CHAPTER 2

A Linguistic-Historical Model
Social Factors in Grammatical Reduction, Imposition, and Adoption

When analyzed with a new approach, the Old Persian inscriptions tell us something important about changes to the constitution of the Persian *people* as a group and their culture during the reign of the Achaemenids. The radical trimming of the inflectional complexity of the Persian language in about one and a half centuries – from the complex inflections of the language of the Bisitun inscription of Darius I in 521 BCE to the much-reduced inflections that peek through the inscriptions of Artaxerxes II (405–358) and Artaxerxes III (r. 358–338) at Susa – points to a *demographic* transformation. This chapter explains how we can determine this through a composite model synthesized for the use of historians from several related subfields of linguistics: historical linguistics, contact linguistics, creole studies, language acquisition research, and studies of bilingualism and grammaticalization. The model, which is based on the combination of observed patterns of language change in population contact and in second-language acquisition, is the main factor that distinguishes the present attempt to account for the evolution of Persian from other attempts, giving new results.[1] As I will show in the subsequent chapter, this model will help us to understand the history of the Persians and their first empire. The same linguistic-historical model will apply, necessarily with a different application and a different outcome, to my account of the origins of New Persian, which is the subject of another book to appear after this one.

This linguistic-historical model is not a tentative hypothesis. Rather, the linguistic phenomena on which it is built have been documented in numerous places and times. The model offers robust inferential explanations and reliable predictions on a par with the predictions offered by

[1] Gerlee and Lundh (2016) provide a clear introduction to the concept of a scientific model and different methods that fall under that category.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

comparative historical linguistics.[2] I claim that it describes universal tendencies in specific socially conditioned varieties of language change.[3] It derives from the distillation of hundreds of articles and books, of which the bibliography here is merely representative, written by researchers whose main interest is not in Persian or other specific languages to the exclusion of others, but in these kinds of general linguistic phenomena. Its bases are recent in scholarship but not new. Those who wish to probe the foundations of the model synthesized here will need to pursue the works listed in the notes, but I hope to have simplified the task of reapplying the synthetic model presented here to other linguistic histories. The study of second-language acquisition, bilingualism, language contact, creoles, and historical linguistics all concern related phenomena, but these subfields of linguistics seem to interact minimally, so they have sometimes generated idiosyncratic jargon differing from linguistic subfield to subfield.[4] Such differences made synthesizing their respective findings a more challenging task, but the good news is that they often independently corroborate each other's findings, too, in a way that can be overlooked when subfields develop separately. The result will probably not satisfy any linguist, as it participates wholly in no subfield of linguistics, being intended to serve linguistic history. For the sake of nonspecialists, I have eschewed some of the jargon entailed in different subfields of linguistics except where I deem technical neologisms useful, in which case I explain the terms.

There are two main parts of this linguistic model, but the two interact enough in their application to linguistic history that I conceive of them as one complex whole that tells us more when they are combined than we can see when they are kept separate. The first part is the most immediately relevant to the problem of the evolution of Old Persian described in the previous chapter. It concerns the social factors triggering grammatical reduction in a language, particularly the loss of inflectional morphology. The second part deals with social factors in linguistic mixture – more precisely the transfer into one language of features from one or more other languages, not always

[2] See Muysken 2017 for a discussion of the use of established sociolinguistic scenarios to illustrate historical linguistic patterns and vice versa.

[3] I use the term universal tendency as Haspelmath proposes (2019: 5), because, while I do not know of any exceptions to the processes posited here, I do not claim that there are no exceptions. This makes the history of Persian studied in this book quite relevant for comparative linguistic purposes. If no exceptions exist, as we may suspect, then this model describes a universal characteristic of human language change under the specified conditions. In that case, the history of Persian studied in this book would be a good example for general linguistics.

[4] E.g., Muysken (2013: 193) observes that, "what is needed is a book in which findings from bilingual processing studies are fully integrated with those from language contact studies."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

a reciprocal process. As concerns the evolution of Old Persian, this second part of the model turns out to be less decisive, but it guides us away from errors in analysis and substantially refines our understanding of the social circumstances that transformed Old Persian. Both parts comprise contributions of innumerable other scholars, but I have not seen them combined and applied to linguistic history as follows in this book.

The composite model employed here is incomplete as it cannot explain or predict all linguistic history, nor is it intended to do so. I do not offer a master theory of all social factors in language change. The model can and should be refined and expanded in the future, integrating new findings and new components. Nevertheless, I believe that the analysis of the history of ancient Persian here, using these tools, resolves an old problem in linguistic history productively. In this chapter, however, I will seldom refer to Persian, the subject of this book. Partly this is deliberately to separate the presentation of the model and its application. Partly it is in the hope that this chapter will be useful by itself as a synthesis of research findings for other linguistic historians interested in correlations between language change and social change but who are not concerned specifically with Persian. Strictly speaking, one does not need to know about ancient Persian to understand this chapter.

Grammatical Reduction: Social Factors in the Loss of Inflection

How does a language lose its inflectional morphology?[5] The findings of various branches of linguistics most relevant to this problem can be approached through five steps, to be explained in order, with basic bibliography at each heading.[6] After explaining these five steps and then illustrating their outcomes with examples from modern circumstances, I will discuss a few more critical factors in the topic of socially induced language change.

(1) Languages vary in their relative grammatical complexity. Some have more complex grammars than others, specifically in inflectional morphology. This is a realization that has faced ideological barriers on two sides.[7]

[5] Sims-Williams and Baerman (2021) offer a subtle typology of grammatical loss in change.
[6] Winford (2003) provides a careful general introduction to the field of contact linguistics.
[7] Kusters 2003; Winford 2003: 52, 217–219; Dahl 2004; DeKeyser 2005; McWhorter 2005: 43–48; Shosted 2006; McWhorter 2007: 21–58; Wray and Grace 2007; Miestamo, Sinnemäki, and Karlsson 2008; Siegel 2008: 18–25; Sampson, Gil, and Trudgill 2009; Trudgill 2010b; McWhorter 2011a: Trudgill 2011; 64–102; Szmrecsanyi and Kortmann 2012; Trudgill 2012; McWhorter 2013b: 415–417; Fenk-Oczlon and Fenk 2014; Baerman, Brown, and Corbett 2015; Dixon 2016: 125–146; McWhorter 2018: 90–109.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Those who learn or teach many languages already know this from direct experience, not as mere intuition. Complexity has many senses, but what I mean by complexity here is straightforward, practical, and descriptive: here, higher complexity in a language means that it is more difficult for adults to learn and to use actively and communicatively.[8] Such complexity can reside in morphology and syntax, lexicon, and other subsystems of a language. Complexity includes grammatical overspecification, structural elaboration, and irregularity.[9] Simplification in a language is an increase of regularity.[10] Inflectional morphological complexity is only one kind of linguistic complexity, but it is by far the most relevant here. It is perhaps the easiest to quantify meaningfully and it is the factor in language complexity that matters more than the rest for adult learners. Complex inflectional morphology burdens the learner with many semantic distinctions to be correlated formally on the spot while using a language in live communication. It is difficult for a new adult speaker of a language to conceive and produce several correctly inflected word forms that correspond with other words' disparate and grammatically heterogeneous forms beyond their phrasal constituents in a sentence, and to pronounce them correctly, spontaneously within a few seconds to keep the listener's attention, and in a way that conveys the intended meaning.[11]

To illustrate such a discrepancy in inflectional complexity between languages using descriptive grammatical terms, compare the ancient Greek verb of Plato's Athens with the verb in ancient Aramaic of Achaemenian Mesopotamia at the same time. Ancient Greek verbs were very complex. Each Greek verb was inflected, potentially, for many hundreds of forms, considering three persons, three numbers (singular, plural, and the rarer dual), seven tenses (present, imperfect, future, aorist, perfect, pluperfect, and future perfect), three voices (denoting verb diathesis, such

[8] Cf. Kusters 2003: 6–9; Trudgill 2009 and 2010b: 311. The nature of difficulty itself is debated in the study of second-language acquisition; see DeKeyser (2005) for clear answers.

[9] These three are emphasized by McWhorter (2011a: 1–2; 2018: 93–99). All of these are hindrances to adult learners. The factors accord inversely with Trudgill's (2010a: 4–5; 2010b: 307–308) characterization of simplification: "the regularization of irregularities," "increase in lexical and morphological transparency," and "the loss of redundancy."

[10] Romaine 1988: 32.

[11] Through various avenues of research into second-language acquisition, Slabakova (2008; 2016: 389–415) proposes that it is precisely inflectional morphology that forms the "bottleneck," or most severe difficulty, for adult learners of a language to traverse. Finkel and Stump (2009: 139–140), seeking a definition of inflectional complexity, posit that "the complexity of an inflectional system is the extent to which it inhibits motivated inferences about the word forms realizing a paradigm's cells." In other words, it is the factor that prevents people who do not know a language well from being able to deduce or guess the right form of a word by analogy or other deductions.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

as transitivity and intransitivity, or a relationship of propriety between the subject and object), and four moods (indicative, optative, subjunctive, imperative). The ancient Greek verb was relatively paradigmatically opaque, with six and sometimes as many as eight frequently unpredictable (irregular) verbal stems to be learned individually for each verb, known as the "principal parts," conveying three different temporal aspects among them.[12] In addition, there is a smaller group of frequently used verbs, called athematic, that require a quite different set of inflectional endings, to be learned and applied only to that class of verbs, with accompanying inflection of the verb stem itself. Moreover – still part of verbal morphology – there are adverbial prefixes or *preverbs* that diversify a verb stem's meanings in often unpredictable ways. This furthers opaque lexicalization, as the meaning of the English word *understand* cannot be deduced from the components *under* and *stand*. Then there are participles formed from every verbal stem, each for a different verbal aspect and diathesis, not to mention other sorts of inflection omitted here.

By contrast, ancient Aramaic had verbs with two verb stems each (one perfective or punctual, one imperfect or potential), many of which have regularly formed factitive and causative variations as well as a regular mediopassive variation. There are also two participles (verbal adjectives, active and mediopassive) that do much verbal work. Almost all verbal roots follow one type (three root consonants), and verb stems are largely predictable from their roots. Aramaic verbs are inflected to agree with subjects according to three persons and, in the second and third persons, two genders. When one of the three radical consonants in a verbal root is a semivowel or another of a few evanescent consonants, some further contractions of syllables occur that are almost entirely regular. The Aramaic verb is not sheer simplicity, but by the measure of verbal morphology alone – to say nothing of the complex noun inflection of Greek and the extremely minimal noun inflection of Aramaic, with two genders and no nominal cases – the grammar of the verb used by a Greek Athenian in the fourth century BCE is exceedingly more complex than that used by their neighbors in contemporary Assyria, then under Persian rule. This is self-evident from the discrepancy in the time it takes to teach the grammar of these two different languages to post-adolescent learners today, as I have done for both in university settings. Why should one natural human language be so much more complex in its morphology than another?

[12] For the concept of paradigmatic transparency, see Finkel and Stump 2009.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

That is a question addressed in recent scholarship, and some of the available answers are summarized and applied here.

Despite the obvious disparity in the inflectional complexity of the grammars of languages, it was for much of the twentieth century, and still is for many today, a widespread assumption that all languages are, rather, equally complex. This has been dubbed the "equal complexity hypothesis."[13] Some linguists and nonspecialists have assumed that morphological simplicity is made up for by higher complexity in another area, such as syntax or phonology. All languages are assumed to be equally complex in an unidentified fixed sum of unmeasured factors between the subsystems of a language.[14] For example, a language with simple verbal morphology should be more complex in its syntax or in some other system of language, to compensate somehow for the unadorned morphology. It is hard to understand why linguists would have assumed this without an attempt at measuring these factors, especially when it is contrary to common experience.[15] Perhaps it arises from a reaction against antiquated moralizing and implicitly racial teleological evolutionary views of linguistic philosophy at the dawn of linguistics. Wilhelm von Humboldt (1767–1835), for example, a leading contributor to the formation of linguistics, held that languages evolved from a primitive isolating type to a more advanced agglutinative type and finally a developed, higher, synthetic inflectional type possible only in a nation of greater "mental power," leading to inspired literature.[16] Perhaps it is due to the more recent tendency of major branches of linguistics to focus on the search for universal grammar and the cognitive underpinnings of all language, worthy objects of attention that happen to distract linguists from what makes languages different in their quest to find what is common to all of them.[17] One can only assume that those who hold a priori that languages are all equally

[13] Szmrecsanyi and Kortmann 2012: 6–7; Oczlon-Fenk and Fenk 2014. See also Trudgill 2012, who calls it the "equicomplexity hypothesis."

[14] References in McWhorter 2001: 127. For a notion of linguists' sensitivity to the issue, see the 221 pages of response by a dozen other linguists, mostly objecting to the argument summarized in McWhorter's title ("The World's Simplest Grammars are Creole Grammars"), in issue 5.2–3 of *Linguistic Typology* (2001) that, in the words of its editor, Frans Plank, "got out of hand."

[15] This has changed in recent years. The more recent debates are about what kinds of phenomena should be measured as complexity in morphology, how they should be measured, how complexity is "stored" mentally, and still other issues. See especially Baerman, Brown, and Corbett 2015, out of the references to step (1).

[16] See Morpurgo Davies 1998: 111–114. The point is explicitly clear in Humboldt's posthumous publication of 1836 (see Humboldt 1999: 88–89).

[17] Kusters 2003: 3–4.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

complex have not pursued working proficiency in or taught different languages of different kinds as adults or studied current research on second-language acquisition, focusing instead on the application of general theories of linguistics to illustrative examples in the quest for universals.[18] By contrast, recent efforts in the linguistic typology have directed renewed attention toward differences among languages. Many studies have definitively demonstrated what adult learners of multiple languages have always known: that some languages really are more complex than others, and they are so according to different criteria of complexity. Complexity is not a given quantity, allocated as a fixed sum to each language, distributed to its different systems in shares having somehow the same total in every language. Some languages have far more grammatical forms to learn and to apply than others, and measurable compensation for simplicity in one area is not necessarily found with complexity in others.[19] Most importantly here, it has been amply demonstrated through studies of second-language acquisition that complexity in some modules or features of a language is more difficult for learners than others. The single biggest challenge for learners is contextual inflectional morphology: the inflection determined by syntactic relations. Purely morphological or inherent inflection, such as the use of different noun classes or declensions to which nouns are assigned without playing a role in syntax or semantics, is also difficult.[20]

[18] Trudgill (2011: 16) even writes, I think half-jokingly, about the doctrine of equicomplexity of languages that, "obviously this was a propaganda ploy that was vital for combating the 'some languages/dialects are primitive/inadequate' view that has been widespread in our society." This would suggest that some linguists knew that the idea was false but adopted it publicly to avoid providing inadvertent support to the notion that some people were inferior by virtue of the minimal inflectional morphology of their language.

[19] Kusters 2003: 9–12; Shosted 2006.

[20] DeKeyser 2005: 5–6: "Difficulty of language form is largely an issue of complexity. . . . [The] problem of L2 users' failing to use morphology, even in comprehension, is so fundamental that it has by itself spawned entire bodies of literature." On the concept of contextual inflection, see Booij 1993 and 1996 and Marzi, Blevins, Booij, and Pirelli 2020: 229–230. Compare Myers-Scotton's "outsider system morphemes" (2006: 269–270), essentially the same concept with different jargon. To explain why these are resistant to transfer from one language to another, she writes (2008: 38), "these morphemes are seen as making the primary arguments in any clause transparent. This insight about their importance in the structuring of clauses helps explain why morphemes of this type are unlikely to be open to transfer or copying in contact situations. They are tied into an agreement system that structures the clause such that undoing the knot is difficult at best." Notably, Myers-Scotton, coming from the subfield of bilingualism studies, with a focus on defining constraints in code-switching, arrived at these results using her own names for the phenomena, and she did so without any reference to the terms "contextual inflection" or the publications of Van Coetsem on language contact, although these other lines of research describe the very same phenomena.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

I have learned in talking with others about this research that it is necessary to address this further, to dispel some common misconceptions. The existence of variation in languages' complexity comes as an unwanted surprise to some because of the long-lived, well-meaning effort to dispel the intuitive, popular, but erroneous connection people draw between complexity of linguistic forms with sophistication of thought. Saying one group speaks a language less complex than another's is sometimes assumed to imply that they are also less complex in their minds or, in a word, stupider. This other, more repugnant, mistake has indeed been used at times to buttress racism and bigotry or beliefs in national linguistic superiority, promoting the false notion that speakers of morphologically simpler languages are simple-minded or unsophisticated, or even that they deserve to be ruled by sophisticated colonizing nations whose languages are complex in a way corresponding to their manner of dominating the world. It is not hard to find examples, particularly from generations long gone by, such as this one from a French military officer's essay on Guinea-Bissau Creole Portuguese, published in 1849: "It is clear that people used to expressing themselves with a rather simple language cannot easily elevate their intelligence to the genius of a European language."[21] Or take the example of William Churchill in 1911, who supposed quite mistakenly of certain South Pacific peoples that "the islanders do not know how to think comparatively" because, in their secondary pidgin English language, comparative forms of adjectives did not exist.[22] It is in the face of these bold errors, rarely if ever offered by serious scholars in the last few generations, that some sensitive observers sidestep the issue by assuming that all languages are not just evaluatively equal as natural phenomena, but *equally complex* – just as the generality of people share the same mental capacities in potential. This is, however, arguing against a repugnant fallacy with a more appealing fallacy. The correct counterargument to such bigotry is that an individual's intelligence (whatever that should mean) has no correlation with the inflectional complexity of that individual's language or languages. As William Greenfield, not a soldier but an aspiring linguist, noted already in 1830 about another creole language, "The human mind is

[21] Bertrand-Bocandé 1849: 73: "On conçoit que des hommes accoutumés à se servir, pour manifester leur pensée, d'un idiome aussi simple, ne purent facilement élever leur intelligence au genie d'une langue européenne." Translation from Holm 2000a: 23. Hock (2021: 728) easily rebuts such views.
[22] Churchill 1911: 27.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

the same in every clime; and, accordingly, we find nearly the same process adopted in the formation of language in every country."[23]

The proposition that all languages are, one way or another, equally complex has, in any case, failed against empirical fact, as some of the studies cited in note 7 in this chapter demonstrate, and as experience in learning and teaching different languages proves. To make this less distasteful, it is useful to recall that intelligence and the complexity of the language one uses are not in fact correlated. Despite the classical scholars who have been heard to boast of the sophisticated complexity of ancient Greek, stating that its complex morphology made it more suitable for philosophy and high civilization, languages with simple inflectional morphology are not more "primitive" and their speakers are not more stupid and less intellectually gifted.[24] Russian inflectional morphology, for example, is clearly more complex than that of Mandarin Chinese or American English or modern Persian, but that does not make it better or worse, or more or less sophisticated, than the others, nor does it make Russians more intelligent and discerning than Chinese Mandarin-speakers or Anglophone Americans or Iranian Persian-speakers.

In one respect, the argument of this book does not require accepting that some languages are simpler than others, even though many Iranic language specialists, cited in the previous chapter, in describing Middle Persian morphology as greatly "reduced" or "simplified," are simply stating the obvious. Perhaps someone will insist, instead, that Middle Persian is just as complex grammatically as Old Persian in some invisible, unmeasurable, counterintuitive way, despite all the estimations of the specialists cited in the previous chapter who did not think twice about asserting that Middle Persian was an exceedingly simple language in its morphology, especially as compared with Old Persian. Such a person will still accept that the two earlier recognized stages of Persian, Old and Middle, are fundamentally

[23] Greenfield 1830: 51; Holm 2000a: 22.

[24] Trudgill puts it well (2011: xvi): "However, after many decades of academic linguists agreeing, and asserting, that there is very definitely no such thing anywhere as a primitive language, it now seems safe to consider this issue in print without this suspicion arising." Ringe and Eska (2013: 55) make a similar point, that "nineteenth-century notions of linguistic 'superiority' have proved to be based on nothing more than Eurocentric bigotry, and observation readily shows that there is no correlation between the structure or typology of a language and the political fortunes or material or mental culture of its speakers." (I would qualify the latter, for indeed the political and material fortunes of the speakers of a language may create circumstances of rapid linguistic change, as I shall explain in the case of Persian.) McWhorter (2013a: 177) reverses the issue, calling into question the positive evaluation of linguistic complexity: "Opaque derivation . . . is cherished by no one as a measure of expressiveness or nuance, and thus a claim of its absence in a language cannot be interpreted as implying its inferiority." Cf. McWhorter 2007: 272–274.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

different, and that Old Persian changed radically before the end of the Achaemenian period, and that inflectional variety was greatly reduced. That striking difference, noted by all specialists, lacks an explanation. Nevertheless, complexity of inflection does matter. This kind of complexity is the critical issue and ought to be recognized because degrees of such complexity can tell us about the history of these languages and the histories of their speakers. It is a feature of diachronic language variation that can be used to identify historical demographic changes, as I will explain now.

(2) Children and adults learn new languages differently. Children naturally learn the grammars of languages to which they are regularly and continually exposed as they receive them, in all their complexity. Adults, by contrast, generally lose their capacity to learn new languages well (in all their inflectional complexity, without an accent, fluently). That is, adult learners of a new language cannot be native speakers. Instead, they usually acquire new languages imperfectly, from the point of view of native speakers, often introducing features from their first language (imposition) and regularizing unpredictable variation in the newly acquired language. They tend to simplify the grammar of the new languages they learn to speak, especially by reducing inflectional morphology.[25]

(3) Mass adult language acquisition produces a grammatically simplified common vernacular. When large enough proportions of a population speaking a language are nonnative adult learners, such as large groups of immigrants or colonists, then the language they have just learned will be restructured because the simplification that they give to it, in whatever degree, becomes normal and regular to the extent that their restructured variety of speech becomes pervasive. Thus, a new variety of an old language, or even an entirely new language, emerges rapidly.[26]

[25] Klein and Perdue 1997; Winford 2003: 222–223; Dahl 2004: 293–295; Birdsong 2006; Myers-Scotton 2006: 323–368; Labov 2007: 349; Slabakova 2008; Matras 2009: 68–86; Trudgill 2010a: 184–192; Trudgill 2010b: 310–313; Trudgill 2011: 33–40; Adone 2012; Ringe and Eska 2013: 32–35; Morgan 2014; Hartshorne, Tenenbaum, and Pinker 2018. The mechanisms of grammatical reduction are addressed later in this chapter. The difference in outcome between first-language acquisition by children and nonnative acquisition by adults is acknowledged generally, but scholarly debate about the phenomenon typically focuses on the search for the reasons behind it and the variation in effects of age on the learning of different modules of language. Those factors are not the object of this study.

[26] Milroy and Milroy 1985; Field 2004; Holm 2004; Wray and Grace 2007; Lupyan and Dale 2010; Trudgill 2011; Bentz and Winter 2013; Ringe and Eska 2013: 45–58.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

(4) Children acquire newly simplified common vernaculars as native languages. When a recently simplified form of language is pervasive in the environment of children, those children will acquire the restructured, inflectionally simplified form of the language natively, and the changes introduced by a previous generation of nonnative speakers then become normal and native. The subsequent generation will learn a restructured, simpler, pervasive variety of speech in the same way as their native, normal language.

These three steps should be explained together even as they are recognized as separate components. The first (2) is a biological factor in language change; the second (3) is social and demographic; the third (4) describes a result of their combination. They constitute an area of widespread new agreement among linguists who study languages affected by population contact. Taken together in this way, they comprise the most important contribution of the field of contact linguistics to the discipline of history. The fifth step will be explained in the next section.

Three Intergenerational Patterns of Language Acquisition

Steps (2), (3), and (4) are most easily understood when we consider patterns of language acquisition between generations. Linguists with these concerns have shown that changes in language structure and complexity correlate with patterns of language acquisition and social relationships.[27] This realization, still relatively new, will have major ramifications for historical linguistics and philology as well as social history.[28] It is not that "language structure mirrors social structure," as some twentieth-century structuralists maintained. That vague and ill-defined attempt at a correlation was mostly abandoned long ago.[29] The relevant pattern is specifically based in patterns of language acquisition over more than one generation. For the sake of

[27] Thurston 1987; Trudgill 2009. Dahl 2004: 296: "this evolution [. . .] is by and large independent of the processes that are usually meant when speaking of 'cultural evolution,' and may even be negatively correlated with the rise of large-scale societies with highly mobile populations. Thus, there is evolution in language, but not in the way that 19th century scholars thought."

[28] Cf. Yakubovich 2008: 210–211: "The study of language contact represents a vigorous branch of modern linguistics, which has received an enormous boost in the last fifty years. The scholars of ancient societies wishing to decode their ethnic history must avail themselves of the achievements in this field." Ringe and Eska (2013) put nonnative acquisition and multilingualism at the fore of their recent introduction to historical linguistics.

[29] For example, Benveniste rightly pointed out that the Russian revolution transformed Russian society, but the Russian language remained largely the same grammatically. It was the lexicon that changed to respond to new needs and politics. See Laplantine 2023, studying and translating a presentation of Benveniste to this effect from 1970. When Lupyan and Dale (2010), however,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

discussion, we can simplify these patterns into three broad types, necessarily ideal types found nowhere purely and exactly as described.

A typical pattern in language acquisition belongs to a mostly monolingual community. I propose to call this *intergenerationally stable* transmission. The morphology of a language in such a community remains mostly stable and changes relatively little when learned normally only by children of native speakers within a closed community. By closed, I mean that they accept few outsiders into their midst for any reason. It could be physical isolation of the linguistic community or the deliberate social exclusion of others. It could be that outsiders have little interest or incentive to join their linguistic community. In any case, children do introduce new, little changes into languages as they learn them, even as native speakers, but they are seldom drastic systemic changes. Older speakers correct children's mistakes. Children become fluent, native speakers of the language before attaining adulthood and most grammatical quirks of early childhood usage, such as leveling of grammatical irregularities, are eliminated as they interact with older users of the language. Unless their language acquisition is interrupted, their first-language acquisition never requires the leveling of complex inherited paradigms. Rather, the children in such a community learn the language of those who care for them, often their parents, and, typically in history, especially their mothers, with a high degree of exact replication, including any degree of linguistic complexity to which they are regularly exposed. The grammars, lexica, and other major features of such languages, used by closed communities, therefore change little over time. When they do change, intergenerationally stable languages normally retain complex features and even grow in complexity slowly over many generations.[30] This sort of language best characterizes populations that have been joined by very few new outsiders over time, populations that are insular, isolated socially or geographically.[31] This pattern seems to be the one historians often assume casually about their historical subjects: language and community are enduringly linked across generations; distinct communities defined by their distinct languages persist without a change in that link for centuries, so one can talk about "Arabs" or "Hungarians" in

declare in an important article that "language structure is partly determined by social structure," they do not define "social structure," nor do they address the long history of discussion of the topic of "social structure." That said, their clear finding that large-scale adult learning of a language correlates with reduced morphological complexity is valid and contributes strongly to the model presented here.

[30] Some communities may deliberately differentiate their languages from those around them to exclude outsiders. See Thurston 1987 and Good 2023.

[31] Milroy and Milroy 1985; Trudgill 2011; Milroy and Llamas 2013.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

the fourteenth century and in the twentieth century as if these communities were intrinsically or essentially the same over seven hundred years; a language correlates with and defines its people; intergenerational changes affecting such peoples are primarily in their external circumstances, not internal and constitutional.

A different situation with a similar outcome sometimes obtains when a population includes many speakers of more than one language over many generations, possibly an entirely bilingual or multilingual community. In this second, *intergenerationally multilingual* transmission pattern, in the long term, the multilingual members of the community will transfer features of one language to another with ease as they all speak the languages in question natively or nearly so. The acceptance and durability of features transferred between languages depends on many factors, not just the kind of linguistic feature in question but also others such as the proportions of multilinguals and the distribution of languages among them, factors pertaining to generational differences, social status and social power, domains of language use, and more. These scenarios can, however, have the effect of adding some kinds of complexity to a language, but such an outcome is not necessary.[32] The expected effect of very long-term, multigenerational transfer of features between the same languages in an area of bilingualism or multilingualism is not the drastic reduction of grammatical forms but the production of a *linguistic area* or *convergence area*, also referred to in English by the German term *Sprachbund*, in which languages of separate origins converge in the patterns of their grammatical phenomena, so that they bear many similarities to one another that their ancestor languages did not exhibit in common.[33] These similarities are called areal features. Because there are different ways in which features are transferred from one language to another (discussed later in this chapter), the linguistic area is an umbrella term for several different phenomena with similar outcomes of linguistic convergence. The assumption, however, is that the people in such a society speak more than one language adequately to communicate, usually because they learned their languages at a young age.

[32] Trudgill 2010b: 309–310; Parkvall, Bakker, and McWhorter 2018: 231.

[33] Hock 1991: 494–515; Winford 2003: 70–74; Hock 2021: 654–724. Ciancaglini (2011: 10) is right that "the term [*Sprachbund*] seems to denote only approximately a whole group of contact phenomena for which a satisfactory historico-linguistic explanation has not been found," but there is reason to be optimistic about the future of areal linguistic studies, as shown by Muysken (2008) and Heine and Kuteva (2005: 172–218). These studies offer useful definitions of linguistic areas and distinctions between different kinds of them. I discuss linguistic areas more in Chapter 4.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

By contrast with these much more frequent circumstances of language use in societies, there is a third and less common situation of *mass nonnative acquisition*. This is when many adults adopt a new language at once. Languages learned by large masses of adult nonnative speakers in a short period tend to become grammatically reduced, particularly in their inflectional morphology. This is because, unlike children, who acquire languages with the fluency of their adult models, adult learners acquire new languages imperfectly with respect to what native speakers hitherto have spoken – with a lexicon more limited than that of natives, with nonnative pronunciation, and with limited command of morphological complexity – as compared with the native usage of the language until that time.[34] They have passed the sensitive period of childhood and early adolescence during which new languages can be learned with native proficiency.[35] The characterization of a nonnative variety as "imperfect" is not a moral evaluation.[36] Imperfect language acquisition by adult learners is normal in historical circumstances of fresh contact between populations with different languages, in which adults hitherto not in contact need to interact. In such conditions, most adult language learners in history did not have access to formal training with carefully designed exercises like those offered today in modern university courses on languages. Most people did not learn languages as educated and literate people tend to learn them today. They did not grow up in a society in which the state instituted mandatory mass schooling in multiple literary languages, as is the case in many countries today. Instead, adult learners in premodern times normally had to pick languages up through casual social exposure, informal instruction, and attentive use with trial and error. They learned languages usually only by necessity or through the ambition to gain a specific advantage. In any case, they usually did so by interacting with other speakers of the language. Under these circumstances, the age of acquisition was especially important. Adult nonnative speakers of a language normally omit its complex features in practice, because those features are cognitively burdensome or they simply have not learned to apply them, and they make do without them, aiming at a functional-level communication recognized by native speakers as foreign.

[34] I have heard claims about individuals who attain perfect fluency in a new language as an adult, with no discernible accent and indistinguishable from native speakers. These would be rare exceptions. If they were not, then we would not have a category of nonnative speaker.

[35] Trudgill 2010b: 313: "Simplification in language contact does not result from nonnative language learning as such, but from post-critical threshold nonnative language learning."

[36] Van Coetsem 2000: 82: "The product of second language acquisition . . . is a language in its own right."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Native speakers who need to interact with adult learners of their own language must tolerate the omission of such complex features. They may even simplify their own speech, too, as an expedient for communication with others who do not know their language adequately. Simplification of one's own language is a normal strategy employed by native speakers to facilitate effective communication in their own language and on their own terms – literally – with nonnative speakers. Nonnative speakers learn to speak and understand simplified forms of speech more easily. A form of language with reduced features, such as a simplified range of inflected noun or verb forms, can become routine and normal in a multilingual social setting in which functional communication is the immediate necessity. When the abbreviated, new form of language acquired by adult learners is passed on to younger generations, as, for example, when a large proportion of a population consists of nonnative speakers, all of whom do without the native speakers' inflectional complexities, then children in their midst acquire simplified, nonnative speech as normal and native. Extensive and rapid mass adult acquisition is probably the *only* major driving force behind extensive morphological simplification and restructuring of any language.[37]

It is through new population contact that nonnative speakers – people who learn a language as an adult – come into being. Individual cases are normal, but scale is an issue. A few strangers from a foreign land will not change the grammar of the local language as it is generally used. But when adult language acquisition occurs on a very large scale, as with, for example, migration of numerous people into a social group new to them, suddenly many of the speakers of one language will become nonnative learners of another, and they may often communicate with each other in their shared nonnative variety. This is especially so when the newcomers originated in communities that do not have another common language. Their knowledge of their new language will usually be of the simplest kind necessary to facilitate daily life. That large proportion of speakers will use a new version

[37] Trudgill 2011, esp. 62–115; McWhorter 2011a: 2–3. Scattered objections to this discovery have not been adequate. For example, Campbell and Poser (2008: 357–362) discount Trudgill's arguments along these lines largely based on a single case study of Hawai'ian phonology. Their argument does not even mention the historical dimension; synchronic description of a language's complexity alone cannot by itself disprove the thesis about past mass nonnative acquisition leading to simplification of grammar. Trudgill himself addresses these criticisms effectively (2011: 55–60). Moreover, the choice of phonology as the criterion was probably fruitless because complexity of phonology does not appear to be a salient issue in cases such as these (Bakker 2009; Bakker and Daval-Markussen 2017: 79–82). It is possible that phonemic inventories are rather correlated with population size (Hay and Bauer 2007).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

of the language, created by the imposition of features of their prior language or languages, giving rise to new pronunciations (foreign accents), abbreviated or neatly analogized paradigms, and regularized or streamlined versions of hitherto irregular grammatical rules and forms. They will misapply or omit word inflection and simplify complexities of the language of the earlier native speakers and they may impose features of their own native languages on their use of the new one. These adult learners will often never have intended to learn the new language fluently, as their goal was not to become a full member of the group of native speakers. In many cases they will not expect ever to be regarded as a full member of the group with which they have come into contact.

The crucial step is that children exposed continually to nonnative speakers will learn from them a restructured form of the language as their own native speech and, in the next generation, they will pass that form of the language on to their own children. In the event of further migrations and contact with still more populations using different languages, newcomers to this linguistic arena will find a simpler variety of the common language available to them. They will more easily learn that variety, if they have a choice and they find it adequate for their aims. It is an easier target with a useful communicative outcome. Within three generations, it is possible for the language in such circumstances to have changed drastically in the direction of inflectional simplification by these processes, to the point that it is mutually incomprehensible with the prior, inflectionally more complex variety of the language. That is, from the old language a new language has arisen on the tongues of newcomers. When children take that new language as their own, it is sustained as a native language to be passed on to the next generation. It is no longer just a strong foreign accent. The morphologically reduced form of the language is now a native tongue.

This kind of language change is not hard to grasp conceptually from ordinary experience. If parents move with their young children to a new locale, where native speakers of the same language speak with a different accent, their children will now learn that accent natively to the extent that they are not confined solely to interaction with their parents and other persons of the same origin. They learn the local accent from those around them. The same phenomenon occurs if the "accent" is, rather, the simplified pronunciation and grammar of foreigners who have moved in around that family and learned the local language nonnatively. The basic process is, therefore, not mysterious, but is easily observable in the world today wherever people are mobile.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

In summary, different social conditions tend to give rise to different kinds of language change. For the purpose of illustration, I distinguished three common ideal patterns relevant to this study. Population changes, migration, and demographic upheavals easily interrupt the typically widespread pattern of intergenerational transmission of unchanged grammar from parents to children. They can transform a language because their intergenerational relationships to each other and to their languages shift. As a consequence, languages acquire lastingly different grammatical characteristics that are determined by social events. Large-scale new population contact between adults in effect strips inflectional morphology from the intercommunal language, especially contextual inflection. Drastic grammatical changes induced by those events are permanent in a language within that population, although any language so affected will undergo further changes by all kinds of processes of language change thereafter. Populations will remain linguistically conservative when characterized by (1) closely knit, multiplex social relationships, in which most members know every other member, and (2) little contact, due to any factor, with outsiders attempting to learn their language. Their languages are likewise, in the long run, the most inflectionally complex. At least they retain inherited complex features such as ornate inflectional morphology.[38] The close-knit, closed population of its users seems slowly and gradually to develop increasingly elaborate forms of speech, marking membership in the closed community. In the jargon of linguistics this is "complexification."[39] Different, but not entirely so, are the common cases of populations having numerous contacts with other groups who speak different languages natively. When children in such situations learn more than one language from youth, they become multilingual persons capable of native-like use in each of those languages. Multilinguals are responsible for the transfer of features between these languages, which may foster simplification, new complexities, or both, depending on many

[38] The pattern of two tendencies was intuited already by Meillet (1912b: 413): "L'extension d'une langue à une grand nombre d'individus nouveaux en banalise les formes; les langues dont l'aspect est le plus singulier sont celles des plus petits groupes sociaux."

[39] McWhorter and Trudgill both suggest that languages used in such social circumstances slowly become more complex over time, free from the social factors that induce reduction. See Trudgill (2010b: 314). A terminological note: the word complexification is ill-formed. It should be complicification, but this neologism is much less likely to be transparent for English-speakers. Labov (2007: 346) calls it "incrementation." Trudgill (2011: 185) claims that, "Linguistic complexity developed in societies of intimates," and that complexification will occur gradually in social groups with "small size, dense social networks, large amounts of shared information, high stability, and low contact" with outsiders.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

factors. They may cause their languages to converge, generating areal features. The third and more distinct pattern occurs when a large population of adults must learn a new language in common at once. Such a language will undergo a potentially drastic degree of morphological reduction. As Wray and Grace put it counterintuitively, languages that many nonnative speakers need to use "will come under pressure to become more learnable by the adult mind" through the elimination of complexity for the sake of the new learners who need to communicate.[40] This is an important pattern of language change arising from the combination of specific social and biological factors: scale of contact and age of acquisition.

Now that I have explained steps (2), (3), and (4), I turn to the fifth step of this part of the linguistic-historical model. It is a caveat.

(5) Languages may be morphologically reduced by this process in *varying degrees*. Reduction of a language's morphology and grammatical restructuring by these intergenerational population changes is not binary: either simplified or not-simplified. Rather, the resulting language may exhibit relative degrees of restructuring and grammatical reduction that vary from instance to instance according to specific historical social factors during the time of large-scale nonnative adult acquisition. Linguists have proposed a few taxonomies of degrees of restructuring in language, but this is work in progress, so we have at our disposal only a limited number of minimally analytical and therefore tentative terms to designate the outcomes of such restructuring. This inhibits our ability to be precise about the character of restructuring in each new historical stage of the evolution of Persian, the subject of this book, but it does not stop us from recognizing the character of partial restructuring that indubitably did occur in a language such as ancient Persian.[41]

To understand this crucial, final step in the model, one must understand the thorny problem of contact languages and especially the *creole*.[42] This will enable one to relate the drastic changes occurring in the history of Persian and other languages to clearly established linguistic categories and living examples of these varieties of language, rather than treating the

[40] Wray and Grace 2007: 557.

[41] Holm 2000b; Neumann-Holzschuh and Schneider 2000; Winford 2000; Winford 2003: 300–301, 254–256; Holm 2004; McWhorter 2005: 19–34; McWhorter 2007: 17–18, 252–276; Holm 2010: 258–259; McWhorter 2014.

[42] Holm 2000a offers a responsible introduction to creoles.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Persian language as a sui generis phenomenon, unlike any other language, following natural laws peculiar to it alone.

"Contact Languages" as Comparanda

Some of the discoveries outlined so far arose from the study of the most striking recent historical examples of languages restructured in the wake of intensive, colonial, population contact. These are languages called pidgins and creoles, inflectionally simplified varieties of other languages used in early modern colonial circumstances, which have become entirely new languages in themselves. To varying degrees such languages can seem to be hybrids through the transfer of features from other languages (language mixing through the transfer of features from one to the other being an issue discussed later in this chapter). Creoles arise by the same natural, social, intergenerational processes just outlined. Because most of the recognized living pidgins and creoles came into existence during the recent age of European maritime empires and colonial domination, and not in the ancient past, their genesis is in several cases well documented. Innumerable publications present research on these drastically restructured varieties of Dutch, English, French, Portuguese, Arabic, and other languages. The term creole comes from Spanish and Portuguese, where it referred to persons born in American colonies created by speakers of those languages. Meaning "created, made, begotten," at first it referred to persons of mixed ancestry in these colonies. As new varieties of colonial languages came to be used by these people, the languages soon also acquired the designation "creole," already by the end of the seventeenth century.[43] From these specific languages named Creole, linguists have generalized a category of language, "creole," which has been applied to similarly simplified varieties of languages with analogous origins. It is important to distinguish creole cultures – the cultures of people called creole – from creole languages, which result from the specific processes under discussion. The term "creolization" thus refers to two different things; this present concern is purely the linguistic one.

The focus on creoles in the modern study of language restructuring through grammatical reduction has had, however, two unintended deleterious effects on related scholarship, each of which requires discussion. First, by basing much of their work on creoles – which do provide conspicuous examples of the restructuring or recasting of languages in distinctive,

[43] Holm 2000a: 16; Winford 2003: 305–306.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

morphologically simpler forms, and of interruptions in the typical intergenerational acquisition of language through social contact – specialists in contact linguistics became mired in ideologically charged debates about the historically particular category of creole itself rather than the phenomena of socially induced language change.[44] It is an emotionally charged debate because the best-known creoles today are radically restructured and morphologically simplified versions of the languages used by European colonial powers. These languages arose as byproducts of social contact often characterized by violence, forced migration, slavery, and oppression, from which descendants of the inhabitants of the colonizing countries continue to benefit, but leaving conditions that have disadvantaged the living descendants of the victims for generations even as they effectively gave rise to new ethnic groups. Humanistic scholars, whose work requires the promotion of cultural understanding, are often interested in defending the dignity of colonized and enslaved peoples and their descendants, and in repelling the disdain and racial bigotry applied unfairly to ethnic groups who had their genesis amid generations of exploitation. In so doing, some have regarded the designation of these colonized peoples' languages, creoles, as the simplest of the world's languages, as adding insult to injury: not only are creole-speakers sometimes treated in appalling, racist ways, but now their languages are deemed "simple"! Because of this, some sensitive defenders of colonized and decolonized peoples have refused to acknowledge that creoles are morphologically simpler, and easier for adults to learn, than their source languages. Naïvely adopting the prejudices of colonizers of bygone times, who disdained simple languages as uncouth or intellectually less capable, they allege that to "exclude" creoles from the rest of languages as belonging to a category of exceptional grammatical simplicity would perpetuate an oppressive colonial discourse and taint this branch of linguistics. This criticism addresses the so-called creole exceptionalism. In so arguing, they fall into the fallacy described under step (1), by assuming that there is any insult at all in declaring that one language is inflectionally simpler than another. They could just as well celebrate the adaptive power of humanity in generating expressive, beautiful languages easily acquired for communication in difficult circumstances, but they do not. Some of them even conflate the views of early colonial scholars of centuries past, who explicitly despised their creole-speaking subjects, with the views clearly expressed by linguists today, whose interest in and fascination with creole-speaking peoples and

[44] See, for example, the objections of Mufwene (1997), DeGraff (2003), and some of the studies collected in Ansaldo, Matthews, and Lim 2007.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

their languages, as human languages, is hard to describe as colonial, and who explicitly state their belief that there are no "primitive" languages.[45] For those interested in radical language restructuring, and in understanding general linguistic phenomena, as these special cases make them especially clear, the debate makes for a messy and deterring obstacle. It distracts from the actual gains of contact linguistics and obscures the arguments in ongoing debates about the specific processes involved in the genesis of creoles.[46] It makes language use a matter of identity for humanistic scholars, who almost universally ignore the critical issue of age of language acquisition, instead regarding a sense of identity as the leading factor in choosing language use.[47]

The second deleterious side-effect of creole studies on this larger problem of linguistics is that the copious attention to creoles created the implicit impression of a binarism separating creoles and noncreoles, "simplified" versus "normal" languages. The languages called creole today are the most outstanding contemporary examples of "contact languages" in that they are inflectionally radically simpler, restructured varieties of languages used by imperial and colonial powers. Meanwhile, their source languages, such as English and French, are still today the primary media used in instruction in European and American universities and indeed by scholars around the world. The distinction between modern creoles and their widely known source languages is therefore especially conspicuous to scholars today, furthering the impression of scholars that a language is

[45] These debates leave interested bystanders like Migge (2019: 862, reviewing McWhorter 2018) wondering about the "absence of hospitality" among those most engaged in the debate. DeGraff, for example, blurs the distinction between folk attitudes toward creole languages with current linguists' stated beliefs about how creoles are exceptional when he claims that linguists "reinforce the belief that Creoles are structurally *inadequate*" (2003: 392; emphasis added). I have not seen any serious linguistic literature from recent times making that claim. Of course, any statement about creoles being "structurally inadequate" would require a demonstration rather than a mere assertion and should be regarded with extreme skepticism. Bakker (2017a: 27) states that "all linguists who have studied creoles agree that creoles are natural languages, with the full range of expressability of the other natural languages of the world." Cf. McWhorter 2007: 272–273; 2018. If creoles are exceptional, then it is not as exceptions to the class of human languages, but as instances of language more exceptionally well characterized by nonnative-learner-induced restructuring. They are exceptional in the sense that the Abkhaz language, for example, is exceptional in respect of its extraordinarily large consonant inventory as compared with those of most other languages.

[46] A less ideologically charged objection to the notion that creoles harbor less complexity has come in the form of studies of creole phonologies. These have demonstrated that such languages tend to have phoneme inventories in the average range of the world's languages. (See, for example, the studies comprised in the first half of Faraclas and Klein 2009.) While these findings are important, they do not address the salient issue of inflectional morphology.

[47] Trudgill 2010a: 184–192, esp. 189: "There is a copious literature . . . which suggests that linguistic accommodation is not driven by social factors such as identity at all but is an automatic consequence of human interaction."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

either creole or not creole, either like or unlike the languages with which they are personally most familiar.[48] Since the nineteenth century, however, and especially in the last thirty years, research has recognized the existence of languages between the status of "normal" change – which is not named by a technical term – and the "creole."[49] These are *partially* restructured languages that vary in degree of inflectional simplification. Only recently has this aspect of the problem come to the fore in concerted discussions, and scholars have begun to search for ways to identify, describe, and even measure *degrees* of simplification or reduction in complexity.[50] Out of necessity, not preference, this book participates in that effort.

Before describing partially restructured and inflectionally reduced languages, in connection with step (5), and explaining what such changes mean or entail, however, it will help to explain more clearly what a creole is supposed to be in the first place. Understanding the linguistic history of ancient Persian requires this orientation, if we are to analyze Achaemenian society correctly in ways that relate the linguistic history of Persian to other documented natural phenomena in language change. This is especially so when recent studies by specialists in Persian have suggested that Persian – particularly New Persian – may have evolved as something like a creole, but they have done so with hardly any investigation into research on creoles.[51] It is necessary to understand the terms one uses.

[48] Humanists, upon learning about creoles, have sometimes become excited to suggest that the language of their academic specialty is a creole, often without understanding the category adequately. These assertions then draw counterarguments easily demonstrating that it is not a creole in the normal sense of the term. These debates would not play out so inconclusively if participants recognized that there are degrees of restructuring. See, for example, the debate on Middle English as a candidate for creole, summarized dismissively by Watts (2011: 83–113). In my view, Middle English is clearly not a creole, but it has also clearly undergone morphological simplification relative to Old English that absolutely necessitates that it be understood through the historical-linguistic model presented here. Scholars of the history of English must explain the history of the demography of English-speakers better to arrive at a convincing solution. Cf. McWhorter 2007: 59–103 for a linguistic analysis of this problem.

[49] Earlier, I proposed to call such "normal" languages "intergenerationally stable" so that we have a name for this phenomenon rather than regarding it as a default condition.

[50] Holm (2004) and McWhorter (2007: 252–265) present the most determined efforts in this direction.

[51] Utas (2006: 246) argues that it is "tempting to see [New Persian] as the result of a process resembling what is known as creolization in modern languages"; later, Utas (2013: 258) backs away slightly from this suggestion, emphasizing that a "great part of the simplification of Old Iranian that final led to the formation of New Persian had obviously already taken place in Middle Persian." Rezakhani (2024: 331) acknowledges that "some" have considered New Persian a creole, but he refuses the term for the discussion of the evolution of New Persian not on analytical grounds but because "creole" "can act as a marker of exceptionalism and indicate a misunderstood social construct," citing Ansaldo, Matthews, and Lim 2007. Yet in 2017 and 2020 he considered New Persian "an almost pidgin form of the language," a "pidginized form of Middle Persian" (Rezakhani 2017; 2020: 263).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Varieties of Contact Languages

Readers already familiar with contact linguistics may wish to skip these substantial sections on the categories of pidgin, creole, bilingual mixed language, and semicreole, as they review what is well known in that subfield of linguistics for specialists in history and for other humanists. Others may find the categories and examples in this section useful for understanding the entire argument of the present work concerning the early history of Persian, and as an orientation to some of the most important issues concerning language restructuring through population contact. It is impossible to grapple with the linguistic history of Persian fully, and to discuss the drastic reduction in its inflectional morphology, without confronting these categories – unless we are to treat Persian as a language unlike any other and not susceptible to the factors involved in change in all other languages.

Practically all languages are shaped in little ways, at least, by their speakers' contact with speakers of other languages, and particularly through the agency of bilingual individuals who replicate features of one of their languages in the other one that they use. The adopted word (usually known as a loanword), transferred from one language into another, is the simplest and most obvious type of such contact-induced change. Some new languages, however, stand out as having arisen precisely because of movements of people or population changes involving speakers of different languages. These are the so-called contact languages: languages profoundly shaped by population contact. By the year 2000, linguists readily acknowledged at least three types of them: pidgin, creole, and bilingual mixed language.[52] Each category entails scholarly debates. I will explain each of the three types just mentioned with specific examples because interested readers familiar with other fields of investigation, such as history and historical linguistics, cannot go far into this topic without encountering the debates about them. These types, as separate categories, are misleading in that they refer to various phenomena that are the outcomes of the same bundle of processes: language acquisition at different times of life, bilingualism, and the transfer of features between languages. The types differ for reasons that have long been difficult to discern because of complex historical social factors not easily measured in each instance. Together, these types still form one of the present bases of research in contact linguistics. Anyone interested in the restructuring of ancient

[52] Thomason 1997a; Holm 2000a: 4–12; Winford 2003: 18–22.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Persian, the subject of this book, will want to know whether Persian at any stage fits into these categories. Because I accept the existence of degrees of inflectional reduction – step (5) in the previous section – I hasten to add that I do not consider any stage of Persian directly attested in premodern history to be definitely a pidgin or a creole as the terms are conventionally employed, but to repeat the critical point: creole/noncreole is not a binary phenomenon. As pidgins and creoles represent extreme cases of restructuring via inflectional reduction and other normal changes, the kind of change to which all languages are potentially susceptible in certain social circumstances, it will help the reader unfamiliar with this topic to know about some outstanding instances of this phenomenon. It will put the partial restructuring of Persian into perspective, as virtually on a scale. This scale will be imprecise. Readers oriented toward humanistic disciplines will not be troubled about such a lack of precision. Extremely complex historical human phenomena that cannot easily be measured or quantified, and for which there are many variables, are the staple food of humanistic investigation. Generally, it is linguists, not historians, who aspire to make a more exact science of their material in the history of language. Historians, whose goals and methods differ from those of linguists, may be satisfied with what I have to offer about the history of languages: robust explanatory models based on biological and social facts, supported by the amply tested methods of historical linguistics and philology, offering numerous closely matching parallel instances, and contributing to a historical narrative with verisimilitude that is supported by all available primary sources. Linguists, too, may appreciate the case I make about Persian, although McWhorter's concise, more synchronic and comparative linguistic argument about Persian probably accords more with their own goals, which usually address language itself as a phenomenon and deal with particulars and instances only instrumentally.[53]

The premise to be accepted is that languages can be restructured and simplified in the ways already explained. Historians are used to these sorts of nonquantitative scalar concepts. With the warning that these are not absolute categories, I shall give examples to illustrate each of the three kinds of contact languages generally countenanced by language contact specialists. To this I will add a fourth type, the semicreole, established in the literature but not widely employed in analysis. In each case, it is critical to observe the social contexts in which these kinds of languages develop. Importantly, I need also to define the terms for the purposes of my own

[53] McWhorter 2007: 138–164.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

discussion, because even specialists in this field have not achieved consensus on the definition of their own primary categories of analysis.

Pidgin

A "pidgin" is a simple language spoken by nobody natively, generated collaboratively and learned by adults for communication primarily with other adult speakers of different languages.[54] Otherwise they cannot understand each other effectively, being limited to gesture, repeated words, and ad hoc jargon.[55] When such jargon (in this technical sense) becomes a linguistic system with a regularly observed, simple grammar, a pidgin has emerged.[56] A pidgin exhibits minimal word inflection or no inflectional morphology at all and begins with a small vocabulary drawn mostly, but not exclusively, from one language in a multilingual environment. The term of art for that main source language is the "lexifier," but pidgins are drastically simpler in inflectional morphology than most other languages including their own lexifiers. At the same time, pidgins often exhibit features of phonology, syntax, or other systems of the first languages of their learners, which they impose on the new pidgin. Sometimes linguists call the latter "substrate" influence, although the term is imprecise and misleading.

Pidgins originate through collaboration in communication. Native speakers of the lexifier deliberately simplify their language to help their adult nonnative speaker interlocutors to understand. This is called "foreigner talk," a term coined on the model of "baby talk."[57] Deliberately simplified foreigner talk is a much easier target of learning to attain for nonnative speakers struggling to understand their foreign interlocutors than the complex natively spoken lexifier. At the same time, a grammatically abbreviated version of a language is a normal outcome of informal adult language learning, as many studies prove, so there are cases in which foreigner talk appears to play no significant role.[58] In any case, for nonnative speakers in

[54] Thomason and Kaufman 1988: 167–171; Holm 2000a: 4–6; Winford 2003: 268–303; Myers-Scotton 2006: 278–280; Siegel 2008; Matras 2009: 277–278.

[55] Hock 1991: 522–523; Winford 2003: 268–269; Hock 2021: 724–727.

[56] Much research on pidgins concerns theories of the process, including the cognitive component, by which the language emerges as a language, not on the social context of its emergence. Because I accept it as a fact that pidgin languages, as defined here, exist, these theories are not addressed directly here. Readers may consult Siegel (2008) for a lucid analysis of the problem.

[57] Ferguson and DeBose 1977. See also Romaine 1988: 72–84 and Hock 2021: 640–642.

[58] Siegel 2008: 26–30. Foreigner talk in pidgin generation is real, however; see the present-day Arabic pidgin examples collected by Avram (2014: 33–37).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

most historical population contact situations, functional efficiency is often, if not usually, a goal preferable to fluency like that of a native. In many cases, native belonging would not be a feasible or even desirable goal for an outsider. Thus, it is reasonable to suppose that one factor in pidgin formation is that its first users had no intention of complete, native-like acquisition of the lexifier.[59] In sum, pidgin formation is a creative, adaptive process for humans in new social environments, developed through the joint agency of adult speakers of different languages.[60] It becomes regular and systematized through mutual use. Because a pidgin is regular, possessing grammatical rules, however simple, it is not possible for a newcomer fluent in the lexifier to speak it through ad hoc simplifications of the lexifier (mere "foreigner talk"). One must learn a pidgin's grammar and conventions. This is not a difficult task, relatively speaking, because it has deliberately been made easy by the necessity of, or interest in, cooperation. Otherwise, it would not work as a reliable system of communication. Pidgins are thus real languages, but they are nobody's first language, and they are not created by a master plan. Along with their creole offspring, they have been proposed to be the inflectionally simplest known languages.[61] That is, indeed, the purpose of their existence.

The linguistic history of a specific pidgin can illustrate this category better than a theoretical sketch. Let us take the example of Police Motu, a pidgin of Papua New Guinea now disappearing. Papua, also known as New Guinea and Irian, is the second-largest island in the world and is also one of the most linguistically diverse countries in the world. An estimated 750 or 860 languages are spoken in Papua New Guinea, the modern country occupying the eastern half of the island, to the point that one can encounter a different language in practically every village and valley.[62] When the British established a regional base for themselves, Port Moresby, on the southeastern coast of Papua in 1873, staking a claim on the large island's territory, population, and resources in competition with the Dutch and the German governments, they inadvertently caused a new pidgin language to develop and propagated it. The people in the immediate area of the settlement of Port Moresby spoke an Austronesian language called Motu. British visitors and other new settlers learned a simplified Motu as

[59] Baker 1990.
[60] Churchill (1911: 1), using the term jargon for pidgin, rightly wrote that "jargon is the speech of necessity."
[61] McWhorter 2001. [62] Lynch 1998: 34, 61–67; Nettle and Romaine 2000: 80–84.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

used by locals to facilitate their communication with them.[63] After Port Moresby became the base for the extension of British control over the surrounding regions of the island, numerous individuals from many different countries in the Pacific and distant continents came to Port Moresby to partake of the new opportunities there and to search for work and resources to exploit. Dutton describes the mix of people consisting of "Chinese, 'Malays,' 'South Sea Islanders,' persons of mixed race, Maltese, Ceylonese, Indians, Filipinos, Europeans, Americans, and a large number of British-Australian origin."[64] The newcomers relied on Motu-speaking locals for necessities, so the pidgin form of Motu found increasingly more speakers among the new arrivals, none of them native speakers of the language. The British government of Papua (1883–1906) created a constabulary to enforce its own claims of government over the island, maintained continuously as a police force under the succeeding Australian government (1906–1975).[65] For this, at first, the British recruited Motu-speaking personnel from the area of their base. The language came to be known as Police Motu from its associations with this organization. As increasingly more people on the island were brought, by hook and by crook, under colonial organization, they found the pidgin Police Motu expedient as a means of communication with authorities and of access to jobs and commodities from the world at large. Representatives of other Papuan communities learned it, even those under British authority who were remote from the Motu-speaking area. Already by 1894, one source reports, Motu had "been carried by retired constables to numerous places where it had never before been heard of."[66] Note that the date of this report means that simplified Motu must have developed, and its use become widely adopted, within ten years of the inception of official British government there. The new language found many more speakers in the decades to follow. In 1928, a government anthropologist noted that the "pidgin-Motuan" "has proved vastly useful," and that "in practice we have made the widest use of Motuan in that simplified (or shall we say degraded?) form which may properly be called pidgin-Motuan."[67] In being shared by many different communities, the pidgin variety of Motu changed and developed in its pronunciation, grammar, and lexicon. Early on it became

[63] Some have claimed that a jargon or pidgin of Motu existed before the arrival of the British, a kind of regularized "foreigner talk" facilitating their communication with their linguistically different neighbors for purposes of coastal trade and cooperation. See Dutton 1997: 13–18.
[64] Dutton 1997: 18. [65] Kituai 1988. [66] Kituai 1998: 107.
[67] Williams 1928: 7–8. The parenthetical remark of moral evaluation is in the original. It illustrates the tendency to stigmatize new, morphologically simplified languages.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

mutually unintelligible with its source and lexifier, the local Motu language from the vicinity of Port Moresby. This was due not just to plentiful words adopted from other languages for colonial novelties, but also, and more importantly, by dropping numerous salient features of Motu grammar and making do without them and using a relatively limited vocabulary. At any time, a few dozen British field officers used it in supervising hundreds of such constables from innumerable different linguistic backgrounds. Through them, Police Motu became a major medium for communication throughout British Papua, easily acquired because of its extreme simplicity.[68] To some extent, it even became a part of the British institutions there, which promoted the use of pidgin Motu while the administrators frowned on pidgin English. For example, uncooperative Papuans put in prison (called *dibura*, or "the dark," in Police Motu) by the British in Port Moresby were expected to learn Police Motu to get by.[69] Some of these prisoners, upon release, went on to work for the constabulary that had recently imprisoned them, carrying the language with them across Papua. In this way, time in prison had taught them a linguistic skill useful for effective British government service. Those who had contact with the British and, subsequently, Australian government became distinguished by this shared language, so the use of Police Motu also signified access to new commodities, technologies, and social authority. This only gave further incentive to Papuans for learning it. British and Australian patrol officers, in charge of the constables, needed a pidgin tongue at least as much as their constables to communicate with them. Recruit Jojoga Yegova recalled that in 1943, in training for the Royal Papuan Constabulary, "Many of us initially used sign language to communicate with each other until we could master Police Motu or New Guinea *Tok Pisin*" (the latter discussed here subsequently).[70] Decisively, many Papuans from different communities who fought alongside each other together with the British in World War II used Police Motu among themselves. Through major shared endeavors, the pidgin Motu became a common link between communities collaborating with the British who could not otherwise communicate directly with one another. In 1963, the authors of an introduction to

[68] On Police Motu generally, see Dutton 1985 and 1997. See also Winford 2003: 297–298, who would rather call it a "simplified language" than a "prototypical" pidgin.

[69] Dutton 1985: 74–78. It was a deliberate policy, at least in the 1890s, to put prisoners, "both at work and in the cell," with others who used Motu, so that any prisoner who spent a few months in the Port Moresby prison spoke "Motuan more or less" (i.e., pidgin Motu). By 1905, prisoners were still "picking up" Motu first, then learning a little English.

[70] Kituai 1998: 179.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Police Motu dubbed it simply "the lingua franca of Papua."[71] A Bible translation in Police Motu appeared in 1964.

Upon the political union of Papua in the south and New Guinea in the north in 1972, and with the independence of Papua New Guinea in 1975, its adherents insisted that Police Motu should be one of the three official languages of the new state. They redubbed it Hiri Motu, "Trade-Expedition Motu," to remove possible negative connotations in the name Police Motu. At that time, there were an estimated 120,000 users of the pidgin as a second language.[72] To this day, Hiri Motu still has a special official status in Papua New Guinea. In reality, Hiri Motu has apparently lost most of its ground to a pidgin based on English, Tok Pisin (from English "Talk Business"), which had long been in use in the north part of the island, New Guinea, since the days of German rule there (1884–1919) and, in gradually divergent varieties, in other Pacific islands to the north and east. Pidgin Motu was useful for much of the twentieth century, but now, as far as reports indicate, only a shrinking, aging population retains it, alongside their separate village languages (including the original Motu) and pidgin English. If Christian missionaries and zealous Papuan users of Hiri Motu had not advocated for it so strongly in the 1960s and 1970s, this pidgin would probably be little documented, and it may have disappeared more quickly. It is, however, likely to die out soon, because it never obtained a substantial number of native speakers and the utility of Tok Pisin (the pidgin of English used in Papua New Guinea) has become greater for younger generations.[73] Again, besides the social circumstances in which they arise, and their relative simplicity, that is a fundamental characteristic feature of pidgins: they have a negligible number of native speakers.[74] By their definition, pidgins will tend to disappear unless they are sustained by a continuous stream of new nonnative learners.[75]

[71] Wurm and Harris 1963: 1. [72] Reinecke 1975: 752.

[73] Romaine (1992: 330) suggests that Hiri Motu is (was?) too regionally marked to become a successful lingua franca, unlike Tok Pisin, which is a relatively "ethnically neutral lingua franca." She also demonstrates (1992: 93–94) a distinct drop-off in the percentage of Hiri Motu-speakers in seven major towns of Papua New Guinea from 1966 to 1971 alone. Hiri Motu seems never to have become current in the northern, formerly German-controlled part of the island, or in the more densely populated highlands that were brought under colonial control later. It may be that the social bases of diffusion for these two pidgins also predisposed them to different careers: relatively few Papuans served in the constabulary, whereas many Papuans went out of their villages to work on plantations, where Tok Pisin found its first speakers.

[74] There are a few exceptional reports about multilingual native speakers of Police Motu. For example, Jack Hides (1906–1938), an explorer of Papua New Guinea, "grew up speaking Police Motu" (Franklin 1989: 99). His father was then the head of the prison in Port Moresby, where Police Motu was the common language.

[75] Myers-Scotton 2006: 280: "Pidgins fade away."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Creole

The second widely accepted category of contact language, and the one that has received the most attention, is the creole.[76] It is also a category used recently by historians rather freely without precision or grounding in the study of actual creoles. This undermines the historical study of creole languages by diluting the utility of the term further than linguists already have done (addressed later). I urge fellow historians to understand the term thoroughly before entering into conjectures and debates about its application.

As already mentioned, the term arose in the eighteenth century to designate a number of languages arising among populations of mixed origins, mainly enslaved and subjugated peoples of Africa and the Americas; subsequently, the term was generalized widely to refer to all sorts of cultural and linguistic products of the mixture of different populations.[77] I do not use the term in the historical sense of the specific social origins of languages bearing the proper name "creole," but rather as a linguistic category. For the purposes of this discussion, I adopt a widely used conventional definition of creole. The definition is simple and well known and serves the present purpose. It is that *a creole is a nativized pidgin*. When children learn a pidgin natively, they begin to develop the pidgin beyond its prior range of uses because it is one of their chief means of verbal expression, if not their only means, and human social needs demand the development of rich and nuanced means of expression. Out of necessity, and possibly by their nature, children accomplish this enrichment in their language acquisition and use. Probably every creole specialist today will quibble about such a definition left unqualified, but most creole specialists do not offer or apply a strict definition to their subject.[78] In the same way, creole specialists disagree with each other constantly about their

[76] The most concise survey of all the issues pertaining to creole studies is provided soberly by Bakker (2017a).

[77] For orientations to the changing significance and contested meaning of the term creole, see Stewart 2007 and McWhorter 2005. Myers-Scotton (2006: 280) explains the sequence clearly: "Before linguists began studying these languages, the term 'creole' was already in use as the term for people who lived in European colonies, especially people of mixed descent. Then, the language varieties that developed in these colonies came to be called creole."

[78] For example, the useful volume *Comparative Creole Syntax* (Holm and Patrick 2007) presents grammatical summaries of eighteen different creole languages without any declared criterion by which they could be so designated. The recognition that these eighteen languages are "creoles" is based on convention and intuition, for each of them is a recently restructured variety of another well-known language retaining a large native speaker population today (Arabic, Assamese, Dutch, English, French, Portuguese, and Spanish). Blasi, Michaelis, and Haspelmath (2017) suggest that not all languages called creole develop from pidgins. Bakker (2004) makes a sound attempt to sort

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

respective definitions.[79] Part of the dispute about the definition of creole is due to the evident confusion among creole specialists about how to begin analysis. Some begin with the conventional designation and seek common denominators in those languages already so called; others begin with a theoretical definition and search the data for examples to match the definition. These methods are at odds and are bound to lead to disagreement.[80] But to say that a creole is a nativized pidgin is still probably the closest thing to a generally accepted working definition.[81] We could call it here by a name other than creole, to sidestep the debate, but it seems absurd to invent a new term for an object about which a substantial subfield has developed, and it would be disingenuous, less meaningful, and even misleading to pretend that we are not talking about the same kind of phenomenon. Until creole specialists agree upon a more refined

the terminology, along with specific claims about the nature of the differences between jargon, pidgin, pidgincreole, and creole.

[79] Blasi, Michaelis, and Haspelmath (2017: 724b) note that, "while many authors of contrasting theoretical stances agree on the existence of a distinctive creole profile, they base their arguments on non-overlapping sets of linguistic features and languages."

[80] Creole specialists debating the validity of the category of creole language should realize a distinction between ideological and analytical terms (although see the exception of Bakker 2017a: 15). "Creole" (the word) did originate as a conventional appellation in specific colonial historical circumstances. Winford rightly notes (2003: 269) that such labels as pidgin, creole, and lingua franca "were first used by lay people or non-specialists before they were adapted as technical labels by linguists." Linguists have attempted to define what the term ought to mean in linguistic research, to endow the term with analytical force, as a term of art, and to use it henceforth in a destigmatized way extricated from its history. The sciences are replete with such terms of art. In this case, it is expedient because the languages under discussion are already mostly known as "creoles." The alternative – to give the phenomenon an entirely different name – has not been attempted successfully. But those wanting creole to serve as an analytical term are opposed by those who deny "creole exceptionalism," with the claim that creole is merely an ideological term and that languages so called are like every other language. In opposing such attempts to define "creole," they endorse the term as ideological and forego analytical precision, which, one supposes, is the reverse of their intended effect. See the wise remarks of Holm on terms like "creole" (2004: xi–xvi), the palpable frustration of McWhorter (2011a: 19–27, 103–119) over the studies edited by Ansaldo, Matthews, and Lim (2007) – see also McWhorter 2013b and, more decisively, McWhorter 2018 – and the discerning take of Bakker 2017a. The opponents of "creole exceptionalism" must explain why mass nonnative acquisition (different from regular childhood acquisition) should *not* matter in a language's history. Meanwhile, a strong empirical basis for the category of creole language has been offered by Bakker, Daval-Markussen, Parkvall, and Plag (2011), Bakker and Daval-Markussen (2017), and Bakker (2017b).

[81] E.g., Romaine 1988; Mühlhäusler 1997: 8–10; Holm 2000a: 6–9; van Coetsem 2000: 185; Thomason 2001: 159–160 ("some creoles are nativized pidgins"); Matras 2009: 278–79; Holm 2010: 256; Webb 2013: 304; Bakker 2017a: 6. There are variant definitions based on differing theories of creole formation, clearly and impartially summarized by Matras (2009: 283–288) and Bakker (2017a: 18–19). One notable outlier, offered by Mufwene, is that creoles form no special category, and that they are merely languages like any other, not simpler than others, and which represent merely a different selection of features from a pool of language features in a speech community. This "feature pool" hypothesis has been refuted; see McWhorter 2012; 2018: 33–62 for clear expositions and rebuttals.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

definition for their subject, this old definition will have to do for noncreolist historians interested in the history of languages in society generally. To define a creole as a nativized pidgin is, however, meaningful beyond that mere statement, because pidgins are morphologically very simple languages, far simpler than their lexifiers, and purposefully so. Therefore, the definition that I use indicates likewise that creoles originate in a similarly morphologically simple state as divergent varieties of an earlier, more complex, language.

Part of the trouble in defining creoles is that they are conspicuous more by the absence of certain characteristics and by difference from their lexifier rather than by common features that arose in them each independently.[82] McWhorter is one of the few creolists to posit an analytical definition for his subject that can be applied as a categorical check. His definition has not been widely accepted, despite its potential utility, and despite the lack of proposed alternatives.[83] He calls his definition the Creole Prototype. By this model one can identify a creole if it has, all at once, three characteristics. His diagnostic characteristics are really the *relative absence* of features typical in languages.

(a) Morphologically: little or no inflectional affixation.
(b) Phonologically: little or no use of tone to distinguish monosyllabic lexical items or morphosyntactic categories.
(c) Semantically: little or no noncompositional combinations of derivational morphemes with roots.[84]

Restated and simplified, the three criteria have to do with word inflection, lexical tone, and opaque word formation. The less a language exhibits these features, the closer it is to McWhorter's diagnostic Creole Prototype. The critical distinction is undefined: how much constitutes "little."[85] Here the Creole Prototype is blurry, but at least we have working criteria for

[82] See Kranich and Breban (2021) on the deceptive inconspicuousness of the loss of features in diachronic linguistics. See also Blasi, Michaelis, and Haspelmath (2017), who show that the category of creole is unlikely to represent a universal default grammar, as once proposed, because of the resemblances of creole grammars to that of their lexifiers.

[83] The neglect seems to be partly because the very idea of a definition of creole came under attack, sometimes in a politically charged way, from a few creole specialists. See McWhorter's discussion (2013b and 2018).

[84] Quoted verbatim from McWhorter 2011a: 39; cf. McWhorter 2005: 9–37. McWhorter 2011b: 111 presents a revised and refined version of this list to account for counterarguments. I have not presented it here because it would require a lengthy explanatory digression.

[85] Cf. Kihm 2012: 658; Thomason 2004: 244. Often creolists assert that creole languages have no inflection, but Bakker (2003) and Holm (2008) point out that some languages conventionally called creole do have minimal inflection. Again, this means that creole/noncreole are not binary states.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

discussion. Notably, the definition is synchronic – that is, it is posited regardless of historical data – but creoles are understood as due to historical social processes. In effect, McWhorter posits that we can infer past social events, even undocumented or prehistoric events, from the grammatical state of a language with these characteristics, without reference to the prior historical state of a language's grammar. This proposal is quite meaningful for the methods of linguistic history.

Recent efforts to determine characteristics special to creoles have achieved important results that refine this model. Computational cladistic methods applied to the large data set of a survey of the features of the world's languages show that the languages already called creoles form a typologically identifiable cluster regardless of their respective lexifiers. It is not that there is any one positive characteristic that is common to all languages regarded as creoles, but there are clear tendencies, such as the use of compounding for word formation and reduplication for grammatical purposes (such as plurality of nouns or repeated action), although these are not universal to creoles. As Bakker and Daval-Markusson explain, "There are no linguistic properties that are unique to creole languages. Phonologically, creoles avoid extremes and are average languages." But then, critically, they further indicate a decisive negative feature universal to the morphologies of the set of modern creoles: "Creoles never inherit contextual inflection from their lexifiers, but they may inherit or develop inherent inflection." Here again, the one thing that all the creoles share is the *absence* of a feature of morphology common to many of the world's languages.[86] This last observation is especially important for a historian or a historical linguist attempting to diagnose the status of an ancient language as possibly having arisen as a creole. McWhorter defines creoles as having "little or no inflectional affixation." It is Bakker and Daval-Markusson who specify the kind of inflection involved. Contextual inflection refers to morphemes that are determined by syntax, such as "person agreement markers in the verb linking to nominal elements, agreement markers in adjectives linking to nouns or case markers in nouns linking to their function in relation to the verb."[87] Studies of language acquisition

[86] Bakker and Daval-Markussen 2017: 83–85. See also Kihm (2003) and Luís (2010), forerunners in this insight. On the distinction between inherent and contextual inflection, see further Marzi, Blevins, Booij, and Pirrelli (2020).

[87] Bakker and Daval-Markussen 2017: 83–85. On this distinction in morphology, see Booij (1993; 1996). Cf. McWhorter 2011b: 96–98. This sort of inflection is very much like if not identical to Myers-Scotton's "outsider system morphemes" in her 4-M model of code-switching (a refinement to her Matrix Language Frame model) (2006: 241–278); cf. Winford (2005: 418–419). Myers-Scotton

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

suggest that contextual inflection is indeed acquired later in the process of learning languages.[88] No language conventionally regarded as a creole has preserved contextual inflection from its parent language or lexifier. This supports the notion that creole languages are a byproduct of an episode of mass nonnative acquisition. Future studies are likely to refine these analyses further, but for a historical linguist or a historian of languages, this criterion is clear and easy to apply in tracing the history of a language from one known stage to another.

For a real example of a creole and its emergence from a pidgin, we can remain in Papua New Guinea, with its high density of local languages. The name Tok Pisin is from English "Talk Pidgin" meaning "pidgin speech," the term pidgin coming ultimately from the English word "business." Tok Pisin is an example of a pidgin variety of English that has become a creole.[89] That is, tens of thousands of people now use this pidgin as a native language; hence, a creole. It is a species of the Melanesian Pidgin English still used in Papua New Guinea, Vanuatu, the Solomon Islands, and other South Pacific islands. Its earliest roots seem to be in an English-based jargon used by British sailors in the Pacific in the late eighteenth century. This was used in a variety of far-flung locations around the ocean for trade and basic communication. In the mid nineteenth century, it was developed into a pidgin among sailors, whalers, and gatherers of sandalwood and sea-slugs for sale in China. Sydney, Australia, a colony founded in 1788, was a major site for the early diffusion of this pidgin English. Its users interacted with locals in many places, from Canton to Samoa, their medium being this simplified English. It was the European plantations of the South Pacific, however, that formed the main sites for the acquisition and subsequent diffusion of pidgin English by Melanesian islanders. Plantation owners in Queensland, Samoa, and elsewhere derived very cheap labor, if

(2008: 22) explains her "outsider system morphemes" with an example: "in German, verbs and prepositions can assign case designations in a clause, but the actual case morpheme (the outsider) appears in elements forming a constituent separate from (i.e., outside) the larger constituent containing the verb or preposition that assigns case." Holm's observations about the presence of inflection in creoles do not cause the "paradigm shift" he supposed (2008) if we distinguish such types of inflection that play syntactic roles of different kinds. None of the features of creole inflection that Holm includes represent contextual inflection bearing agreement with other constituents of a sentence outside of its own phrase, such as case marking on nouns or plural agreement on verbs.

[88] Booij 1996: 11–12. See Slabakova 2008: 100–112 on the bottleneck of "functional morphology" in second-language acquisition, although her focus is limited to the mismatch of morphological inflection and semantics between languages as a hurdle, and not variations in inflectional complexity.

[89] On its development, see Mühlhäusler 1983; Holm 1988–1989: 2.526–551; Crowley 1990; Romaine 1992: 31–54; Mühläusler 2003: 5–8; Winford 2003: 291–293; Tryon and Charpentier 2004.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

not forced labor, by contracting or kidnapping workers from various Melanesian islands for periods of several years at a time. The labor exchange lasted for decades. These workers, taken abroad from their home islands, rarely shared a common language among themselves, by the employers' design, so they learned the pidgin English made available by their employers in the context of their daily work for those years and, eventually, when some of them returned to their homes, they brought it with them. It remained for them a medium for access to goods and opportunities beyond their immediate locale. Samoan plantations were one of the main sites for the development of Melanesian pidgin in the last quarter of the nineteenth century. The Samoan plantations were run by Germans, where speakers of pidgin English had little linguistic contact with native speakers of English. This allowed the pidgin to develop independently of any native variety of English. The pidgin was adopted further in German-controlled New Guinea (1884–1919), the north side of the island in which British Papua was situated, where new plantations received some of the very same laborers, originally from different islands, who had been employed in Samoa, along with still more imported laborers from other islands. After World War I and the end of German government there, Melanesian pidgin English was adopted more and more widely in villages as a means of access to goods and opportunities outside the village. Christian missionaries also used it to reach polyglot locals, contributing to its standardization. Numerous inhabitants of villages in Papua New Guinea, each of whom spoke a home language with a small number of other native speakers, called a Tok Ples, from English "[the] talk [or speech, of one's home] place [or village]," or "local language," moved to bigger towns, many of them colonial foundations, and to plantations, and mingled with inhabitants of other regions with whom they had little or no contact before. Tok Pisin served, and still serves, to create an ever-broadening speech community in such sites. People who move from the linguistically diverse villages of Papua New Guinea to the towns find Tok Pisin expedient as a common language. When members of different language communities in Papua New Guinea form families and procreate, the common language of their home is sometimes Tok Pisin, the one language held in common by both parents. Children then learn it as their first language of home use. Even when no parents use Tok Pisin as a first language, it can displace the local languages because it represents, to its young speakers, a path to meet their aspirations as inhabitants of the independent state of Papua New Guinea. Villages now witness generation gaps in which the youth are deterred from using the language of their

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

elders, preferring the supralocal Tok Pisin instead.[90] In this way Tok Pisin has become a creole, a pidgin acquired as a native language, in these environments, and shows striking innovations in its usage. For many speakers outside of the cities, however, it is still technically a pidgin, a secondary language not spoken natively. Tok Pisin spoken natively does show new features and developments as it is augmented to serve the communicative needs of those who use it as their dominant language and for every basic human need.[91] Moreover, native speakers of Tok Pisin who learn English, perhaps in school, creatively import nonpidgin English words and features into their dominant language. This is one variety of "decreolization" (discussed here later), the term used for the gradual assimilation of a creole to its lexifier. From the conventional definitions of pidgin and creole, it is evident that pidgin and creole stages of development can coexist, and that one language can be both at the same time for different speakers.[92] Tok Pisin is a pidgin that has also become a creole for many thousands of native speakers.[93]

Linguists have documented many creoles in the world, most of which have arisen as an indirect result of European colonization of the last several centuries and are, hence, young languages with specific, identifiable, sometimes documented, relatively short periods of genesis.[94] What is special about creoles is that, because they are nativized pidgins, and pidgins are drastically simpler in contextual inflection than most other languages, the young creole languages are simpler in the same way, too. Note that pidginization is not the root of all morphological simplification in a language. It is rather nonnative acquisition that appears to be the major cause of such change.[95] Pidginization represents only an unusual but natural occurrence giving rise to radically restructured, and hence especially conspicuous, new languages.[96]

[90] Kulick 1992 and 2019. [91] Smith 2002.

[92] Baker (1997: 92) regards this as a problem for the definition of the terms. I see no problem in it, but rather a useful distinction.

[93] Bakker (2002) has proposed the name "pidgincreole" for such pidgins that acquire native speaker populations, with corresponding effects on the pidgin language. Another common term for the same thing is "expanded pidgin" (Siegel 2008: 4).

[94] Holm 1988–1989.

[95] In a study published posthumously (2017: 160–161), Arends (1952–2005) meticulously showed that, in the genesis of Suriname creoles in the period roughly 1650–1750, "for every locally-born child acquiring the creole as a first language there were around ten African-born adults – for whom learning the creole was a matter of second language acquisition." In this case, these local "creoles" were for a long time, for most of their users, what we would call a pidgin. Arends makes it clear that adult learning is what propagated it not all at once in a single episode but over several generations.

[96] McWhorter 2005 and 2007.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Bilingual Mixed Language

Initially dubbed "intertwined language," this has been proposed as a third kind of contact language, acknowledged as relatively rare. It was claimed as distinct only in the 1990s, beginning with the research of Peter Bakker.[97] The characteristic feature of a bilingual mixed language is that it emerges among children of parents who have two different languages; one language, typically the mother's, provides the grammatical matrix, or the verbs and verbal grammar, while the other language, usually the father's, provides much of the lexicon, or the nouns and nominal inflection. Both components retain most of the grammatical complexity of the two intertwined languages proper to their originally separate systems. Bakker attributes the genesis of such languages to the identity formation of teenaged offspring of immigrant men who marry local women.[98] "These multi-ethnolects," or intertwined languages, Bakker argues, "are used by second-generation immigrants and symbolize their distinctness in society but, at the same time, their connections to both the new land and the parents' country of origin." The available evidence suggests that intertwined languages have their genesis with the formation of a separate ethnic group born of this mixture of populations "within a very short time frame of a few decades."[99] Such languages arise when fully bilingual young people deliberately used elements of the language of their mother and of their father together in the same utterances. Specifically, they have used very large quantities of their father's nominal vocabulary in their mother tongue.

An outstanding example of bilingual mixed language is Michif, used in the prairie lands of Canada, spoken today by several hundred people.[100] It is characterized by Cree verbs and syntax while the great majority of the nouns are from French. Fur hunters and traders from Quebec and Acadia from the seventeenth century onward took Cree-speaking brides as they moved westward in North American land. The mixed-offspring people are known as Métis ("Mixed"), and some of their offspring generated Michif. Bakker argues that what began as an "in-group language" or linguistic code used by a young generation of mixed ancestry to define itself and to express their common identity as mixed became, after a generation, nativized when acquired as native by subsequent children. This resembles the process by

[97] Bakker 1997; Matras 2009: 288–306; Bakker 2012: 182–183; Meakins 2013; Hock 2021: 749–757.
[98] Cf. Muysken 2013: 202: "The most complex types of code-mixing have been documented with adolescents and young adults. . . . Most cases of complex code-mixing have been recorded with second-generation speakers in immigrant communities."
[99] Bakker 2012: 180. [100] Bakker 1997; Bakker and Papen 1997.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

which a pidgin becomes a creole when acquired by children, but it does not entail significant morphological simplification. Both intertwined languages and creoles are native tongues, but the two have different characteristics. Bilingual mixed languages may remain highly complex in their morphology, for they did not originate as pidgins. Bakker argues that bilingual mixed languages "are associated with very specific social events," such as the intermarriage of immigrant men with local women, within the memory of the first speakers of such languages.[101] Many speakers of Michif have known neither French nor Cree, but only the younger, mixed language formerly comprising parts of both.

Although other examples of this sort of mixed language have been proposed,[102] not all linguists accept the bilingual mixed language as a truly separate category. Winford analyzes intertwined languages simply as extreme cases of lexical borrowing by bilinguals whose shared dominant language is the mother's tongue.[103] Versteegh makes similar observations and is ready to reject the category for this reason. As he points out, Michif and other "mixed languages" are a regular possible outcome of relatively extreme but normal code-switching and word-borrowing by bilinguals.[104] Therefore, they are not a separate species of contact language due to a "new mechanism of language change." Michif represents, therefore, lexical borrowing and transfer between languages used by bilinguals that is unusual only in its extent. After a period of such mixing, the new mixture became permanent and normal. When such extensive code-switching (also known as code-mixing) becomes a fixed norm among children, they can grow up speaking a "mixed language." Michif would therefore be a special variety of Cree with a massive importation of adopted French words brought in by children at one formative stage in the history of a bilingual community.[105]

If pidgins are just the genesis-stage of creoles, and mixed languages are the result of the transfer of a copious lexicon from one language into a matrix language by young bilinguals, we are therefore apparently dealing with a single bundle of language phenomena having different names for different stages of development and different outcomes. The main variable between each of these three is the age of the speakers' language acquisition over more than one generation. Creoles are more likely to be documented than pidgins, as they are more likely to survive by the defining fact that they have acquired native

[101] Bakker 2012: 182. [102] Bakker and Mous 1994. [103] Winford 2005: 396–402.
[104] Although today the expression "code-switching" is often used simply to mean signaling competence or participation in more than one culture, it technically refers to embedding words or phrases from one language in the matrix of another. See Hock 2021: 637–640.
[105] Versteegh 2017. Cf. Winford 2003: 168–207; Hock 2021: 757.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

speakers. Bilingual mixed languages are relatively rare and arise in special bilingual social groups in which mothers and fathers have different languages. In these cases, language acquisition in childhood plays the main role, rather than adult language acquisition, and accordingly morphological complexity can be maintained. But it is the creoles that have attracted the overwhelming attention of linguists specializing in language contact as they are undeniably restructured, independent, young varieties of other, more morphologically complex, older languages, associated with a group that acquired it in childhood. They are therefore exemplary for theories of the outcome of language contact. The great attention paid to pidgins and their creole descendants has led, however, to a predominant, implicit, binary notion of "normal" languages versus contact languages that are primarily creoles. As I have emphasized, this notion of two states of language is mistaken.

Semicreoles: Partially Restructured Languages

As already explained under (5), plenty of research has demonstrated that there are *degrees of restructuring* toward morphological simplification in languages. Between the outcomes of pidgin/creole formation and the far more ordinary and continuous intergenerationally stable situation of large-scale language acquisition of their parents' language by children everywhere, a variety of degrees of restructuring can occur in different demographic conditions of language contact. New languages arise that are distinctly simpler than their lexifiers, retaining, for example, some minimal noun or verb inflection, but do not seem to have emerged from a pidgin. In view of such languages, the notion of a strict binary opposition of normal language and pidgin/creole should disappear. Some languages not typically regarded as creoles seem very much like creoles in their morphological simplicity, without being quite so reduced as, for example, Tok Pisin vis-à-vis English. These are reduced versions of other languages that have distinctly been transformed through a social event: an episode of large-scale nonnative acquisition of the language (steps (2)–(4) in the model introduced earlier in this chapter).

Creole specialists have long accepted that degrees of creolization are possible, although they disagree about the taxonomy or the factors by which those degrees emerge.[106] Once one accepts that degrees of restructuring exist, one

[106] Some creolists use terms like basilectal, mesolectal, and acrolectal, or degrees of "indigenization," to describe the qualitative degree of difference or distance of a creole from its lexifier (Winford 2003: 254–256, 313, 355). Others have occasionally referred to the most radically restructured creoles as "deep creoles," as if far advanced into the process of restructuring, again signifying the existence of degrees. The term decreolization designates a creole's gradual assimilation of more features of its

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

may attempt to define those relative degrees and to explicate the factors giving rise to variation in degrees of restructuring. Here we arrive at a frontier of research in contact linguistics. Specialists in contact linguistics have not yet formed a consensus on the terms to be used to describe different degrees of restructuring. There are several hindrances to such an agreement. A major one is the matter of assessing morphological complexity and simplicity. Because degrees of restructuring are expressed at least partly in terms of complexity and simplicity, the baseline for measuring change depends on the complexity of the prior state of the language now restructured, as compared with the simplicity of the restructured variety. But the "starting point" of complexity before restructuring varies from language to language in ways hard to measure or compare, and it is likewise conditioned by its own special prior population history. A partially grammatically reduced variety of modern Russian, for example, could, in theory, retain more complex morphology than the literary English of the present, while seeming rather simplified from a Russian point of view. The relativity of changes to a prior state that varies from instance to instance inhibits the development of a general scale of restructuring. Discovering methods to measure complexity, determining the kinds of complexity that should be measured, and the relative significance of the different kinds are unresolved issues on this frontier in linguistics. Still further, the demographic scenarios involved in the partial restructuring of languages vary according to many social factors that are often not discernible either to historians or linguists.

Linguists already employ at least five different terms for partially restructured languages:

(1) new contact vernacular
(2) indigenized variety[107]

On the surface, these first two terms cautiously express ambivalence about the problem of relative simplicity, even though their proponents do acknowledge relative simplicity and complexity. They do no more than to designate a new common variety of an old language as the product of language contact or as having been adopted by a new population. As

lexifier, where the creole coexists socially with its lexifier, again indicating degrees of relative outcomes in restructuring.

[107] Siegel 2013: 517–519; earlier Siegel (1997: 119–125) proposed the term "indigenized variety" with three specific features in mind. (1) "The existence of a continuum of socially and situationally marked varieties, similar to a creole continuum, ranging from a basilect (furthest from standard native varieties) to an acrolect (closest to standard native varieties)." (2) "Formal simplicity in comparison to other varieties [of the language] (especially in the basilect), found in the reduction of categories . . . optional morphological marking of nouns and verbs and the regularization of paradigms." (3) "Substrate influence in lexicon, morphology, and syntax, as well as in pragmatics."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

technical expressions, they merely acknowledge the existence of a new variety of an old language. "Indigenized," as a term by itself, indicates only that the new language appeared when adopted by an "indigenous" population.

(3) partially restructured vernacular

This term also reflects some ambivalence about degrees of complexity, but "restructured" implies fundamental differences between the grammars of the older language and the new variety of language.

(4) creoloid[108]
(5) semicreole[109]

These last two terms have the advantage of concision. More importantly, they directly connect the new, restructured variety of language with the phenomenon of linguistic creolization. They both thereby acknowledge a class of languages called creoles (something not every linguist is willing to do, for reasons already explained), and posit a variety of language that is somehow between creoles and noncreoles. The two terms are in effect synonyms. Both refer to a language that was grammatically reduced and simplified through an episode of mass adult acquisition but that did *not* emerge from a much simpler pidgin. In other words, in its complexity, relative to its lexifier, it is somewhere between the highly reduced variety characteristic of a former pidgin and that of a "normal" language that never arose from a pidgin (at least for many prior generations). Normal, here, of course, refers not to a characteristic of the language as a language but to its social environment. It means just that the great majority of the language's learners have been children for many generations, which is a normal situation for most languages. One must choose between synonyms for the sake of consistency. Creoloid simply means "creole-like," suggesting merely that it is similar to creoles without being identical to them.[110] As a term, semicreole makes the stronger commitment to an explanation in the terms described here: it refers to languages that have undergone processes similar to those leading to creoles – specifically the formation of a new, simple variety of a language in an episode

[108] Platt 1975 and 1978; Trudgill 2003: 31; Swann, Deumert, Lillis, and Mesthrie 2004: 62.
[109] Holm 2002; Swann, Deumert, Lillis, and Mesthrie 2004: 274–275.
[110] The term was proposed by Platt in 1975 and 1978, who wrote, "It would seem that a new concept needs to be established for certain speech varieties which are, strictly speaking, neither creoles nor part of a post-creole continuum but which share many features with them. Unlike creoles, these varieties did not develop from a pidgin but by some other process." Trudgill (2011: 67–68) appears to be the only active user of this term. By creoloid he means a language that did not emerge from a pidgin but that shows "relatively undramatic" simplification and hybridity ("admixture").

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

of mass adult acquisition – but have not attained all of the characteristics of creole, because it did not emerge from an entirely nonnative pidgin.[111] *Semicreole* is therefore the term that I will adopt here. It has the advantage of being associated with the two most cogent efforts to describe and explain partially restructured languages.[112] It is not that I insist on this term for the future, but I do not wish for now to invent a sixth term for the phenomenon of a language that has been grammatically reduced through mass nonnative acquisition without ever having been a pidgin. We need terminological consistency and not the ad hoc proliferation of descriptive explanatory terms for the same phenomena. Trudgill's definition of a creoloid stands also for the semicreole:[113] "A language which, as a result of language contact, has experienced simplification and admixture, but has not undergone the reduction associated with full pidginisation (nor, therefore, the expansion associated with creolisation). Such a language will resemble in its linguistic characteristics a creole which has undergone decreolisation, but will be different in its history." Simply put, a semicreole looks like a creole in its morphological and lexical simplification but it is not so severely reduced and it shows no sign of a pidgin past. Holm has analyzed the process of semicreole formation in two ways. The first is to tease apart the layered linguistic changes involved in partial restructuring through a plausible model. This breaks the bundle of features in a semicreole down into distinct formative linguistic processes acknowledged in studies of language contact.[114] For the formation of a semicreole is not a uniform, single linguistic change, but it depends on many component processes each of which is variable in its extent. Holm's second

[111] Holm (2004: 6) notes that Schuchardt was the first to use the term ("halbkreolisch") in 1889.

[112] Holm 2000a: 9–12; Holm 2004: xi–xvi; McWhorter 2007: 252–276. Interestingly, Thomason (2001: 195) has come to hold that "the category 'semi-creole' doesn't really exist," referring to an earlier article of hers in which she rather makes the term useful, writing that "A semi-creole is not a class in itself, but is merely a label that recognizes the possibility of degrees of disruption of a continuous transmission process . . . It seems to me that placing such borderline cases on a continuum between the clear noncreoles and the clear creoles is preferable to putting them in classes for which precise characterizations and definite classificatory criteria are lacking" (Thomason 1997b: 84). But the notion of "clear creoles" is not so clear-cut, nor is that of noncreoles. The term semicreole is useful for precisely the reason that Thomason indicates. It recognizes the possibility of degrees of restructuring, in episodes of mass nonnative acquisition, and when the term semicreole is used now in contact linguistics, that is generally how it is used. Watts' objection (2011: 112) to the term creoloid – that the term "only shifts the problem away from looking at language contact as a highly complex area of concern in sociolinguistics" – makes little sense. It is rather that terms like creoloid acknowledge a class of languages that demand more explanation.

[113] Trudgill 2003: 31. Trudgill adds that a semicreole should always be mutually comprehensible with its lexifier, but I do not hold this as necessary.

[114] Holm (2001; 2004; 2013: 258–259) proposes fives component steps in semicreole formation: dialect leveling, language drift (simplification), imperfect language shift with transfer of features between languages, borrowing from short-lived pidgin varieties, and reintroduction of features of the lexifier.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

analysis of semicreole formation, like that of Parkvall, is demographic.[115] It shows that the *proportion* of native speakers and nonnative speakers in the environment of language contact is critical in the formation of a partially restructured variety of a language. His compilation of population data from Dutch, English, Spanish, and Portuguese colonies of the eighteenth century – with the rapid increase of Western European maritime colony formation – reveals a correlation of the proportion of the population of Europeans among colonized peoples to the degree of restructuring in the subsequent language of the population. Specifically, where European colonists made up only about 6–8 percent of the population of the colonies, and the rest of the population were locals and imports from elsewhere, a radically restructured creole of the European language developed (such as Negerhollands Creole Dutch, Jamaican Creole English, and Papiamentu Creole Spanish). These creoles bear only superficial resemblance to their lexifier, being essentially entirely new languages. They have all been called creoles initially on the basis of the intuition that they represent similar phenomena. But Holm found that where the Europeans made up 32–59 percent of the population of their colonies, *partially* restructured varieties of the language were the outcome, and he found no evidence for a pervasive pidgin stage (Afrikaans, African American Vernacular English, Brazilian Vernacular Portuguese, and Nonstandard Caribbean Spanish).[116] These are Holm's *semicreoles*. Their morphologies are noticeably reduced, although they retain some older inflectional patterns, and they often show some kind of "hybridity" in the sense that features have been imported from other languages used in the environment of their formation that are not found in the lexifier. Table 2.1 reproduces exactly the contents of Holm's chart on this point.[117]

The first three rows show the proportion of "whites" in these colonial settings where creoles developed. The rest are sites of semicreole formation.

Independently, Trudgill estimates that a proportion of 50 percent nonnative speakers visiting a site would reach "a threshold level at which some aspects of the nonnative variety" of a language "could transfer to the native."[118] The reason for the difference in outcome is not difficult to understand. The children of speakers of a pidgin who have very little access to native speakers of the language on which the pidgin is based will elaborate that pidgin into a creole with minimal exposure to the unrestructured variety of the language as spoken by a native speaker. They will perpetuate the pidgin speech as a native tongue without importing features

[115] Parkvall 2000. [116] Holm 2004: 70–71; 135–137. [117] Holm 2004: 71.
[118] Trudgill 2011: 57–58.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Table 2.1 *"Estimated proportion of whites in various societies in the late eighteenth century" (from Holm 2004)*

Colony	Developing language	Percent
Virgin Islands	Negerhollands Creole Dutch	c. 6
Jamaica	Jamaican Creole English	c. 8
(Rural) Curaçao	Papiamentu Creole Spanish/Portuguese	c. 7
Virginia	African American English	c. 59
Brazil	Brazilian Vernacular Portuguese	c. 32
Cape Colony	Afrikaans	c. 47
Cuba		c. 56
Puerto Rico	Nonstandard Caribbean Spanish	c. 45
Santo Domingo		c. 34

from the lexifier, because the lexifier is not regularly available to them and therefore unfamiliar. In situations where proportionately more native speakers of the target language are available, and in which contact with them is to that extent more regular or intimate, they will have more exposure to nonrestructured speech models in direct transmission, and they will be that much more habituated to the grammatical usages and morphology of the older native speakers. This finding is strongly supported by other, independent research by Bentz and Winter, demonstrating a striking inverse correlation between the number of noun cases in languages of the world today and the proportion of nonnative speakers of those languages. That is, the more adult foreigners learn a given language, the fewer noun cases it is likely to have, an extraordinary finding based on a very large sample of the world's languages.[119] In short, the proportion of native speakers to nonnative speakers seems to be a decisive factor in the degree of restructuring of a language. This happens to cohere with the long-standing theoretical model of pidgin and creole formation, which posits access to native speakers, or the lack thereof, as critical.

The clearest, and perhaps paradigmatic, example of a language regarded as a semicreole is Afrikaans, the divergent form of Dutch that emerged in the polyethnic setting of colonial South Africa.[120] The genesis of Afrikaans offers many close parallels with that of Middle Persian, as I will show. Dutch colonists in Capetown, which they founded in 1652, created an environment in which Dutch was the main medium of communication

[119] Bentz and Winter 2013.
[120] Holm 1988: 2.338–352; Thomason and Kaufman 1988: 254–256; de Kleine 1997; Holm 2000b: 27–29; Trudgill 2011: 68 ("creoloid").

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

but in which a large proportion – as many as two-thirds – of the population spoke other languages natively. As Trudgill explains,[121]

> Dutch was acquired by very large numbers of nonnative speakers, initially speakers of the indigenous Khoi and San languages, and then by speakers of Bantu languages, and immigrant speakers of German, French, the Scandinavian languages, Portuguese, Malagasy, Malay, Indonesian languages, and others. There was much racial mixing and intermarriage, as the current "Coloured" population of South Africa shows. Wives in particular were quite likely to be native speakers of some language other than Dutch. This led to a situation where the nature of transmission became crucial. So many people initially acquired and used Dutch as an L2 [nonnative second language] that, in later generations, children were born who became, as it were, native speakers of nonnative Dutch. A nonnative variety became nativized and in time the dominant, and subsequently only, variety.

The result was a new, divergent species of Dutch with reduced grammatical paradigms, no grammatical gender except in pronouns, and plenty of words adopted from Malay, Khoi, French, German, and other sources.[122] Languages partially restructured in a local setting may eventually spread far and wide. Afrikaans began as a rapidly restructured variety of Dutch in a setting with only hundreds of speakers in southern Africa in the late seventeenth century. From that beginning in 1652, there was a labor shortage relative to the needs and aspirations of the colony. Workers were imported continuously from Europe, Africa, and Southeast Asia. Adult locals began to learn the Dutch of the founders who ran the colony, but they learned it as adults with limited exposure to the language. One generation later, in 1685, "a colonial official reported that the settlers' children were picking up the broken Dutch of the non-whites."[123] The children of Dutch-speaking immigrants were thus learning a different, restructured form of Dutch within a single generation. This was the beginning of the formation of Afrikaans. The "non-whites" just mentioned included slaves, servants, and the nurses of the "white" children, who were of different origins (local and imported) and who learned Dutch partly from other nonnative adult learners. At this point the settlement population still amounted to no more than several hundred, including large

[121] Trudgill 2011: 58–60.
[122] As Raidt (1995: 133) describes it, "The main changes that led to the genesis of Afrikaans occurred in the noun and verb systems as a result of morphological reduction, simplification, and regularisation."
[123] Holm 2004: 42, citing Valkhoff.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

proportions of both Africans and non-Dutch Europeans, especially French Huguenots, who also had to learn Dutch as adults. The Dutch settlers were mostly men, so they often took mates who did not know Dutch from childhood. The non-Dutch mothers of their children passed on a nonnative adult learner's Dutch language to their children. Archival research by Edith Raidt has demonstrated that women played a pivotal role in the formation of Afrikaans.[124] By 1700, in just two generations, many of the features that distinguish Afrikaans from Dutch language were already established, such as the elimination of case from the pronominal system and the collapse of the present and perfect tenses.[125] By 1800, the grammar of Afrikaans was essentially what it is today and its use had spread hundreds of kilometers beyond the Capetown colony.[126] More than two centuries later still, in the wake of a long-lived political regime in which "white" speakers of Afrikaans enjoyed a disproportionate share of power and wealth in South Africa, there are more than 6 million native speakers of Afrikaans, in a form of Dutch exhibiting many of the changes already present by 1700, during the lifetime of the founders of Capetown. Centuries later, the language still bears the characteristics it acquired during the demographic upheaval of colony formation long before.

Afrikaans is not a creole, because it did not originate directly from a pidgin, as far as the evidence shows, although speakers of a pidgin variety of Dutch probably played a role in its formation. It is also more complex in its morphology than many creole languages, although languages already classed as creoles do exhibit varying degrees of morphological complexity.[127] Afrikaans does, however, resemble creoles in its simplification relative to its lexifier and the social circumstances in which it diverged from Dutch. It represents a class of languages that have become distinctly simpler through mass adult acquisition but not through the extreme reformulation and novelty of a pidgin. It is, in short, a semicreole, bearing this useful name for a type of "contact language."

Semicreole is not the prettiest term in the scholarly lexicon. It does require some explanation. I have found that newcomers to the topic understand the phenomenon of simplification of contextual inflectional morphology most quickly, however, if they learn about creole languages first. This is one reason that I have retained the term semicreole for the

[124] Raidt 1993 and 1995. [125] Holm 2004: 42–43.
[126] Ponelis (1993: 1–67) provides detailed information on demography and language in the development of Afrikaans. See also Combrink 1978: 71, 88n14.
[127] Holm and Patrick (2007) present a volume of summaries of creole syntax, which illustrates variation in complexity.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

present analysis. It would be strange to ignore the phenomenon of the creole in explaining what happens to languages that are radically shorn of inflection. The study of creole languages led the way to this kind of analysis, and this term follows from that.

State and Process in Contact Linguistics and Linguistic History

Calling a language creole or semicreole refers to its present state as an outcome of a prior social process.[128] We can call the corresponding historically prior processes *creolization* and *semicreolization*, respectively. Linguists usually address creolization through analytical and descriptive linguistic approaches rather than through social history because their general goal is understanding language as such, not specific events. For linguistic history, by contrast, the process, as an event or series of events, matters more than the resulting state, for history narrates events. To say that history entails diachronic analysis is practically tautological, but the ramifications of this distinction are easy to overlook when we apply the model. The different goals of these research enterprises – the nature of language, on the one hand, for linguists, and the determination and understanding of events of the past on the other, for historians – generate a methodological discrepancy between the two when it comes to studying so-called contact languages as historical subjects. The terms we use matter here. Those concerned with modern languages of recent origin may be satisfied with nearly synchronic categories describing states such as creole and semicreole. Linguistic historians, however, particularly when concerned with premodern languages, will typically be concerned with languages over longer durations. In these cases, the names designating different kinds of contact languages as belonging to a state become much less useful. Designations like "semicreole" imply that specific processes of striking socially induced change had already occurred in a language prior to the time addressed in a language, but the conditions for such processes do not persist forever, and rarely – one would expect – for more than several generations. It is probably adequate to say that a creole five or six generations old is still a creole, especially relative to its extant lexifier, but for how long is it meaningful for an ancient language that has undergone creolization to be known as a creole? I do not imagine that five hundred years from

[128] Cf. van Coetsem 2000: 185: "Indeed, the specificity of the creole resides primarily not in its nature, but in its genesis." This is unlike a pidgin, which lasts only as long as it mostly has nonnative speakers.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

now it will be useful to describe any future language descendant of Tok Pisin as a creole language, although the creolization in its past will have been one of the most extraordinary determinant events in the history of that language. A language can undergo semicreolization once, itself a process requiring more than one generation of language acquisition, leaving the symptoms of inflectional reduction for all subsequent generations of that language's speakers, long after the social factors of population contact and nonnative language acquisition that made it a semicreole have ceased. But for how many centuries thereafter is such a language to retain the designation "semicreole" after undergoing semicreolization? There is no clear answer to this question because the terms describe outcome states rather than events.

In short, the linguistic historian must beware of using basically synchronic categories for diachronic narrative purposes because the languages will often outlive the meaning of the states designated by those terms. Pidgin, creole, semicreole: these are all terms descriptive of a state of a language consequent to demographic upheavals in a relatively recent past that initiated grammatical reduction. That is, they all refer implicitly to prior intergenerational processes of language change connected with large-scale nonnative language acquisition. Synchronic terms such as these are not useful to describe outcomes the effects of which persist for ten or twenty or forty generations, when a language's later speakers are mostly very remote from the demographic upheavals that shaped their language, despite the clear permanent effects of those past events on their language. The linguistic historian cannot insist, therefore, on these terms as permanent designations of a language that has undergone such contact-induced transformations in the distant past. It is not meaningful to say that a creole remains a creole for a thousand years, even if it remains relatively unadorned with inflectional morphology for that entire time. It is therefore better for the linguistic historian to speak of languages *undergoing the process* of creolization or semicreolization – or other such processes that may be identified – rather than *being* creoles or semicreoles. Afrikaans can now be said just to be a language very closely related to Dutch, or a variety of Dutch, rather than a semicreole of Dutch. Although it was formed through semicreolization, that formative process was long ago.

There is another discrepancy between the methods of contact linguistics and historical approaches to language. Specialists in contact languages have largely ignored the traditional methods of comparative historical linguistics. This seems to be in part because historical linguistics assumes a genealogical framework that creole specialists have struggled to accept,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

given their different theories about creole-formation. When creolists have emphasized that creoles are "hybrid languages" with grammatical, and not just lexical, genealogies in more than one language, this would appear to pose a challenge to the established methods of historical linguistics, which assume that each language has one ancestor.[129] A related problem is that language genealogies discovered by the methods of historical linguistics depend on shared innovations rather than losses of features, but pidgins, creoles, and semicreoles are most conspicuous in the loss of features – more precisely, by the nonreplication of the contextual inflection – of their lexifiers. For these reasons, historical linguistics does not and perhaps cannot account well for those languages that are mostly the vehicles of nonnative speakers,[130] although the attempt should be made. This helps to explain the reason for the gap between historical linguistics and creole studies. The gap may, however, be filled partly by linguistic history, paying special attention to population history.

Reduction and the Transfer of Features as Concomitant Factors

Now it is necessary to introduce the second part of the linguistic-historical model, which is not a matter of intergenerational patterns of language acquisition, but rather of multilingual usage. Other kinds of changes appear when more than one language is used in the same circumstances by the same people. Understanding the processes by which features are transferred from one language to another requires attention both to the multilingual agent, who enacts this transfer, and to the larger social context of such multilinguals. Linguistic transfer, often construed as language mixture, is another undeniably real kind of socially conditioned language change beyond the sorts of change addressed descriptively by the traditional comparative method of historical linguistics, strictly speaking. Historical linguists have used these obvious kinds of features in their diagnoses of language history since the nineteenth century, but they have not incorporated more recent research findings on the transfer of features between languages.

As explained, Holm employed the category of semicreole to account for languages that have been restructured to some degree in ways resembling

[129] Already by 1921 the historical linguist Meillet responded to the early creolist Schuchard over this question. See Meillet (1921: 76–109) in his essays "La problème de la parenté des langues" and "Les parentés de langues." Still recently, Blasi, Michaelis, and Haspelmath (2017: 723), for example, hold that "creoles have more than one language in their ancestry."

[130] Cf. Kranich and Breban 2021.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Table 2.2 *McWhorter's grid of nine types, with some of his examples*[131]

	Mixture: lexicon only	Mixture: lexicon and syntax	Mixture: lexicon, syntax, phonology, morphology
No simplification	Loanwords in any language	Romanian in Balkan linguistic area	Media Lengua
Moderate simplification	NCSL: English, Persian	Semicreole: Afrikaans	Shaba Swahili
Extreme simplification: creoles	Hawai'ian Creole English	Most creoles	"Deep creoles": Tok Pisin

creolization but without qualifying as a creole. This posits a taxonomy of distinctions in three relative degrees: *unrestructured*: *semicreole*: *creole*. McWhorter's treatment of the semicreole goes further than Holm's in giving close attention to this dimension of mixture, the degree to which a language incorporates features from other languages. To achieve a finer typology, he crosses these three qualitative degrees of grammatical reduction, or simplification – none, some, much – with three degrees of mixture of features from other languages – lexicon only, lexicon and syntax, or all subsystems including phonology and morphology – to create a grid of nine types.[132] He then populates these nine proposed types with representative examples of languages familiar to researchers in language contact. A bilingual mixed language might show no simplification but a maximum of "hybridity," whereas a "creole" like Hawai'ian Creole English is extensively simplified but minimally mixed. Other creole specialists appear not to have taken up the challenge of testing it or even addressing this typology. In what follows I discuss it and, ultimately, reject it as it is, while agreeing with the importance of the issues it exemplifies.

Among these nine, reproduced in Table 2.2, McWhorter proposes a new category he calls NCSL, or "Nonhybrid Conventionalized Second Language." For him, this is a close cousin of the semicreole, and the NCSL is especially relevant to the present investigation of Persian because New Persian is one of McWhorter's five main examples of his NCSL. Although his study of Persian takes into consideration changes in prior stages of Persian (Old and Middle) somewhat, the data he uses are

[131] McWhorter 2007: 254. [132] McWhorter 2007: 254.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

mostly synchronic and modern, comparing the grammar of modern New Persian with that of its closest living modern Iranic relatives to further the case that NCSLs exist.[133] (The NCSLs include some of the world's most-used languages, such as modern English and Malay.) In this taxonomy, what distinguishes an NCSL from a semicreole like Afrikaans is its "Nonhybrid" characteristic. That is, the NCSL was a language acquired nonnatively (as a second language) by many people at once and became a conventional language, and so it is equivalent to a semicreole in its extensive, but not extreme, grammatical reduction, but it is also different in that it exhibits hybridity of lexicon only, not hybridity of phonology, morphology, or syntax. By contrast, McWhorter's semicreole category exhibits hybridity in its syntax as well as lexicon. (By hybrid, the sense here is just that features are transferred from one language to another, not something more essential or profound.)

It is already a challenge to use a tripartite division of languages into creole, semicreole, and unrestructured or "normal." Three degrees of simplification call for a qualitative assessment relative to a prior state differing in each case, making the model fundamentally scalar but its application comparative and therefore slippery and less analytical.[134] To use the ninefold typology is more difficult still, because one must assign a language to one of the categories based on rough comparison with other languages while taking the transfer of features from one or more languages into the new language into account. "Hybridity" or mixture is also hard to measure, but for different reasons. Identifying genuine instances of the transfer of features between languages in context, as opposed to "internal" grammatical change, is not a trivial matter, though methods for doing so have been suggested.[135] Another potential problem here is that the greater the degree of simplification from the lexifier, the more difficult it may be to determine mixture, because greater reduction implies fewer distinct features, which are one basis for identifying mixture. Moreover, degrees of reduction are relative, as already emphasized. Languages that were already simple in morphology before restructuring are not likely to show striking degrees of morphological simplification when they are restructured: there is not as much to lose. The contribution of McWhorter's taxonomy of "contact languages" is that it acknowledges and insists on the existence of degrees of restructuring, formalizes them

[133] McWhorter 2007: 266: "the placement of languages in the chart is based on their synchronic character."

[134] Kihm 2012: 658. [135] Heine and Kuteva 2005: 21–34; Bowern 2013: 424–426.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

for discussion and analysis, and separates the issue of mixture from reduction of grammatical features while requiring that it enter the discussion

Human Agency in the Transfer of Features and Grammatical Reduction: Three Syndromes

Positing degrees of mixture crossed with degrees of morphological reduction demands that the linguistic model I am using here incorporate findings from other research avenues in contact linguistics and the related field of research on bilingualism. Other researchers have demonstrated languages are not mixed just at random when they "come into contact," but that the transfer of features from one language is specifically conditioned, just as morphological reduction in a language depends on specific variables in the circumstances of social contact and age of language acquisition. The analysis of mixture – more precisely the transfer of features from one language to another – is more complex and entails factors that undermine the utility of the optimistic nine-part grid just examined.

Bilingual or multilingual individuals are the agents of transfer of features from one language to another,[136] but multilingualism is of different varieties in different individuals and operates differently in different settings. One decisive factor is variation in the individual's proficiency in different languages, in turn connected with many other factors such as the individual's age of acquisition, frequency of use, and social domains of use of different languages. A second factor is variation in distribution of the languages spoken by different members of a society. A third is variation in the relative extent and intensity of use of different languages in a society at large. The kinds of crosslinguistic transfer that occur depend on variation in these factors and probably others too. Paying close attention to these complex distinctions can, however, help historians to diagnose aspects of the social history of a language's speakers. As I will explain, the agency of speakers of different languages can sometimes be discerned through the directions of transfer of features from language to language. This tells historians more about the history of the speakers than the mere observation that features were transferred from one language to another as a consequence of population contact. It can also diagnose different kinds of

[136] Ringe and Eska (2013: 60–63) consider the limited agency of monolinguals in the transfer of features from another language into their own.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

social contact and relationships between speakers of different languages. The transfer of features from one language to another can thus tell us much about the social circumstances of language change. In what follows here, I employ Frans Van Coetsem's (1919–2002) insights into the transfer of features between languages, which focus on the agency of individual speakers. These are supported generally by the research of others, and I build on them further.[137]

One of the critical factors here is called "linguistic dominance."[138] This refers not to social dominance of a language or of its speakers over others but to the language that is the psycholinguistic default for a user, usually the one in which the user is most proficient in a domain of language use. This is the speaker's "dominant language."[139] Although there are apparently exceptions for those who grow up from early childhood with regular exposure to more than one language and become equally fluent in two or even more languages from childhood, most multilingual people have one dominant language, usually the common language of the home early in life, with variable degrees of proficiency in one or more other languages.[140] They may have varying proficiency in different languages according to different, specific domains or contexts of use, such as when one language is used with family and another at work outside of the home.

[137] Van Coetsem 1988; 1995; 2000. Van Coetsem created an idiosyncratic terminology and adopted an opaque style of presentation to explain his insights. I have avoided using some of his terms like "SL agentivity" to reduce obfuscation especially for the sake of nonspecialists. Many scholars in contact linguistics have ignored or overlooked his work; his obscure style of presentation may have contributed to that. Van Coetsem's presentation has, however, been clarified by Winford (2005; 2013: 170–172) and Butts (2016: 13–20, also discussing important antecedents of Van Coetsem's model). The latter applies it with important results to the relationship between Syriac and Greek. Henkelman (2011: 588–595) and van Bladel (2021, addressing Henkelman's application) use Van Coetsem's model to consider the history of contact between Elamite and Persian.

[138] Van Coetsem 1995: 70: "A bilingual speaker . . . is linguistically dominant in the language in which he is most proficient and most fluent (which is not necessarily his first or native language)." Cf. Winford 2005: 376–377.

[139] Cf. Myers-Scotton (2006: 295–296), who emphasizes that "bilinguals rarely have equal facility with both languages."

[140] Petersen 1988; Van Coetsem 1995: 70–71 and 2000: 32, 52; Winford 2005: 376–377; Myers-Scotton 2006: 3, 295. Van Coetsem (1995: 72; 2000: 93–94) allows the possibility of "neutralization" of linguistic dominance, in which the speaker appears to be dominant in more than one language, or at least dominance cannot be discerned because of near-equivalent proficiency in more than one language. Van Coetsem does not discuss differences between early and late bilingualism (acquiring both languages before the age of seven or afterwards), or, differently, between simultaneous bilingualism, when a child learns more than one language from infancy, and successive or sequential bilingualism, when a child is exposed to a second language from a young age, but after already gaining proficiency in a prior, first language; such differences may matter (Myers-Scotton 2006: 291, 328–329).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

The way in which a bilingual person transfers features from one language to another correlates to the speaker's linguistic dominance in the sense just described. Van Coetsem identified different syndromes (as I shall call them) of transfer rooted in the agency of the multilingual individuals who are the living media of such transfer. The two to which he gives the most attention are *adoption* and *imposition*. Speakers of more than one language freely *adopt* (or borrow) so-called loanwords into their dominant language from other languages they speak, which they incorporate into the grammatical matrix of their dominant language, but they can seldom impose and propagate words from their dominant language into another language used by a community that does not need or accept those foreign words.[141] They likewise seldom transfer grammatical features such as word inflection into their dominant language from another language, because they are completely fluent in their dominant language, and so are not motivated, by themselves, to borrow foreign inflectional patterns (though this can happen in some specific social situations, to be discussed). By contrast, speakers of more than one language are prone to *impose* features of the pronunciation (phonology) and usage (such as syntax) of their dominant language onto another language with which they are less proficient, or which they are just learning. They may even reduce unfamiliar grammatical features, especially in inflectional morphology, of their nondominant language(s), in which, by definition, they are less proficient. Their lower proficiency, be it great or slight, is the obvious reason behind imposition; they make up for their lack of proficiency in the one language with patterns imposed from their default, dominant language.[142] In summary, if we pay close attention to the agency of the multilingual individual, we see that features from different modules of a language are transferred differently according to the speaker's language dominance. The typical pattern of transfer is that bilinguals *adopt* "loanwords" from their second language into their dominant language (L2→L1 lexical transfer) and they tend to *impose* features of phonology and syntax from their dominant language onto their nondominant language (L1→L2 phonological and syntactic transfer), which appears as imperfect usage from the point of view of its highly fluent speakers.[143]

[141] For this reason, the expression adopted word or borrowed word is preferable to the common term loanword.

[142] Winford 2003: 217–218.

[143] I do not discuss here the concept of the "scale of adoptability" (Haugen 1950) or "stability gradient" (Van Coetsem 1988: 25–45; Van Coetsem 2000: 58–62; Winford 2005: 337; Butts 2016: 18–19), with its antecedents in Whitney 1881, as the matter is explicable without this concept.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Given these two syndromes, the motive for simplification or reduction of the morphology of a language by adult learners, discussed earlier in this chapter, becomes clearer, and a third syndrome emerges. A nonnative speaker who does not know inflectional forms of the target language cannot use them effectively; one has no motive to impose morphology from one's dominant language (L1) onto the target language (L2) because it would hinder communication to use inflectional forms unknown to the L2-speaking audience. Foreign morphology will not be understood by native speakers of the L2.[144] For example, a Russian speaking German to German-speakers would only cause confusion by adding Russian noun case endings to German nouns. Nonnative speakers cannot avoid phonology or syntax – speaking words requires phonology, and using more than one word entails word order and hence syntax – but they can try to communicate without the appropriate inflections of words, *omitting* inflectional morphology when it is not familiar. This can be regarded as a species of imposition not because it must occur in the company of imposition, but because its agent is the speaker whose dominant language is a different one – typically, the nonnative speaker. Nevertheless, this pattern is distinct from other kinds of imposition, because it entails the omission of features in L2 rather than the exportation of features from L1.[145] It often does not involve the transfer of clearly identifiable features from their dominant language.[146] In short, not two but three different kinds of linguistic change result from the agency of the bilingual speaker for whom, in most cases, one language is the dominant language. Put in less precise terms for the sake of concision, hybridity of lexicon occurs largely through *adoption*: taking features, usually words, from a second language into one's dominant language. That is, adoption is enacted by native speakers of the borrowing language (L1). Typically, words adopted from another language will be

[144] Muysken (2013: 207) summarizes earlier research: "In L2 learning, there is no transfer of L1 morphology. L1 affixes never appear on L2 lexical items in interlanguage, and a rare L1 lexical item may carry L1 affixes, but then they tend to form part of an unanalyzed whole." That said, Muysken (2013: 200) also notes that in morphological imposition ("insertion" in his usage) in code-mixing, "the most common case involves nominal plural elements." Not surprisingly, this example is a matter of inherent not contextual inflection.

[145] Foreigner talk is an exception to this pattern, whereby the L1-speaker omits features to be accessible to someone who does not know those features.

[146] Van Coetsem (1995: 76): "reduction is a clear sign of SL agentivity [i.e., imposition]." See Van Coetsem 1988; 1995; 2000; Johanson 2000: 165–166; Winford 2005. This model, based on considerations of the transfer of linguistic features from the point of view of contact linguistics, is supported by research on code-switching among bilinguals, specifically the Dominant Language Hypothesis and the Matrix Language Frame theories. See Petersen 1988 and Myers-Scotton 2006: 242–278.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

inflected according to the morphological rules of the new recipient language. Hybridity in phonology and syntax tend to occur by *imposition* of features from the speaker's dominant language (L1) onto the speaker's nondominant language (L2). Reduction of inflection, or *omission*, as I dub the third syndrome here, tends likewise to be due to the agency of nonnative speakers.[147] It is therefore not a matter of degree of mixture or hybridity along an axis involving lexicon > then lexicon and syntax > then lexicon, syntax, and phonology, as discussed just now when reflecting on McWhorter's grid of types. This conception of degrees of mixture does not work. Rather, varieties of transfer of features in language change differ due to a combination of the individual speakers' variable proficiency in languages, their agency in language use, and the subsystem of a language affected. Typically, it is simply omission that leads to the reduction of a language's morphology. Where the omission becomes conventional and pervasive, there appear the striking effects seen in the recognized pidgins, creoles, and semicreoles, with their reduced inflectional morphologies relative to their lexifiers.

There is yet another critical variable factor that goes beyond the individual bilingual person's agency. That is the distribution of bilingualism in the individual's society at large, which acts as a condition for the individual speaker's communicative needs. Where people communicating together are multilingual in the same languages at once, they can meaningfully transfer features between the shared languages as a tool of their communication. When they use features from two or more languages at once, this is known as code-switching or code-mixing.[148] Bilinguals who share the same languages have these expanded means of convenient intercommunication at their disposal. Another phenomenon that depends on the distribution of bilingualism in the society at large has been discussed: when a large group of speakers of various languages acquires a new common language as adults, they will show poor proficiency from the point of view of native speakers, and they may even generate a pidgin or semicreole, which may in turn include features imposed and propagated from one of the participants' languages.

Then there are common cases in which speakers dominant in one language all know a second language well, whereas speakers whose dominant language is the second language do not know the other one. These are

[147] Van Coetsem (2000: 181–182) calls this "reduction," but that strictly refers the outcome. I prefer to call the act itself "omission."

[148] Myers-Scotton 2006: 239–241; Gardner-Chloros 2010; Muysken 2013.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

known as situations of unilateral bilingualism (or "nonreciprocal bilingualism," "unidirectional bilingualism").[149] In this scenario, the monolinguals typically are or are felt to be socially superior in some way, or they speak a language so widely known that they have little need of another language. That is, native speakers of a high-status or high-utility language can expect others to learn their language, not the other way around – so the native speakers of the high-status language do not learn the other language. By contrast, the bilinguals in this scenario sometimes are or feel themselves to be socially inferior or marginal, or they belong to a distinct "minority" community with a language not known by outsiders. In such cases of unilateral bilingualism, borrowing may go far beyond the normal adoption of words from other languages but may involve the transfer of any grammatical feature of a language, even basic patterns of syntax and morphology, into their dominant language. For an example of this more extensive kind of borrowing (L2→L1 transfer), most speakers of Basque in Spain also speak Spanish, whereas those in Spain for whom Spanish is the dominant language seldom speak Basque at all. Basque-speakers may, and do, freely adopt all kinds of Spanish features into their Basque speech – not just adopted words but adapted grammatical constructions and whole phrases – and they can expect fellow Basque-speakers to understand patterns imported from Spanish, because they speak Spanish, too. But Spanish-speakers will not import Basque materials in the same way because generally they do not know the Basque language.

These tendencies in the transfer of features from one language to another, which are rooted in the individual's language dominance and personal agency but are conditioned by the same factors operating on a larger, social scale, turn linguistic mixture and the transfer of features into an important diagnostic tool for linguistic history. The kind of mixture – the result of linguistic transfer – may help to identify the primary agents in the formation of a new contact language. The three positions on McWhorter's hybridity spectrum are "lexicon only," "lexicon and syntax," and "lexicon, syntax, morphology, (and) phonology." Van Coetsem's model calls into question the distinction of the last two types on this spectrum by showing that there is more than one process giving rise to hybridity and these processes affect different modules of a language. For example, imposition of one's dominant-language phonology will often not be accompanied by hybridity of syntax and morphology; it is just a foreign

[149] Hock 1991: 492; Winford 2003: 64, "one-way bilingualism"; Myers-Scotton 2006: 48–49; Matras 2009: 57–60. Cf. Heine and Kuteva 2005: 237–238; van Bladel 2021: 458.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

accent. Studies of the transfer of features between languages show that the language dominance of participating speakers plays a decisive role in deciding which kind of transfer of features will arise in each language. Reduction in inflectional morphology arises with omission in nonnative use, and this will become conventional in episodes of large-scale adult acquisition of a language. The proportion of nonnative speakers in that episode matters, with the outcomes of pidgin/creole, semicreole, or less pronounced changes. Different kinds of mixture still offer a useful index for explicating a language's social history because it sheds light on human agency. In the next chapter of this book, I will apply both criteria – reduction of inflectional morphology and outcomes of transfer of features between languages – to the analysis of the history of ancient Persian and its speakers. Nevertheless, I will use Holm's concept of semicreole, which does not link that term to hybridity of lexicon and syntax alone, rather than that of McWhorter's, which does. For my analysis, semicreoles constitute a class of languages that have become distinctly simpler in inflectional morphology through mass adult acquisition but not solely from a pidgin origin, regardless of degrees or types of mixture. As emphasized, I think it is more appropriate to discuss language changes as processes, and not to insist on modern terms describing resultant states. Therefore, it is more important for linguistic history to identify episodes of semicreolization than to label a language synchronically as a semicreole, even if the process and the concept are almost identical.

Decreolization

Creole specialists have long recognized linguistic effects due to the ongoing contact of speakers of creoles with speakers of their language's lexifier. Sometimes native speakers of a creole also learn their language's lexifier with high or partial fluency. For example, colonial governments sometimes have instituted language policies to replace creole varieties of the colonial government's language by formal education in the lexifier. The goal is to give access to opportunities through the prestigious and internationally used high-status languages of literary use; the consequence is demoting and reducing the use of the creole and, typically, stigmatizing it. Such efforts have had mixed results due to many factors. The important point here is that bilingual speakers of a creole and its lexifier can and do transfer features between the lexifier and the creole that diverged from it previously, fostering a new mixture of lexifier and already constituted creole. The creole acquires some of the lexifier's characteristics and converges more

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

with it, adding, for example, some of the contextual inflection shorn during pidginization. Creole specialists call this process "decreolization." In the jargon of creole studies, the product of decreolization is called an "acrolect" of the "basilect" creole.[150] As is evident, this phenomenon makes the analysis of the history of creolization muddier. The process seems to reverse itself, but it is due to specific circumstances.

Terms like acrolect and basilect are descriptive and do not account for the agency of the bilingual speaker. That agency is not fixed or implied by these names. Linguistic decreolization can be due to the agency of a speaker whose dominant language is the lexifier, in which case the speaker imposes features of the unrestructured lexifier on the creole. Decreolization can also be due to the agency of a speaker whose dominant language is the creole, who turns to the lexifier to import words for novel goods or technical terminology that did not exist in the prior pidgin, or who adopts features of the lexifier to attract the social status that goes with it. In either of these ways, the morphology of the lexifier can be adopted or imposed in the creole. Because the absence of inherited contextual inflection is a criterion for identifying creoles, restoration of such inflection by contact with the lexifier would tend to obscure the pidgin or creole origin of a language.

The theory of decreolization should apply to semicreoles as well as creoles. These possibilities must be considered in the application of this model to historical instances.

The Pace of Language Change and "Modern" Ancient Languages: A Note for Historical Linguistics

One of the ramifications of the entire foregoing discussion is counterintuitive. When drastic demographic changes occur, languages may change rapidly, too, with the attrition, loss, or new emergence of grammatical structures, phonology, vocabulary, and other features evident within just a few generations, as few as two or three.[151] A new variety of the language, as acquired imperfectly by adults and then natively from them by the children among them, can become notably reduced or changed as compared with that of those children's grandparents and great-grandparents. What is counterintuitive is the speed of the change. Historical linguists have typically fabricated their own rough, ad hoc estimates for the time between

[150] Romaine 1988: 158–161; Winford 2003: 314, 354; Hock 2021: 744–745.

[151] Holm (2000a: 3) cites the early creole researcher Reinecke (1937: 6), who stated that "Languages can be observed taking form within a man's lifetime." See Baker 1995 for examples of this chronology, and Daval-Markussen and Bakker 2017: 133–136 for a summary discussion of this issue.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

poorly documented or prehistoric stages of a language, but without any evidence.[152] They often guess at periods of several centuries to account for changes in grammatical structure on the assumption that language change is a matter of a very gradual trickle of cumulative effects, like the slow accretion of stalactites in a cave in which water seeps through limestone. When literary standard languages, acquired through schooling, predominate in the record for centuries, the normal assumption is that the vernacular, spoken varieties change gradually in the undocumented background. Rather, it is now clear that abrupt population contact and the ensuing social relations are the primary determinant of language change toward morphological simplicity. The term "abrupt" matters because it refers to the sudden need of adults to learn the language, not necessarily to long-term change over many generations. Language learning by children, in a second generation of population contact, does not lead to extreme morphological simplification. The ongoing entry of foreigners into a society, over more than one generation, however, should entrench the effects of morphological simplification, through a condition of continuously renewed abrupt contact by people who need to learn the new language as adults. While all languages do change slowly and gradually, languages can also change drastically and rapidly in the midst of rapid and drastic population change.[153] Upheavals in populations can act on the elaborate inflections of a language as a hatchet may carve down or truncate an ornate stalactite that has grown slowly by dribbles over centuries or millennia. The signs of those upheavals are permanent, although the reduced stalactite may grow new structures over subsequent centuries and millennia. Instances of languages being restructured radically within three generations of colonial contact are well documented, the most typical being the languages dubbed pidgin and creole that have arisen in recent centuries, for which documentation is sometimes sufficiently ample to illustrate their geneseis. This fact challenges the estimates of some traditional historical linguists that have posited stages of language development over long prehistoric centuries. We should expect that centuries in social

[152] This resembles the neglect of population histories by specialists in creoles and contact linguistics. Cf. the critical remark by Yakubovich (2008: 205) about Indo-Europeanists who "use . . . intuitions or makeshift theories as their main guides for drawing sociolinguistic implications from the available data" when many studies of language contact are available to provide a framework for understanding.

[153] Trudgill 2011: 2–13. Already Van Name (1869–1870: 123) remarked that in the case of creole languages, "two or three generations have sufficed for a complete transformation" from their lexifier, due to the "greater violence of the forces at work," as distinct from the slow gradual change found in more familiar cases.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

isolation will contribute to the intergenerational replication of a language's archaic features, but that reduction takes place in abrupt episodes.

It is worth remarking, by way of an example not irrelevant to the ancient history of Persian, that this applies to the current estimates for the date of the ancient, orally transmitted Zoroastrian scriptures known as the Avesta, composed in the ancient Iranic language called Avestan, a close relative of Old Persian. Specialists have regularly posited very ancient dates for the composition of the Avestan corpus, in its two linguistic strata called Young and Old, pushing it back even as far as the early or middle second millennium BCE, a thousand years before the Bisitun inscription or more.[154] The argument is nowhere explicit, but the idea seems to be that such an "archaic" kind of Indo-European language must be very ancient on the assumption that significant language change requires long periods. The research on language change in population contact presented here suggests rather that the complexities of Avestan grammatical inflection merely reflect the speech of a community joined by very few outsiders since time of their ancestors' separation from the speakers of Old Indic, whenever that separation occurred. The complexity of Avestan morphology and its similarity to that of the reconstructed Indo-European ancestor *by itself* tells us nothing about its age. Perhaps the Old Avestan Gathas were composed as long ago as 1500 BCE, as some specialists have conjectured; perhaps they were even composed as late as the time of the conqueror Cyrus II of Anshan, as others have supposed. If our only evidence is the language of the Avestan texts themselves, we are simply in the dark about their age without a theory about the history of the speakers of the Avestan language as a population. Other historical evidence, beyond the grammar of the language, is required to make such an estimate. In the case of Avestan, the quotation of Old Avestan texts within Young Avestan texts provides a clinching argument for chronological order that happens to match our expectations in the language's grammatical development. But seemingly archaic, grammatically complex languages can persist for centuries without reductions in complexity if foreigners never learn that language. Indeed, as already said, such languages may become gradually more complex, with innovations, while spoken in relative social isolation. In short, philologists are mostly guessing, almost at random, when they estimate the age of ancient texts solely on the basis of relative states of grammatical change and development, unless they know something about the prehistory of the population of a language's speakers or unless

[154] E.g., Boyce 1992: 45; Skjærvø 1994: 201–202; 1999b: 6–7; and 2009a: 46.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

something stated in the texts provides a chronological anchor.[155] We cannot be certain of such things without a population history, which is usually not possible for prehistoric societies. That said, the grammatical characteristics of a language may tell us about the prior population history of its speakers.

The foregoing synthesis dispels a second misconception sometimes appearing in the study of the ancient Iranic languages: that simpler languages belong to a "modern" type. In Chapter 4, we will encounter scholars (such as Meillet, Reichelt, and Paul) who have casually regarded simplified ancient Iranic languages like Middle Persian as appearing "modern." As Trudgill has indicated, however, scholars familiar with European languages that have been affected by simplification due to long-term high rates of adult second-language learning "have been tempted to regard" simplification in language over time "as a diachronic universal."[156] This gives the impression that "modern" languages – languages of societies participating vigorously in various kinds of modernity – are generally simpler. Simplification would seem thus to be a universal process into which every language would eventually slide among peoples who adopt "modernity." As is evident by now, this assumption entails numerous problems. Not least among them is the multivalence of the word "modern," which connotes different things in different areas of research. Strictly speaking, the combination of the principles just discovered means that the relative simplicity or complexity of a language cannot tell us how old a language is or how "modern," but rather what kind of social changes its speaker population has undergone.[157] That is, nevertheless, an extraordinary breakthrough for linguistic history.

Exoterogeny and Esoterogeny

In this chapter I have tried to define terms of analysis for future efforts in linguistic history. Most of these terms have been adopted from specialized niches of inquiry in linguistics for the use of historians. I have avoided

[155] Kellens (2000: 2) argues that "dating [the Avesta's] constituent parts can be attempted only by approximation, on the basis of linguistic arguments alone." We can be more specific, by saying that this means only the relative dating of the Avesta's constituent parts with respect to each other, not dating according to an established era.

[156] Trudgill 2010b: 313; see also 2012.

[157] It must be acknowledged, however, with Trudgill (2011: 185–189), that the conditions of mass migration with which we associate modernity, eroding the insularity of dense social networks, do correlate with the loss of linguistic complexity. Only in this restricted sense could the term "modern" be applied to this phenomenon; even then it is potentially misleading.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

proposing new jargon when so many terms are in use. There are two other terms for which I foresee some utility that I will introduce now, both coined by William G. Thurston in 1987 and used occasionally by other linguists since then.[158] Thurston dubbed languages without an appreciable number of nonnative learners *esoteric* languages, and those with many extracommunal users or nonnative learners *exoteric*. From these uses of those words, he created two neologisms: *esoterogeny* and *exoterogeny*. That is, changes in a language induced by native speakers who live in a relatively closed social network can be described as esoterogenic, in that they are enacted by the linguistic community's insiders, all of whom learned the language from childhood. Thurston's concern in creating these terms was with the deliberate efforts of speakers of a language in a closed community to differentiate their language from that of their neighbors.[159] By contrast, changes induced by those coming from outside of a linguistic community are "exoterogenic." These terms may help linguistic historians within certain limits that I will now discuss.

Earlier, I indicated that we lack a common term for languages that are transmitted "normally" between generations without the interruptions in perfect acquisition by nonnative-speaker outsiders. I proposed calling that process *intergenerationally stable transmission*. Esoterogeny, as coined by Thurston, unfortunately does not adequately describe such transmission; it describes rather the coming into being of new features induced by socially closed communities for their linguistic self-differentiation. Furthermore, in the humanities, to call such languages "esoteric" may give the odd impression that they are used for secret rituals or occult phenomena, because the word esoteric already carries such connotations. Nevertheless, I think the term exoterogeny may have some utility for linguistic history. I acknowledged that it is practically impossible to evaluate the degree of a language's restructuring toward morphological simplicity apart from the useful degrees of *creole*: *semicreole*: *unrestructured*. It may be difficult for linguistic historians to decide whether a language attested in history fits one of these categories because its genesis may be unattested. As Watts rightly warns, "the term 'creole' should not be bent to such an extent that it can be pressed into service to define all the results of all language contact situations."[160] We can use the term exoterogeny as an umbrella term to describe the process by which creoles, semicreoles, and

[158] Thurston 1987; McWhorter 2007: 11; Good 2023.
[159] Good (2023: 391–394) claims that this was deliberate in the case of the Cameroonian language Fang.
[160] Watts 2011: 95. Watts summarizes an argument by Görlach (1986) to argue that Middle English could not be a creole, yet the latter concludes by comparing Middle English with Afrikaans, the paradigmatic example of a semicreole.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

similar languages come into existence without having to specify whether the language underwent a pidgin stage or not. If a language has undergone a phase of exoterogeny, it means that it became distinct from its ancestor through nonnative use. Thurston defined the term thus: "Exoterogeny is the process of simplifying an esoteric language to create a register that is more easily acquired by outsiders."[161] This definition does not specify speakers' deliberate intention. We are dealing here with the same kind of process that gives rise to pidgins, creoles, and semicreoles, as Thurston indicated when he defined the term. Yet exoterogeny becomes imprecise when we consider that areal features (discussed later) may all be regarded as exoterogenic. Nevertheless, exoterogeny is a term that may bear some utility. For those squeamish about terms like semicreole – despite its clear meaning – "exoterogeny" may provide a refuge.

Ramifications

The synthesis of linguistic research presented here demonstrates strikingly that the linguistic description of a language at a single historical stage permits some inferences about the prior history of the population that spoke it. This is not entirely news for historical linguists, who have long mined ancient languages for linguistic borrowings such as "loanwords" signifying social contact of different kinds. The lexicon of adopted words alone, for example, can signify domains of social contact, as, for example, when terms used for trade, or government, or religion are the ones that are adopted. But the model synthesized here shows that still more can be discerned than mere contact. Even lexical borrowing can indicate more than just the domains of contact. Such phenomena may illustrate the social relationships of those in contact due to the role of different kinds of speaker-agency in the transfer of features. Moreover, when the inflectional morphology of a language appears to have been noticeably reduced in a short period, we should assume that some demographic upheaval has occurred among speakers of the language at some time in the past, and that contact-induced restructuring through mass adult acquisition is behind that. Even when a language is documented at one time only, with no record of its prior history, if its grammar is highly simple relative to its cognates, or even relative to most languages, we may suppose that some such event took place in its undocumented past.[162]

[161] Thurston 1987: 150.

[162] This is the main thesis of McWhorter in his work of 2007. McWhorter further suggests (2011a: 17–18), optimistically, that "Grammatical analyticity can be of use in reconstructing population movements for which there is otherwise only fragmentary evidence, or even in directing researchers as to where to look for signs of population movements as yet uninvestigated," and that conspicuous

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004

Fundamentally, it is necessary for linguistic historians to pay close attention to the age of acquisition of a language by historical subjects. This is a diachronic factor that varies according to specific social settings. It is the historian, not the linguist per se, that can perform this analysis best.

Language change due to historical social and demographic factors has fallen between several areas of scholarly investigation. The sound and highly effective methods of traditional historical linguistics do not take social factors much into account. Those methods, which are based on description, are not designed to account for such factors. Contact linguistics and creole studies have, by contrast, with several notable exceptions, lacked historians able to elucidate population histories convincingly in support of their hypotheses, leading to inconclusive debates. When combined, historical research and linguistic models will lead to new results useful to historians, as I will show in the next chapter.

simplicity in a language "is even a scientifically appropriate cause for supposing such rupture [in language transmission] even if other evidence proves unrecoverable."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:38:13, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009727709.004