# Introduction

# 1

# Language Contact: What a Rich and Intellectually Stimulating History since the Late Nineteenth Century!

Salikoko S. Mufwene & Anna María Escobar

## 1. Preliminaries

In linguistics today, stating that one does contact linguistics draws no more curiosity than saying that one studies, for instance, sign language. Anybody working on some aspect(s) of language contact should know that the statement is not informative enough, because, as we show in Section 2, there is a wide range of topics that are subsumed by this umbrella label. Literally, as well explained by Weinreich (1953) and Cohen (1956), *language contact* is used in reference to the coexistence of languages in the mind of a speaker/signer and/or to their coexistence in a social or geographical space.

To be sure, no contact really takes place in the latter cases if segregation is so rigid that nobody in the coexistent populations of speakers learns the other ethnolinguistic group's language. In reality, from a macroecological perspective, language contact at the population level obtains when there is at least one individual that learns the other group's language and can spread its features in his/her group. Usually more than one "dispersing individual" are involved, driven by some reason to interact with members of the other community, though many other members of the relevant populations may not engage in such interactions. What is relevant is that language contact at the population level presupposes language contact at the level of at least some individuals. There are numerous population structures in which, owing to differences in economic and/or political powers between the coexistent ethnolinguistic groups, at least some members of one or both groups are expected to learn the other

community's language, such as in European colonies during the last few centuries (e.g., those who worked as intermediaries, interpreters, or colonial auxiliaries; see Samarin 1989, Lawrence, Osborn & Roberts 2006, Mufwene 2008, 2020a, Van den Avenne 2017).

Contact linguistics is about various facets and consequences of the coexistence of languages in individual speakers' minds and in particular populations. Instances of these include language shift at the level of either the individual or a population. A consequence of this may be language endangerment and loss (LEL), if most or all the members of the relevant population speaking a particular heritage language stop using it and thus converge in shifting to one that they either consider more advantageous to them or are forced by various ecological pressures to speak more often (e.g., Fishman 1971, 1991, Mufwene 2020b). The outcome of this convergence of behaviors can be compared to the emergence of new norms in a particular language in a contact ecology, when speakers converge in selecting variants (forms or structures) that come from another language they or some of them also use.

Such feature adoptions have generally been referred to as transfers or borrowings, depending on the scholar's perspective. On the other hand, the convergences also produce norms even in monolingual populations, out of the contacts of different dialects or simply out of the contact of different idiolects, as speakers negotiate tacitly about features of their respective varieties. One can thus conclude that the fundamental contact is enabled by interacting individuals, at the level of idiolects, native and/or non-native; whatever else happens at the population level is produced by convergence (Mufwene 2001). Labov (1972, 2001) explains this well in showing how outliers sometimes bring into their social networks linguistic features they learned in another network. It takes a central or influential network member that the outlier interacts primarily with to copy the feature from him/her and spread it to other members (Eckert 1989). That is, the latter copy it from the central or influential member, or from other copiers, rather than from the outlier (now the "dispersing individual") who brought it to their network. This is basically how change occurs, with some individuals innovating or introducing (different features of course) and others copying and spreading them (Croft 2000, Mufwene 2008, Fagyal et al. 2010).[1] Thus, we can argue later that contact is an important actuator of language change.

This approach can also help us explain why some language varieties are less, or more, "focused" than others (using LePage & Tabouret-Keller's 1985 terminology). For example, foreign-workers' interlanguages differ from pidgins, at least the expanded ones that we know today, because they lack communal norms, although the speakers produce some of the same

[1] Fagyal et al. (2010) show that "dispersing individuals" store "a great diversity of variants," some of which can spread in their interaction networks and constitute change.

features in their interlanguages (Perdue & Klein 1992, Perdue 1995). The reason for this state of affairs is that (allowing some over-simplification) several migrant workers live in segregated groups in which they socialize in either their heritage language or some other language of their countries of origin while their children learn the local language from native speakers. Interlanguages remain individual-speaker phenomena, which the migrant workers (adults) produce only when they interact with people that do not speak the language(s) they are more competent in. By contrast, pidgins are, according to the received doctrine, communal varieties arising primarily from the regular interactions of non-native speakers with their trading partners and among themselves in the trade language to which they have had limited exposure.

Through their interactions, speakers influence each other – accommodating each other on different features – and converge toward some group or communal norms (despite some natural variation in the emergent system). The ecologies in which the migrant workers' interlanguages emerge (Pfaff 1981, see below) are not conducive to the emergence of communal norms, as they do not form a community with their changing interlocutors with whom they communicate in the host country's language. Although the occasional accommodations made by fluent speakers to their deviations from the target language – a kind of foreigner talk at the workplace – may reinforce the departures at the individual level, they do not at some communal level. Communal norms emerge and can give rise to new varieties in contact ecologies in which the speakers interact or socialize frequently with one another, accommodating each other's forms and structures.

As much as some literature on (naturalistic) L2 acquisition has focused on interlanguages, Mufwene (2010) concluded that this scholarship can only make a limited contribution to research on the emergence of creoles, because it is focused on individuals, whereas genetic creolistics deals with communal norms as the outcome of convergence. The same can be said of contact dialects. To be sure, one can learn about transfer and substrate influence (such as about possible trajectories of linguistic influence) but not about how some substrate elements attested in some interlanguages have converged into substrate influence in a creole, while some others do not.[2]

Consistent with Braj Kachru's (1985) distinction between the "Inner," "Outer," and "Expanded Circles" of World Englishes, the presence of communal norms helps distinguish the Englishes of the Outer and Expanded Circles. Because English generally functions as an official language in the Outer Circle – which consists of former British exploitation colonies – and the national elite have embraced it as an emblem of their

[2] This regards the complex dynamics of competition and selection from a communal feature pool (Mufwene 2001, 2002) that the ecologies which generate the migrant workers' interlanguages do not produce.

social status and often socialize in it, national or regional norms have emerged in different parts of the world. This practice has generated what outsiders can identify as Indian, Singaporean, South African, and Nigerian norms, among other varieties misidentified as "indigenized Englishes."[3] These norms are the outcomes of competition and selection in their respective communal (national or regional) feature pools, in which individual speakers accommodate each other and converge toward shared local, national, or regional features (Mufwene 2001). The use of English in the Expanded Circle, on the other hand, does not lead to the emergence of national or regional norms. For instance, Japanese or German speakers of English in, respectively, Japan or Germany do not normally speak English among themselves.[4]

The study of language contact at the population level entails an additional level of complexity that has usually been overlooked in the literature, which we all should be aware of, although this dimension is generally also missing from this Handbook. The omission appears to be a legacy of the beginnings of linguistics, with historical and genetic linguistics, in which languages were conceived of as organisms, therefore without internal, inter-idiolectal variation (without their social life). It is also a consequence of the coexistence of the research area with theoretical linguistics, so central to linguistics to date, where it is generally assumed (in fact since de Saussure 1916) that communal features are shared by all individual speakers. Numerous grammars have been published that are based on work with one or a couple of consultants! We have generally not compared idiolectal systems with their communal counterparts, which are typically constructs of convenience, in order to check, for example, whether they are isomorphic in their details. Such isomorphism would be contrary to what the study of interlanguages suggests. What is the status of features that remain idiosyncratic of particular idiolects and are not shared by all the members of the community of practice?

[3] Identifying them as "indigenized Englishes" is quite biased, because American and Australian Englishes have indigenized as well, in the sense of being rooted in the new ecologies where they have evolved (Mufwene 2009). Note also that national categories are inaccurate constructs of convenience; those who reside in the relevant countries make more distinctions. For instance, there are South African Black, Colored, and Indian Englishes, which are distinct from South African White English, which is associated with White South Africans who are L1 speakers (Mesthrie 2017). In polities such as the USA, the Natives distinguish between dialects and sociolects, while in many Outer Circle countries, one can recognize variation associated with particular regions or ethnolinguistic backgrounds. That is, norms allow variation – from inter-idiolectal to social and regional – in ways that cannot be determined for Englishes of the Expanding Circle, despite obvious similarities among the idiolects of speakers who share the same vernaculars.

[4] To be sure, now that multinational corporations have important offices staffed with senior personnel from different places around the world and also hold virtual meetings at that senior administration level, people from the same national background *do* get to interact with each other in English. However, these are not the kinds of interactions that would produce national norms. Besides, whoever has observed interactions at such meetings or workplaces will notice that people from the same national backgrounds will talk to each other privately in their national language.

Despite progress in the scholarship on L2 acquisition, studies of bilingualism and multilingualism are dominated by those focused on the population level.[5] Perhaps because urban youth "stylects" (see Section 3) have been treated as discourse phenomena (from the perspective of performance), the discussion of the relation between variation at the idiolectal level and variation at the communal level has not come up. We hope to make more sense of (some of) these questions in the next section, where we argue that research on language contact has evolved from being marginal in linguistics to acquiring a central position, with contact at the population or the individual level acknowledged as an actuator of language change. Thus, the challenge has been to articulate the extent to which contact dynamics (dis)favor the emergence of new communal norms. These apply to the emergence of contact varieties – including creoles, pidgins, other mixed varieties, new ethnolects, and interlanguages – and evidently also to convergence and divergence between languages. The social histories of populations in contact, including the roles that multilingual speakers play in their respective ecologies, draw attention to the complexity of language change and of the social aspects of linguistic behavior. In the next section, we provide a selective synopsis of how language contact as a research area has evolved since the nineteenth century.

## 2. A Historical Survey

The study of language contact has been part of linguistics since the late nineteenth century, at variance with what is now known as genetic linguistics. The latter was then preoccupied with the uniparental representation of language speciation in especially the Indo-European family, and indeed also with the classification of languages into families and subfamilies such as Romance and West-Germanic. The comparative method of the nineteenth-century Neogrammarians has remained the backbone of the research area to date, though it is now enriched with computational modeling techniques. The Stammbaum, now also identified as cladogram, continues to serve as the proud demonstration of the success of that particular scholarship.[6] Lexical borrowings were excluded from the comparisons. No question was asked as to why language contact was considered a spoiler of what otherwise should be perfect, recursive speciation processes corresponding to different time depths. Deviations from the regularities of change revealed by the comparative method, such as "sound laws," were accepted

[5] There is indeed well-established scholarship on child bilingualism, which has also contributed to the important "Talk Bank" called CHILDES. However, as can be noticed from publications such as Yip & Matthews (2007) and Silva-Corvalán (2014), the scholarship informs the reader about the development of bilingualism in individual speakers, even siblings in these specific cases, but not about the emergence of bilingual norms in a community.

[6] More recent modeling techniques include splitgraphs for neighboring languages that represent contact zones of unrelated languages, such as in the Amazonian Basin (Epps & Michael 2017).

as exceptions, as long as they were not numerous enough to question the accuracy of the relevant analyses.

The inconvenience of language contact became more obvious in the late nineteenth century, when European philologists and other precursors of (genetic) linguists noticed how English, French, and Portuguese – and of course also Dutch and Spanish – had metamorphosed into creoles and pidgins (interpreted then as broken languages) in, respectively, the European plantation settlement colonies around the Atlantic and in the Indian Oceans and in trade colonies of West Africa and the Pacific. Both kinds of new language varieties were outcomes of how differently the non-European speakers of non-Indo-European languages reproduced the Indo-European languages they had learned. Contact was obviously the reason; and the question of the role of language contact in language speciation needed to be addressed.

However, the social ideologies of the time, which assumed purity of species and languages, disavowed these new colonial varieties of Western European languages as impure "bastard tongues" – as Derek Bickerton provocatively also referred to them in the title of his 2008 book – or aberrations that were exceptional in the way they had emerged. With some exceptions, such as Hugo Schuchardt (see below), those who described them attributed the structural changes to the mental and anatomical inferiority of their presumably less evolved speakers.[7] This is the explanation then advanced by Adam (1883), Baissac (1880), Vinson (1882, 1888), Bertrand-Bocandé (1849), and Gonzales (1922). Then, genetic linguists apparently treated language contact as irrelevant to language speciation, generally ignoring Schuchardt's (1882) position that creoles and pidgins suggest instead how language diversification actually occurs. More specifically, Schuchardt attributed the current structures of modern languages to what Roger Lass (1997) calls "imperfect replication" by successive generations of their speakers, under contact conditions. Notwithstanding some oversimplification of the actual restructuring process on our part, substrate influence during language shift, for example, accounts for how Vulgar Latin had evolved into the Romance languages, with the "reproductions" incrementally diverging from the lexifier, as in the case of creoles.

Ignoring Schuchardt, genetic linguists dismissed creoles and pidgins by fiat as "mixed languages" and therefore anomalies by contrast to European varieties of Indo-European languages, all the way into the first half of the twentieth century.[8] Likewise, Hjelmslev's (1938) contention that all languages are mixed to some extent got negligible attention. To date, there is

[7] Although opposed to slavery, advocating thinking of races as biological species and noting that non-Europeans expressed emotions in the same ways as their European counterparts, Charles Darwin nonetheless characterized Africans and Native Americans as less evolved than Europeans in *The descent of man* (1871).

[8] Whitney (1881, cited by Appel & Muysken 1987) invoked borrowings to account for deviations from the findings of the comparative method. He actually may have been the first to think of, in Haugen's (1950) terminology, a "scale of adoptability" from nouns to grammatical structures. However, we could also see the influence of this genetic linguistics

still a category of "mixed languages," whose definition varies according to author. For instance, Thomason (2001) offers a less restrictive interpretation, which includes creoles – unlike some creolists who do not fit them in this category – although she is still at variance with Hjelmslev. Language contact is the reason why it is still widely assumed that creoles do not belong in the same genetic families as their lexifiers (Thomason & Kaufman 1988).[9] One must wonder why no natural account has been presented for why, for instance, all the Indo-European languages do not have the same morphosyntactic template that they would have inherited from Proto-Indo-European and the grammars of the Romance languages do not represent a continuous rectilinear or unilinear evolution from that of Vulgar Latin. Yakov Malkiel (1978) is among the exceptions in invoking contact to explain differences between Western and Eastern Romance languages.

The first half of the twentieth century saw the development of structuralism in Europe, with Ferdinand de Saussure (1916) showing a greater interest in synchrony than in diachrony and developing more focus on structure/system (*langue* in French) than on speech (*parole* in French), although he situated much of the actuation of language change in the latter. The same period also saw the development of descriptivism in Anglophone North America, also mainly focused on synchrony, which was prompted largely by the endangerment of Native American languages. Anticipating that they would soon vanish, North American linguists endeavored to describe these languages.

The linguists were also in a situation where, unlike their European counterparts, they could find so many non-Indo-European languages within their own borders and were sensitive to the negative impact that the dominant European languages brought by the colonizers exerted on their vitality. Notwithstanding this, Sapir (1921) acknowledged "individual and communal variation in English" (p. 157), as well as "linguistic interinfluencing" (chapter 9) between the languages of neighboring populations, where influence tended to be in one direction, from the most dominant culture to the other. Bloomfield (1933: 476) also noted that speakers adapt their speech habits to those of their interlocutors, where the conditioning factors

tradition on Chomskyan linguistics in the 1980s, in the CORE/PERIPHERY distinction, in which, strangely, the "core" applies to the largest part consisting of regularities in the grammar of a language, while the "periphery" includes exceptions (or what Haugen 1950 calls "residual structural irregularities"), which are associated with language contact and other vicissitudes in the history of the relevant language. This is justified with, for instance, irregular noun plurals in English, as in *children*, *oxen*, *stimuli*, *corpora*, and *criteria*, with the latter three reflecting a history of borrowings from, in this case, Latin and Greek. By contrast, the first two reflect retentions from Old English and Middle English, in which the suffix *-en* was an alternative to *-(e)s* (as in *oxes* vs. *oxen*). Frequency has helped the forms survive the general trend of nominal plural with *-(e)s*.

[9] This position is disputed by Posner (1983, 1985, 1996) and Trask (1996), who claim that French, Spanish, and Portuguese creoles are the newest Romance languages. According to Mufwene (2001, 2005, 2009, 2020a), what matters the most is that they are natural offspring of their lexifiers and have not evolved exceptionally (see also DeGraff 2005).

are the "density of lines of communication" and "the relative prestige of [the] social groups" (p. 345) (Fagyal et al. 2010: 263).

Both structuralist and descriptivist approaches to language yielded several descriptions of non-European languages, with the "system" as its object. No significant study of language contact emerged until the middle of the twentieth century, about the same time Noam Chomsky (1957) would revolutionize the field with generative linguistics. Although a turning point in modern linguistics, this new trend has also remained focused on structure/system (a static framework, Bailey 1973: 34–5) and reiterated commitment to study competence and the faculty of language (or Universal Grammar – corresponding to de Saussure's *faculté de langage*) over performance (corresponding to de Saussure's *parole*). The generative approach helped cement the view that internal factors were sufficient to achieve rigor in linguistic explanations.

It was at this critical moment, in the mid-twentieth century, that Charles Ferguson, Joshua Fishman, Einar Haugen, and Uriel Weinreich, among others, regenerated interest in language contact, albeit in different though related ways. While very much interested in the Chomskyan Revolution, Uriel Weinreich revived the study of bilingualism, identifying the individual, more specifically the speaker's mind, as the locus of language contact, especially through his study of interference. He proposed distinguishing between three types of bilingualism: coordinate, compound, and subordinate. This terminology has since been used in studies of both second language acquisition and societal bilingualism. The research, which culminated in his book *Languages in contact* (1953), would also establish the connection between language contact and language change (a dynamic framework, Bailey 1973: 34–5), embedding both in the social context of speakers' interactions and highlighting "the interrelations between language history and culture history" (Bleaman 2017). This is reflected in the following famous statement of his: "In speech, interference is like sand carried by a stream; in language, it is the sedimented sand deposited on the bottom of the lake. The two phases of interference should be distinguished" (Weinreich 1953: 11).

The distinction Weinreich made between individual and societal bilingualism led him to pay more attention to various social factors that govern the functional distribution of the coexistent languages in their speakers' lives, viz., which communicative domains are associated exclusively or primarily with which particular language (see also Ferguson 1959). Indirectly, he also advocated for what we are calling "communal norms" when he wrote: the "impact of interference phenomena on the norms of a language may be greater if the contact occurs through groups of bilinguals" (Weinreich 1953: 3).

In collaboration with Marvin Herzog and William Labov, he drew attention to the actuation of language change, situating this not only in language-internal variation but also in language contact, a facet of the

social context of language practice. He characterized this approach to language change as speaker-oriented. Anticipating one interpretation of uniformitarianism in linguistics today, their seminal essay, titled "Empirical foundations for a theory of language change" (1968), argues that "the same social and linguistic factors that condition linguistic variation today also account for historical change," as paraphrased by Bleaman (2017). Perhaps this explains partly why variationist sociolinguists like talking about "language variation and change" and even have a journal named this way.

However, the phrase is problematic from an evolutionary perspective, as variation provides materials on which evolution works – in the case of language, through interactions among idiolects, i.e., inter-idiolectal contacts. However, variation is not necessarily indicative of change in progress. Nonetheless, Weinreich, Labov & Herzog (1968) were correct in arguing that "finding order and structure" in "linguistic heterogeneity" (which they also characterized as "orderly heterogeneity," p. 100) can say something about the "course of [language] development" (p. 99). The order was to be found in the "social and stylistic determinants" of heterogeneity that regulate what speakers do.

Einar Haugen developed an interest in bilingualism among Norwegian immigrants to the USA and its impacts on their heritage language. Embedding (American) Norwegian and its speakers in the population structure in which English is the dominant and economically more powerful language, he went on to study its attrition. This was at a time when, as noted by Baran (2018), European immigrants who arrived after the American Revolution were not immediately recognized as American and all people of European descent were not yet united under the label of "White Americans." With his *The Norwegian language in America: A study in bilingual behavior* (1953), Haugen pioneered the study of language endangerment and loss (LEL) in North America. In his subsequent book *Bilingualism in the Americas* (1956), he distinguished languages depending on their social history: native, colonial, immigrant, or creolized.

Haugen also distinguished what he considered true language loss from language mixing. The former is produced by language shift, while the latter is a consequence of attrition in the heritage language and represents language change. Rejecting the nineteenth-century ideology of language purity, he would have thus disagreed with those creolists according to whom creoles coexisting with their lexifiers are dying by "decreolization" (e.g., Hazaël-Massieux 1999), i.e., the process by which speakers substitute acrolectal features for basilectal ones. (We return to this topic below.) Note that both language shift and language mixing can be identified in, for instance, Nancy Dorian's (1981, 2010) work on the obsolescence of Sutherland Gaelic and certainly also in the endangerment and loss of the languages of immigrants from other nations in European settlement colonies since the fifteenth century, where the language of a particular nation