do communal patterns emerge from individual speakers’ behaviors?<sup>30</sup> Changes typically apply in non-uniform fashions, affecting idiolects differentially and applying only to some sub-groupings of the larger population. Thus, Labov (2001) reports that, overall, African Americans have not participated in the Northern Cities Shift in the United States. In this particular case, population structure provides a useful ecological explanation, bringing to bear the fact that, as observed by Labov (see also

<sup>28</sup> Labov (2001:514) captures something similar with his “Golden Age Principle,” according to which “At some time in the past, language was in a state of perfection,” consistent with the common reaction among purists that one’s language is degenerating or decaying.

<sup>29</sup> Saussure (1916) may have been a little mistaken in analogizing a language, interpreted as a system, to a chess game, in which players follow communal conventions. Regarding languages, the putative conventions are more like emergent patterns (i.e., partial regularities emerging to the analyst) than something whose existence is independent of the speaker’s knowledge. Those partial regularities do not entail that a person who acquires a language naturalistically learns particular rules. Linguistic regularities are like beauty, in the eyes of the beholder. As McCawley (1976) points out, it is not evident that the generality of the rules by which speakers putatively operate coincides with those that the linguist seeks or claims to capture. This observation follows from the fact that a speaker or a naturalistic language learner is not a linguist. He/She does not process the materials from which his/her competence gradually develops in the same way as the linguist who gathers a (significant) corpus of (varying) data, analyzes the corpus, and seeks to capture the broadest generalization(s) applicable to the data that interest him/her.

<sup>30</sup> This question is connected to the disputable assumption among linguists that native speakers of the same variety use the same grammar (identically represented in all their minds) to process and produce utterances (Mufwene 1992b).

Wolfram 2000), Americans are sensitive to race distinctions. In fact, most Americans tend to socialize along race/ethnic lines and maintain a strong sense of ethnic identity. While most Whites would not like to be associated with their stereotypes of African American linguistic peculiarities, most African Americans, at least those associated with African American English, would not like the opposite evolution either. According to Wolfram (2000), young rural African Americans prefer to identify with urban African American vernacular English (see also Cukor-Avila & Bailey 1996), considering the conservative variety of their parents too similar to White English.<sup>31</sup> White and African American speakers would thus prefer to follow evolutionary trends within their respective ethnolects, ignoring or only following with interest those occurring in the other(s). The analogy with virology is clear, as viruses tend to be transmitted through interacting hosts.

We are thus also in a position where we can justifiably speculate on divergence phenomena. Could a population's determination not to be assimilated by, or confused with, another population account for cases where two varieties such as European American and African American English varieties evolve in divergent directions, even in the Southern states where over two centuries of intimate cohabitation predated the last century of segregated life? Isn't this what the literature on the divergence of White and African American English varieties is ultimately telling us, thus indirectly underscoring the significance of population structure as an ecological factor in evolution?

## 5. Conclusions

It is evident that studies of language evolution stand to gain a great deal by adopting a population genetics approach, one that acknowledges that populations are groupings of individuals, that the agency of many diachronic processes lies in the activities of individual members, that such processes need not be uniform across populations, and that the activities are largely determined by the ecologies in which the individual agents evolve. I submit that languages or dialects are species of idiolects, based on family resemblance. They share with viral species several properties that are relevant to understanding evolution; therefore the biological evolution model that linguists should consider for inspiration appears to lie in virology and/or epidemiology but not in animal biology.

Languages are like viruses essentially in that they are both parasitic species whose lives depend on the activities of their hosts, thus also on the latter's patterns of social interaction. However, it is equally important not to overlook some important differences between them, as these can account for their respective evolutionary peculiarities. Unlike

<sup>31</sup> This social attitude accounts for the resilience of African American English, as stigmatized as it is, because it also functions as a marker of identity. It also accounts for why the Ocracoke brogue, spoken by White islanders on the coast of North Carolina, is endangered by the vernaculars of white immigrants from the neighboring mainland (Wolfram & Schilling-Estes 1995), whereas Gullah, the creole vernacular spoken by African Americans in coastal South Carolina, has so far been endangered by the exodus of its speakers, not by the White mainlanders' "invasion" (Mufwene 1997). White mainlanders have mixed with White Islanders but not with coastal African Americans. In coastal South Carolina and Georgia, the residential communities remain racially segregated, just like many parts of American cities.

viruses, idiolects as linguistic organisms do not have anything corresponding to a biological genotype, simply because, if one can at all claim that they are begotten through the communicative practices of their speakers, their structures develop piecemeal. They develop incrementally as individuals learn to communicate and grow to produce increasingly complex utterances.<sup>32</sup> On the other hand, a virus starts its life with a fully-structured genotype (by gene recombination at conception) and has an incubation stage. This peculiarity is worth bearing in mind, although, like an idiolect (even in its mature stage), it can change its ontogenetic structure several times over, in a Lamarckian style, and it can be modified when it spreads to other hosts.

On the other hand, a virus may have started its life outside its current host, whereas an idiolect may not – only its features can. An idiolect's host is also its maker. Unlike the genes of a biological organism, the features of an idiolect cannot be transmitted – literally, in the form of a copy – from one host to another. They are typically copied with modification, under competition with similar inputs, by the learners, and they are recombined into new “systems” with some inter-idiolectal variation. Thus, no two speakers produce the same sound in physically and acoustically identical ways and no two speakers have exactly the same range of denotational and connotational meanings, as well as pragmatic constraints, associated with the same words. Cases of misunderstandings between speakers of the same dialect also suggest that no two idiolectal “systems” are identical, though speakers more familiar with each other have an easier time understanding each other. These observations follow from the fact that every speaker differs physiologically from other speakers and is thus equipped with a different configuration of speech organs.<sup>33</sup> Moreover, from a social ecological perspective, no two speakers have had identical experiences of social interaction, hence of being exposed to identical samples of speech or signing, which have influenced their language “acquisition” processes.

Still, despite these differences, there are similarities striking enough to justify the approach advocated in this essay. Polyploidy suggests that gene selection and recombination apply to viruses in more or less the same ways feature selection and recombination apply to idiolects, gradually in the latter case, but subject to ecological constraints in both cases. Idiolects are more likely to be influenced by those of speakers that their hosts/makers have interacted the most frequently with. Speakers accommodate each other, minimizing chances of being misunderstood. Likewise, viruses in a population are more likely to share a lot of genetic materials when their hosts socialize with each other than when they do not.

However, even under such conditions, it is still informative to understand under what more specific conditions a particular gene or feature becomes dominant or recessive, while the host has actually received, or been exposed to, all of them. How does selection work and what particular factors constrain it? Linguists can invoke markedness, or some

<sup>32</sup> By puberty, the cognitive infrastructure that enables this linguistic development is assumed to have reached an acceptable level of maturity, and the “linguistic system” – as inferred by the linguist – is considered full-fledged, although the vocabulary will continue to grow and more stylistic variation will emerge.

<sup>33</sup> The physiology of a speaker is part of the immediate external ecology of language “acquisition.” The variation alluded to here has to do with factors such as the following: specific shape and size of the oral cavity, length and width of the throat, size and length of the nasal cavity, alignment of teeth, and width and thickness of the tongue.

optimality constraints, to explain why some variants are (dis)favored at the level of both idiolects and languages or dialects. While a wide range of ecological factors can account for selection of particular genes for dominance within a biological population, it is not clear to me what factors can be invoked to account for a similar process at the level of (the conception of) organisms. Comparisons such as undertaken in this essay suggest that linguistics and biology can very well inspire each other in addressing evolutionary issues.<sup>34</sup> Questions regarding the specific ways in which selection proceeds at different levels are, I hope, among those that future interdisciplinary research in biological and language evolution will explore.

### References

- Adam, Lucien. 1882. *Les classifications, l'objet, la méthode, les conclusions de la linguistique*. Paris VII.
- Adam, Lucien. 1883. *Les idiomes négro-aryens et malayo aryens: essai d'hybridologie linguistique*. Paris: Maisonneuve.
- Becker, Karl Ferdinand 1833. *Das Wort in seiner Organischen Verwandlung*. Frankfurt am Main: Joh. Chris. Harmannsche Buchhandlung.
- Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma.
- Bickerton, Derek. 1984. The language bioprogram hypothesis and second language acquisition. In *Language universals and second language acquisition*, ed. by William E. Rutherford, 141–161. Amsterdam: John Benjamins.
- Bickerton, Derek. 1990. *Language and species*. Chicago: University of Chicago Press.
- Bopp, Franz. 1833. *Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Gothischen und Deutschen*. Berlin.
- Chaudenson, Robert. 2001. *Creolization of language and culture*. London: Routledge.
- Chaudenson, Robert. 2003. *La créolisation: théorie, applications, implications*. Paris: L'Harmattan.
- Chomsky, Noam. 2000. *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.
- Cukor-Avila, Patricia & Guy Bailey. 1996. The spread of urban AAVE: A case study. In *Sociolinguistic variation: Data, theory, and analysis*, ed. by Jennifer Arnold, Renée Blake, Brad Davidson, Scott Schwenter, & Julie Solomon, 469–485. Stanford, CA: CSLI Publications.
- Darwin, Charles. 1859. *On the origin of species*. London: John Murray.
- Dawkins, Richard. 1976. *The selfish gene*. Oxford: Oxford University Press.
- DeGraff, Michel. 1999a. Creolization, language change, and language acquisition: A prolegomenon. In *Language creation and language change: Creolization, diachrony, and development*, ed. by Michel DeGraff, 1–46. Cambridge, MA: MIT Press.
- DeGraff, Michel. 1999b. Creolization, language change, and language acquisition: An epilogue. In *Language creation and language change: Creolization, diachrony, and development*, ed. by Michel DeGraff, 473–543. Cambridge, MA: MIT Press.
- DeGraff, Michel. 2003. Against creole exceptionalism. Discussion note. *Language* 79.391–410.

<sup>34</sup> This is a position consistent with part of the conclusions of Hull (2002:27), when he states, “Of course the flow of information may just as well go in the other direction. Biologists may find the solutions that linguists have suggested useful in biology.”

- DeGraff, Michel. 2005. On creole genesis and language acquisition: Some methodological and theoretical preliminaries. To appear in the *Journal of Pidgin and Creole Languages*.
- Filppula, M., J. Klemola, and H. Pitkänen, eds., 2002. *The Celtic Roots of English*, Joensuu.
- Fracchia, Joseph & R. C. Lewontin. 1999. Does culture evolve? *History and Theory: Studies in the Philosophy of History* 38, no. 4, 52–78.
- Goodman, Morris. 1993. African substratum: Some cautionary words. In *Africanisms in Afro-American language varieties*, ed. by Salikoko S Mufwene, 64–73. Athens: University of Georgia Press.
- Goodnight, Charles J. & Michael J. Wade. 1999. The ongoing synthesis: A reply to Coyne, Barton, and Turelli. *Evolution* 54.317–324.
- Hanski, Ilkka. 1996. Metapopulation ecology. In *population dynamics in ecological space and time*, ed. by Olin E. Rhodes, Jr., Ronald K. Chesser, & Michael H. Smith, 13–43. Chicago: University of Chicago Press.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hey, Jody. 2001. *Genes, categories, and species: The evolutionary and cognitive causes of the species problem*. Oxford: Oxford University Press.
- Hock, Hans Henrich & Brian D. Joseph. 1996. *Language history, language change, and language relationship*. Berlin: Mouton de Gruyter.
- Hoenigswald, Henry M. 1990. Does language grow on trees? Ancestry, descent, regularity. *Proc. Am.Phil.Soc.* 134: 10–18
- Hull, David L. 1988. *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.
- Hull, D. L. 2002. Species, languages and the comparative method. *Selection* 3,1.17–28.
- Koerner, Konrad ed. 1983. *Linguistics and evolutionary theory: Three essays by August Schleicher, Ernst Haeckel, and Wilhelm Bleek*, with an introduction by J. Peter Maher. Amsterdam: John Benjamins.
- Kretzschmar, William & Susan Tamasi. 2003. Distributional foundations for a theory of language change. *World Englishes* 22.377–401.
- Labov, William. 2001. *Principles of linguistic change: Social factors*. Malden, MA: Blackwell.
- Lewontin, R. C. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1.1–18.
- Lieberman, Philip. 1991. Review of Bickerton 1990. *American Anthropologist* 93.998–1000.
- Lieberman, Philip. 2002. On the nature of evolution and the neural bases of human language. *Yearbook of Physical Anthropology* 45.36–62.
- Maher, J. Peter 1983. Introduction. In *Linguistics and evolutionary theory: Three essays by August Schleicher, Ernst Haeckel, and Wilhelm Bleek* (Konrad Koerner, ed.) xvii–xxxii. Amsterdam, John Benjamins.
- Mayr, Ernst. 2001. *What evolution is*. New York: Basic Books.
- McCawley, James D. 1976. Some ideas not to live by. *Die neueren Sprachen* 75. 151–65.
- McMahon, April. 1994. *Understanding language change*. Cambridge: Cambridge University Press.
- McNeill, David. 1992. Review of Bickerton 1990. *International Journal of Primatology* 13.575–583.
- Montgomery, Michael. 1995. The koinéization of colonial American English. In *Sociolinguistic studies and language planning: Proceedings of the XVIth Annual Meeting of the Atlantic Provinces Linguistic Association*, ed. by Catherine Philipponneau, 309–331. Moncton, NB: Centre de Recherche en Linguistique Appliquée.

- Mufwene, Salikoko S. 1991. Language genesis and human evolution. Review article on Bickerton 1990. *Diachronica* 8.239–254.
- Mufwene, Salikoko S. 1992a. The reviewer responds. *Diachronica* 9.149–151.
- Mufwene, Salikoko S. 1992b. Why grammars are not monolithic.” In *The joy of grammar: A festschrift in honor of James D. McCawley*, ed. by Diane Brentari, Gary Larson, & Lynn A. MacLeod, 225–50. Amsterdam: John Benjamins.
- Mufwene, Salikoko S. 1997. The ecology of Gullah’s survival. *American Speech* 72.69–83.
- Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, Salikoko S. 2003a. Genetic linguistics and genetic creolistics. Short note. *Journal of Pidgin and Creole Languages* 18.273–288.
- Mufwene, Salikoko S. 2003b. The shared ancestry of African-American and American White southern Englishes: Some speculations dictated by history. In *English in the southern United States*, ed. by Stephen J. Nagle & Sara L. Sanders, 64–81. Cambridge: Cambridge University Press.
- Mufwene, Salikoko S. 2004a. Language birth and death. *Annual Review of Anthropology* 33.201–222.
- Mufwene, Salikoko S. 2004b. Multilingualism in linguistic history: Creolization and indigenization. In *Handbook of bilingualism*, ed. by Tej Bhatia & William Richie, 460–488. Malden, MA: Blackwell.
- O’Hara, Robert J. 1994. Evolutionary history and the species problem. *American Zoology* 34: 12–22
- Paul, Hermann. 1891. *Principles of the history of language*. Translated from the second edition of the original by H. A. Strong. London: Longmans, Green, and Co.
- Pinker, Steven. Review of Bickerton 1990. *Language* 68.375–382.
- Radick, G. 2002. Darwin on language and selection. *Selection* 3, 1.7–16.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*, ed. by Charles Bally & Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot (Trans. Wade Baskin. 1966. *Course in general linguistics*. New York: McGraw-Hill Co.)
- Schleicher, August. 1863. *Die darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Haeckel*. Weimar: Böhlau.
- Schleicher, August. 1869. *Darwinism tested by the science of language*. Translated by Alexander V. W. Bickers. London: J.C. Hotten. Reprinted in Koerner, ed., 1983.
- Sober, Elliott. 1984. *The nature of selection: Evolutionary theory in theory and focus*. Chicago: University of Chicago Press.
- Sperber, Dan. 1996. *La contagion des idées*. Paris: Odile Jacob.
- Thomason, Sarah G. 2001. *Language contact: An introduction*. Washington, DC: Georgetown University Press.
- Tristram, H.L.C., 1999. *How Celtic is Standard English?* St. Petersburg: Nauka.
- Tristram, H.L.C., ed., 2000. *The Celtic Englishes II*, Heidelberg: Carl Winter.
- Vennemann, Theo. 2001. Atlantis Semitica: Structural contact features in Celtic and English. In *Historical Linguistics 1999: Selected papers from the 14<sup>th</sup> International Conference on Historical Linguistics, Vancouver, 9–13 August 1999*, ed. by Laurel Brinton, 351–369. Amsterdam: John Benjamins.
- Vinson, Julien. 1882. Créole. In *Dictionnaire des sciences anthropologiques et ethnologiques*. Paris.
- Vinson, Julien. 1888. La linguistique. *La grande encyclopédie* 22.286–296. Paris.

- Weinreich, Uriel. 1953. *Languages in Contact: Findings and problems*. New York: Linguistic Circle of New York.
- Weinreich, Uriel, William Labov, & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for historical linguistics*, ed. by W. P. Lehman & Yakov Malkiel, 97–195. Austin: University of Texas Press.
- Wimsatt, William C. 1999. Genes, memes and cultural heredity. *Biology and Philosophy* 14.279–310.
- Wolfram, Walt. 2000. On the construction of vernacular dialect norms. In *Papers from the 36<sup>th</sup> Meeting of the Chicago Linguistic Society*. Vol. 2: *The Panels*, ed. by Arika Okrent & John Boyle, 335–358. Chicago: Chicago Linguistic Society.
- Wolfram, Walt & Natalie Schilling Estes. 1995. Moribund dialects and the endangerment canon: The case of the Ocracoke Brogue. *Language* 71.696–721.