The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/105702

Please be advised that this information was generated on 2020-10-24 and may be subject to change.
Creoles and the notion of simplicity in human languages
by Claude Hagège

1. Introduction
In the following comments, I will examine the notion of complexity (Section 2), the metric proposed by McWhorter to measure it (Section 3), the limitations of the comparison between creoles and older languages (Section 4), and the extent to which creoles may be said to be simple (Section 5). In conclusion, I will suggest a criterion for characterizing creoles as distinct from older languages (Section 6).

2. The notion of complexity
2.1. Complexity and Universal Grammar
According to McWhorter, a complex language is one which, if compared to a simpler one, contains more “overt signalling of [...] distinctions beyond communicative necessity” (Abstract). McWhorter’s purpose is not to examine how languages other than creoles differ among themselves with regard to these distinctions – although this would also be an interesting study. He simply says that creoles are on the lowest level of the complexity scale. Consequently, since, in McWhorter’s view, they are “unobscured by the results of millennia of [...] drift which make Universal Grammar such a challenge to glean in older languages” (Section 5), creoles are the most direct illustrations of Universal Grammar (UG).

If this notion involves universals of the Greenbergian kind, then it would mean that creoles contain more universal features than older languages. But McWhorter’s reference to works such as Seuren & Wekker (1986) indicates that what is referred to here is Universal Grammar defined in Chomsky (1981a, 1981b) as a system of principles some of which are rigidly fixed and hence
invariant while others are not fully fixed: they specify dimensions along which individual languages may differ. These dimensions, which represent a belated recognition of the importance of typology in the history of the Chomskyan paradigm (it was initially rejected: cf. Hagège 1981: 65, Note 26), are called parameters. Since the parameters define the range of crosslinguistic variation, language acquisition will consist of setting the parameters for UG.

What is McWhorter’s position on this issue?

2.2. Complexity and parameter setting

Given that UG is definitionally associated with unmarked parameter settings, it is not easy to see what exactly McWhorter means when he denies any similarity between the latter and his own hypothesis (Section 5.2) while at the same time maintaining that creoles most clearly reflect UG. Furthermore, McWhorter says (Section 1) that his “program dovetails with Bickerton’s Language Bio-program Hypothesis” – a framework which, as early as in 1981, considered creoles as manifestations of a core structure for human language. Thus, for McWhorter as well as for Bickerton, language complexity depends on the degree of distance from this core structure. But since McWhorter does not adopt the method of parameter settings even though the latter is implied by UG, we are confronted with the risk of running into circularity: on the one hand, UG is defined in terms of traits found in creoles, and on the other hand, creoles are declared to be closer to UG than other languages. However, there is a way out: despite the closeness of his and Bickerton’s positions, McWhorter proposes that his framework contains a novel feature inasmuch as it is more typological (Sections 1 and 5.3).

In order to measure the degrees of relative typological complexity of languages, one needs a metric. Let us examine McWhorter’s metric.

3. The metric of complexity

3.1. Number of marked phonemes

From among the four diagnostics of grammatical complexity that are proposed (Section 2.4.3), the least controversial seems to be the number of marked phonemes, defined as those encountered less frequently in the world’s languages. I agree that a phonemic inventory is more complex if it has more marked members.

3.2. The notion of rule

However, I do have doubts about the syntactic diagnostic: “a syntax is more complex than another to the extent that it requires the processing of more rules”. I fail to see what the empirical reality of a rule might be. Despite
the deep-rooted use of the notion of rule in linguistics dominant since 1965, the status of rules as experimentally demonstrable facts has not yet been substantiated (cf. Hagège 1981: 78–81).

3.3. Verbal constructions

As far as the overall complexity of a grammar is concerned, one may wonder what McWhorter means by “fine-grained semantic and/or pragmatic distinctions”. Let us examine the examples he provides as illustrations of the syntax of verbs. He says that Koasati yókpa ‘to love’, which requires an instrumental prefix st-, has a “semantically opaque instrumental government” and is therefore “akin to Germanic and Romance which also have rather arbitrary cases of verb government” such as dative-marked ich glaube ihm in German and accusative-marked je l’écoute in French. The reader has no choice but to conclude that these structures, strange though it may appear, seem “opaque” or “arbitrary” to McWhorter only because they differ from their English equivalents I believe him and I’m listening to him! McWhorter does not mention that, with this kind of reasoning, to in listen to might seem “arbitrary” to a native speaker of French! A native speaker of English would also find “arbitrariness” in Japanese verbs requiring a relator with their nominal complement: the postposition ni, e.g., Tadao-ni au ‘to meet Tadao’, Tadao-ni tsuku ‘to touch Tadao’ or ‘follow Tadao closely’ or ‘refer to Tadao’. I cite these as a small sample of the many counterexamples to the claim that Japanese “does not require storing a subset of verbs with SEMANTICALLY ARBITRARY case government specifications”. But even if this were true of Japanese, calling case government phenomena of languages other than English “semantically arbitrary” amounts to claiming that the most natural criterion of non-opacity and non-arbitrariness for government is the one suggested by the facts of English!

3.4. Inflectional morphology

Thus, McWhorter’s third diagnostic of grammatical complexity – verbal constructions – is not supported by fully convincing arguments. What about the fourth one: inflectional morphology? Examining morphophonemics and suppletions as well as what he calls “declensional and arbitrary allomorphy”, McWhorter says that the various inflectional strategies of languages like Latin or Russian “must be learned and stored with the root”. This applies to native speakers, I presume, since the case of foreign learners is irrelevant here. The learning process by native speakers is by no means difficult: McWhorter says that his “metric takes as a given that all languages are acquired with ease by native learners” (Section 2.4.2). But if so, then in exactly what respect can we say that the examples of inflectional morphology studied in this passage are more complex than the facts found in languages without inflectional morphology?
4. Limitations of the comparison between creoles and older languages

4.1. Saramaccan on one side, Tsez, Lahu, Maori on the other: simple vs. complex?

In Section 3, McWhorter compares Saramaccan with Tsez in order to show that even a three-centuries-old creole like the former is less complex than an older language such as Northeast Caucasian Tsez. In Sections 4.1 and 4.2, he compares Saramaccan with Lahu, this time to show that of two analytic languages, the older one is the more complex. Next (Section 4.3), he compares Saramaccan with Maori with the aim of demonstrating that even a language which lacks inflection but which, unlike Lahu, lacks tone as well is more complex than a creole. Overall, the basic intent underlying such comparisons would meet with the approval of most linguists interested in typology, including myself. However, when we examine the details, the picture changes: What can we learn about creoles when they are compared with other languages in terms of traits that are limited to certain parts of the world where their occurrence may be due no less to areal diffusion than to genetic kinship? I will give a single example: resultative compounds.

4.2. Resultative compounds

Citing Matisoff (1973: 207-208), McWhorter mentions verbal concatenations such as those found in Lahu, e.g., tú ‘to kindle’ + tò? ‘to catch fire’ > tú tò? ‘to catch fire’ (Section 4.1). He notes that in addition to its use in front of verb compounds or monosyllabic verbs, the negator mâ can also appear BETWEEN the two elements of the compound, yielding, in this particular example, tú mâ tò? ‘does not catch fire’. In actuality, this structure is not specific to some verbs only: in addition to the case illustrated here where the second verb (tò?) appears after the verb tú, there are many other binomial verbal constructions in Lahu. Their general meaning is resultative (cf. Matisoff 1973: 208) and with negation, it is potential-negative (e.g., c¯a ‘feed’ + mâ NEG + câ ‘eat’ > c¯a mâ câ ‘is unable to eat’). This is like in most other Sino-Tibetan languages as well as in Austro-Asiatic and Thai, where these structures, generally called “resultative compounds”, have been recognized for some time (for Chinese, cf. Hagège 1975: 145–154; for Hmong, Vietnamese, Thai, and Khmer, cf. Bisang 1992: 50, 229–230, 299–301, 338–339, 400–402, and Hagège 1994).

Thus, one can hardly speak of “exceptional negator placement” in Lahu: resultative compounds are widespread in at least one language family. However, they are scarce elsewhere: I know of no analogous cases other than some types of verb phrases in Indo-Aryan and Dravidian languages. Why should we expect creoles to possess such infrequent phenomena and what is the point of comparing Saramaccan with Lahu in this respect?
5. Creoles as simple languages

5.1. On “complex” features in three creoles

5.1.1. Pidgins, creoles, and the rapid rate of their formation. Given that, as I have tried to show, the metric proposed by McWhorter and the comparisons he makes between creoles and older languages leave room for reservations, we now need to turn to the task of exploring the notion of simplicity as applied to creoles before proposing a simplicity metric alternative to McWhorter’s.

McWhorter maintains that “the identification of scattered exceptions in various creoles to the general tendency [he has] identified does not constitute a refutation of [his] argument” (Section 4.2). Thus, I will refrain from engaging in guerilla warfare by citing counterevidence that would suggest a qualification of the notion of simplicity as applied to creoles. I only wish to point out that the way creoles have been formed has often resulted in certain traits which, given the youth of creoles stressed by McWhorter throughout his article, should not have appeared so fast.

I will draw my examples from three languages, of which only one, Bislama, is mentioned by McWhorter. It is listed at the end of the article among the 19 natural (creole) languages in which none of the features – ergativity, gender marking, and many others – occur that are found in a number of older languages. My second example will be Guadeloupean Creole which, for some reason, is not mentioned in the list. Guadeloupean Creole is closely related to Martiniquan Creole – which is on McWhorter’s list – but it is also different from it, if only because the circumstances of colonization and the origins of the populations were not the same (cf. Cérol 1991: 48). The third language I would like to introduce here, Sango, is not mentioned in McWhorter’s article at all. Like Bislama, it is a pidgin; cf. Samarin (2000: 320–321), where it is assigned this status though it fails to meet the requirements of Sebba’s fifteen defining features of pidgins (Sebba 1997). So are Tok Pisin, Solomon Island Pijin, and others, which are included in McWhorter’s list although pidgins, according to McWhorter, are “universally agreed to be rudimentary codes not fulfilling the needs of full language” (Section 2.3). I am not sure there is universal agreement between creolists on this matter. Sango (along with French) is an official language in the Central African Republic. It is widely used in newspapers and in public life, just like Bislama in Vanuatu. While as of now, these pidgins do not have native speakers, they might soon acquire them, and thus become creoles.

5.1.2. Sango. Sango possesses five tones and thus invalidates the claim that, with the exception of Saramaccan, there is no creole with lexically contrastive or morphosyntactic tone (Section 6). Sango tones are not the same as those which characterize the vowels or syllables of corresponding words in the sub-
strate: a Yakoma-Ngbandi language spoken along the banks of the Mbomu river. Since Sango tones are far from being mechanically inherited, the process by which they have been formed is remarkable in its speed (cf. the chapter entitled “The Creole laboratory” in Hagège 1990: 16–26; cf. also Hagège 1993: 128–130 on “Pidginogenesis as a rapid process”). This leads us to qualifying the claim that creoles have not had enough time to undergo changes that would result in complexities known from older languages.

5.1.3. **Bislama.** Bislama possesses an inclusive plural and an inclusive dual pronoun – *yumi* and *yumitu* (or *yumitufela*), respectively – in opposition to the exclusive plural *mifela* and dual *mitufela* (cf. Charpentier 1979: 307-309). Although McWhorter does not mention this trait among those that are absent from the 19 creoles of his list, it could be considered to be “incidental to basic communication” since most creoles do not have it. But in actuality it is not incidental. In Bislama as in Koliveu, Ninde, Navwiese, and other Melanesian languages which served as substrates, knowing who belongs and who does not belong to the close-knit group tied together by social solidarity is culturally essential.

5.1.4. **Guadeloupean Creole.** What McWhorter writes about the last five languages on his list (Haitian, Mauritian, Seychellois, Martiniquan, and French Guianese Creoles) – to wit, that they “have heavily borrowed from the French lexicon” (Footnote 20) – also applies to Guadeloupean. However, I do not consider this fact sufficient for assigning a special status to this language. Guadeloupean has a good deal of derivational morphology, e.g., *kok* ‘penis’ – *koke* ‘to make love’, *modi* ‘cursed’ – *modisyon* ‘curse’, and many other examples, which, in strictly morphological terms, can be compared with pairs such as Russian imperfective *pisat’ – perfective *napisat’* ‘to write’. These forms are mentioned by McWhorter as illustrations of inflectional strategies but in fact they illustrate derivational processes (even though the imperfective/perfective polarity belongs to the verbal paradigms of Russian and other Slavic languages). What such facts show is that the simplicity of creoles is not an invariant and monolithic notion. Some creoles have already begun to evolve at a striking speed towards a more complex stage.

5.2. **On intonation as an ever forgotten, and nevertheless essential, linguistic phenomenon**

It is remarkable that in McWhorter’s article nowhere is there any mention of linguistic phenomena other than those whose material expression is segmental. He is not the only linguist to manifest this attitude. Most linguists work and reason about human languages in a way which I cannot but attribute to the
pressure of written representations: segmental phenomena are, or can easily be, written whereas intonational contours are ordinarily not represented in writing except in works whose very focus is on suprasegmental phenomena. However, the importance of intonation in morphosyntax cannot be neglected by any means (cf. Hagège 1986: 23–24, 53–54; 1990: 40, 66, 83–84, 177–178, etc.). The pressure of written representation is, to some extent, unconscious. Yet professional linguists are perfectly aware that, in all languages of the world, an intonational difference is often the only phenomenon that distinguishes two utterances, with the prosodic curve only vaguely suggested by punctuation in languages that have a tradition of writing. An example is English

\begin{align*}
I & \text{ know that: } \\
\text{he came yesterday} & \text{ vs. } \\
I & \text{ know that he came yesterday,
\end{align*}

where the commonly proposed historical derivation of the complementizer \textit{that} from the demonstrative \textit{that} would not be conceivable were it not for the fact that in the first utterance, there is a sharp melodic boundary between \textit{that} and \textit{he came yesterday}, as \textit{he came yesterday} is new information, whereas in the second utterance, \textit{I know}, which is the new information, is clearly distinguished from \textit{that he came yesterday} (which is generally but not necessarily uttered in a lower register, so that the only reliable distinction is the melodic one).

Since intonation is a universal way of distinguishing linguistic utterances, one may hypothesize that, in the course of the history of human languages, additional means, namely segmental devices, were more and more resorted to without them superseding intonation. It is true that the syntax of creoles is less rich than that of older languages as far as SEGMENTAL MORPHOSYNTACTIC MEANS are concerned. But would it not be worthwhile to undertake a thorough examination of the use creoles make of intonation in morphosyntax? Given that creoles have not yet had enough time to develop many segmentally-coded grammatical distinctions, do they perhaps have a wealth of prosodic means that render them richer than older languages in this respect? I am not proposing that the answer is yes, but it seems to me that the question deserves further research.

5.3. \textit{Creoles and the linguistic cycle}

McWhorter does not exclude the possibility that creoles might become more complex than they are now should they undergo a series of changes during long periods of time. I would like to point out that this applies to every human language provided one gives up a linear view of linguistic evolution. Linguistic evolution is cyclic, not linear (cf. Hagège 1993: Chapters 4 ["The presence of Language Builders in creologenesis"] and 5 ["Language Builders and the linguistic cycle"]). As a consequence, many languages were simpler at an earlier stage and they may be complex today because they are in a complex phase of their history. Conversely, languages whose words are monosyllabic and invariable may have been inflectional in a remote period of their development (e.g.,
Commentary on McWhorter: Claude Hagège

Chinese, cf. Hagège 1993: 159). Does it follow that it should be possible to find languages that are older than creoles and yet simpler because they are in a simplicity phase of their history?

But what is a “simple” language? The final part of my comments will answer this question by proposing an alternative criterion of simplicity.

6. Dominant traits and recessive traits

The first diagnostic of McWhorter’s metric (cf. Section 3.1 above) makes use of the notion of markedness, understood not in terms of parameter setting but as a trait which is less frequent in the world’s languages. Given this concept of markedness, I propose to distinguish between two types of traits whose labels I will borrow from genetic biology with some modification of their meanings. The third of Gregor Mendel’s laws states that, given a pair of different genes coming from the two parents, one characteristic will appear in the resulting organism while the other will be latent. The former is called DOMINANT and the latter RECESSIVE. Let us call those linguistic traits DOMINANT which are the most frequent in the world’s languages, and let us call the least frequent ones RECESSIVE. My hypothesis is that creoles will generally have more dominant traits than other languages (cf. Hagège 1986: 119–124). This claim is not unrelated to the notions of “basic communication” and “communicative necessity”, used by McWhorter, but it seems to me that it is less impressionistic and more objective.

A set of at least twenty features pertaining to all levels – phonemic, morphological, syntactic, and lexical – should be established and measured in terms of degree of diffusion among human languages. Creoles would then be compared with older languages on this basis. Evidential marking, obviative marking, and most of the features listed by McWhorter in Sections 1 and 6 (but not all: SVO, in this sense, is more marked than SOV; cf. Hagège 1986: 59) are recessive features (cf. Hagège 1986: 40, 109–112, and passim, where percentages are given based on a sample of more than 700 languages). It might turn out that several creoles contain recessive features. For example, the Guadeloupean Creole focalizing strategy which reiterates the same word first as a focused noun and then as a verbal predicate, as in se dokte k li dokte [(it) is doctor that he doctors] ‘he is a really good doctor!’ (further examples in Cérol 1991: 83–84) is a recessive feature also found in the African substrates and in other African languages. (For example, Yoruba can relativize a verb by reiterating it as a derived verbal noun; cf. Hagège 1975: 109.)

Compositional iconicity found in several creoles including Tok Pisin (e.g., gut – nogut ‘good – bad’; cf. Êpêra (a Choco language of northwest Colombia) pîpî-wi, in Hagège 1997) is perhaps, surprisingly, also a recessive trait. On the other hand, older languages contain dominant features which are not found...
Creoles and the notion of simplicity in human languages

in most creoles. But the overall result of this investigation will be, in all likelihood, that there are more dominant features in creoles and fewer recessive ones than in older languages.

7. Conclusion

In conclusion, it should be clear that I do not deny the assertion that creole grammars are the world’s simplest grammars. Instead, I propose to qualify the claim somewhat on the basis of an alternative kind of typological inquiry and to test it by using an alternative definition of the notion of simplicity.

Received: 18 April 2001 Revised: 7 May 2001

Correspondence address: Collège de France, 11, place Marcelin Berthelot, 75005 Paris, France; e-mail: claude.hagege@free.fr

References

Simple and transparent
by Pieter A. M. Seuren

It would be difficult for me not to appreciate McWhorter’s thesis that creole languages distinguish themselves by the relative simplicity of their grammars as a function of their relative youth, since that is what is proposed in Seuren & Wekker (1986; reprinted as Chapter 19 in Seuren 2001). There we put forward the hypothesis that the structural properties of early creole languages are largely the result of the necessity, imposed by forced migration and forced labour, to be maximally simple and easy to learn, and that even the oldest creole languages have not had the time required for them to develop the more complex grammatical machineries that characterize older languages. One welcomes McWhorter’s attempt at making the notion of grammatical simplicity (complexity) more explicit so as to create a better testing ground for the hypothesis.

In McWhorter’s words (Section 2.4.3): “The guiding intuition is that an area of grammar is more complex than the same area in another grammar to the extent that it encompasses more overt distinctions and/or rules than another grammar.” Sensibly trying to remain as theory-neutral as possible, McWhorter applies this “guiding intuition” to (i) phonology (number of marked elements), (ii) syntax (number of rules), (iii) semantics (number of distinctions), and (iv) morphology (makes for complexity generally). He might have referred to Seuren & Wekker (1986: 64–66), where the following is said:

[A] maximization of S[emantic] T[ransparency] involves three strategies for grammars: (1) maximal uniformity of treatment of semantic categories, (2) minimal reliance on rules or rule types that are highly language-particular, and (3) minimal processing. Or, to put it briefly, UNIFORMITY, UNIVERSALITY, and SIMPLICITY. […] Given [the] strategy [of uniformity] one will expect few arbitrary grammatical distinctions, as with grammatical gender or conjugational idiosyncrasies, or with derivational processes in morphology. Moreover, one will expect a uniform strategy for arranging verbs and their arguments (subject, object, indirect object). Thus, rules that bring about variations in the order of subject (S), verb (V), object (O), and, though less crucially, indirect object (IO), will be untypical of creole languages. […] [Universalit] renders morphology essentially alien to creole languages, since whatever universals enable the growth of a morphological system in a language leave ample room for a multitude of often haphazard variations.

In fact, one might be tempted to say that McWhorter’s attempt at closing in on the notion of complexity hardly contains anything that goes beyond what is already proposed in Seuren & Wekker (1986). He contributes examples and illustrations, always useful, of course, but hardly any new notions or insights.
In this context one is surprised to find (Section 5.1) that McWhorter summarily dismisses Seuren & Wekker (1986), placing this publication in the “one form – one meaning” camp, even though these authors say explicitly (1986: 63, 66):

In the light of what we know today about semantic structures and semantic elements, it appears to be entirely unreasonable to think in terms of a condition of one-to-one mapping. Semantic structures are inevitably much richer than linguistic surface structures, mainly because they must be fully explicit and fully unambiguous. […] There is, as yet, no generally accepted theory of semantic structure. Agreement in this area does not go beyond the almost trivial condition that the language of semantics must contain the formal means for a logical calculus on analyses in terms of predicates and their arguments, plus quantifiers and logical connectives. […] All semantic theories agree that something like predicate calculus, with bound variables and the rest, must determine the structure of semantic analyses.

Consequently, Seuren & Wekker (1986: 64) propose that “any theory of S[emantic] T[ransparency] will have to be formulated in terms of a grammar that defines the mapping relations between surface structures and semantic analyses”.

Had McWhorter drawn the consequences from this uncontroversial and theory-neutral point of view, his proposals concerning syntactic and semantic complexity could have been more specific and better motivated, and thus more interesting, than they have turned out to be. He might, for one thing, have followed Seuren & Wekker (1986), when they discuss scope relations in semantic analyses. As the scope-bearing elements in logical structure are “lowered” into the lexical matrix-S, the elements with larger scope tend to stay to the left of those with smaller scope (compare Not many trains are comfortable versus Many trains are not comfortable). This left-to-right correspondence constraint is, however, totally disregarded in morphology, which makes Seuren & Wekker (1986: 68) observe:

This, again, leads to a ban on morphology, since, as is well known, morphological processing of lowered elements leads to a massive violation of this scope-order correspondence, and thus requires a great deal more cognitive processing than is needed for sentences with a regular scope-order correspondence. In this light it is not surprising to find that in many creole languages verbal tenses, modalities and aspects are expressed by means of preverbal particles, and not by morphological means, as in the majority of more advanced languages. The occurrence of such particles is a direct reflex of their semantic scope after lowering.

1. McWhorter’s discussion of the “one form – one meaning” constraint is, moreover, less than satisfactory. It is left without any further specification, and is, in fact, considered exclusively in terms of lexicalisation, a topic hardly touched upon in Seuren & Wekker (1986).
McWhorter might, in any case, have discussed these and similar points made in Seuren & Wekker (1986), as they are directly relevant to his concerns.

McWhorter argues against the semantic transparency hypothesis saying that “at the end of the day it is unclear that creoles are ‘semantically transparent’ overall to any greater extent than certain older languages” (Section 5.1). This may be so for “CERTAIN older languages” (under a proper definition of semantic transparency), but, like McWhorter’s complexity thesis, this is not what the transparency hypothesis is about. Seuren & Wekker might just as well say (though they will not): “at the end of the day it is unclear that creoles are ‘grammatically simple’ overall to any greater extent than certain older languages”, and use that as an argument against the complexity thesis. McWhorter’s claim is (Section 6):

[I]f all of the world’s languages could be ranked on a scale of complexity, there would be a delineable subset beginning at the “simplicity” end and continuing towards the “complexity” one all [members] of which were creoles.

This does not imply that all creoles are below the alleged cut-off point. All it says is that the simplest languages, those at the bottom end of the scale, are all creoles, and that creoles generally tend to gravitate towards that bottom end. This is the statement of a tendency. Since the transparency hypothesis states a corresponding tendency: “We now put forward the idea that creole languages are linguistically characterized by a tendency to maximize S[emantic] T[ransparency]” (Seuren & Wekker 1986: 64). McWhorter’s argument against this hypothesis cuts no ice at all. And if it did, it would equally affect his own complexity thesis.

It seems useful, in this context, to consider a suggestion (made in Seuren 1996: 344) to the effect that each language has associated with it a so-called SEMANTIC QUESTIONNAIRE that has to be “filled in” by any speaker of the language before any sentence can be formulated. Thus, to stay with McWhorter’s own example (1), when a Kikongo speaker wants to express the proposition ‘Past [I buy a goat]’ (s)he must specify first whether the ‘Past’ is recent, not so recent or remote, or else no grammatical sentence is possible. An English or Japanese speaker does not have to do that, though they must specify whether the situation described by ‘I buy a goat’ is to be located in the past or not (besides a few more specific questions for each of these two languages in particular). A Malay speaker does not even have to specify that, though (s)he may do so, if (s)he wishes. For an English sentence in the present tense it must be specified whether or not the state or action described in the matrix proposition is of transient duration; if so, the progressive form must be used, if not, the bare present tense will do. In some languages it must be specified whether the proposition is based on hearsay or comes under the speaker’s direct responsibility, or whether it is generic/habitual or not, etc. etc. Classifier lan-
languages want speakers preparing a (plural) NP to determine the broad cognitive category the object or objects spoken about are deemed to belong to. 2

Drawing up precise semantic questionnaires for given languages would not only be of great use to semantic theory, it is also quite feasible. If this were done for a sufficient number of creole and non-creole languages, significant differences might show up: one might find a statistically relevant tendency for the creole languages to minimize the questionnaires. Such a result would make the notion of semantic complexity operational and would take the theory of creole languages, in this respect, beyond the merely putative or impressionistic.

A further useful notion, in this respect, is that of SECONDARY CONSTRUCTION, as specified in Stassen (1985). “A secondary construction [...] is grammaticalized only in older or more advanced languages, but its semantic content is expressed by normal creative and ad hoc means in younger or less advanced languages” (Seuren & Wekker 1986: 66). Examples are comparative constructions (dealt with in Stassen 1985), conditionals, concessives, absolute participials (this being said ...), statements of price, credit, debt, and the like, constructions like the more the better, as soon as, anything but, more and more, etc. etc. It would seem that older languages, in particular those spoken in a rich social and cultural setting, can afford the luxury of multiplying the number of chapters in their grammars by introducing more and more such secondary constructions. The prediction is that creole languages will have very few of them.

What creole languages will, on the whole, lack is the “luxuries” that established languages can afford more easily. “Such ‘luxuries’ seem to be provoked by social differences within the speech community, where speakers of higher rank have an interest in making their speech hard to imitate. But in a situation of incipient creolization, such luxuries will hardly be found” (Seuren & Wekker 1986: 68). It is as with cutlery: in some families there is a piece of cutlery for almost every specific purpose, but when they go on a camping trip they will do with the simplest of forks, knives, and spoons. Creole languages are still on a camping trip.

Received: 23 March 2001   Max Planck Institute for Psycholinguistics

Correspondence address: Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands; e-mail: pieter.seuren@mpi.nl

Acknowledgment: This work was supported by a grant from the Max-Planck-Institut für evolutionäre Anthropologie in Leipzig.

---

2. The notion of semantic questionnaires makes it immediately clear that machine translation is not possible without taking into account the wider cognitive context of each utterance, which is unrealistic, given the resources available.
Commentary on McWhorter: Jacques Arends

References


Simple grammars, complex languages
by Jacques Arends

John McWhorter’s claim that “the world’s simplest grammars are creole grammars” is not new. It was made some thirty years ago when Saramaccan, the creole adduced by McWhorter to argue his case, featured as the world’s “least complex” language in the 1971 edition of the *Guinness Book of World Records* (Price & Price 1991: xii). While this claim – not founded on any serious evidence – was removed from later editions of that work, it now reappears – based, it should be noted, on more informed linguistic considerations – in the pages of *Linguistic Typology* as part of McWhorter’s theory of the “Creole Prototype”. According to this theory, creoles are simpler than other languages because, having evolved from an affixless pidgin in a relatively recent past, they have not existed long enough to have developed the kind of complexities found in non-creole languages.

As is readily acknowledged by McWhorter, the issue of grammatical complexity is not a simple one, if only because of the problems inherent in its measurement. However, since his aim is to compare creole and non-creole languages in terms of grammatical complexity, he cannot escape from proposing some kind of complexity metric. Although his selection of the four diagnostics of complexity is based on his assumption that they will “arouse the least possible controversy from as wide a spectrum as possible of linguists” (Section 2.4.3), they are by no means uncontroversial. First of all, they are all strictly quantitative, i.e., they are all of the type “more is more complex”. A grammar is judged to be more complex if it has more (marked) phonemes, more tones, more syntactic rules, more grammatically expressed semantic and/or pragmatic distinctions, more morphophonemic rules, more cases of suppletion, allomorphy, agreement. Qualitative aspects of complexity, such as the internal complexity of the rules themselves, are not taken into account. Second, the metric is strongly biased towards grammatical “building blocks”, i.e., elements such as phonemes and tones, paying little attention to the processes to which these elements are subjected. As an example of the latter one could think of the complex phonological contraction rules operating in some creoles, e.g., within the
tense/mood/aspect system of Sranan. Finally, the issue of the relative weight of the number of rules as opposed to their internal complexity is not addressed. To take a rather artificial example: Which grammar is more complex, the one with \( n \) rules, each of complexity \( C \), or the one with \( 2n \) rules, each of complexity \( C/2 \)? Although questions such as this cannot be easily answered, they should at least be addressed in a paper whose central thesis rests entirely on the concept of grammatical complexity.

Another problem is constituted by the fact that not all creoles adduced by McWhorter as cases of grammatical simplicity have been described in sufficient detail to allow such a far-reaching claim. Even for Saramaccan, his pièce de résistance, no comprehensive grammar is available. That is why McWhorter’s conclusions are necessarily based on only half a dozen partial studies, supplemented with data gathered from his informants. With regard to the latter, unfortunately no information is provided, neither about the time, place, and manner of their collection nor about the informants, which makes it difficult to assess their reliability. This is important as informant data play a substantial role in McWhorter’s argumentation.

With the exception of Baba Malay, all creoles adduced by McWhorter to support his case belong to the category of European-lexicon creoles, with English, French, and Portuguese taking care of the lion’s share (eight, six, and five, respectively). In most cases, the selected creoles belong to closely related language clusters, such as the Suriname Creoles (Saramaccan, Sranan, Ndyuka), the Gulf of Guinea Creoles (São Tomense, Principense, Anonbonese, Angolar), and Melanesian Pidgin English (Tok Pisin, Bislama, Solomon Islands Pijin). This reduces the number of languages for which the simplicity claim is made to a mere handful. Apart from the fact that Spanish-lexicon creoles are completely absent from McWhorter’s list, the question is whether the same result would be obtained if creoles with non-European lexical donors, which do not always conform to the classical creole type (Thomason 1997), would have been included.

Some questions may also be asked with regard to the features selected by McWhorter for the comparison of Saramaccan, on the one hand, and Tsez, Lahu, and Maori, on the other. One wonders what would happen if presumably more complex phenomena such as tense/mood/aspect or adpositions were included in the comparison. Although these topics have not been investigated in any detail for Saramaccan, work on its sister language Sranan, such as Winford’s (2000a, 2000b) detailed studies of the tense/mood/aspect system, suggests that the situation may turn out to be much more complex than would appear at first sight.

As McWhorter is well aware, his Prototype Theory bears some remarkable similarities to Bickerton’s Bioprogram Theory. In one of its versions, Bickerton (1984: 178) claimed that the bulk of Saramaccan grammar could be cap-
tured in less than ten rules of syntax – a simple grammar if ever there was one! Another similarity is that both theories are, crucially, based on the assumption that creoles develop out of pidgins. Although this idea has a long history in creole linguistics, the fact of the matter is that, with only few exceptions, the existence – let alone the structural make-up – of these alleged pre-creole pidgins is undocumented. In other words, the pidgin ancestry of creoles is based on theoretical motivations rather than on empirical fact. This (over)reliance on theoretical considerations points to a third similarity between the two theories: their value resides more in their capacity to generate discussion (not a bad thing for a theory, of course) than in their contribution to deepening our understanding of creole languages as complex empirical phenomena.

Received: 2 April 2001

Correspondence address: Algemene Taalwetenschap, Universiteit van Amsterdam, Spuistraat 210, 1012 VT Amsterdam, The Netherlands; e-mail: j.arends@hum.uva.nl

References


The complexities of arguing about complexity

by Wouter Kusters and Pieter Muysken

John McWhorter is to be commended for keeping the issue of complexity of creoles on the research agenda, and putting it squarely in the arena of linguistic typology and language change, where it belongs. Particularly insightful is the explicit comparison between a creole and an analytical language like Lahu. However, our task here is to raise critical points rather than simply concur, and there are a number of issues that come to the fore.

First, the issue of time. Languages with histories of thousands of years of uninterrupted transmission and ongoing elaboration are contrasted with re-
The complexities of arguing about complexity

cent languages (creoles). However, for the vast majority of the world’s languages we simply do not know how long ago they emerged and took on their present form. Current work by one of us (Muysken) suggests that, e.g., the past two millennia were certainly ones of upheaval and frequent restructuring for many languages of the Andean/Amazonian fringe, highly complex with McWhorter’s measures. Furthermore, there is no independent evidence on how long it takes for a language to become complex: is there a crucial difference between 300 and 600 years here? In the absence of such evidence, it is vacuous to make claims about development. Furthermore, it is questionable to contrast historical objects such as creoles with other languages viewed a-historically.

A second issue is the source of complexity in language, which McWhorter, following earlier work by Bickerton, attributes to “baroque accretion”. McWhorter’s perspective is a functionalist one: language is there to convey meaning and complex features of individual languages which do not contribute to this are historical accidents. This perspective leaves much to be desired. Carstairs-MacCarthy (1999) draws attention to the fact that many languages have complex systems of inflection classes, which show remarkable persistence over time (cf. also Küsters in preparation). According to this author, such classes have a basis in the evolutionary basis of our basic language capacities. One need not agree with Carstairs-MacCarthy’s explanations, but the very stability and vitality of these systems runs counter to the idea that they are simply the debris of history.

A third issue was raised by Lightfoot (1979) in a review of Li (ed.) (1977), in terms of Ebeling’s Law. David Lightfoot attributed to the Dutch slavist Carl Ebeling the observation that the less we know about a language, the more regular its phoneme system appears. The sad fact is that we know very little about most creole languages. One of the few good descriptions of a creole is Kouwenberg’s (1991) study of the moribund Berbice Dutch Creole, of which only the rudiments were still available at the time of description. Certainly there is no description of Saramaccan available with anything like the level of detail of Matisoff’s (1973) grammar of Lahu. Any time we do work on Saramaccan or on its sister language Sranan, new facts are discovered. Recent work on the supposedly well-described Hawaiian Creole English tense/mood/aspect system by Veelupilai (in preparation) is unearthing completely new facts about the language. Altogether, creole languages were often assumed to be instantly known by observers in the colonial era and theoretical linguists in the post-colonial era, due to their European lexicon and simple root shapes. This has stood in the way of serious description. Methodologically, Ebeling’s Law would imply that we compare two systems which are equally well-described.

Given the eloquently phrased contrastive analysis by McWhorter of Saramaccan, Tsez, and Lahu, where could the complexity of creoles lie? Clearly
Commentary on McWhorter: Wouter Kusters and Pieter Muysken

not in the phonological shapes and morphological endings (although there is some of that, too), but in the complex phrasal compounding and in the combinatorics of roots. As in English, it is precisely the absence of morphology on roots that makes the system complex.

This last point, lexical complexity, leads us to a final issue: the vague and hybrid nature of McWhorter’s complexity measures: half theoretical, half empirical; half system-oriented, half item-oriented. This ambivalence can be illustrated with a number of examples.

First, Germanic verb-second order is hard to acquire for L2 learners, but perfectly straightforward for L1 learners (Clahsen & Muysken 1986). While the facts are complex when viewed from the outside, theoretical grammarians have formulated perfectly simple and elegant rules to describe it. Complexity lies to some extent in the nature of the description. A description intended to grasp the acquisition process of an L1 learner leads to a different notion of complexity than a description for the L2 acquisition process, or the description of a language structure itself. These distinctions are blurred by McWhorter’s intuitive notion.

Second, and similarly, McWhorter claims that Lahu is more complex than Saramaccan in having an overt marker linking causatives to their non-causative counterparts. However, the reverse could also be argued, and has indeed been argued by Muhlhausler (1974): lexical rules simplify a lexicon.

Third, McWhorter’s notion of complexity is not sufficient when comparing, for instance, the following phonological inventories: /p t k pʰ tʰ kʰ b d g bʰ dʰ gʰ/, which can be analysed with the help of four binary distinctive features, and /p t k s d kʰ/, which needs five binary distinctive features. Are the number of distinctive features decisive here or the number of phonemes? McWhorter’s reluctance to base his notion of complexity on language processing leaves such questions insoluble.

Thus McWhorter’s notion of complexity remains half-baked. We think the reason for this is that it is not sufficiently modular. Many linguists would argue that a language is the product (among other things) of the interaction of a lexicon and various rule systems. This perspective would lead us to see where complexity lies in the trade-off between these subsystems. In McWhorter’s undifferentiated perspective, this remains unclear.

It would be nice if there were independent tests for the complexity of a language, e.g., if it were easy for a L2 learner to pass as a native speaker (a measure rejected by McWhorter). On this score, creole languages do not seem encouraging for McWhorter. We do not know whether anybody has claimed to come near to sounding like a native speaker of Saramaccan, but for Sranan even many Surinamese (let alone non-Surinamese) are criticized as non-fluent. Similarly for Papiamentu: only a handful of non-native speakers are grudgingly acknowledged as speaking decent (mind you, not fluent) Papiamentu.
The complexities of arguing about complexity

In conclusion, McWhorter has successfully put the notion of complexity in language back on the research agenda. However, his notion of complexity remains too vague and is intended to cover too much of a grammar to be of much use. It does not tell whether a language will be difficult for an L1 learner or an L2 learner, neither does it evaluate language on anything more than an intuitive level.

Received: 24 April 2001
Revised: 29 May 2001

Correspondence addresses: (Kusters) Leiden Centre for Linguistics (ULCL), Universiteit Leiden, Postbus 9515, 2300 RA Leiden, The Netherlands; e-mail: w.kusters@let.leidenuniv.nl; (Muysken) Algemene Taalwetenschap en Dialectkunde (ATD), Katholieke Universiteit Nijmegen, Postbus 9103, 6500 HD Nijmegen, The Netherlands; e-mail: p.c.muysken@let.leidenuniv.nl

References
1. Introduction

McWhorter’s target article makes two very strong claims. The first is stated in its title: “The world’s simplest grammars are creole grammars”. The second appears to come as a consequence of the first: “Creole grammars constitute a synchronically identifiable class”. This second claim builds on McWhorter’s (1998: 790) earlier claim according to which creole languages constitute a “synchronically definable typological class”. In this paper, I provide an alternative way of addressing the issues of the alleged simplicity of these languages and of the alleged similarity between them. The following three questions will be discussed in turn. First, what do creole languages really have in common? Second, why do creoles tend to be isolating? Third, why do creoles tend to look simpler? These questions will be addressed from the point of view of the relexification account of creole genesis advocated in Lefebvre (1998) and the references cited therein. Section 5 weighs apparent simplicity against hidden complexity. Section 6 considers McWhorter’s hypothesised creole typological features in light of the previous discussion. It will be shown that the features proposed by McWhorter as identifying creoles are derivable from a sound theory of how creoles come about. Section 7 concludes the paper.

2. What is it that is similar among creole languages?

One way of addressing the issue of the similarity between creole languages is to ask whether they form a typological class. In my view, they do not. In the paragraphs that follow, I substantiate this claim on the basis of the research on creole genesis that I have been conducting over the last twenty five years.

When we started this research, our basic assumption was that it should be possible to account for the formation of creole languages in terms of the same processes that are at work in language genesis and language change in general; that is: relexification, a cognitive process that has been shown to play a role in the formation of mixed languages (see, e.g., Media Lengua, see Muysken 1981); reanalysis, a major process in linguistic change (see, e.g., Heine & Reh 1984); and dialect levelling, a process that has been shown to take place when dialects of the same languages come into contact (see, e.g., Trudgill 1986).

Our hypothesis (see, e.g., Lefebvre & Kaye (eds.) (1986), Lefebvre & Lumsden (1989, 1994a, b), Lefebvre (1998), and the references cited therein) was that the creators of a creole, adult native speakers of substratum languages, use
What you see is not always what you get

187

Figure 1. Relexification (Lefebvre & Lumsden 1994a, b)

the properties of their native lexicons, the parametric values and the semantic interpretation rules of their native grammars in creating the creole. On this hypothesis, the bulk of a creole’s lexical entries is created by the process of relexification. Two other processes, fed by the output of relexification, dialect levelling and reanalysis, also play a role in the development of the creole.1

Relexification applies in creole genesis when speakers of several substratum languages are targeting the same superstratum language (Lefebvre & Lumsden 1994a). The process was first defined by Muysken (1981: 61): “Given the concept of lexical entry, relexification can be defined as the process of vocabulary substitution in which the only information adopted from the target language in the lexical entry is the phonological representation.” In Lefebvre & Lumsden (1994a, b), the process has been represented as a two step process: copy and relabel, involving the specifications in Figure 1. (For a detailed description of the representation in Figure 1, see Lefebvre 1998: 16–18.)

In a lexicon that is in the process of being relexified, each lexical entry acquires a second phonological representation that is derived from the lexifier language. As is shown in (1), following Mous’s (1994) proposal, at a certain point in the process, each lexical entry has two phonological representations.

(1) \[
\begin{bmatrix}
\text{[phonology]}_i \\
\text{[semantic feature]}_k \\
\text{[syntactic feature]}_n
\end{bmatrix} \quad \text{or} \quad \begin{bmatrix}
\text{[phonology]}_j \\
\text{[semantic feature]}_k \\
\text{[syntactic feature]}_n
\end{bmatrix}
\]

In the history of an early creole community, at some point, the substratum languages cease to be spoken. The original phonological representations are no more used. Consequently, they eventually disappear from the lexicon. The new lexical entries thus have the semantic and syntactic properties of the original

1. For an extensive discussion on these two processes and on how they interact with relexification in creole genesis, see Lefebvre (1998) and the references cited therein, and Lefebvre (2001a).
Commentary on McWhorter: Claire Lefebvre

ones, and phonological representations derived from a phonetic string in the superstratum language. This is represented in (2).

(2) 
\[
\begin{array}{c}
\text{[phonology]}_f \\
\text{[semantic feature]}_k \\
\text{[syntactic feature]}_n \\
\end{array}
\]

The nature of the process of relexification predicts that the creole lexical entries will have phonological representations derived from the superstratum language and syntactic and semantic properties derived from the substratum languages. 2

The hypothesis was tested on the basis of Haitian Creole. The test of the hypothesis consisted in a detailed comparison of the lexicon and grammar of Haitian Creole with those of its contributing languages: French, its superstratum language, and Fongbe, one of its substratum languages. 3 The details of the threeway comparison are extensively reported on in Lefebvre (1998). The results of the linguistic test show that, to a large extent, the hypothesis is supported by the data. In the paragraphs that follow, I provide an overview of the results of this comparative study, so as to provide the reader with some background information for the discussion that follows on the issues at stake in this article. The nominal structure, the tense, mood and aspect markers, the parameters and the verb doubling phenomena will be discussed in turn.

The data in (3), from Lefebvre (1998: 78), provide an overview of French nominal structure. They show that, in this language, the definite determiner, the possessive, and the demonstrative determiners all precede the head noun, and that there can be only one of these per noun phrase (3a). Singular and plural forms are contrasted in (3b) showing that plural is encoded in a bound morpheme in French.

(3) a. *le mon ce crabe
   DET POSS DEM crab
b. le/mon/ce crabe b', les/mes/ces crabe
   ‘the/my/this crab’  ‘the/my/these crabs’

2. As has been pointed out in Lefebvre & Lumsden (1994a), this account of creole genesis is a further development of the second language acquisition theory of creole genesis (see, e.g., Schumann 1978, Andersen 1980, Thomason & Kaufman 1988, etc.); it is claimed that, in creole genesis, involving situations where there is little access to the superstratum language, the process of relexification is used by speakers of the substratum languages as the main tool for acquiring a second language: the superstratum language.

3. Due to various constraints, we had to limit the detailed study of the substratum languages of Haitian to one language. On the basis of non-linguistic factors, Fongbe was chosen as the substratum language to be studied in detail. In no way does this methodological choice entail that Haitian is Fongbe relexified. For a thorough discussion of the methodology of the research and the validity of the linguistic test, see Lefebvre (1998: 52–77) and the references cited therein.
Table 1. Properties of definite determiners in French, Haitian, and Fongone (= (11) in Lefebvre 2001b)

<table>
<thead>
<tr>
<th></th>
<th>French [+definite] determiner</th>
<th>Haitian/Fongbe [+definite] determiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-nominal</td>
<td></td>
<td>Post-nominal</td>
</tr>
<tr>
<td>marked for gender and number</td>
<td></td>
<td>unmarked for gender and number</td>
</tr>
<tr>
<td>allomorphs: <em>lelllesl</em></td>
<td></td>
<td>allomorphs: <em>la, a, an, nan, lant</em> , <em>mn</em></td>
</tr>
<tr>
<td>anaphoric and cataphoric</td>
<td></td>
<td>anaphoric</td>
</tr>
<tr>
<td>partitive <em>dudes</em></td>
<td></td>
<td>no partitive</td>
</tr>
<tr>
<td>obligatory with generic or mass nouns</td>
<td></td>
<td>impossible with generic or mass nouns</td>
</tr>
<tr>
<td>no bare NPs</td>
<td></td>
<td>bare NPs</td>
</tr>
<tr>
<td>*Det [relative clause] N</td>
<td></td>
<td>N [relative clause] Det</td>
</tr>
</tbody>
</table>

In Haitian (4a) and Fongbe (4b) the determiners all follow the head noun. In both languages, a possessor phrase, a demonstrative term, the definite determiner, and the plural marker may all co-occur within the same nominal structure. In both languages, the plural marker is an independent morpheme.

(4) a. *krab* [mwen Ø] *sa a yo*
    b. *âs *mn [nye lO] *èlO *â *lè*

    crab me GEN DEM DET PL
    ‘these/those crabs of mine (in question/that we know of)’ (Lefebvre 1998: 78)

The Haitian and Fongbe nominal structures thus contrast in the same way with the French nominal structure with respect to word order, co-occurrence restrictions of determiners, and with respect to whether the plural marker is a free (in Haitian and Fongbe) or a bound (in French) morpheme.

Furthermore, with the exception of their phonological representations, the properties of the definite determiners are the same in Haitian and in Fongbe; these properties contrast in a systematic way with those of the French definite determiner. These contrastive properties are summarised in Table 1 based on the detailed description in Lefebvre (1998: 79–84).

Moreover, the definite determiners involved in the Haitian (5a) and Fongbe (5b) nominal structures also play a crucial role in the clause structure of these two languages. For an extensive discussion of these facts, see Lefebvre (1998: 219–247).

(5) a. Li *rive a*
    b. *È* wá 5

    ‘He has arrived.’ (as expected/as we knew he would)

The definite determiner plays no role in the structure of French clauses.
A Haitian (6a) or Fongbe (6b) nominal structure may contain a noun followed by the plural marker only. In such a case, the structure is interpreted as definite.

(6) a. krab yo
   b. àś̀mn ̀ř
      crab PL
      ‘the crabs’, *(some) crabs’ (Lefebvre 1994a: (31))

Comparable constructions are impossible in French.

The following data show that Haitian (7a) and Fongbe (7b) both allow for bare NPs.

(7) a. M’ ̅achte krab.
    b. N’ ̅xɔ àś̀mn.
       I buy crab
       ‘I bought (some) crabs.’ (Lefebvre 1994a: (32))

Bare NPs are not allowed in French.

In both Haitian (8a) and Fongbe (8b), when the definite determiner and the plural marker co-occur within the same nominal structure, the definite determiner must precede the plural marker.

(8) a. krab la yo / *krab yo a
    b. àś̀mn ́s ̀ř / *àś̀mn ̀ř ́s
       crab DET PL crab PL DET
       ‘the crabs (in question)’ (Lefebvre 1994a: (33))

Finally, in both languages, there is variation among speakers with respect to the possibility of co-occurrence of the determiner and the plural marker. Crucially, the patterns of variation are the same in both languages. Two slightly different grammars have been reported on in the literature. They are summarised in Table 2.

The French and Haitian paradigms of deictic terms are also strikingly different, whereas the Haitian and Fongbe paradigms of deictic terms are strikingly parallel. Due to space limitations, suffice it to say here that, while French has eleven deictic terms that can be involved in the nominal structure, Haitian and Fongbe have two each: sa and sîla (Haitian) and (é)lô and (é)nî (Fongbe). In Lefebvre (1997, 1998: 89–101), it is extensively argued that the properties of the two Haitian terms are not the same as those of the French lexical entries which were the source of the phonological representation of the Haitian ones (ça and cela/celui-là, respectively); it is further extensively argued that the two Haitian terms do have the same distributional and syntactic properties as the Fongbe corresponding ones. Furthermore, in Lefebvre (2001a), it is shown...
What you see is not always what you get

Table 2. Possibility of co-occurrence of determiner and plural markers in two different grammars of Haitian and Fongbe

<table>
<thead>
<tr>
<th>Haitian</th>
<th>Fongbe</th>
</tr>
</thead>
</table>

that in both Haitian and Fongbe, there are three semantic interpretation patterns (identified below as G1, G2, and G3, where *α* is a variable that can take the values + or −) for the pairs of deictic terms. These are shown in (9a) for Haitian and (9b) for Fongbe. Crucially, these patterns are identical for both languages.

(9) a. G1 *sa* [+proximate] *sila* [−proximate]
    G2 *sa* [*α* proximate] *sila* [−proximate]
    G3 *sa* [*α* proximate] *sila* [*α* proximate]

b. G1 *(é)n*5 [+proximate] *(é)nē* [−proximate]
    G2 *(é)n*5 [*α* proximate] *(é)nē* [−proximate]
    G3 *(é)n*5 [*α* proximate] *(é)nē* [*α* proximate]
    (Sources: G1: Anonymous 1983, Segurola 1963, and my own fieldnotes; G2: Lefebvre 1997; G3: my own fieldnotes.)

The data discussed in (3)–(11) show the remarkable parallel that exists between the nominal structures of Haitian and Fongbe. As is extensively argued in Lefebvre (1998: 89-101, 2001a), the extraordinary similarity that exists between the functional categories of the Haitian and Fongbe nominal structures follow from relexification.

In Haitian Creole, the verb of a finite clause is invariant. In French, however, the verb of a finite clause obligatorily bears inflectional morphology encoding tense, mood, aspect, and person and number. None of the verbal morphology found in French has made its way into Haitian. Haitian follows the pattern
Table 3. The inventory of TMA markers in Haitian (H) and in Fongbe (F) (Lefebvre 1996: 281, (115))

<table>
<thead>
<tr>
<th>ANTERIOR</th>
<th>IRREALIS</th>
<th>NON-COMPLETE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past/Past perfect</td>
<td>Definite future</td>
<td>Habitual</td>
</tr>
<tr>
<td>H</td>
<td>F</td>
<td>H</td>
</tr>
<tr>
<td>te</td>
<td>kò</td>
<td>ap</td>
</tr>
</tbody>
</table>

Indefinite future
| H | F |
| a-va | ná-wá |

Subjunctive
| H | F |
| pou | ní |

of its West African (non-Bantu) substratum languages in having invariant bare verbs. In both Haitian and Fongbe, temporal relationships, mood, and aspect are encoded by means of markers occurring between the subject and the verb. The inventory of the TMA markers of Haitian is quite parallel to that found in Fongbe; see Bentolila (1971), Lefebvre (1996, 1998: 11–140). This is shown in Table 3.4

Both languages have a marker which encodes anteriority. Both lexically distinguish between definite and indefinite future. The definite future markers are used to convey the speaker’s attitude that the event referred to by the clause will definitely take place in the near future. By contrast, the indefinite future markers are used to convey the speaker’s opinion that the event referred to by the clause might eventually or potentially take place at an undetermined point in the future. The fact that speakers of Haitian distinguish between definite and indefinite future is widely documented in the literature (cf. Valdman 1970, 1978, Spears 1990, and the references therein). For Fongbe, this distinction is pointed out in Anonymous (1983: V, 3). Both languages have a marker glossed

---

4. The inventory of TMA markers in Haitian and Fongbe is established in Lefebvre (1996) on the basis of syntactic tests which set the preverbal markers apart from modal and aspectual verbs. First, they all occur between the subject and the verb. Second, preverbal markers occurring in the same column in Table 3 are mutually exclusive, showing that they are in a paradigmatic relationship. Third, while modal verbs do allow for deletion of their VP complement, preverbal markers do not (for Haitian, see Koopman & Lefebvre 1982, Magloire-Holly 1982, Spears 1990; for Fongbe, see Lefebvre 1996). Fourth, most of the preverbal markers in Table 3 have no meaning outside of the TMA system. Finally, the TMA markers may combine to form complex tenses (Lefebvre 1996).
as “subjunctive” for convenience. This term subsumes the three meanings of \textit{pou} and \textit{nî} respectively: both may be interpreted as ‘must’, ‘should’, or ‘may’. Both languages have a form which encodes imperfective aspect. As can be seen in Table 3, there is a one-to-one correspondence between the preverbal markers in the two languages, except that Fangbe has one encoding the habitual aspect, and Haitian does not. In Lefebvre (1998: 111–140) it is argued that, while the phonological representation of the tense, mood, and aspect markers of Haitian are derived from French phonetic strings, their semantic and syntactic properties follow the details of the corresponding substratum lexical entries.

At the beginning of this paper, as part of the hypothesis on creole genesis, it was stated that the creators of a creole use the parametric options of their own grammar to assign a value to the parameters of the language that they are creating. The hypothesis predicts that, where the parametric values of the substratum and superstratum differ, the creole should have the same parametric value as the substratum languages. With one exception (discussed below), the three-way comparison in Lefebvre (1998: 349–374) supports this general hypothesis. As can be seen in Table 4, at the time the research was conducted, parameters were formulated in terms of correlations between the availability of functional categories and a related syntactic phenomenon. As has been pointed out in Lefebvre (1998: 387), the parametric options set in the creole are the result of its creators’ reproducing the properties of the functional categories of their own lexicons through relexification. The correlations discussed in Lefebvre (1998: 349–374) are summarised in Table 4 (for each parameter, the proposer of the correlation is mentioned within square brackets). As can be seen, the parametric options of Haitian systematically contrast with those of French and follow those of substratum languages of the type of Fangbe.

Koopman (1986) observes that other subsets of data, which can also be formulated in terms of parametric options, show similar behaviour. For example, she remarks that in Haitian, as in West African languages, headless and infinitival relative clauses are not available. This contrasts with French, where both types of relative clauses are available. Koopman further points out that, in contrast to French, where the set of phenomena referred to as quantifier float is available, Haitian and West African languages lack such phenomena.

There is one exception to the general pattern reported on in this section: whereas both French and Fangbe are null subject languages, Haitian is not. In recent literature, it has been proposed that languages with syntactic clitics should be considered null subject languages; see, e.g., Jaeggli (1984), Hulk (1986), Roberge (1990). Both Fangbe and French have syntactic clitics, but Haitian does not (cf. Lefebvre 1998: 148–157 and the references therein). Since syntactic clitics did not make their way into the creole, as will be further discussed below, the value of the null subject parameter had to be reset (cf. Lefebvre 1998: 349–351).
Table 4. Comparison of the parametric options in the three languages under comparison (Lefebvre 1998: 387, Table 13.7)

<table>
<thead>
<tr>
<th>Availability of</th>
<th>Fongbe</th>
<th>Haitian</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb raising to infl (correlates with inflectional morphology on the verb) [Pollock 1989]</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Serial verbs (correlates with lack of derivational and inflectional morphology) [Baker 1991, Muysken 1988]</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Double-object constructions (correlates with availability of Genitive case in nominal structures) [Johnson 1991]</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Negative quantifiers as NPs (correlates with availability of bare NPs) [Déprez 1999]</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Verb-doubling phenomena (correlates with the properties of the determiner system) [Lefebvre 1998: 363–374]</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>

Verb-doubling phenomena involve four constructions which contain what looks like an exact copy of the predicate (here exemplified by *wá/rive* ‘to arrive’): temporal adverbial, as in (10), causal adverbial, as in (11), factive clauses, as in (12), and the predicate cleft construction, as in (13). (Examples (a) are from Fongbe, (b) from Haitian; see also (1)–(4) in Lefebvre 1994b.)

(10) a. *Wá Jan wá (trôlô) bô Màrí yî.*
    b. *Rive Jan rive (epî) Mari pati.*

    'As soon as John arrived, Mary left.'

    b. *Rive Jan rive Mari pati.*

    'Because John arrived, Mary left.'

(12) a. *Wá ḍê ḍê Jan wá ý vîvî nû nû*
    b. *OP John arrive DET make(-happy) for mother tîn.

    his
What you see is not always what you get

195

b. Rive Ø Jan rive a, fé manman li kòntan.
arrive OP John arrive DET make mother his happy
‘The fact that John arrived made his mother happy.’

(13)  a. Wá wè Jan wá.

b. Se rive Jan rive.

It is a well known fact that, while verb-doubling phenomena are attested in Haitian and in West African languages, they are not attested in French (see Koopman 1986, Lefebvre 1998, and the references cited therein). Moreover, as is demonstrated in Lefebvre (1998: 363–374), the properties of the verb-doubling constructions in both Haitian and Fongbe are strikingly similar.

The overview of data pertaining to major subsystems of their grammars shows that Haitian Creole shares major properties with its substratum languages. Data showing that Haitian Creole lexical entries reproduce the semantic divisions of their substratum languages, in spite of the fact that their phonological representations are derived from French phonetic matrices, may be found in Lefebvre (1998, 1999), in Lumsden (1999), and in the references cited in these publications. In Lefebvre (1998: 248–301), it is shown that, to a great extent, the syntactic properties of Haitian verbs also correspond to those of the substratum languages rather than to those of French. The inventory and the properties of the Haitian derivational affixes are also argued to be extremely similar to those of the substratum languages rather than to those of the lexifier language (see Lefebvre 1998: 303–333, and the references cited therein). The principles governing the concatenation of words into compounds in Haitian also appear to follow the substratum languages (see Brousseau 1988, 1989, Lefebvre 1998: 334–349).

It thus appears that the Haitian lexicon manifests the semantic and syntactic properties of its substratum languages. Similarly, Haitian reproduces the principles of concatenation and the parametric values of its substratum languages. Abstracting away from the phonological representations of the Haitian lexical entries, it appears that Haitian Creole manifests the typological features of the Gbe (here represented by Fongbe) and other West African languages, the substratum languages, rather than those of French, the lexifier language.

Now, if relexification has played a central role in the formation of Haitian Creole, it is logical to hypothesise that this cognitive process has also played a major role in the formation of other creole languages. By hypothesis, then, these other creoles would also reproduce the properties of their substratum languages. In his comparison of Solomons Pidgin with its source languages, English, the lexifier language, and Kwaio, an Austronesian substratum language,
Commentary on McWhorter: Claire Lefebvre

Keesing (1988) shows extensively that Solomons Pidgin does reproduce the properties of its Austronesian substratum languages.

Keesing (1988: 1–2) writes:

Sitting on a Solomon Islands mountain in 1977, reading Derek Bickerton’s review article on “Pidgin and Creole Studies” (1976), I was led to think more seriously than I ever had about the history and structure of Solomon Islands Pidgin. I had earlier been struck, when I had learned Solomon Pidgin in the 1960s through the medium of Kwaio, an indigenous language I already spoke fluently, that this learning task mainly required learning Pidgin equivalents of Kwaio morphemes. The syntax of Solomon Pidgin was essentially the same as the syntax of Kwaio, although somewhat simpler and lacking some of the surface marking; in most constructions, there was a virtual morpheme-by-morpheme correspondence between Kwaio and Pidgin. (This was not just an odd local process of calquing: the Pidgin I was learning in terms of Kwaio was spoken with only minor variations throughout the southeastern and central Solomons, although it was everywhere adapted to local phonologies.) Although most of the Pidgin lexical forms were ultimately derived from English, I found this largely irrelevant to my language-learning task. The semantic categories they labeled corresponded to Kwaio ones, not English ones; grammatical morphemes corresponded to Kwaio ones, not English ones. Thus semantically Pidgin *dae* corresponded directly to Kwaio *mae* ‘be dead, die, be comatose, be extinguished,’ not to English “die.” Pidgin *baebae* corresponded to the Kwaio marker of future/nonaccomplished mode, *ta-*, not to English “by and by”.

Keesing accounts for the linguistic situation he describes in terms of calquing. That is, the substratum speakers of Solomons Pidgin calque the properties of their native languages (e.g., Kwaio) when speaking the pidgin. The type of calquing that Keesing describes corresponds to the definition of relexification given at the beginning of this section.

Keesing (1988) documents the fact that calquing of the substratum properties can be observed throughout the lexicon of Solomons Pidgin. He shows that its pronominal system is quite similar to that of the complex system of the substratum languages in distinguishing singular, dual, and plural, inclusive and exclusive 1st person plural, etc. He argues that the Tense/Mood/Aspect system of Solomons Pidgin reproduces the idiosyncrasies of the system of the substratum languages. As Keesing (1988: 215) puts it: “In fact, the entire set of Kwaio particles marking the time-frame of the verb, some of which are preverbal and some postverbal, correspond in their Solomons Pidgin usage to a set of particles derived from English but carrying exactly the same import as the Kwaio particles, and placed in exactly the same slots”. Keesing further shows that, as is the case in the substratum languages, Solomons Pidgin has a predicate marker. The same pattern is also found in interrogative constructions, relative clauses, etc. In short, Keesing provides extensive evidence that, while
What you see is not always what you get

the phonological representations of Solomons Pidgin lexical entries are derived from English phonetic strings, the properties of these lexical entries do not correspond entirely to those of English lexical entries; he convincingly demonstrates that the properties of the Solomons Pidgin lexical entries do, however, correspond to those of its substratum languages, including functional category lexical entries. Compare (14a) from Solomons Pidgin with (14b) from Kwaio (data from Keesing 1988: 214).

(14) a. Olketa bae-i go.
    FP(them) FUT-SRP(3PL) go
    ‘They will go.’

b. Gila ta-la leka.
    FP(them) FUT-SRP(they) go
    ‘They will go.’

While olketa in the pidgin derives its phonological representation from the English expression *all together*, it has the meaning and uses of the substratum strong personal pronoun *gila* ‘them’. While bae in the pidgin derives its phonological representation from a reduced form of the English expression *by and by*, its meaning and uses correspond to the substratum lexical entry *ta-* a future marker. As in the substratum language, the future marker of the pidgin is marked for a 3rd person pronominal form. This pidgin form is derived from the English *he*, but it does not share the uses of the form it is phonologically derived from; it does, however, share the properties of the substratum forms.

So, the relexification account of creole genesis predicts that Atlantic creoles will reproduce the properties of their West African substratum languages, while Pacific ones will reproduce those of their Austronesian substratum languages. Atlantic and Pacific creoles are thus expected to differ in the same areas of lexicon and grammar as West African and Austronesian languages do among themselves. For example, while the pronominal system of Solomons Pidgin reproduces the singular, dual, plural inclusive and exclusive 1st person distinctions of its substratum languages, as was mentioned above, Haitian also reproduces the particularities of its substratum languages. Consider the paradigm of personal pronouns in Table 5. While French has six forms, Haitian has only five. Like in Fongbe, in Haitian, the same form serves as both 1st and 2nd person plural (for further discussion, see Lefebvre 1998: 141–143). Likewise, verb-doubling phenomena of the type in (10)–(13) are only found in those creoles for which the substratum languages have them. Thus, while Atlantic creoles have these constructions, inherited from their substratum languages, Pacific creoles do not have them because their substratum languages do not have them. Consequently, on the relexification account of creole genesis, creole languages cannot be argued to be typologically similar. Rather, what
Commentary on McWhorter: Claire Lefebvre

Table 5. French, Haitian, and Fongbe personal pronouns (Brousseau 1995: 18, from Waldman et al. 1981)

<table>
<thead>
<tr>
<th></th>
<th>French</th>
<th>Haitian</th>
<th>Fongbe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>moi</td>
<td>‘I, me’</td>
<td>nyè</td>
</tr>
<tr>
<td>2SG</td>
<td>toi</td>
<td>‘you (SG)’</td>
<td>‘you (SG)’</td>
</tr>
<tr>
<td>3SG</td>
<td>luielle</td>
<td>‘he, she, it’</td>
<td>li</td>
</tr>
<tr>
<td>1PL</td>
<td>nous</td>
<td>‘we, us’</td>
<td>‘we, us, you (PL)’</td>
</tr>
<tr>
<td>2PL</td>
<td>vous</td>
<td>‘you (PL)’</td>
<td>‘we, us, you (PL)’</td>
</tr>
<tr>
<td>3PL</td>
<td>euxelles</td>
<td>‘they, them’</td>
<td>yo</td>
</tr>
</tbody>
</table>

appears to unite creoles of different geographical areas is the main process – relexification – by which they come about.

In spite of this rather categorial conclusion, there is, nonetheless, one feature that creole languages appear to share: it is the fact that they tend to be isolating languages. I now turn to the discussion of this point.

3. Why do creole languages tend to be isolating?

The observation that creoles tend to be isolating languages goes back to Schuchardt (see the collection (1979)) and Hesseling (1933: xvi). It is also found in Hagège (1985: 39). Mufwene (1986, 1990, 1991) shows that this tendency appears to hold even when the contributing languages are not isolating ones. For example, he documents the fact that Kituba, a creole that has emerged almost exclusively from contact among agglutinative Bantu languages, is an isolating language. “Kituba has selected Kikongo’s seemingly marked periphrastic alternative over the more common and apparently unmarked agglutinating system” (Mufwene 1990: 12). More recently, McWhorter (1998: 792) has proposed that lack of inflectional morphology is a feature of the creole prototype.

How does the relexification account of creole genesis handle the fact that creoles tend to be isolating? The answer to this question lies in the way that functional category lexical entries acquire a label in creole genesis. According to Lefebvre & Lumsden (1994a, b), this is achieved in one of two ways. First, since the creators of a radical creole do not identify the functional categories of the superstratum language, because they do not have enough exposure to the language, they do not relabel the functional category lexical entries of their own lexicon on the basis of those of the superstratum language; rather, they relabel them on the basis of MAJOR-category lexemes (e.g., nouns, adjectives, verbs, adverbs, and adpositions) of the superstratum language. For example, the postnominal definite determiner of the Haitian substratum languages in (4) has been relabelled on the basis of a French postnominal adverb (see Lefebvre 1998: 78–84). Likewise, the tense, mood, and aspect markers of the substratum
languages of Haitian in Table 2 have been relexified on the basis of French periphrastic expressions (Lefebvre 1996, 1998: 111–140).

The absence of syntactic pronominal clitics in a creole whose contributing languages all have syntactic pronominal clitics (as is the case of Haitian, as we saw above) can also be argued to follow from this perspective. The following scenario is proposed in Brousseau (1995) for Haitian, which does not have syntactic pronominal clitics in spite of the fact that both its superstatum and substratum languages have them. Brousseau hypothesises that the creators of Haitian relexified the clitics of their own lexicon using French strong personal pronouns. Note that these French forms were also used to relabel the lexical entries copied from the strong pronouns. So, on this hypothesis, the copied lexical entries of all three Fongbe 1st person singular pronominal forms were relabelled on the basis of French moi, yielding mwen in Haitian, as is shown in (15) (where the syntactic features [+/- argument] stand for strong and clitic forms, respectively).

(15) Fongbe Haitian
a. [1st], [−plural], [+argument] nyè mwen
b. [1st], [−plural], [−argument], [+nominative] in mwen
c. [1st], [−plural], [−argument], [−nominative] mi mwen

Consequently, in the incipient creole, there would be three homophonous forms for the 1st person singular pronominal lexical entries. The availability, in the incipient creole, of the lexical entries in (15) would enable the creators of Haitian who had both strong and weak pronominal forms in their original lexicons to reproduce these forms in the creole. However, using the same superstratum string to relabel several lexical entries copied from the substratum language(s) yielded redundancy in the newly created lexicon. Brousseau (1995) thus further hypothesises that the three homophonous lexical entries in (15) were reduced to one, with their common features, yielding a single Haitian lexical entry unspecified for the features [α argument], where α is a variable that can take the values + or −, and [α nominative], where α is a variable that can take the values + or −. The reduced lexical entry is thus: /mwen/: [1st], [−plural]. The fact that this lexical entry is underspecified for the feature [α argument] also enabled the creators of Haitian, who had both strong and weak pronominal forms in their original lexicons to produce these forms while speaking the creole. Whether the first generation of Haitian native speakers was exposed to the data in (15) or to the reduced lexical entry just mentioned, they had no clue, however, for distinguishing between strong and weak forms on the basis of the data. Presumably, they observed the same form in all contexts where a pronominal was used by the adult population. Furthermore, Brousseau (1995) points out that the context par excellence where the clitic and the strong forms were distinguished in terms of word order in the original grammar – that
Commentary on McWhorter: Claire Lefebvre

is, in nominalisations – had been abandoned in the early creole. It is thus reasonable to conclude that the first generation of Haitian native speakers could not deduce the availability of syntactic clitics on the basis of the data that they were exposed to. Still according to Brousseau (1995), the first generation of Haitian native speakers presumably interpreted these data as: /mwen/: [1st], [−plural], [+argument].

Thus, in modern Haitian, there are no syntactic clitics. The fact that syntactic clitics did not enter the creole can thus be derived from how relabelling is hypothesised to proceed in the case of functional category lexical entries in creole genesis.

A second way by which a functional-category lexeme can acquire a label in creole genesis is through reanalysis. As will be seen in the next section, under specific circumstances, it may happen that such a lexeme cannot be relabelled at the time relexification is taking place. In this case, the copied lexical entry is assigned a null form, represented by Ø in Figure 1. As has been proposed in Lefebvre & Lumsden (1992, 1994b), a functional category lexical entry that has been assigned a null form at relabelling may be signalled by a periphrastic expression. For example, a lexical entry having a temporal/aspectual meaning but a null phonological representation may be signalled by the use of an adverb with a similar meaning. The periphrastic expression may later become the phonological representation of the lexical entry initially assigned a null form, through the process of reanalysis. Such cases are reported in the literature (for an example from Tok Pisin, see Sankoff 1991).

The fact that creoles are generally isolating languages thus follows from the relexification account of creole genesis described above. Since the functional-category lexemes of creole languages derive their phonological forms from major-category lexemes in the superstratum language, or from reanalysis, and since these categories are typically free morphemes, it follows that creoles will tend to be isolating languages (see Lefebvre & Lumsden 1994a, b).

5. Mufwene (1989: 124) accounts for the isolating character of creoles by appealing to the notion of salience: “With regard to the issue made here, viz., explaining why periphrasis is generally preferred to inflections in PC’s, I submit that salience should do.” The proposal advocated in our research is somewhat similar to Mufwene’s, for MAJOR categories may be viewed as “salient” when compared with MINOR categories. As is pointed out in Lefebvre (1998: 48), however, the proposal in Lefebvre & Lumsden (1994a) is more specific, since it links the observed facts to the processes that generate them, namely, relexification and reanalysis, and to the linguistic material on which these processes apply in creole genesis.
4. Why do creole languages look simpler?

McWhorter, in this issue, states that creoles are simpler than both their lexifier and substrate languages. On the relexification account of creole genesis assumed here, the issue of the alleged simplicity of creoles can only be taken up in terms of a comparison of a creole with its substratum languages. This is thus the methodology that I will adopt in addressing the question at stake in this section. Are creole languages really simpler than their substratum languages? Or do they just happen to “look” simpler? In the paragraphs that follow, I present a way of looking at the data that support the second alternative.

In my view, creoles only look simpler than their substratum languages. And the fact that they look simpler than their substratum languages lies in the fact that, due to constraints associated with the process of relexification, there are more covert lexical entries in creoles than there are in their substratum languages. By covert lexical entry, I mean a lexical entry that is required by Universal Grammar but that is phonologically null. In practical terms, this means that such a lexical entry has a syntactic function that can be argued for, but that it is not pronounced. A case in point would be the accusative case in English. This case is required by Universal Grammar (see Chomsky 1981), but in English it is covert, unlike in other languages, such as Quechua, that have overt case morphology. Another relevant example involves the optional pronunciation of the complementiser that in English, as in John said Ø he would come. There is a consensus in the literature that, when the complementiser that is not pronounced in a sentence of the aforementioned type, the syntactic position is nonetheless filled by the features of this complementiser, and the covert complementiser plays a syntactic role in the structure of the clause. In light of this preliminary discussion, I now turn to the discussion of phonologically null lexical entries produced at the time the process of relexification is taking place in creole genesis. It will be shown that phonologically null lexical entries are the results of constraints involved in relabelling.

According to Muysken (1981: 62), relexification is semantically driven: “For relexification to occur, the semantic representations of source and target language entries must partially overlap; otherwise, the two entries would never be associated with each other. Other features of the two entries may, but need not, be associated with each other.” In Lefebvre (1998: 17), I take the position that, in relexification, copying may apply to all lexical entries and that it is relabelling that is semantically driven. Thus, only those functional categories that have some semantic content (e.g., determiners, demonstrative terms, etc.) may be assigned a new label during relexification. Functional categories that have no semantic content (e.g., case markers, operators, etc.) are copied but they are not relabelled; they are phonologically null or covert; they are represented by zero in Figure 1. Practically speaking, this means that these lexical
Commentary on McWhorter: Claire Lefebvre

entries are not pronounced. As is pointed out in Lefebvre (1998: 17–18), the claim that functional categories may be assigned a null form at relabelling is independently motivated by the fact that, in natural languages, functional categories required by Universal Grammar are not always spelled out, as we saw above on the basis of data drawn from English.

In this respect, consider the Haitian and Fongbe nominal structures in (4), reproduced as (16) for convenience.

(16) a. krab [mwen Ø] sa a yo
   crab me GEN DEM DET PL
   ‘these/those crabs of mine (in question/that we know of)’

The possessive phrase that follows the head noun of these structures is comprised of a pronoun and a case marker. The case marker is overt in Fongbe but covert in Haitian (see Lumsden 1991). Since case markers do not have semantic content, the Fongbe case marker could not be relabelled, and thus, the copied lexical entry from this substratum case marker was assigned a phonologically null form at relabelling. On the basis of syntactic tests, Brousseau & Lumsden (1992) argue that Fongbe tɔn has the properties of Genitive case (=’s in English) rather than those of Objective case (=of in English). Lumsden (1991) argues that the Haitian possessive phrase has the same properties as the corresponding Fongbe one. He thus identifies the phonologically null case marker as Genitive.

Another example of a functional category that could not be relabelled because it does not have semantic content involves the operator found in relative and factive clauses. This operator is lexical in Fongbe (17a) but it is covert in Haitian (17b), as is illustrated in (17) involving factive clauses.

(17) a. Wá ɖeè Jan wá ʒ …
   b. Rive Ø Jan rive a …
   arrive OP John arrive DET
   ‘The fact that John arrived …’ (=3 in Lefebvre 1994b)

The properties of the Fongbe operator are extensively discussed in Kinyalolo (1993) and in Collins (1994). In Lefebvre (1998: 203–205), it is argued that the Haitian null form in (17) has syntactic features that are manifested in the syntax of the construction, and that these features parallel those of Fongbe ɖeè.

The two sets of data presented above illustrate cases where a phonologically null lexical entry in the creole results from the constraint that relabelling is semantically driven. There is another constraint that is involved in the process and that may also yield phonologically null lexical entries in the creole.
As was mentioned earlier, in Lefebvre & Lumsden (1994a) it is proposed that functional categories of the substratum languages that have some semantic content are relabelled on the basis of major category lexical items of the superstratum language that have some semantics in common and similar distributional properties. Relabelling is thus constrained by what the superstratum language has to offer in terms of appropriate phonetic strings to relabel a copied lexical entry. If no appropriate form is found, the copied lexical entry remains covert, that is without a label. This proposal accounts for differences observed between creoles formed from the same substratum languages but different superstrata.

For example, French based creoles of the Atlantic were able to reproduce the postnominal determiner of their substratum languages (see (4)) because French has an adverbial form that has the appropriate properties to relabel the copied lexical entry. Saramaccan, an English based creole with the same substratum languages as Haitian (cf. Smith 1987), however, was not able to reproduce its substratum languages’ postnominal determiner because English does not have an appropriate form to relabel the copied lexical entry. On the other hand the lexical -self anaphor of the substratum Gbe languages was reproduced in the English- and Dutch-based creoles because these superstratum languages have a -self anaphor: Berbice Dutch -zelfu (from Dutch -zelf; Robertson 1993: 307); Gullah -self (from English -self; Mufwene 1992: 169); Saramaccan -seéi (from English -self; Veenstra 1996: 43). Since French does not have a -self anaphor, French based creoles have a covert lexical entry in this case, as is illustrated in Figure 2.6 Thus, phonologically null lexical entries in a creole may result from the fact that the superstratum language does not have an appropriate form to relabel a copied lexical entry.

As can be seen from the above examples, there are more covert forms in the creole than there are in the substratum languages. As has been pointed out in Lumsden (1995), this makes the creole lexicons look “simpler” than the original lexicons. Furthermore, lexical entries that are not required by Universal Grammar, and that cannot be relabelled due to either one of the two constraints discussed above, may simply be abandoned. This is the case, for example, of the logophoric pronoun of the Haitian substratum Gbe languages (see Lefebvre 1998: 147). A logophoric pronoun is a pronoun that has no independent reference. Because they are not semantically independent, logophoric pronouns cannot be relabelled. There are no arguments that would support an analysis according to which there would be a null logophoric pronoun in modern

6. For extensive discussion of these facts, see Lefebvre (1998: 159–171). The idea that the lexical entry copied from the substratum -self anaphor could have been assigned a phonologically null representation at relabelling is attributable to John Lumsden (research seminar, Fall 1993). The further development of this idea is mine.
5. Apparent simplicity and hidden complexity

In this section, I would like to call the reader’s attention to some semantic interpretative facts showing that “what you see is not always what you get” and that “what you see is sometimes simpler than what you in fact get”.

Consider the predicate cleft construction in (13), reproduced as (18).

(18) a. Wá wì Jan wá.
   b. Se rive Jan rive.
   ‘It is ARRIVING that John did.’ (not, e.g., leave)

As is extensively shown in Lefebvre (1990), in this construction the clefted constituent may be assigned an interpretation that goes beyond what is actually found in the clefted phrase. For example, the clefted constituent in (19) may be assigned three different contrastive interpretations: one bearing on V, another one bearing on the VP, and a last one bearing on the internal argument of the verb.

(19) Se manje Jan manje pen an
    ‘It is EATING the bread that John did.’ (not, e.g., throw it away)
    ‘It is EATING THE BREAD that John did.’ (not, e.g., wash the dishes)
    ‘It is eating THE BREAD that John did.’ (not, e.g., eat the apple)
    (Lefebvre 1990: (44))

The example in (20) presents similar focus ambiguities.
Finally, when the affected argument of a verb has been clefted, the contrastive interpretation of the cleft constituent bears either on the noun phrase or on the whole VP. This is illustrated in (21).

(21) Se pen an Jan manje.
    it-is bread DET John eat
    ‘It is the BREAD that John ate.’ (not, e.g., the apple)
    ‘It is EATING THE BREAD that John did.’ (not, e.g., wash the dishes)
    (Lefebvre 1990: (53))

The semantic interpretation facts in (18)–(21) are not directly accessible from the surface structures and they require semantic rules of interpretation that do far more than just establishing a one-to-one correspondence between the surface structures and their interpretations. Larson & Lefebvre (1991) analyse these facts in terms of quantification of events. These facts, and others of the same type that are discussed in Lefebvre (1998, in press), show that some Haitian grammatical properties are more complex and certainly more opaque than a “simplicity” approach to creole languages would lead one to believe.

6. McWhorter’s list revisited

McWhorter, in his target article, provides a list of fourteen features that he claims will never be found in a creole language. As he puts it: “Crucially: One would find a great many of the above features in the lexifier and substrate languages that were spoken by the creators of these creoles”. In the theory of creole genesis advocated in the previous sections of this paper, the creators of a creole do not have enough access to the superstratum language to learn the functional categories of that superstratum language. Thus, on this approach, the lexifier language is not pertinent to explain the absence, in creoles, of the list of items (almost all related to functional categories) provided by McWhorter. Only the substratum languages are pertinent for the discussion of this list. So, in the paragraphs that follow, I discuss McWhorter’s list with respect to the substratum languages of Haitian, mainly Fongbe, and occasionally, other West African languages. Then, I propose a global evaluation of these features.

None of the substratum languages of Haitian have ergative case; but even if they did, on the theory of relexification outlined in this paper, ergative case
would have been assigned a null form at relabelling, and thus, it would not be visible in the incipient creole. Gbe languages do have evidential markers (see Lefebvre & Brousseau to appear). As is shown in Lefebvre (1998: 213–217), Haitian has a subset of those. Inalienably possessed objects must appear in the Genitive case in Fongbe (see Lefebvre & Brousseau to appear). As we saw in (4), due to the semantic constraint on relabelling, Genitive case is covert in Haitian and so there is no way to tell whether inalienably possessed objects occur in the Genitive or in the Objective case. Fongbe does not have switch reference, inverse nor obviative marking. If it did, it is unlikely that these morphemes would have made their way into Haitian because, provided that they have enough semantics to be relabelled, there may not be any appropriate French phonetic string to relabel the substratum morphemes. Fongbe, like the other Gbe languages, does not have verb raising to INFL (see Table 4), and thus it does not manifest verb-second phenomena, nor the syntactic asymmetries between matrix and subordinate clauses that go with them. As we saw in Table 4, verb raising to INFL (and eventually to a higher position in the syntactic tree) correlates with inflecional morphology on the verb. Haitian (see Table 4) follows the pattern of its substratum Gbe languages with respect to this parametric option: neither have inflectional morphology on the verb. For subjunctive marking, Gbe languages encode this mood by means of a preverbal marker. As we saw in Table 3, this preverbal marker was reproduced in Haitian by relexification. Gbe languages present a few cases of syntactic clitic movement (see Lefebvre & Brousseau to appear). As was shown in Section 4, syntactic clitics are not reproduced in a creole as a consequence of how relabelling proceeds in the case of functional-category lexemes. It follows that, unless a creole develops syntactic clitics, clitic movement will not be found in incipient creoles.

McWhorter claims that creoles will manifest only an SVO word order. As is shown in Lefebvre & Brousseau (to appear), Fongbe manifests a surface word order that is SVO in some contexts (mainly finite clauses) and OVS in others (mainly nominal and nominalised structures). In Lefebvre & Lumsden (1992), it has been proposed that word order in a creole is established in the following way. Because the creators of the creole are aiming to reproduce the superstratum sequences they are exposed to and since they are able to identify the major category lexical entries, the word order of major category lexical items and major constituents in the creole will follow that of the lexifier language. However, because the creators of a creole do not have enough exposure to the superstratum language, they cannot identify its functional-category lexemes; when they relexify the functional-category lexemes of their native lexicons, they keep their original directionality properties. Hence, these items are predicted to have the same word order as in the substratum languages. The Haitian data presented in Lefebvre (1998) show that this hypothesis is borne
out. The data in Table 3 and in the nominal structures in (4) constitute examples in point. (For further discussion of this issue, see Lefebvre 1998: 388–390.) Additional evidence for this claim comes from Berbice Dutch. Kouwenberg (1992) reports that Eastern Ijo, Berbice Dutch’s main substratum language, is underlingly an OV language. Dutch, the lexifier language, is also underlingly an OV language. Berbice Dutch itself is a VO language. Kouwenberg explains this situation as follows. In Dutch simple clauses, the verb moves to INFL such that, at surface structure, Dutch simple sentences exhibit the order SVO. According to Kouwenberg, the creators of Berbice Dutch perceived this order and hence established SVO for the creole.7

Gbe languages used to have noun classes; the latter are attested by frozen forms in the modern varieties. Whether these noun class prefixes were still productive at the time Haitian Creole was formed is unknown to me. The fact that Haitian Creole does not have noun class prefixes, however, suggests that they were probably no longer productive in Gbe at the relevant time. This claim is supported by the fact that all the productive morphology of Gbe has been reproduced in the creole, as is extensively demonstrated in Lefebvre (1998: 303–334, and the references cited therein). Finally, while Fongbe has phonological tones, Haitian Creole does not (see Cadely 1994).

This terminates the discussion of each feature in McWhorter’s list on the basis of Haitian and its substratum languages. I now turn to a more global evaluation of the facts discussed above.

I begin with the facts that are not in agreement with McWhorter’s claim. The Haitian data involving the subjunctive and the evidential markers constitute counterexamples to McWhorter’s claim. The fact that Haitian does not manifest verb-second phenomena is irrelevant to its being a creole. As we saw earlier, the availability of this option in a particular grammar correlates with the availability of inflectional morphology in that particular grammar. Other languages, not identified as creoles, lack inflectional morphology, and hence, verb movement to INFL and, in some cases, to COMP. Chinese is a case in point. In turn, the presence of this feature in McWhorter’s list is in contradiction with the property that he claims characterises the items in his list: “Crucially, none of these factors require inflectional morphology for their occurrence in a grammar, and thus their absence is not an epiphenomenon of isolating typology” (McWhorter, this issue). Consequently, the features in this first group should be dropped from McWhorter’s list of features that are excluded from creoles.

---

7. As noted by Kouwenberg (1992), however, Berbice Dutch has postpositions. This should come as no surprise since Dutch also has postpositions. The fact that Saramaccan has postpositions (cf. Müysken 1987) when its English lexifier language does not, however, constitutes a counterexample to the proposal in Lefebvre & Lumsden (1992) on how word order is established in an incipient creole. The latter data require further investigation.
A second group of features comprises those that can be derived from the relexification account of creole genesis outlined in the previous sections. For example, all the features that are related to case marking (that is, lack of ergative case, lack of overt genitive case, and lack of particular case distinction for inalienable possession) are derivable from the semantic constraint on relabelling; on this constraint, case markers are not relabelled in relexification, and thus, they are predicted not to be overt in the incipient creole. Likewise, under the condition that they have enough semantics to be eligible for relabelling, markers or morphology involving switch, obviative, or inverse reference cannot be relabelled for lack of appropriate material in the superstratum language. Similarly, the lack of clitic movement in creoles follows directly from the lack of syntactic clitics in these languages. As we saw in Section 4, the lack of syntactic clitics in a creole is derivable from the way relabelling is hypothesised to proceed in creole genesis. The fact that creoles are SVO is also derivable from the proposal concerning how word order is established in creole genesis contexts, even in cases where contributing languages are SOV. So, all the features in this second group are derivable from a sound theory of how creole languages come about.

Finally, the absence of tones in creoles may be due to the mixed character of these languages. For example, as is argued in Brousseau (in preparation), the accentual system of Haitian represents a principled compromise between the tonal system of its Gbe substratum languages and the extremely simple accentual system of French. Likewise, the phonological system of Haitian represents a principled compromise between the phonology of its contributing languages. This is just like the relexified lexical entries which represent a principled compromise between the properties of the substratum lexical entries and those of the superstratum language (see Figure 1 and (2)).

On the basis of this global evaluation of McWhorter’s list, my conclusion is the following. The first group of features should be removed from the list because they do not stand in the face of the counterexamples that have been presented on the basis of Haitian. The second and third groups of features should be retained. These are the ones that can be derived from the relexification account of creole genesis presented in earlier sections of this paper.

7. Conclusion

The central thesis advocated in this paper is that creole languages are not so “simple” as they may look on the surface. It was shown that the creators of creoles are adult native speakers who use the properties of their own lexicons and grammars in creating the creole. The bulk of a creole’s lexicon is thus created through the process of relexification. This account of creole genesis predicts that creoles reproduce the properties of their substratum languages, in such a
way that creoles from different geographical areas will manifest the same type of differences among themselves as their respective substratum languages do. It was proposed that what appears to unite creoles of all geographical areas is the main process – relexification – by which they come about. In spite of this strong conclusion, it was shown that creoles appear to share at least one feature in the fact that they tend to be isolating languages. It was argued that this property of creoles follows from the relexification account of creole genesis. Regarding the issue of simplicity per se, it was shown that, due to constraints associated with the process of relexification – the fact that relabelling is generally semantically constrained, and the fact that relabelling is, in particular, constrained by what the superstratum has to offer to relabel a copied lexical entry – there are more covert lexical entries in creoles than there are in their substratum languages. This makes creoles “look simpler” than the original lexicons. Semantic interpretation data were presented showing that apparent simplicity may hide effective complexity. Finally, the evaluation of the features proposed by McWhorter to be absent from creoles led to an interesting conclusion. Putting aside the few features that should be removed from the list, it is possible to derive the other ones from a theory of creole genesis based on the major process that is a work in creole formation. A list that comprises features that seem heterogeneous at first glance thus acquires some motivation when related to the process that creates the languages under discussion in this special issue.

Received: 11 May 2001 Université du Québec à Montréal Revised 28 June 2001 Max-Planck-Institut für evolutionäre Anthropologie

Correspondence address: Département de linguistique, Université du Québec à Montréal, C.P. 8888, succ. Centre-Ville, Montréal, Québec H3C 3P8, Canada; e-mail: lefebvre.claire@uqam.ca

Acknowledgements: The content of this paper builds on a public debate on issues in creole studies between myself and John McWhorter. The debate took place March 22, 2001 at the University of California at Berkeley. I want to thank Suzan Ervin-Tripp, Dan Slobin, and all the participants to this debate for their comments and insightful questions. I want to thank Bernard Comrie, David Gil, and Frans Plank for their comments on a first version of this paper. Thanks to Andrée Bélanger who helped format it. The research underlying this paper was financed by CRSH, FCAR, and FIR-UQAM.

In the course of this research, I had several major collaborators whom I wish to thank for their contribution at some point or other: Jonathan Kaye, Diane Massam, John Lumsden, Anne-Marie Brousseau.

References

Commentary on McWhorter: Claire Lefebvre


Commentary on McWhorter: Claire Lefebvre

On the origin of creoles

A Cartesian critique of Neo-Darwinian linguistics

by Michel DeGraff

1. Neo-Darwinian creolistics: Whence and whereto? A sketch

The main goal of this essay is to constructively deconstruct the age-old myth that “the world’s simplest grammars are creole grammars” and to demystify the methodological (mis)practices that underlie this myth and its corollaries throughout creole studies and beyond.

I start with some notes on historiography and methodology, connecting certain trends in 20th- and 21st-century creolistics to outdated (quasi) Darwinian concepts in early 19th-century comparative-historical linguistics. Then I move to linguistics per se, inspecting the empirical and theoretical bases of creolists’ foundational assumptions about creole diachrony and synchrony. This will (re-)establish the epistemological limits of certain key terms in creole studies, including “pidgin(ization)”, “creole/creolization”, “young” vs. “old”, “simple” vs. “(most) complex”, etc. I will argue that these terms, although per-
haps useful as a-theoretical heuristics and as sociohistorical approximations, cannot serve as theoretically-grounded linguistic-structural taxa: Given Universal Grammar and its Cartesian-Uniformitarian foundations (Chomsky 1966, 1981, 1986, 1995, etc.), there cannot be any invariant and sui generis set of structures and processes that fall under the labels “creole” and “creolization.” Assuming Universal Grammar, creolization reduces at the individual level to the same sort of cognitive processes that underlie idiolect formation through language change, and so are creole languages aprioristically undistinguishable from non-creole languages – that is, there is no synchronic creole typology that excludes non-creole languages. As I show below, such claims go against the grain of the most ancient and the (still) most prevalent dogma in creole studies.

1.1. Historiography and epistemology: From Schleicher to Popper

My essay is best introduced by the following quotes from Schleicher (1863), Saint-Quentin (1872), and Adam (1883) on simple(st) grammars and from Ost-hoff & Brugman (1878), Foucault (1972), and Popper (1965) on methodology. The first two sets of quotes ((1)–(2)) are words from the past and the last three ((3)–(5)) are words of caution for the future, and all five are relevant to linguists’ time-honored search for simplest grammars. These quotes speak for themselves, and eloquently so.

1.1.1. (Pre-)Darwinian linguistics: Schleicher on the “Tree of Language”

(1) a. THE RULES NOW, WHICH DARWIN LAYS DOWN WITH REGARD TO THE SPECIES OF ANIMALS AND PLANTS, ARE EQUALLY APPLICABLE TO THE ORGANISM OF LANGUAGES, THAT IS TO SAY, AS FAR AS THE MAIN FEATURES [of Darwin’s theory] ARE CONCERNED. (Schleicher 1863 [1983: 30], emphasis in original; also see 1983: 16–17)

b. The construction of all languages points to this, that the eldest forms were in reality alike or similar; and those less complex forms are preserved in some idioms of the simplest kind, as, for example, Chinese. […] In this remote stage of the life of speech, there is consequently no distinction […] between verbs and nouns; there is neither declension nor conjugation.1

(Schleicher 1863 [1983: 51])

---

1. Darwin himself was somewhat ambivalent about the use of morphosyntax as a measure of complexity cum perfection; see, e.g., Darwin (1871: Chapter 2, 61–62).
c. [L]anguage is of significance not only for the elaboration of a scientific [i.e., taxonomic] systematization of humanity, but also for the evolutionary history of man. […] Now language has revealed itself to science as something that has evolved very gradually […] The comparative anatomy of languages shows that the more highly organized languages evolved very gradually out of simpler language organisms, probably in the course of very long time spans. 

(Schleicher 1863 [1983: 79])

1.1.2. Neo-Darwinian creolistics: Saint-Quentin and Adam on simple(st) languages

(2) a. [Creole grammar] is, therefore, a spontaneous product of the human mind, freed from any kind of intellectual culture. […] When one studies its structure, one is so very surprised, so very charmed by its rigor and simplicity that one wonders if the creative genius of the most knowledgeable linguists would have been able to give birth to anything that so completely reaches its goal, that imposes so little strain on memory and that calls for so little effort from those with limited intelligence. An in-depth analysis has convinced me of something that seems paradoxical. Namely: if one wanted to create ab ovo an all-purpose language that would allow, after only a few days of study, a clear and consistent exchange of simple ideas, one would not be able to adopt more logical and more productive structures than those found in creole grammar.

(Saint-Quentin 1872 [1989: 40–41]; my translation)

b. [Cayenne Creole] grammar […] is nothing but the grammar that is common to the languages of Guinée. The latter we can call langues naturelles as opposed to langues cultivées. For the botanist, plants that are naturelles are superior to plants that are cultivées to the extent that the former are pristine products that are free of intentional adulteration. Likewise, for the linguist, the speech of peoples considered primitive has primacy over the speech of civilized peoples: the former is closer to the sort of grammatical instincts of which children’s utterances reveal processes that are simple, logical and fast. […] [Cayenne Creole] grammar is more naturelle than that of Sanskrit, Latin, and French. But this grammar did not spontaneously emerge in Guy-
ane; it was imported from Africa.

(Adam 1883: 4–5; my translation)²

In reality, the Malagasy slaves [who created Mauritian Creole] have brought along to Mauritius their native grammar, but not the forms that I have previously mentioned [inflectional gender marking, inflectional plural marking on verbs, the *avoir* (‘to have’) auxiliary, the verbal copula]. These forms are the product of an evolution that has not happened in the Polynesian and Melanesian languages. In dealing with such forms, Malagasy speakers kept their native grammar […] In Mauritius, this native grammar reasserted its influence.

(Adam 1883: 7; my translation)

1.1.3. **On Cartesian-Uniformitarian linguistics: The Neogrammarians**

(3) a. These [methodological] principles are based on a two-fold concept, whose truth is immediately obvious: first, that language is not a thing which leads a life of its own outside of and above human beings, but that it has its true existence only in the individual, and hence that all changes in the life of a language can only proceed from the individual speaker; and second, that the mental and physical activity of man must have been at all times essentially the same when he acquired a language inherited from his ancestors and reproduced and modified the speech forms which had been absorbed into his consciousness.

(Osthoff & Brugman 1878 [1967: 204])

b. If someone could once and for all manage to get rid of these generally harmful expressions “youth” and “old age” of languages! These and many others in themselves quite innocent grammatical terms have so far been almost exclusively a curse, hardly a blessing. For the child who was born in Greece in the Homeric age, who became aware of the speech forms of his linguistic community by hearing them, and who then reproduced them in order to make himself understood by his fellow men – for that child were these speech forms ancient?

(Osthoff & Brugman 1878 [1967: 205–206])
1.1.4. **Foucault on “chimera and reverie” in linguistics**

(4) Suffice it to recall that the quest for primitive language, a perfectly acceptable theme up to the eighteenth century, was enough, in the second half of the nineteenth century, to throw any discourse into, I hesitate to say error, but into a world of chimera and reverie – into pure and simple linguistic monstrosity.

(Foucault 1971 [1972: 223])

1.1.5. **Popper on (criticisms of) myths as science**

(5) Thus science must begin with myths, and with [...] the critical discussion of myths, and of magical techniques and practices. [...] The critical attitude, the tradition of free discussion of theories with the aim of discovering their weak spots so that they may be improved upon, is the attitude of reasonableness, of rationality. It makes far-reaching use of both verbal argument and observation – of observation in the interest of argument, however.

(Popper 1965: 50)

1.2. **Schleicherian roots of Language and route to progress: From isolating to agglutinative to inflectional/fusional**

Taken together, the quotes in (1)–(5) suggest that the search for “the world’s simplest grammars” has been going on for quite a while, and so has the dogma that “the world’s simplest grammars are creole grammars”. This search for “simplest grammars” is part of a larger search for, and larger myths about, the origins and evolution of our species. This search seems driven by an apparently innate drive to (in Schleicher’s phrase) “classify humanity”.

Schleicher’s *Glottik* is linguistics as natural history. By definition, its morphological taxonomy of languages qua organisms is regulated by universal natural laws and reducible to the genealogical mapping of their teleological development toward “idioms of higher organization” (Schleicher 1850, 1863, 1865, etc.).

In the intellectual climate of early 19th century, Schleicher’s pre-Darwinian language-as-organism evolutionary approach belonged to the “normal science” of his period (“normal” in the Kuhnian sense). In retrospect, Schleicher’s *Glottik* can be viewed as one of the central myths that gave vigor, status, and popularity to comparative-historical linguistics; see Hoenigswald & Wiener (eds.) (1987) and Alter (1999) for comprehensive overviews and, specially, for an array of positions similar to Schleicher’s through much of the 19th century and beyond.

At the core of Schleicher’s language-as-organism evolutionary hypothesis was a then-attractive congruence between, on the one hand, the simple-to-
complex Darwinian evolution of biological organisms from “one-celled organisms” to “higher living beings” and, on the other hand, the postulated historical progression of languages, through variation and subspeciation, from isolating to agglutinative to inflectional/fusional (Schleicher 1863 [1983: 50–60], also Schleicher 1850 [1852: 6–13]). In the Schleicherian organic school, linguistic evolution, on a par with biological evolution, was to be modeled by a “tree of life”, a *Stammbaumtheorie* – a genealogical (“family” tree) diagram depicting phylogenetic relationships, with (at most) one parent for each daughter node.\(^3\)

Schleicher tries to establish his morphology-as-biology congruence by, inter alia, metaphorically taking the “simple cell” and the “simple root” as “the common primitive forms” of biological and linguistic evolution, respectively (Schleicher 1863 [1983: 55]). More explicitly, Schleicher takes “the radical elements [i.e., isolating root morphemes] as the CELLS of [prototypical primitive] speech” (1863 [1983: 53], emphasis in original) and posits an isolating proto-language – one made up exclusively of affixless roots – at the evolutionary beginning of each language phylum. The proto-language’s monomorphemic words are the linguistic analogues of the “one-celled organisms” at the roots of biological evolution. Thus, Chinese with its tendency toward isolating morphology is most primitive while Sanskrit with its inflectional/fusional morphology is most advanced since, according to Schleicher and others, inflectional/fusional morphology marks the highest degree of complexity and perfection (see (1b) and (2b); also see Schleicher 1850). It is thus that morphology – inflectional morphology, in particular – has long served as the chief measure of evolutionary progress and/or structural complexity (but see Note 1).\(^4\) Related measures are found in 20th- and 21st-century linguistics, as in the works of, e.g., Jespersen (1922: 233–234), Whinnom (1971: 109–110), Samarin (1980: 221), Seuren & Wékker (1986), Bickerton (1988: 274–276), Comrie (1992: 208–209), Seuren (1998: 292–293), McWhorter (1998, 2000a, b, this volume),

---


On the origin of creoles

Time (i.e., language age) is a critical factor in Schleicherian models of linguistic complexity. In such genealogical-cum-teleological approaches to the linguistic *systema naturae*, time is built-in as a prerequisite for the development of complexity in both biological and linguistic evolution: complexity qua “higher organization” takes “very long time spans” to evolve (see (1c)). Thus, the following three related propositions:

(6) a. Complex species (i.e., those more advanced in the evolutionary hierarchy, connoting greater perfection) must be relatively old.5

b. Young languages and the early ancestors of old languages are necessarily simple (i.e., more primitive in the evolutionary hierarchy).

c. The youngest/earliest languages are the simplest (i.e., most primitive).

In the domain of Language, the hypothetical original species – Schleicher’s ultimate *Ursprachen* – have had no time to evolve any complexity; thus they

5. Beyond old age, languages in Schleicherian linguistics enter into a stage of “senility” in which their inflection (if any) becomes moribund. Schleicher calls this degenerative phase “historical life” as opposed to “language evolution”, with the former being a “retrograde metamorphosis” of the latter. In other words, “history” follows, and undoes the effect of, “evolution”. And Schleicher adds that “retrograde metamorphosis” in morphology (e.g., the reduction of morphological complexity) is proportional to the “historicity” of the corresponding people. As one case study, Schleicher compares English with Icelandic. English speakers have been more historically active than Icelandic speakers: the latter remain relatively immune from language contact in continental Europe. Therefore morphology is more simplified in English than in Icelandic. (See Schleicher 1850 [1852: 23–30]; also see Maher 1983: xxviii–xxix for discussion of Schleicher’s views on “historicity” and Trudgill 1989 and references therein for sociolinguistic interpretations of “historicity” sans “senility”.)

As of Humboldt (1836 [1988: 203–213]), he takes “the wearing-down of inflections [which] is an undeniable fact” as a consequence of “the mind’s progress”, alongside “vernacularization” and language contact (e.g., “foreign immigrations”). Given Schleicher’s view above, the interesting, if surprising, observation here is Humboldt’s (1836 [1988: 205]) remark that morphological decay results from “the mind’s progress”: “The more mature the mind feels itself to be, the more boldly it works in combinations of its own, and the more confidently it casts away the bridges that language constructs for the understanding.” For Humboldt, this “more mature” genius, whose maturity is due to inter alia the intellectual development made possible by inflection, can cleverly decide to replace synthetic structures (e.g., affixes for nominal case and verbal tense) with analytic structures (e.g., prepositions and preverbal auxiliaries), thus promoting semantic transparency and ease of articulation (Humboldt 1836 [1988: 206]).

Taking Schleicher’s and Humboldt’s teleological-genealogical programs to their logical consequences, one must then claim that Prototypical Creoles, with their (alleged) “lack of inflection” (see (11)), are either most “senile”/“retrograde” or most “mature” in revealing “the mind’s utmost progress”(1).
are located at the prototypical primitive stage (i.e., the isolating/affixless stage) while “languages of a higher organization [e.g., Indo-Germanic] have arisen from simpler forms, through a process of gradual development” (Schleicher 1863 [1983: 50]). It is in this vein that Schleicher (1865 [1983: 79]) takes language to be “of significance not only for the elaboration of a scientific [i.e., taxonomic] systematization of humanity, but also for the evolutionary history of man” (cf. Note 4).

1.3. A foundational myth: Creoles as contemporary Ursprachen

“Systematization of humanity” via linguistic structures and the (alleged) lack thereof is also found at the very inception of creolistics and throughout its existence. Creolists’ own “classic” systematization is based on the age-old orthodoxy that creole morphology is EXTRAORDINARILY simple or simplified from a diachronic and/or synchronic perspective.6

In the 17th to 19th centuries, such orthodoxy had explicitly “race”-based underpinnings. One taken-for-granted piece of “normal science” revolved around the notion that non-whites were inferior human beings, and so was non-whites’ speech considered inferior to whites’ speech. For candidate (perhaps ambiguous) illustrations of such beliefs, see the works of early creolists such as Pelleprat (1655), Saint-Quentin (1872) (see (2a)), Baissac (1880), Adam (1883) (see (2b)), along with 18th/19th-century dictionary and encyclopedia entries for the word “creole” (see, e.g., Pierre Larousse’s 1869 Grand dictionnaire universel du XIX siècle and Vinson’s entry in the 1889 Dictionnaire des sciences anthropologiques). In such works, the distinctive features of creoles were not due to socio-historical factors only, but they were also taken to reflect the inferiority of their (non-European) speakers. The latter were deemed to be cognitively unable to master the “complexities” of European languages. Per this orthodoxy, “primitive” people spoke “primitive” languages (see Section 3.1; cf. (2) and Note 35).

1.4. Toward Cartesian-Uniformitarian creolistics

At the turn of the 19th century the Neogrammarians, while adopting Schleicher’s seminal insights in comparative-historical methodology (e.g., with re-
On the origin of creoles

221

spect to the reconstruction of untested proto-forms), fought hard to get rid of his excessively organic and teleological metaphors. The Neogrammarians’ stated goal was to understand Language and language change through the study of variation among (related) idiolects and through the study of universal laws (e.g., sound-change laws) that are ultimately rooted in the psychology and physiology of individual speakers. This was a shift of interest from the reconstruction of Ursprachen from archives to the analysis of contemporary idiolects in vivo (i.e., as manifested in individual speech). For Neogrammarians (see (3)), grammars live in speakers’ minds, not in society; thus, the study of individual grammars as manifestations of a “psychical [i.e., psychological] organism” should take epistemological priority over “Historical Grammar” (i.e., “descriptive grammars of different periods [… tacked together”): only the former is truly “scientific”; see Paul (1890 [1970: xxxvii, xliii, 1–19]). As Osthoff & Brugman (1878 [1967: 198]) put it, what needed correction is the methodology whereby “[I]languages were indeed investigated most eagerly, but the man who speaks, much too little”.

Continuing this “Cartesian” (i.e., universalist and mentalist) trend into the 20th century, Boas and his students, including Sapir, assumed both the existence of language universals, the “psychological reality” of idiolects, and the independence of phenotypes, language and culture across humanity; see, e.g., Boas (1911: 11), Sapir (1921: ix, 207–220). Both Boas and Sapir reacted strongly against the sort of Romantic ethnocentrism inherent in Schleicherian correlations between morphology, age, and complexity. About the (non-)rapport between linguistic morphological types and cultural evolution, Sapir (1921: 219) wrote:

(7) [A]ll attempts to connect particular types of linguistic morphology with certain correlated stages of cultural development are vain. Rightly understood, such correlations are rubbish. […] Both simple and complex types of language of an indefinite number of varieties may be found spoken at any desired level of cultural advance. When it comes to linguistic form, Plato walks with the Macedonian swineherd, Confucius with the head-hunting savage of Assam.

Sapir’s view is robustly supported by crosslinguistic evidence both within and across genetic phyla, among languages of similar and dissimilar time depths (e.g., English, Icelandic, German, Chinese, Wampanoag, Kivunjo, and Nicaraguan Sign Language). In a related observation, Thomason (1980: 361) pointedly notes that inflectional morphology can vary greatly within single genetic groupings, as in Indo-European:

(8) [T]he general pattern of development from flexional to isolating morphology is well known. But this progression is much more advanced
in some branches of the family than others, and the least elaborate inflectional systems in modern Indo-European languages – those of the western European languages, say – bear little resemblance to the most elaborate systems, such as noun declension in most Balto-Slavic languages. It is not just a matter of a reduction in the number of cases or gender distinctions; often, the categories themselves have changed.

Sapir’s and Thomason’s observations are worth keeping in mind throughout the discussion in this paper.

The Schleicherian genealogical classification of language by age cum complexity loses further grounding with the advent of Chomsky’s “Cartesian Linguistics” in the second half of the 20th century (see, e.g., Chomsky 1966, 1981, 1986, 1995). As in the Neogrammarian dogma in (3), grammars are inherently parts of human biology, not autonomous living organisms that undergo birth, age, senility, and death independently of their speakers. Generative linguistics’ objects of study are, in Cartesian mode, internal properties of individual minds (i.e., mental grammars qua I(NTERNAL)-LANGUAGES). Thus, generative linguistics is “internalist biolinguistic inquiry” (Chomsky 1995: 1–11, 2001: 41–42) and is intrinsically Uniformitarian (cf. Descartes’s assumption that “reason is by nature equal in all men” and Descartes’s notion of knowledge as mental representations; see Chomsky 1966 for relevant discussion).

Per current assumptions in biology, the basic morphology of the human brain is uniform across the species. It has thus become more difficult, if not impossible, to theoretically correlate biological evolution with crosslinguistic variation, unless one adopts a quasi-Lamarckian view of language change whereby crosslinguistic typology can be reduced to genetic variation across human groupings. In this quasi-Lamarckian scenario, language-specific structures (e.g., isolating vs. agglutinative vs. inflectional/fusional morphology) would be correlated with variations in the human genetic blueprint. This is not a likely scenario given current results in language acquisition and biology: linguistic structures do not evolve and are not transmitted like DNA.7

7. Any congruence between genetic and linguistic evolution is mediated via geographical isolation, population displacements, and the like, not by any biological causal relationship. See Hoensgawal & Wiener (eds.) (1987), Bateman et al. (1990), and Cavalli-Sforza (2000: Chapters 5, 6) for some comparison of (the mechanisms underlying) genetic and linguistic evolution. Current Anthropology (1990, vol. 31, numbers 1–4) offers diverging opinions on the methodology of such comparisons. But the point remains that Stammbaumtheorie, even if useful for approximating population displacements through time, is at the very best only in rough correlation with genetic transmission. (Also see Note 18.)
1.5. Back to Schleicher(ian creolistics)'s Ursprachen

Notwithstanding current advances in Cartesian and Uniformitarian (bio)linguistics, there is a sense in which Schleicher’s approach has survived the 19th century into the 20th and now 21st century, albeit under new theoretical guise. While modern linguistics is making steady advances in its exploration of our intrinsically human and species-uniform Universal Grammar (UG), “creolistics” has kept up, and even revived, early 19th-century notions of language evolution. Indeed, creolistics is perhaps the only field where the search for a genealogical and typological class of “simplest grammars” is still at the center of contemporary research. It is thus that certain trends in creolistics are reviving Schleicher’s Glottik with creole languages as the new class of youngest, thus structurally simplest, linguistic species. In this modern Glottik, creole languages are living specimens of Ursprachen, i.e., contemporary proto-languages – “the world’s only instantiation of spoken language having been ‘born again’” in McWhorter’s (Section 2.3) evangelical phrase.

However, an empirically and theoretically grounded implementation of the claim that the world’s simplest and/or most optimal grammars are creole grammars still constitutes the holy grail of creolists through an unbroken lineage of research programs (see Note 6). In the 20th and 21st centuries, starting with, e.g., Jespersen (1922) and up to McWhorter’s target article in this volume (henceforth WSG³) linguists have used morphosyntax to try and identify sufficient and necessary conditions that would make creoles deeply special in a structural and synchronic sense. Perhaps unsurprisingly, many of these contemporary observations on creole morphosyntax are quite reminiscent of those encountered in 17- to 19th-century texts; see DeGraff (2001b: 88–98) for further details. For example, consider the statement that “creoles are natural languages reborn from a radical reduction of their source languages into makeshift jargon” (WSG: 144). This statement finds direct antecedents throughout creole studies, from its very inception (see Note 6).

An explicit quantification for the alleged maximal simplicity of creole grammars is offered in WSG. Even though it makes no reference to Schleicher, McWhorter’s recent work (1998, 2000a, b, WSG) is, of late, the most sustained and perhaps most widely read effort to articulate a structural basis for Schleicherian linguistics, with the aim of categorizing languages according to some explicit complexity hierarchy. McWhorter tacitly assumes the Schleicherian dogma whereby linguistic typology must, at all costs, include a genealogically and structurally well-defined class of “simplest grammars”. In his

8. When making reference to (parts of) WSG, I will explicitly use “WSG”, possibly followed by (sub)section or page number. Otherwise, (sub)section numbers without further indication refer to my own commentary on WSG.
revamped complexity scheme, Prototypical Creoles are the new Ursprachen, languages created ab ovo from virtually “ground zero” complexity (WSG: Section 4.4). This is somewhat reminiscent of Saint-Quentin’s structural claims in (2a). Thus, McWhorter’s hierarchy continues the Schleicherian tradition of putting certain languages (here Prototypical Creoles) in a deeply-special class of linguistic neonates – contemporary fossils of Language at its evolutionary incipience. In a nutshell, the argument is that “because so much of a grammar’s complexity results from the operation of random accretion over time, creoles display less complexity than the rest of the world’s natural grammars” (WSG: 133).

The central assumptions here are (i) that creole languages are markedly younger languages than non-creoles, and (ii) that this age difference is linguistically measurable and significant, contra Osthoff & Brugman’s (1878) admonition in (3a). These assumptions are related to THE foundational claims in creole studies, namely the oft-repeated dualist statement that creoles are “non-genetic” languages that emerge via an abnormal “break in transmission” whereas non-creole languages gradually evolve “genetically” via “normal transmission” (the modern locus classicus for this claim is Thomason & Kaufman 1988: 8–12, 206, and passim; see Section 3.3 for discussion). In the classic creole-genesis scenarios, creole youth stems from the “pidgin-to-creole life cycle” as signaled by the concomitant morphological bottleneck (see references in Section 1.2 and in Note 6). In the most recent exponent of this dogma, creoles are “born as pidgins, and thus stripped of almost all features unnecessary to communication” (WSG: Abstract).

1.6. Toward Cartesian creolistics (redux): A guide for “learning by debunking”

As will become obvious through the development of the present critique, current research on creoles as contemporary Ursprachen (as, e.g., in WSG) presents us with a “modern” collage of pre- and neo-Darwinian claims about language evolution. The antecedents of such claims go back to the structural-cum-genealogical speculations of, e.g., Schleicher (1863) (see (1)) and Saint-Quentin (1872) (see (2a)). Thus, a full-fledged Popperian (see (5)) critique of (neo-)Schleicherian-cum-Quentinian creolistics can proceed via a close examination of the proposal in WSG about age-complexity correlations. This proposal conveniently provides us with an updated adaptation of linguistic-genealogical arguments that have run virtually uninterrupted through the past two centuries. A critique of this proposal will, I hope, clear up the scene for empirically-responsible and theoretically-grounded Cartesian-cum-Uniformitarian creolistics (i.e., the sort of creolistics that does not assume any a priori
fundamental structural distinction between creoles and non-creoles). As we will see below, modern creolists’ measures for age-complexity correlations – like their early 19th-century intellectual antecedents – amount to an empirically, theoretically, and logically flawed view of creole formation and language change. Bearing in mind that “science must begin with the criticism of myths”, these re-formulations of Schleicherian linguistics provide a valuable point of departure for learning from our mistakes and for advancing our knowledge (in Popperian mode; see (5)). My critique will thus exemplify “learning by debunking” (in Gould’s 1996: 351–353 terminology). Indeed I share in Umberto Eco’s optimism in his book *Serendipities* (1998: iix) where he so describes various brands of “lunatic” linguistics (cf. Foucault’s allusion to “chimera and reverie” in (4)): “[E]ven the most lunatic experiments can produce strange side effects, stimulating research that proves perhaps less amusing but scientifically more serious”. Here are the two “serendipities” to be derived from the discussion of neo-Darwinian (or neo-Schleicherian) creolistics below: (i) establish the epistemological limits of the terms “pidgins”, “creoles”, “young” vs. “old”, “simple(st)” vs. “(most) complex” as linguistic-structural and historical-phylogenetic taxa; and (ii) promote the study of language contact and its ubiquitous outcomes (including “creole genesis”) in a Cartesian and Uniformitarian analytical framework.

Fundamental to neo-Schleicherian arguments about creoles’ lack of complexity are the two notions “pidgins” and “features unnecessary to [basic] communication”. These notions still remain ill-defined. I will argue in Section 4 that there can hardly be a core set of basic-communication features that defines the structural essence of ALL pidgins. As for the exclusively-creole combination of structural features claimed to be “predictable from the history of creoles in pidginization”, this typology will be argued to be deeply problematic, on both empirical and theoretical grounds (see Section 2 and references in Note 12). Problematic as well, at least from a linguistic-theoretical standpoint, is the notion “age of languages”: I will argue that thus far no well-defined and

9. See DeGraff (2001b: 98–99, in preparation) for some of the sociological costs that are associated with the dogma that creoles are structurally distinct from “regular”/”normal” languages (cf. Note 6). Such dogma has, inter alia, undermined the role that creole languages should play in the education of creole speakers and in the exercise of their human rights. With respect to Haiti, see critiques in P. Dejean (1989, 1993), Y. Dejean (1975, 1993, 1999b, forthcoming).

10. Darwin himself (1871: Chapter 21, 385) is worth quoting in that respect: “False facts are highly injurious to the progress of science, for they often long endure; but false views, if supported by some evidence, do little harm, for every one takes a salutary pleasure in proving their falseness; and when this is done, one path towards error is closed and the road to truth is often at the same time opened.” Darwin may have been overly optimistic. For the sort of harm that can be caused by false views in the human sciences, see, e.g., Gould (1996); cf. Note 51.
independent (i.e., testable and non-circular) LINGUISTIC metric can objectively measure the age of languages.

Yet it is the age of languages that in creole studies is often taken (in a manner reminiscent of early 19th-century – Schleicherian – linguistics) as one crucial factor leading to complexity differentials between creoles and non-creoles. For neo-Schleicherians, creoles’ lesser complexity is yet another “predictable result of their youth” (see, e.g., WSG: Section 1), notwithstanding the fact that there is still no reliable litmus test for young languages as a linguistic class (see (7)–(8); also see Sections 3 and 6).

What about “complexity” per se? Here too creolists enlist terminology and assumptions that will be shown to be ill-defined and empirically- and theoretically-controversial in fundamental ways. For examples, the “complexity metric” in Section 2.4 of WSG is theoretically peculiar: it is based on an arbitrary list of superficial linguistic features with no psychologically-relevant, theoretically-grounded, or independently-motivated unifying basis. In fact, the conceptual foundations of this metric are either left undefined (along with, e.g., “communicative necessity”, “basic [human] communication”, and language “age”) and/or are made largely incompatible or irrelevant to what we (seem to) know about linguistic typology, historical linguistics, language acquisition, language processing, and theoretical linguistics (cf. WSG’s interpretations of “complexity”, “Universal Grammar”, etc.). Furthermore, the choice and testing of creolists’ complexity metrics is often circular, tendentious and empirically flawed (see Sections 5 and 6).

Chomsky once wrote that “linguistic theory must be constructed with explicit and precise definitions and operational tests” (1957: 233). In the absence of empirically- and theoretically-grounded definitions of “creoles”, language “youth”, “pidgins”, “basic [human] communication”, “complexity”, and so on, what are we to make of the proposition that “the world’s simplest grammars are creole grammars” because creoles, and only creoles, are “born again” languages? I take it that, in the absence of “explicit and precise definitions and operational tests” this most simplistic proposition can provide us with only a mismeasure of creole languages and, indeed, a mismeasure of creole SPEAKERS if one assumes a Cartesian (or Humboldtian) approach whereby languages are properties of minds.11

In what follows, I will document the theoretical, empirical, and bibliographical lapses that undermine the recent (and not-so-recent) claims that “the world’s simplest grammars are creole grammars”. In particular I will review the foun-

---

11. Cf. Humboldt’s “inner linguistic sense” which is intimately related to the “genius” (e.g., the “mental capacities”) of particular peoples at particular evolutionary stages; see Corcoran (2001) for one recent discussion in the context of creole-genesis scenarios; also see Humboldt’s remarks on the evolution of inflection in Note 5.
ational assumptions upon which traditional and recent claims about creole complexity (or lack thereof) rest. These assumptions touch on the following questions, which I address in turn:

(i) How are “creoles” defined? (Section 2)
(ii) How does one measure the “youth” of languages? (Section 3)
(iii) Can we determine the features “necessary to basic human communication”? What is the relationship between “basic human communication” and “simplification” in “pidgins”? Does Universal Grammar define “basic communication” requirements? (Section 4)
(iv) What is the epistemological status of complexity metrics in creole genesis scenarios? (Section 5)
(v) Given (non-)answers to all of the above, (how) can we falsify the claim that “a subset of creole languages display less overall grammatical complexity than older languages”? (Section 6)

2. Defining “creole”

What’s in a name? Long ago, Francis Bacon (1620 [1994: 64–65]) warned us that that there “are names for things that do not exist (for just as there are things without names because they have never been seen)” and that certain “obscure and deep seated [terminology] is derived from an incorrect and unskilled abstraction”.

In 1878, the Neogrammarian manifesto also warned linguists against “[believing] that they have then fathomed the essence of the phenomena when they have devised a name for the thing” (Osthoff & Brugman 1878 [1967: 202]).

What exactly are creolists from Saint-Quentin to McWhorter comparing when they make claims to the effect that “the world’s simplest grammars are creole grammars”? What is the exact scientific purpose of this comparison? The postulation of some structural essence common to all “creole” languages is central to the present discussion as it has been from the beginning of creole studies. That this structural essence is the outcome of “simplification” is another age-old tenet in creole studies. As Chaudenson (1994: 41) notes, “the idea that creoles are simplified versions of their European ancestors is as old as creole studies”.

Yet it has often been argued in Uniformitarian (non-dualist) fashion that “creolization is a sociohistorical, not a structural, process” (Mufwene 2000a).
In this view, creoles cannot be distinguished a priori from non-creoles on strictly synchronic structural grounds. I myself have now adopted a strictly language-external definition whereby “creole” is a sociohistorical attribute that connotes the results of particular types of (abrupt) language contact marked by exacerbated social distance cum power imbalance; see DeGraff (1999a, b) and references therein for overviews. I also take it, as a working hypothesis within the generative framework I work in, that:

(10) [C]reoles [as mental entities, i.e., I(nternal)-languages] are no more and no less than the result of extraordinary external factors coupled with ordinary internal factors […] [Within mentalistic approaches to language creation and language change,] THE NOTION OF “CRE-OLIZATION” AS A UNITARY AND DISTINCT LINGUISTIC PHENOMENON EVAPORATES. (DeGraff 1999b: 477; emphases added)

The Cartesian-Uniformitarian position in creole studies has long been opposed by the dualist orthodoxy that assumes a strong form of “creole exceptionalism”. Per this dualism/exceptionalism, creole languages – thus creole speakers – are deeply special, with genealogical and structural properties that are fundamentally distinct from their non-creole counterparts. This is grosso modo the position of neo-Schleicherian creolists who posit both a diachronic exceptionalism (“pidgin ancestry” and creole “youth”) and a synchronic exceptionalism (a “Creole Prototype”). These distinctions are straightforwardly Schleicherian in the rapport they try to establish between complexity and time: “creoles are the world’s only instantiation of spoken language having been ‘born again’, when speakers expanded pidgins” (WSG: Section 2.3), and creoles’ exceptional youth (i.e., their exclusive “pidgin ancestry”) entails that “the world’s simplest grammars are creole grammars”. As in Schleicher’s evolutionist scenario, it is assumed that “much of a grammar’s complexity results from the operations of random accretion over time” (WSG: Section 2.3). However, these genealogical and structural distinctions lack theoretical and empirical substance.

2.1. Haitian Creole: An anti-prototype Prototypical Creole

On the synchronic structural front, the Creole Prototype is equated to the combination of the following traits (WSG: Note 1):

On the origin of creoles

(11)  
  a. lack of “inflectional affixation”;
  b. lack of “tone distinguishing monosyllabic lexical items or encoding morphosyntactic distinctions”;
  c. lack of “opaque lexicalization of derivation–root combinations”.

The assumption is that inflection, tone, and opaque derivation–root lexicalizations increase complexity. As of the lack of complexity-inducing morphological processes in the Creole Prototype, it is assumed as a consequence of the hypothetical pidgin-to-creole cycle, its concomitant morphological bottleneck, and the correlates thereof vis-à-vis the ab ovo growth of creole affixes. See the following quotes from WSG: Section 5.2:

(12)  
  a. When language “begins anew” amidst pidginization, the linguistic vehicle consistently lacks affixation entirely or exhibits it only minimally, with affixes developing only slowly even when the pidgin is creolized.
  b. Observed and documented processes of language change make it clear that the main source of affixes is erstwhile free morphemes.

There is already a plethora of data and observations that, taken together, invalidate the empirical and theoretical claims in (11)–(12) about creole morphogenesis. Yet, in creolistics and in other human sciences, complexity rankings and genealogical scenarios are no simple matters, scientifically and sociologically (see, e.g., Gould 1996 for relevant caveats). Thus, the need to address the claims in (11)–(12) from ground-zero up, starting with the theoretically and empirically central aspects of the Creole Prototype.

To start with, let’s take Haitian Creole (hereafter HC). This sociohistorically prototypical plantation creole has been argued in DeGraff (2001b: 69–88) to manifest both “inflectional affixation” and “opaque lexicalization of derivation–root combinations”, contra the predictions in (11). Moreover virtually all HC affixes are etymologically related to French affixes (i.e., virtually


14. Creolists often pay lip service to the age-old dogma that “morphology [is] essentially alien to creole languages” (Seuren & Wécker 1986: 66) while their very data illustrate robust patterns of both inflectional and derivational morphology, including opaque lexicalizations. In fact, some of the relevant data in these works suggest that certain creoles may well have more affixes than certain non-creoles, including a subset of the creole’s ancestor languages; e.g., many affixes in Haitian Creole have no counterparts in the Fongbe substrate; see DeGraff (2001b: 58–69) contra Lefebvre (1998: Chapter 10).
none of these affixes result from grammaticalization), thus shedding doubt on the claims in (12).

Well-documented facts of Haiti’s sociohistory and demographics teach us that HC should count as a bona fide creole, even a “most creole of creoles”, one whose historical conditions “have been perfect for the preservation of a basilectal creole” (McWhorter 1998: 809, 812; 2000b: 206). But, in creole studies, even established historical facts can be tinkered with to fit the scenario du jour: After robust HC data were advanced as counterexamples to pro-prototype claims (see DeGraff 2001b), HC’s privileged status as “basilectal creole” and “most creole of creoles” got revoked. HC is now taken to “not exemplify the Creole Prototype in the purest possible form” because of alleged “contact over the centuries with French” (WSG: 143).

Historically, linguistic interaction in colonial Haiti between Europeans and Africans was by far the most intense at the onset of contact, with French structures having had the most influence in the formation of (proto-)HC quite early on, in late 17th through early 18th century. Thereafter, contact with French speakers was greatly reduced after the sugar boom in the middle of the 18th century: the labor needs of expanding sugar plantations led to a drastic increase in the arrival rates of Africans. Before the sugar boom, the colony was still made up of mostly small homesteads – the société d’habitation – many of which subsequently and gradually gave way to the brutal segregation of the plantation economy – the société de plantation. (See Baker & Corne 1982, Chaudenson 1992, Chaudenson & Mufwene 2001, and Singler 1996 for an overview of the historical and demographic details and for pointers to the relevant literature.)

After the independence battles of 1791–1803, the French presence was virtually eliminated. There have always been, and still exist, various degrees of contact between HC and French, specially at the higher echelons of society, with concomitant contact phenomena in both languages as naturally expected. However, Haiti’s history has allowed post-genesis HC to evolve in relative isolation from its socially-remote, albeit prestigious, lexifier. Haiti is the only New World plantation society that eliminated most of its “lexifier” population through war. From the société de plantation onward, French has been especially remote from (most of) the monolingual peasantry, Haiti’s numerical majority. More generally, the vast majority of contemporary Haitians in Haiti are monolingual creolophones with relatively little contact with other languages (one notable exception is the region alongside the Haitiano-Dominican border where HC speakers are regularly exposed to Spanish varieties; see DeGraff 2001b: 68 for one linguistic-structural consequence of this contact in the domain of morphology). Such a degree of linguistic isolation is quite unlike the situation in other creole-speaking communities in the Caribbean. These well-known sociohistorical and demographic facts explain why Haiti does not
offer the same sort of “creole continuum” that is found in the former English colonies of the Caribbean. Unlike (say) the English-lexicon Caribbean creoles, HC by and large (i.e., the majority of HC speakers) did not “remain in contact with [the] lexifier” (contra WSG: Note 17). It thus seems unlikely that the bulk of HC morphology would have been created post-genesis via late borrowings from French (e.g., after the elimination in 1803 of the majority of potential “lenders”).

Furthermore there is no evidence that there ever was an earlier stage of (proto-)HC where all creole varieties were uniformly devoid of all affixes. Such an affixless stage is even less likely considering that most of the HC lexicon is etymologically French and that HC speakers – like any other human speakers in our documentable past – are, in principle, able to extract morphological (e.g., affixal) information from stored patterns in their lexicon. In fact, whatever mental capacities would enable late-borrowing(-cum-restructuration) of French affixes in the course of HC’s post-genesis diachrony would have

15. McWhorter himself seems well-aware of these sociohistorical facts, when needed for his theorizing elsewhere. For example, these facts play a key role in his Afrogenesis speculations (see Appendix B). Witness the following four assumptions about the genesis of plantation creoles (McWhorter 2000b: 200–207): (i) “the lexifier was available to all slaves not only during the société d’habitation phases but even later, during the plantation stage”; (ii) “adult slaves were capable of obtaining a viable second-language register of the lexifier”; (iii) “plantation-born children were even better situated to acquire the local standard than their parents”; (iv) “[t]he intimate conditions within which blacks and whites lived in sociétés d’habitation would have made acquisition of the lexifier even more compelling”; (v) “in a society like Haiti, where French speakers were ousted early in the colony’s history, conditions have been perfect for the preservation of a basilectal creole”. Assumption (i) seems controversial for well-documented sociohistorical and demographic reasons, specially at the plantation stage: given marronage, social segregation, low European-to-African ratios, etc., not ALL slaves could be exposed to, and learn, the lexifier; see Pelleprat (1655), Girod-Chantrans (1785), Moreau de Saint-Méry (1797), Descourtilz (1809), etc. But assumptions (ii)–(v) seem much less controversial, independently of the use to which they are put in McWhorter’s “Afrogenesis Hypothesis”; see, e.g., Baker & Corne (1982), Chaudenson (1992), Chaudenson & Mufwene (2001), Singler (1996), Mufwene (2001) for related facts.

In the case of Haiti, propositions (ii)–(iv) straightforwardly entail that many early HC speakers had a “diglossic competence between an L2 variety of the lexifier and a [...] pidgin” and that both the founding slaves and plantation-born children could acquire “the local standard” (or, at least, a variety thereof). If so, then one can reasonably argue that “diglossic” early HC speakers were in a position to adopt and adapt affixes and other patterns from their “L2 variety of the lexifier” in order to structurally expand the emerging “creole”. This renders unlikely any scenario whereby incorporation of French-derived affixes into HC is necessarily a late post-genesis phenomenon (cf. Appendix A). The linguistic evidence points to the same conclusion: affixes in all varieties of HC – including any contemporary “preservation of a basilectal creole” (see (v) above) – have cognates in French affixes; see, e.g., Fattier (1998), DeGraff (2001b), contra Lefebvre (1998).

(On the empirical and theoretical status of McWhorter’s (2000b) “Afrogenesis Hypothesis” and the “imported pidgins” therein, see Appendix B.)
also enabled the adoption/adaptation of similar French affixes at the earliest stages of (proto-)HC. This is especially so given the numerical preponderance of French over Africans and the lesser racial segregation at the early stages of contact. In other words, the sociolinguistic context was more learner-friendly at the onset of contact – during the société d’habitation – than later on – during the société de plantation (see Note 15). No known (psycho- or socio-)linguistic principle could have forced all the speakers in the (early) contact situation to systematically ignore all morphological patterns in the available lexifier varieties (also see Appendix A).

Similar remarks apply to the diachrony of “opaque lexicalizations of derivation–root combinations” in HC – throughout I am assuming Uniformitarianism (e.g., that innate mental capacities for Language have remained uniform across the species in the past few millennia and across socioeconomic contexts; see (3)). Indeed, if HC speakers could create such morphology via “borrowing” through POST-genesis (reduced) contact with French speakers, then no known (psycho)linguistic constraint would prevent the creators of HC from creating this morphology at the ONSET of contact (e.g., during the société d’habitation, when the demographic and sociolinguistic factors were MORE favorable for the acquisition-cum-reanalysis of such forms via relatively HIGHER exposure to French speakers and, thus, more intimate familiarity with the lexifier). I must note again that, as far as we can tell, the bulk of the HC lexicon (both roots and affixes) has always been etymologically related to French (see Section 3.3). This means that creole creators from the get-go massively adopted (at the very least) French lexical forms, including opaque lexicalizations.

Thus, we can conclude that HC – a bona fide prototypical creole in the sociohistorical sense – with its inflectional and derivational morphology and its opaque lexicalizations represents a robust counterexample to classic neo-Schleicherian claims about creole morphogenesis via an affixless pidgin (also see Appendix A).

2.2. The Creole Prototype is an anti-Saussurean artificial language

On the theoretical front, given the very nature of the lexicon (in the Bloomfieldian sense) as the repository of Saussurean arbitrariness and given now-standard results in (psycho)linguistics, there seems to be no reason to expect “opaque lexicalization of derivation–root combinations” to take millennia to develop.16 Indeed, it can be reasonably argued that “lexicalization entails that

---

16. The claim about the absence of creole “opaque lexicalizations” is not new. Seuren & Wekker (1986: 66–68) take “idiosyncratic exceptions”, “highly specialized lexical items”, and “richer expressive means” to constitute a class of lexical “luxuries” that are exclusive to non-creoles (i.e., to “older or more advanced” languages). See DeGraff (2001b: 88–98) for related orthodoxies in creole studies, some of which go back to the 17th century.
the form is no longer generatable via metaphorical inference and now requires storage as an independent form” (WSG: Note 10). In this statement, “lexicalization” is clearly defined, in Cartesian fashion, in terms of mental processes in the heads of individual speakers, namely “metaphorical inference” and lexical “storage”. The latter is what is normally (and crucially) involved in the creation of “opaque lexicalizations” as Saussurean (thus, holistic) signs in the speaker’s mental lexicon. To also claim that such a mental process necessarily requires “millennia” to unfurl (McWhorter 1998: 792–793, 798, 812, etc.) leads to a scenario in which individual speakers can live for millennia – not a likely scenario.

A more realistic and theoretically-grounded alternative is to assume that opaque lexicalizations as properties of I-languages (thus, of individual minds, in concert with communal conventionalizations and cultural transmissions) do not require millennia to develop: like other creole lexical items, opaque lexicalizations can be innovated in creole grammars or they can be “inherited” from (and via exposure to speakers of) the source languages; see Fattier (1998), DeGraff (2001b: 76–82). Regarding the “inheritance” of opaque lexicalizations, a pro-prototype creolist could try and argue that such “inherited” lexicalizations should not count for testing the prediction in (11c) because they were not developed over millennia by the creole speakers themselves. But this argument is circular since (Prototypical) Creoles are a priori defined as young languages by pro-prototype creolists. The argument is fallacious in yet another way: many bona fide opaque lexicalizations in, e.g., contemporary English were not innovated by the current generation of English speakers – these lexicalizations were “inherited” by modern English speakers.17

17. (Neo-)Schleicherian linguistics makes recurrent appeal to “very long time spans” as necessary conditions for the development of structural complexity beyond the requirements of basic/primitive communication. Witness Schleicher’s pronouncement in (1c) and contemporary beliefs that complexity requires “millennia” to develop its “above and beyond UG” overspecifications (see, e.g., WSG: Sections 1, 2.3, 4.2, 5). It thus seems that (neo-)Schleicherian linguistics, unlike Cartesian-Uniformitarian linguistics (see, e.g., (3a)), gives epistemological priority to some sociohistorically-oriented reification of E(xternal)-languages, at the expense of mental grammars qua I(nternal)-languages. This, in turn, leads to a confusing conflation of mental processes and historical factors (the latter can be considered “historical accidents”; see Chomsky 1995: 6–7, 11). If human biology (UG, say) sets boundaries on the shapes of all natural languages (including the crosslinguistic inventory of what WSG: Section 2.3 calls “‘ornamental’ elaborations”), then no language-specific properties (and no language-specific “ornament”) need “millennia” to develop. This also applies to opaque lexicalizations as discussed in the main text and to the other linguistic “ornaments” in (18) that are considered “incidental to basic communication”. Thus “‘ornamental’ elaborations” cannot be excluded a priori from creole languages. Here too the ill-defined linguistic distinction “old” vs. “young” becomes a methodological trompe-l’œil (see (3b)). The main text elaborates on this E/I-language distinction in resolving the old-versus-young issue; also see DeGraff (1999a: 8–9).
2.3. **Empirical and methodological considerations**

On the strictly-empirical front, tone, inflection, and opaque lexicalizations have been documented in a variety of creole languages; see, e.g., the observations and references in Muysken & Law (2001) and Muysken (in press). Muysken & Law (2001: 49) make the important observation that:

\[ \text{(13) There is no doubt that in many core creoles there is very little inflectional affixation. This is also to be expected since in the second language acquisition of the European colonial languages, during the early stages of the process of creole formation, inflection is often lost. However, the contributing superstrate and substrate languages were not very rich in their inflection either, and in several cases (Berbice Dutch Creole, Papiamentu, Cape Verdean) we do get some inflection. If we take creolization in typologically very different languages [...] there is simplification and regularization of inflection, but not loss of inflection. The absence of inflection is the singlemost frequently noted supposedly typological feature of creoles, and indeed may be the way many people identify a language as a creole. However, it may be the accidental by-result, from a scholarly point of view, of the limited typological spread in the languages contributing to the prototypical creoles.} \]

In a similar vein, Givón (1979: 20–21) remarks that the “reduction of inflections” in Caribbean creoles is as expected given the inflectional profiles of the substratum. From this perspective, the synchronic structural criteria in (11) and the concomitant diachronic assumptions in (12) are empirically and theoretically flawed. In the following section, I address, inter alia, the assumptions in (12). These assumptions crucially enter in the theoretical basis of (neo-) Schleicherian genealogical creolistics.

3. **Creoles, how “old” are you?**

3.1. **Creoles as “born again” languages?**

The widespread consensus across time, across space, and across theories in creolistics and beyond is that creoles are young languages – linguistic neonates that embody an evolutionary prior stage in relation to non-creole languages; see DeGraff (2001b) for an overview. Girod-Chantrans (1785 [1980]) found HC to be “nothing but French back in infancy”. Adam (1883: 3) reports that “[i]n Europe, creole speech is universally considered an infantile jargon” (also see (2b)). Jespersen (1922: 228) wrote that creole creators spoke “as if their minds were as innocent of grammar as those of very small babies [...] [thus, creoles’] inevitable naïveté and [...] childlike simplicity [...]”. In Hall’s (1962)
pidgin-to-creole life cycle, creole languages are the one exception to the principle of ‘‘normal’ language [being] handed down from generation to generation’. Bickerton’s Language Biogrogram Hypothesis turns creole speakers into linguistic “Adams and Eves” (in Richard Price’s terminology, as cited in Corcoran 2001). Seuren & Wekker (1986: 66, 68) and Seuren (1998: 292) contrast languages that are “older or more advanced” and “sophisticated” to creoles, which are “younger or less advanced” – “beginning” – languages (see Note 16). Thomason & Kaufman (1988: 8–12, 206, 211, etc.) consider creoles as non-“genetic” (i.e., “parentless”) languages that, unlike “genetic” languages like English, have evolved via some kind of “abnormal [break in] transmission” (see Note 22). WSG: 131 follows suit: “[A]ll [of the world’s natural languages] trace back tens of millennia. […] with one exception […]: […] creole languages have, by definition, existed only for several centuries at the most. The oldest known creoles today […] trace back to the late fifteenth century”.

In quasi-Schleicherian mode (see (1)) and against Neogrammarian warnings (see (3)), the notions “younger” vs. “older” language often form the cornerstones of dualist claims about complexity differentials between creoles and non-creoles (see Note 17). For example, the proposition that “the world’s simplest grammars are creole grammars” is argued to be “a predictable and, in the end, rather unremarkable result of the recent origins of creole languages” (WSG: 162). More explicitly (WSG: 132):

(14) “Tens of millennia of drift would leave all grammars existing during that timespan equal in terms of the amount of complexity accreted beyond the bounds of the genetic specification for language. This stipulation predicts, then, that one subset of the world’s natural languages, creoles, would differ from the rest of the world’s natural languages in displaying less of this kind of needless complexity.”

3.2. Why/how do we “age” languages?

There is one fundamental question that is evaded throughout: How does one scientifically measure the “age” of languages for the (presumably) scientific purpose of correlating language age with language complexity?

Individual speakers (be they creole-speaking or not) do not live through millennia. Neither do their respective “I(nternal)-languages” qua Cartesian mental grammars. So the statement that non-creole languages – and only non-creole languages – have existed for “tens of millennia” surely cannot refer to I-languages. Such “tens of millennia” estimates must apply to some extensional notion of language – perhaps some (reified succession of) socially-determined “E(xternal)-languages” or communal languages (see (3) and Note 17 above on relevant methodological caveats, going back to the Neogrammarians).
And still we are faced with the fundamental question: When do E-languages (or some communal/sociohistorical reification thereof) commence? Under certain definitions and for certain well-defined purposes, language age can certainly serve as a useful heuristic in various disciplines, including archeology, ethnography, anthropology, historical linguistics, etc. For example, one can say that, in a sociohistorical sense, HC is a “young” language as it marks the identity of a newly-created community – a community that did not exist before the 17th/18th century (I thank Heliana Mello for an enlightening discussion of this point). In a somewhat related sense, varieties of (say) Indian English and the corresponding (speech) communities may also count as “young”. But notice that this sort of youth is not necessarily correlated with increased structural simplicity: in some ways, “young” Indian English may even be more complex than the Queen’s “old” English (consider, say, the phonology of retroflex stops in Indian English). Similar remarks could, in principle, apply to any “old”– “new” language pair, as with, e.g., European and Brazilian Portuguese.

Notions such as language birth, age, and death are also assumed implicitly and a-theoretically when we use terms such as “Proto-Indo-European”, “Latin”, “Old French”, “Middle French”, “Modern French”, etc., as classificatory devices. But, notwithstanding the popularity and sophistication of Stammbaumtheorie qua “Tree of Language” (cf. Darwin’s Tree of Life), old vs. new linguistic species cannot be discriminated by any measure that looks like biological genetic criteria (e.g., DNA, interfertility). There is no clear notion whereby E-languages can be taken to reproduce like living organisms. Neither do we have clear linguistic-structural analogues for the DNA sequences that have now become so handy in tracing biological phylogenesis (see Section 1.4 above about the (im)possibility of quasi-Lamarckian linguistics). 18

There are a number of fascinating sociological factors vis-à-vis why, when, and how certain languages start being (perceived as) “new” languages with birth certificates that distinguish them from their relatives. As a facile illustration, one can compare the status of Danish, Norwegian, and Swedish (or Span-

---

18. On a par with the linguistic criteria, the biological criteria for phylogenetic descent are still debated (e.g., on structural-cum-philosophical grounds: for example, is genetic overlap more important than interfertility in determining species boundaries? cf. chimps vs. humans, and lions vs. tigers vs. tigons/tigers). Yet, the genetic criteria for the phylogeny and identity of biological species are much better understood than their counterparts in historical-comparative linguistics: the biological criteria are grounded in increasingly sophisticated knowledge about genetics and evolutionary biology. It is now possible to try and retrace phylogenetic lines as far back as hundreds of millennia (e.g., back to our ancestral Eve; see Cavalli-Sforza 2000). This is quite unlike genetic linguistics (see the references in Note 3 for an array of challenges to Stammbaumtheorie; also see Note 7). Most linguists still have no clue about Eve’s native language, notwithstanding speculations that “the first language had no affixes”, which would make Eve THE Prototypical (Ur-)Creole speaker! (WSG: Section 5.2; also see references in Note 4).
ish, Portuguese, and Italian; or Serbian and Croatian) as “distinct” languages vs. that of Chinese as “one” language. Suffice it to say that these distinctions are more relevant for issues of identity (and) politics than for strictly linguistic (typological) matters.

Another example will drive the point home (in a metaphorical and literal, if not statistically-representative, sense): In New York, I once saw a sign that advertised (in Haitian Creole!): *Isit nou pale Franse* ‘Here we speak French’ (I have updated the sign’s orthography to fit the official HC phonemic-spelling norms; see Y. Dejean 1980). The HC-as-French sign had been proudly displayed by a Haitian employee next to a Spanish sign advertising *Aquí se habla Español*.

One may well chuckle at this story and argue that it is surely not representative of Haitians’ meta-linguistic attitudes, but we still need to ask: What are the precise linguistic-structural criteria – the operational typological threshold – that would classify, e.g., Missouri French and Cajun French as bona fide varieties of French while HC is usually not so classified? Perceived notions of (non-)distinctness (in this case, between French and historically-related varieties) may be quite useful (or harmful) for creating and promoting stereotypes, political identities, and community boundaries, but they seem to have little to do with linguistic typology per se. It is thus not so surprising that the perception of separateness between a creole and its source languages is not uniform across all creole speakers; see Winford (1994: 45–48) and Mühleisen (2000: 84–92) for recent discussions.

For typological and sociological reasons (e.g., regarding creole-based education), I myself consider my native HC to be “distinct” from its Romance and Niger-Congo ancestors. For linguistic-typological reasons, one may well consider some variety of modern continental French (as spoken by Jean-Yves Pollock, say) to be distinct from its Latin and Old and Middle French ancestors as well as to its American relatives (Québec French, Missouri French, Cajun French, etc.); and so does Modern English fall into distinct varieties that in turn are distinct from Proto-Germanic and Old and Middle English varieties. In a somewhat related vein, I can also envisage, for methodological reasons, that children’s early grammars are “distinct” from the grammars of their models – the older peers and caretakers that provide children with Primary Linguistic Data (see, e.g., Rizzi 1999 and references therein for discussions of children’s “grammatical invention”).

If our discussion of age-related complexity differentials (and the concomitant claim about “the world’s simplest grammars”) is to advance in a scientifically viable manner, we need an independent, theoretically grounded measure for measuring the birth and age of (new vs. old) languages in some linguistically relevant fashion that is impervious to our (often tacit) preconceptions, otherwise creolistics may well become the world’s most simplistic science. Recall
Chomsky’s (1957: 233) statement that “linguistic theory must be constructed with explicit and precise definitions and operational tests”. Such explicitness and precision is even more urgent when dealing with languages that have generally been stigmatized from the very moment that they were identified and “baptized”.

In the absence of such measure (and given the discussion in Section 2; see Note 19), can we rely on our intuition to discriminate newborn from multi-millenarian languages? The question is not so simple.

Let’s get back to the case of Jean-Yves Pollock as a speaker of Modern French. What exactly does it mean to say that (Modern) French is an old language? Here we must be terminologically picky in order to try and make sense of this question whose presuppositions are infected with metaphors that contradict one another. Do the idiolects of Pollock and his peers underlie an E-language that has existed for tens of millennia? If so, his (E-)French would actually include, and be the continuation of, a very long ancestry, including at least Middle French, Old French, Latin, etc. and all the varieties in between. And so would Spanish, Portuguese, Italian, Romanian, etc., be continuations (i.e., Schleicherian growths) of Latin and its descendants along the Stammbaum lines that connect Latin to the corresponding Romance variety. In other words, if each Romance language qua old language has, by definition, existed for “tens of millennia”, then Romance speakers all speak a selfsame continuation of their common ancestor (say, Latin), which, in turn, implies that all Romance languages constitute the selfsame multi-millenarian language, by transitivity. Not a satisfying result, at least not for Romance linguists who have done much work to isolate robust parametric differences within Romance.

3.3. Creoles as multi-millenarian morphosyntactically wrinkled neonates

Let us ask again: Can linguistic typology help us decide whether a given language has existed for “tens of millennia” or is “born again” and “begins anew”? Can neo-Schleicherian (i.e., genealogical) creolistics constructively engage linguistic theory in any scientific way (e.g., in any way that resembles how evolutionary biology engage (phylo)genetic analysis)?

Take morphosyntax. Can relatively well-understood typological properties (e.g., lexical semantics, derivational morphology, underlying word-order, scrambling, nominal case morphology, presence of definite articles) be used as genetic tracers for writing up the “birth certificate” of languages? Let’s assume so (only for the sake of argument), and let’s take these typological properties as the analogues of, say, DNA sequences in the dating of biological

---

19. Recall from Section 2 that the structural measures in (11)–(12) misdiagnose creole languages. So we cannot take the criteria therein as a litmus test for a creole typology (if any).
On the origin of creoles

species (but see Notes 18 and 20 and Appendix B). Then, along the aforementioned typological dimensions, HC can be argued to be more similar to Modern French than Modern French is to Old French or to Latin. Thus, on the counts of both perceived similarity (in the eyes of certain creole speakers like that HC speaker in the Brooklyn store mentioned in Section 3.2) and typological closeness (along certain typological variables), HC seems no younger (or no older) than Modern French.

What about morphology? It can be straightforwardly argued, as in Section 2.1 (also see Fattier 1998 and DeGraff 2001b, contra Lefebvre 1998), that most HC affixes historically derive from French affixes. Here the lexifier-creole similarities greatly exceed, both in cardinality and in systematicity, the sparse correspondences that have been used to argue for the “Afrogenesis” of French-lexicon creoles such as HC and Mauritian Creole from a single 17th-century West African pidgin ancestor (see Appendix B). In the case of HC, the lexifier-creole morphological continuity is not at all surprising: most of the HC lexicon is etymologically related to French. In turn, the French lexicon itself is mostly derived from Latin – with French emerging through language contact as occasioned by Roman imperialist conquests (compare with creole genesis in the context of Europe’s imperialist conquests in Africa and the Americas).

Thus, the very morphology and history of HC (as compared to the morphology and history of, say, French) challenges the exclusive “born again” or “recent origins” status bestowed on radical creoles as a class by neo-Schleicherian creolists. As far as I can understand neo-Schleicherian techniques for linguistic phylogenetic analysis and for identifying language birth (but see Notes 18 and 20 and Appendices A and B), HC’s etymological longevity and unifor-

20. This subset of typological features – lexical semantics, derivational morphology, underlying word-order, scrambling, nominal case morphology, presence of definite articles – is tenden-
tiously chosen to illustrate (superficial) morphosyntactic similarities across creoles and their respective lexifiers and (superficial) morphosyntactic dis-similarities within accepted Stamm-
baumtheorie phyla. Given the complex nature of Language (in an information-theoretic sense, within our current best theories) and given the concomitant vastness of typological diversity (which is somewhat still uncharted), it should be possible to handpick arbitrary sets of (su-
uperficial) morphosyntactic features along which any two languages will appear similar or dis-similar (see Appendix B). As Chomsky (1986: 151) reminds us, “even languages that have separated only recently may differ in a cluster of properties, something that has been observed in comparative studies”. One such comparative study is Meillet (1929) where it is observed that “neo-Latin languages fall into a typological class that is quite remote from the structural type represented by Latin” (1929 [1938: 80], my translation; also see Meillet 1912 [1926: 148]). (For additional caveats on the use of morphosyntactic comparisons in phylogenetic linguistics, see Note 18; also see Thomason’s caveat in (8).)
mity and its morphosyntactic inheritances-cum-restructurations should be the exclusive province of multi-millenarian languages.21

Recall the classic “pidgin-to-creole life-cycle” scenarios whereby pidginization creates a radical bottleneck for lexical and morphological development, thus forcing creoles to emerge from affixless ancestors (see Note 6; also see (12) and discussion in Section 4.2 below). A related and more general assumption is that “affixation […] emerges from the grammaticalization, reanalysis, or reinterpretation of material which was not originally inflectional” (WSG: Section 5.2). Taken together, these assumptions entail that affixes in allegedly young languages such as HC must, in general, emerge via the grammaticalization of erstwhile free morphemes (see (12)). These scenarios are robustly contradicted by HC where, from genesis onward, almost all affixes have had, and still have, cognates in French affixes – which in turn often have cognates in Latin morphology. Etymologically these cognates diagnose millennia of seemingly unbroken transmission, quite an ancient pedigree for the morphology of a “most creole of creoles”. Indeed, there is no documented stage in HC diachrony where the language was affixless or with most affixes derived from “erstwhile free morphemes” or with most affixes derived from outside of French (also see Section 2 and, especially, Appendix A).

In this respect, Goodman (1964: 26–28, 122–124) gives a variety of HC examples in the nominal and verbal morphosyntactic domains that suggest that “the initial speakers were exposed to a French which was virtually as complex inflectionally […] as is standard French”. Putting aside the fact that “standard French” (specially back then) was an artificially constructed language with few, if any, native speakers, Goodman’s contention seems somewhat extreme, for at least two reasons: (i) most whites in colonial Haiti, and even in France, were more likely to be illiterate speakers of rural “patois” – Langue d’Oc, Langue d’Oïl, Norman French, etc. – than fluent speakers of standard/literate French (see, e.g., Chaudenson 1995: 18, Chaudenson & Mufwene 2001: 151–153); and (ii) not all creole creators were exposed to the same (non-native approximations of) native French varieties: right from the onset of contact, there must have existed a continuum of contact varieties, which were subsequently modulated through sociolinguistic factors into later varieties, including those known to us today (see, e.g., Alleyne 1971, Mufwene 2001; also see Note 15 and Appendices A and B).

21. In this regard, it is instructive to contrast HC – a “born again” language? – with English – a so-called “tens of millennia”-old Germanic language. In the latter, the majority of affixes are etymologically non-Germanic; this apparently is also true for the rest of the English lexicon which has been estimated to be 65% non-Germanic. Furthermore, English diachrony and the genesis of HC exhibit comparable morphosyntactic differentials; see, e.g., DeGraff (1997, 2000, 2001a, forthcoming).
Nonetheless Goodman’s point about HC’s morphological complexity vis-à-vis its lexifier is valid to the extent that the available archival and comparative evidence suggests a robust degree of etymological continuity, which in turn disconfirms the radical morphological bottleneck posited by the pidgin-to-creole scenarios. Alleyne (1971: 172–174) makes a similar point, subject to similar caveats, when he gives linguistic evidence from French- and English-lexicon creoles that their lexifiers “in their full morphological systems where used in the contact situation”. In the same vein, Mufwene (2000b: 9) writes that “to the extent that English pidgins and creoles, as well as indigenized Englishes, can ultimately be traced back to Old English, they all have a long history”. Thus, as Goodman, Alleyne, Mufwene, and many others have argued before, there seems to be little, if any, evidence that creole genesis must prototypically proceed via “a radical reduction of [the] source languages into makeshift jargons” (cf. WSG: 144).

To recapitulate: Language age has long been taken as the crucial factor that determines level of complexity – this is in keeping with Schleicher’s intuition about the genealogy of morphology (see Section 1.2). Schleicherian linguistics takes for granted the existence of some independent, precise and operational “language dating” algorithm for genealogical/phylogenetic analysis. Yet the language-dating heuristics that have thus far been used to diagnose language youth (e.g., pidgin-to-creole symptoms such as development of new affixes via grammaticalization) simply fail to account for robust data in HC. The latter’s morphology is incompatible with its postulated ancestry in some hypothetical affixless pidgin. HC, as a sociohistorically prototypical creole language, manifests multi-millenarian morphological wrinkles.

At this stage, this reader is left begging what sorts of criteria are tacitly applied in creolists’ genealogical heuristics. Thus far, it looks like we are dealing with either some arbitrary (perhaps sociologically motivated but unstated) presuppositions and/or some circular argument. The circularity would go something like this: Creole languages are “new” because they are creole languages whereas non-creole languages are “old” because they are not creoles. 22

22. Adopting Thomason & Kaufman’s (1988) model, one could equate their “genetic” languages with WSG’s “old” languages and their “non-genetic” languages with WSG’s “young” languages. For Thomason & Kaufman creole languages emerge “non-genetically” through some abnormal “break in transmission” whereas non-creole languages gradually evolve “genetically” via “normal transmission”. Here is Thomason & Kaufman’s litmus test for distinguishing “genetic” from “non-genetic” languages: “[I]f transmission has been interrupted, then there should be […] a lack of correspondence among the various subsystems of the language, most probably between the lexicon as a whole and the grammar as a whole” (Thomason & Kaufman 1988: 11; also see pp. 8–12, 206, 211, etc.). Thomason & Kaufman’s litmus test is challenged by the same sort of epistemological-methodological and empirical-structural problems already mentioned in the main text. Their structural criterion is not given any operational
Can creolists’ theoretical elaborations on the concept “pidgin(ization)” and its import in the “pidgin-to-creole life-cycle” get us out of this conundrum?

4. On “pidgins”, “simplification”, and “basic communication”

4.1. Epistemological issues: Vagueness, circularity, falsifiability, etc.

One time-honored tradition in creole studies views pidgins and the protocreoles they gave birth to as paragons of “basic [human] communication” with near-zero complexity. The precursors of this view go back to, e.g., Saint-Quentin (1872) and Schuchardt (1914); see (2a) and (19). We find similar views in 21st-century neo-Schleicherian creolistics (WSG: 126).23

(15) [C]reole creators, in creating the pidgin that later developed into a creole, strongly tended to eschew traits from their native languages which were incidental to basic communication, and that such traits were therefore absent in the natural languages that the pidgins were transformed into.

I myself don’t trust my intuitions on the elusive notion “basic communication”, and specially not so in the teleological-functional context of (15) and its congeners in creole genesis scenarios. Above all, I don’t know what the evolutionary and structural correlates of basic communication are. Has UG specifically evolved to perfectly implement basic communication? Is UG a, or THE, perfect (post-)pidgin grammar? What are the morphosyntactic requirements of “basic communication”? Are syntactic categories like N(P)s, V(P)s, C(P)s, etc., “functionally central”? What about XP-movement and other structural transformations? Couldn’t “basic communication” do without them on a par with, e.g., formal languages? Ditto with respect to (abstract) Case marking and testable measure. In the particular case of HC, lexical and grammatical correspondences between the creole and its lexifier show no greater amount of discrepancies than their counterparts in French and English diachrony. In fact, in certain domains (e.g., lexicon, morphology, underlying word order, and nominal inflection), the discrepancies in French and English diachrony may even be greater than in HC and Jamaican Creole diachrony, respectively (abstracting away from the rate at which discrepancies qua innovations propagate through the speech community). Similar non-creole, yet (perhaps?) “significant”, discrepancies are also found in Indo-European and elsewhere (see, e.g., Thomason 1980 as cited in (8)). Pending an objective and falsifiable measure of “significant discrepancy” (what is Thomason & Kaufman’s “significant” threshold?), certain lexical-vs.-grammatical “discrepancies” in non-creole diachrony seem as “significant” as in creole genesis. See Mufwene (1998, 2000a, b, 2001) and DeGraff (2001a, b, forthcoming) for further details on the theoretical abnormality of abnormal transmission in creole genesis (also see Notes 20 and 21).

23. The statement in (15) controversially equates “creole creators” with those that “creat[ed] the pidgin”. Besides, such equation is anachronistic: if (as claimed in (15)) creoles are “the natural languages that the pidgins were transformed into”, then pidgin creation necessarily precedes creole creation.
and the $X^0$-vs.-XP distinction. Is there a maximum number of thematic roles per verb in “basic communication”? What is the shape of the lexicon in “basic communication”? What is the minimal phonemic inventory required by “basic communication”? Etc., etc. (Cf. Note 35.)

From what I gather in the UG literature (e.g., from *Linguistic Inquiry* articles), there is not much there that can be straightforwardly related to “basic communication” needs. And it even seems that there are sociolinguists interested in language change who, like generativists, are quite skeptical about teleological functionalism. For example, Labov (1994: Chapter 19) bears the title “The overestimation of functionalism”. In the next chapter on “The maintenance of meaning”, Labov concludes (1994: 598):

(16) A good many theories of language put forward recently would explain language structure as the result of the intentions of the speaker to communicate meaning to the listener. There is a part of language behavior that is subject to conscious control, to deliberate choice, to purposeful and reflective behavior. But as far as I can see, it is not a major part of the language faculty, and it has relatively little influence on the long-range of language structure.

Given such widespread, plus my own, skepticism about teleological explanations for language structure, I had hoped to find a clear structural definition of “basic communication” in 21st-century Schleicherian writings. After all, “basic communication” is the linchpin of creole genesis scenarios based on age-complexity correlations. Yet “basic communication” is thus far left without any explicit and operational criteria. No independent algorithm is provided to derive the denotation of this term and its linguistic profile. This is (methodo)logically debilitating. In absence of independent criteria for “basic communication”, Schleicherian creolists’ arguments risk circularity of the following sort:

(17) a. “Basic communication” comprises whatever structural properties make up (many? most? all?) creole languages.

b. Conversely, out of the vast array of superficial crosslinguistic distinctions, let’s (arbitrarily?) select a quite small inventory of features that happen to not exist in (many? most? all?) creoles and let’s make these features “incidental to basic communication”. In other words, features that are absent in (many? most? all?) creoles are not “basic” to communication.

(18) Features “incidental to basic communication” include “ergativity, grammaticalized evidential marking, inalienable possessive marking, switch-reference marking, inverse marking, obviative marking,
Commentary on McWhorter: Michel DeGraff

‘dummy’ verbs, syntactic asymmetries between matrix and subordinate clauses, grammaticalized subjunctive marking, verb-second, clitic movement, any pragmatically neutral word order but SVO, noun class or grammatical gender marking (analytic or affixal), [...] lexically contrastive or morphosyntactic tone beyond a few isolated cases”, and “lexicalized derivation–root combinations”

(WSG: Section 6 and Note 20)

The reasoning in (17) is the theoretical essence of neo-Schleicherian creolistics while the list in (18) is explicitly offered as a negative litmus test for basic communication. Yet, no independent justification and theoretical argument is advanced to explain why the features in (18) are “incidental to basic communication”. Given the vast array of superficial crosslinguistic distinctions, why the ad hoc list of 15 features in (18) and not others? For example, why should ergativity, but not accusativity, be dispensable in basic communication? Pending answers to these and other questions above, it seems to me that appealing to some arbitrary list of scattered features to derive creoles’ simplicity via a “basic communication” pidgin that lacks such features runs immediately into theoretical trouble (e.g., circularity, theoretical vacuity, and unfalsifiability).

4.2. On the making of “pidgins”

What are “pidgins” and how do they emerge? One thing that we seem to know, based on a variety of comparative evidence, is that pidgins (be they “early”/“reduced” or “extended”/“expanded”) cannot be uniformly reduced to some sort of lowest-common-denominator “basic communication” natural language: in standard descriptions, “early” and “extended” pidgins fall at opposite ends of the structural and functional continua.

About “EARLY pidgins”, one common observation is that they often emerge as reduced communication systems used in restricted and specialized contexts (e.g., for sporadic limited exchanges outside of one’s speech community). The prototypical, if controversial, definition for (early) pidgins is that they arise as “makeshift adaptations, reduced in structure and use, no one’s first language” (Hymes 1971a: 3; also see Schuchardt 1909, Jespersen 1922: Chapter 12, Bloomfield 1933: 472–473, Hall 1962: 151–153, etc.; see references in Section 1.2 and Note 6). Bickerton (1999: 49), for one, considers early pidgins to be “reduced well below the minimum required by natural languages”. It may even be argued that early pidgins may be unlike native languages to the extent that early pidgins’ (lack of) structure seems to fall outside the formal boundaries for natural languages as set by UG; see DeGraff (1996b, 1999b: 499–500) for some discussion. It seems then that the “drastically reduced linguistic structure and lexicon” of early pidgins with restricted functions is the result of “the very first stage of rudimentary language learning” (Hall 1962: 151–153). These
structural and functional restrictions immediately disqualify early pidgins as candidates for viable full-fledged systems for human communication. If – and this is a big “If” – “viable full-fledged […] human communication” entails the ability to encode for transmission the expressive needs of normal human beings across a functional range of topics. Something along the lines of this assumption is adopted in WSG: Section 2.3, where it is claimed that pidgins in the pre-expansion stage are “universally agreed to be rudimentary codes not fulfilling the needs of full language”.

As for “extended pidgins” (e.g., in Melanesia), they seem to function beyond and above “basic” communicative needs: these pidgins incorporate various structural properties that are as formal (i.e., not “functionally central”) as that of any full-fledged natural language (see, e.g., Hall 1962: 154–155, Keesing 1998, 1991, Siegel 1999; below I illustrate non-“basic” features in pidgins).

It thus appears that, as in other cases of (I-)language creation by adults, the making of pidgins leads to distinct looking results depending on contingent sociohistorical specifics. In Hymes’s (1971b: 69) words, “the characteristics found in development to, and of, a pidgin admit of degrees. […] pidgins and pidginization are instances par excellence of variable adaptation of means to an audience and situation”.

Cartesian-Uniformitarian methodology (see, e.g., (3)) invites us to sort out historical processes and the external entities they create (“E-pidgins”, say – social entities) from psychological processes and their concomitant individual-level creations (“I(ternal)-pidgins”, say – mental entities); see Paul (1890), Andersen (1983), and Siegel (1999) for useful overviews and methodological caveats (also see Note 17). My (null) working hypothesis is that the making of I-pidgins (I-pidginization, if you will) enlists cognitive processes that commonly unfurl, not just in situations of abrupt and/or limited language contact, but also in the various instances of second language acquisition in “ordinary” contexts of language contact.

In this Cartesian (i.e., mentalist and internal) perspective, I(NTERNA L)-pidgins are not sui generis: they are the internal linguistic states – (transient or crystallized) interlanguages, if you will – in which adult language learners (qua second-language creators in need of a lingua franca) routinely find themselves. When viewed externally a stabilized (E-)pidgin, on a par with a communal (E-)language with native speakers, is a reification and conventionalization of the creations of individual speakers interacting in specific sociohistorical contexts, with their particular linguistic ecologies and their particular communicative requirements. From that perspective, the fact noted above that pidgins in distinct sociohistorical matrices may widely differ from one another (i.e., that pidginization and pidgins “admit of degrees”, in Hymes’s words) can be naturally and constructively related to, inter alia, the observed variability in
the outputs of second-language acquisition (e.g., with respect to the structural profiles and functional characteristics of adult learners’ interlanguages and fossilizations thereof); see DeGraff (1999b: 479–508) for one overview and some references.

This Cartesian “I-pidgin(ization)” working hypothesis seems heretical to the field at large (see Siegel 1999 for one recent overview of diverging perspectives). Indeed, one (now familiar) truism in creolistics is that pidginization is a sui generis process that eschews (virtually) all morphology; see Jespersen (1922), Hjelmslev (1938), Bickerton (1988), Seuren & Wekker (1986), Seuren (1998), McWhorter (1998, 2000a, b, WSG), etc.). Per this truism, pidiggins are uniformly (and, for some, teleologically) designed ab ovo as simplest languages. The corollaries of this truism – the pidgin-to-creole life-cycle and its concomitant morphological bottleneck – constitute received wisdom in language-contact and historical-linguistics textbooks. By definition, the pidgin-to-creole cycle is exclusive to creole formation and is radically different from processes underlying the diachrony of non-creole languages. It is noteworthy that this scenario is still part of the communis opinio in creole studies notwithstanding the fact that the “classic” pidgin-to-creole litmus test fails on representative creoles, including HC (see Section 2 above and also Alleyne 1971). At this point, a bit of critical historiography is in order, before exploring modern exponents of the pidgin-to-creole cycle.

The view that pidginization entails a morphological bottleneck – a “stripping” of language-particular morphology – is already found in Schuchardt’s description of “the creole before [it] become[s] the native language of the majority” (1914 [1980: 91]):

(19) For the master and the slave it was simply a matter of mutual comprehension. The master stripped off from the European language everything that was peculiar to it, the slave suppressed everything in it that was distinctive. They met on the middle ground.

Schuchardt’s hunch echoes through much of contemporary creolistics. We still find allusion to “stripping” as in Bickerton’s (1988: 272–278) claims that “a sharp, and in some cases quite radical, reduction in the structural properties of the original target language was an essential prerequisite for new language formation” and that such reduction entails that, in the formation of radical creoles, “the target’s bound morphology [is] stripped […] thoroughly”.

More recently, Schuchardt’s “simpl[e] matter of mutual comprehension” has been linked (as in (20a)) to “basic communication” and its “functionally central” features (see, e.g., (15) and (20b)). Here, one ill-defined term – Schuchardt’s “middle ground” – is replaced in WSG by another ill-defined term – “basic communication” – which, in turn, is based on some ill-defined
mechanics for “stripping away virtually all of a language’s complexity (as defined in WSG), such that the complexity emerging in a creole is arising essentially from ground zero”. And both the early-20th and the early-21st century scenarios involve “deliberate design” – somewhat reminiscent of “naive or teleological design” as criticized by Labov (1994: Chapters 19–20), see (16) and also Paul (1890 [1970: xlv–xlvi]):

(20) a. “[P]idgins [are] stripped of almost all features unnecessary to communication” (WSG: Abstract)

b. “[P]idgins are] communication vehicles deliberately designed to eschew all but the functionally central” (WSG: Section 2.3)

In (19), Schuchardt takes some intuitive impression of “basic variety” cum “baby/foreigner talk” (cf. Bloomfield 1933: 472, Ferguson 1971, 1975, 1981) to a structural extreme. He assumes that speakers in contact situations can systematically suppress structures that are “peculiar”/“distinctive” to their respective native languages in order to create an “emergency language” for “mutual comprehension”. But this entails that, whenever speakers of (say) languages X and Y need an “emergency language” for “mutual communication”, X speakers can correctly decide which of their native structures will “meet with [Y speakers’] total incomprehension”, and vice versa. And this would be why, in the European-African “emergency language”, the Europeans eliminated European affixes (e.g., plural-marking suffixes such as English -s) while the Africans, in analogous fashion, suppressed the expression of African affixes (e.g., plural-marking prefix such as Duala ma-); see Schuchardt (1914 [1980: 91–92]).

I find Schuchardt’s claims in (19) and its modern implementations (see paragraphs below) theoretically and empirically challenging, even if seemingly common-sensical. In particular, I don’t understand the psycholinguistics of finding the “middle ground”. Ferguson (1971, 1975, 1981) seems right that every native speaker can resort to some recognizable and negotiable simplified register for speaking to linguistically handicapped foreigners. Yet, notwithstanding the broad tendencies identified by Ferguson and others toward universals of simplification (see, e.g., contributions to Clyne (ed.) 1981), foreigner talk doesn’t seem to constitute a crosslinguistically well-behaved “simplest” structural type. Ferguson himself (1971: 146, 148; 1981) shies away from positing (simplification in) “foreigner talk” as a general, absolute, and sui generis process in pidginization. Instead, regarding simplification, Ferguson cautiously notes:

(21) It is […] clear that relative simplicities [e.g., smaller lexicon, less morphology and allomorphy] occur under a wide variety of circumstances, such as in pidginization, normal diachronic change, language
Commentary on McWhorter: Michel DeGraff

acquisition, language pathology and register variation, although the
details differ from one set of circumstances to another. It is also clear
that simplification is rarely if ever the exclusive diagnostic charac-
teristic of a particular language/variety/register in comparison with
another [...].

It may well be the case that all contact situations entail simplification (how-
ever defined – e.g., as reduction in structural irregularities) to some noticeable
extent. It may also be the case that simplification can happen without (large-
scale) language contact (I return to this below). 24 However, speakers engaged
in language contact are neither telepaths, nor (psycho)linguists, nor fluent in
each other’s languages. Therefore they cannot systematically decide what
in their native speech should unambiguously count as “peculiar”/“distinctive”
to the foreigner’s ear. Furthermore, in deciding what’s “peculiar”/“distinctive”,
the “middle ground” creators must, strangely enough, abstract away from pho-
netics – the language-particular component that is most accessible to the for-
eigner’s ear. This is much easier said than done. For now, the (psycho-)linguistics of negotiating this “middle ground” strikes me as quite mysterious.
One (perhaps less mysterious) alternative then is to posit that speakers of any
pair of languages \(X\) and \(Y\) know in advance what the “middle ground” ought to
be, independently of any contrastive analysis of \(X\) vs. \(Y\). A cognitive prerequi-
site for successfully establishing this “middle ground” is that speakers of \(X\) and
\(Y\) (and of all other languages) share a universal set of hardwired instructions for
finding this “middle ground” as “simply a matter of mutual comprehension”.
(See Clyne (ed.) 1981 for bibliographies on foreigner talk; also see Section 4.4
on simplification in pidginization.)

One possible non-innatist, teleological answer to the puzzle of Schuchardt’s “middle ground” equates the latter to a pidgin qua (near-)perfect “ba-
cic communication” system (see (20)). This hypothetical pidgin is built al-
most exclusively on “functional central” features (i.e., it is “stripped off of
almost all features unnecessary to communication”). However, it is still not
clear how speakers (or linguists for that matter) can “deliberately” sort out
between “functionally central” vs. communicatively “unnecessary” linguistic
properties. Deciding what is “functionally central”, and why, remains an ever-
elusive task that has long frustrated expert linguists who are deliberately tack-
ling this problem in the leisure of their research offices, thus the lively debates
in functional linguistics toward the discovery of deep-seated, (i.e., non ad hoc)

24. See, e.g., Meillet (1919 [1926: 201]), Weinreich (1953: Section 2.3), Givón (1979: 20–
22), Thomason (1980: 361–362; see quote in (8)), Trudgill (1989: 228–229), Chambers
comparative case-studies across creole and non-creole diachrony.
correlations between function and structure. (Also see Mufwene 2000a: 72–76 for sociohistorical arguments against the teleological view of pidginization and creolization.)

There is one camp though where something like “basic communication” – qua universal set of INNATE instructions for finding the “middle ground” – has been proposed. Recently, Klein & Perdue (1997) have proposed that all second language learners go through a stable and universal “basic variety” (BV) stage. This BV, although not mentioned in WSG, somewhat looks like “basic communication” in WSG, at least in spirit: BV is a well-defined I-language, a predetermined state of the language faculty. The BV’s prototypical features are: no inflectional morphology, no grammatical morphemes, $(NP_1)\rightarrow V \rightarrow NP_2$ order, tense-marking via adverbials, no movement, no complex hierarchical structure, etc. (Klein & Perdue 1997: 311–326, 332, 336). These “organizational principles” are genetically wired via some sort of “core UG”, thus relatively independent of the native and target languages. Here is how Klein & Perdue locate BV vis-à-vis UG – compare (22a) with (20), (22b) with the “contingent accumulation of ‘ornamental’ elaboration that older grammars drag along with them” in WSG: Section 2.3, and (22c) with (29):

\begin{enumerate}
\item \textbf{a.} [T]he human language capacity provides us with the potential to process very complex structures but does not force us to do so. […] [BV] is simple and still extremely functional. (Klein & Perdue 1997: 302)
\item \textbf{b.} [F]ully fledged natural languages are but elaborations of this BV. They add some specific devices, such as inflectional morphology or focus constructions; they also add some decoration, pleasant to the ear, hard to learn, but faithfully handed down from one generation to the next. But essentially, they build on the same organizational principles. (Klein & Perdue 1997: 304)
\item \textbf{c.} [T]he BV simply and directly reflects the necessary, rather than the more accidental, properties of the human language capacity. (Klein & Perdue 1997: 304)
\end{enumerate}

Not only is Klein & Perdue’s BV controversial among second-language researchers, on both empirical and theoretical grounds (see articles in Jordens (ed.) 1997), but more to the point its postulated structure is quite unlike what we see across pidgins (see Section 4.3). That pidgins do not instantiate a uniform structural template is also recognized by Klein & Perdue (1997: 340). Furthermore Klein & Perdue’s “basic variety” is expressively handicapped and lacks some of the structural characteristics (e.g., complex hierarchical structure) associated with full-fledged languages (Klein & Perdue 1997: 302, 333). Thus BV is not a good candidate for an expressively adequate “basic communication” system. Plus the ban on (complex) recursion in BV does make it
an unlikely candidate for a prototypical ("simplest") human language. In fact BV would require more, not fewer, constraints to prevent Merge from creating complex embeddings; this makes BV quite im-"perfect" as an I-language (pace Klein & Perdue’s 1997: 337 claim that BV is a “perfect” I-language). If BV is as “perfect” as claimed by Klein & Perdue, then the ban on complex embeddings in early interlanguage is not a strictly I-language phenomenon rooted in some core parameter-setting; instead such ban must be a side-effect of rather superficial online production strategies that ease the processing burden of the non-native learner at the beginning of acquisition.

4.3. *Are pidgins designed ab ovo from “ground zero” complexity?*

The exclusively teleological-functional definition of pidgins – as speech that “eschew[s] all but the functionally central” (see (20)) – is incompatible with a variety of well-documented facts from psycholinguistics and contact linguistics.

On the empirical front, alongside robust evidence for various kinds of simplification in pidginization, there are well-documented pidgin structures that are “inherited” from (some of) the source languages (see below). In other words, these structures were not “eschewed” from the creole creators’ native languages, whether or not these structures were "functionally central". The retention and reanalysis of source-language structures in pidginization (and creolization) is not surprising given what psycholinguistics and sociolinguistics have taught us about language transfer in second-language acquisition and about the dynamics of contact linguistics; see Weinreich (1953), Labov (1994, 2001), Mufwene (1990, 2000a, b, 2001), and DeGraff (1996b, 1999b, d) for some overview and further relevant comments. Besides, the fact that certain pidgin structures are absent in many “old” languages across phyla and across time suggests that not all pidgin structures are required for basic communication. This too is unsurprising. The sociolinguistic specifics of each instance of language contact are contingent on history. It is thus tautological that pidgins’ source languages (and other relevant sociolinguistic factors in language contact) vary across time and space. Pidgins will thus “inherit” (and re-analyze) selected patterns from the languages in contact. Since it cannot be the case that every such “inheritance” exists in every (“old”) language, these “inheritances” cannot all be taken as “functionally central” to “basic communication”, lest many “old” languages are communicatively dysfunctional.

Counterexamples to (20) go as far back as Schuchardt (1914), if not earlier. Schuchardt documents widespread substrate influence on the developing “emergency language” at all levels of structure: syntax, lexicon, lexical semantics, proverbs, etc. For Schuchardt, such transfers are quite natural (1914
On the origin of creoles

251

[1980: 93]). **“[T]he slaves spoke the creole not only with the Whites but also among themselves while their mother tongue was still in existence, the latter being moreover constantly revived to some extent by the continual immigration from Africa.”**

This rudimentary sketch of the socio- and psycho-linguistics of language contact in creole genesis has since been confirmed and refined across a wide variety of cases. It has now been painstakingly documented that pidgins are pregnant with (reanalyzed) structures from the languages originally in contact, alongside structural innovations. Such massive transfers, restructurations, and innovations give rise to an array of syntactic options for any given semantic function within and across pidgins. Unsurprisingly distinct pidgins select distinct functions for morphosyntactic marking and their morphosyntactic options often enter into competition for the expression of similar functions.

Now, one could well try and argue that whatever source-language properties survive pidginization and make it into the creole must be “functionally central” properties that are “necessary to human [basic] communication”. In other words, given evidence for admixture in creoles (see, e.g., references in Notes 26 and 31), if creoles begin anew at virtually ground zero, then it must be the case that language transfers do not introduce (substantial) complexity in creole genesis.

Now, one could well try and argue that whatever source-language properties survive pidginization and make it into the creole must be “functionally central” properties that are “necessary to human [basic] communication”. In other words, given evidence for admixture in creoles (see, e.g., references in Notes 26 and 31), if creoles begin anew at virtually ground zero, then it must be the case that language transfers do not introduce (substantial) complexity in creole genesis.

Witness the following quote from WSG: Section 4.4:

(23) Creole languages are unique in having emerged under conditions which occasioned the especial circumstance of stripping away virtually all of a language’s complexity (as defined in this paper), such that the complexity emerging in a creole is arising essentially from ground

---

25. For Schuchardt (1914 [1980: 91]), “the African languages exert a pace-setting influence […] not after the creole had already become the native language of the majority; nor yet when it was created as an emergency language”. In contemporary creolistics terminology, we can translate Schuchardt as saying that substrate influence via transfer is most felt sometime during the (more stable) pidgin phase, after the initial emergence of the “emergency language” (i.e., the jargon or early-pidgin), but before it acquires native speakers.


27. A somewhat related, though distinct, argument has been explicitly advanced by Adam (1883: 4–5); see (2b). For Adam it is the substratum that limits the complexity of the “hybrid” (read “relexified”) creole. This is in opposition to Saint-Quentin (1872) where creole simplicity is “a spontaneous product of the human mind, freed from any kind of intellectual culture”; see (2a) and Note 35. See DeGraff (2001b: 90–98, 106: Note 7) for some discussion.
zero, rather than alongside the results of tens of thousands of years of other accretions.

This postulated “stripping away” entails that pidginization systematically filters out from the languages in contact “the results of tens of thousands of other accretions”, allowing the retention of only features that are “functionally central” (cf. WSG: Notes 11, 13). However, given the right ecology, admixtures do carry along “incidental” features from the languages in contact. This carry-over of source-language features is favored by, e.g., relative homogeneity of (some of) the languages in contact, relative exposure to these source languages, the socio-psychological profiles of the speakers in contact, and other sociolinguistic incidental factors (see the references in Note 26, specially works by Alleyne, Mufwene, Siegel, Singler, and Thomason & Kaufman). Such factors may even include the whims of “one [pidgin] speaker”; see Nichols’s (1986: 240) speculations about the grammaticalization of evidential marking in Chinese Pidgin Russian (see (24b)). No matter how they get carried over into the emergent contact language, these admixtures cannot all be subsumed under the (still elusive) category of “functionally central” properties. Indeed these admixtures and concomitant restructurations-cum-innovations do not seem required by basic communication; they even include some of the “incidental” features in (18). To wit, the very preliminary sample in (24) (also see the discussion in Section 6.4).

(24)  

a. Capeverdean Creole has object clitics and inflectional number marking (Baptista 1997: 262, 2001); Baptista (2001) also documents suffixal number marking in a variety of contact languages.

b. Chinese Pidgin Russian has one grammaticalized evidential marking (Nichols 1986; also see (25)).

c. Fanagalo Pidgin has noun classifiers (Heine 1978: 223).28

d. Kituba exhibits a periphrastic aspectual marker preceding the verb and a bound tense marker suffixing onto the verb (Mufwene 1997: 179).

e. The Lingua Franca exhibits inflectional gender marking and agreement (Muusse & Arends forthcoming).

28. It is interesting to note that Fanagalo, which Heine (1978: 228) considers “an extreme case of pidginization”, manifests six nominal class prefixes – more than Tsez (cf. WSG: Section 3.1.2) and more than (most) contemporary Germanic and Romance languages, all of which should count as much “older” than Fanagalo. This is as expected given the linguistic ecology surrounding Fanagalo’s genesis and the accidental aspects of language creation. This also illustrates the theoretical futility and empirical vacuousness of neo-Schleicherian creolistics.
On the origin of creoles

253


g. Ndjuka-Trio Pidgin exhibits OsV, in addition to soV and soV; and in WH-questions, the WH-object can occur either before or after the subject (Huttar & Velantić 1997: 105–108). 29

h. Nubi Arabic has lexically and morphosyntactically contrasting tone (Heine 1982: 26, 41–43, etc.).

i. Palenquero has object clitics (Schwegler & Green forthcoming).

j. Palenquero has (at least) three strategies for sentential negation: preverbal negation, discontinuous double negation (with simultaneous pre- and postverbal marking), and postverbal negation (Schwegler & Green forthcoming).

k. Papiamentu manifests both tone and stress (Rivera-Castillo 1998).

l. Solomons Pijin has a postverbal (and non-affixal) aspcetual marker that exists alongside other tense/mood/aspect markers that are preverbal (Keeseing 1991: 325–330).

m. Taimyr Peninsula Russian-based Pidgin has predominating SOV order (Wurm 1996: 86–87).

What this means from a superficial (E-language) perspective is that pidginization and creolization appear to drag, restructure, and (re-)create arbitrary amounts of so called “long-ago”/“baroque”? “random” accretions, thus perpetuating, and adding to, whatever complexity may have already existed in prior diachronic cycles. (Also see Note 26 and Muysken & Law’s 2001: 49 caveat, cited above in (13).)

In a related vein, the sheer structural diversity of pidgins refutes the claim that pidgins are “communication vehicles deliberately designed to eschew all but the functionally central” (see (20)). 30 Pidgins’ structural diversity includes

---

29. Notwithstanding McWhorter’s (WSG: Section 5.2) disclaimers separating his own notion of complexity from other linguists’ claims about (word-order) markedness, SVO is listed in WSG: Section 6 as one of the items that make up the structural profile of “the world’s simplest grammars” (see the arbitrary list of “complex” features in (18)). Thus, along this particular word-order dimension and given the superficial metric in WSG, Ndjuka-Trio Pidgin (with, e.g., soV, soV and OsV) is both more marked and more complex than “old” languages that are straightforwardly SVO. Note that the Ndjuka-Trio word-order alternations are surely not functionally central to basic communication given their absence in many presumably-functional “old” languages.

30. Do the counterexamples in (24) really disconfirm neo-Schleicherian proposals? WSG: 151 offers a ready-made, but misleading, answer: “The identification of scattered exceptions in various creoles to the general tendency I have identified does not constitute a refutation of my argument”. My counter-reply is simply based on the fact that this caveat is not heeded in WSG. For now, the features that are characterized as “incidental to basic communication”
patterns that are unfamiliar to speakers of “old” languages like English and French. Commenting on such diversity while sketching the “exotic” character of Chinese Pidgin Russian, Thomason & Kaufman (1988: 191) sum up this picture much better than I can, so I quote at length:

(25) Like the other pidgins described in [the section on diversity in pidgin structures], Chinese Pidgin Russian has features that are unusual among the better-known pidgins and creoles with European lexical sources: SOV word order, postpositions as well as preposition(s), V NEG word order, and a few inflectional and derivational affixes. […] None of these features could be predicted as the result of the operation of universal structural tendencies alone, because the suffixes represent marked constructions, and the word order features are different from the ones found in other contact languages. The presence of both preposition(s) and postpositions is itself rather highly marked in universal terms. […] We should emphasize, finally, that the examples given in this section do not by any means exhaust the instances of pidgin structures that are not promising candidates for simplified lexical source language features or features of universal grammar. […] Our goal here has been to

(e.g., in (18)) seem even more “scattered” than my counterexamples. And so are the features of the Creole Prototype in (11). For example, what (if any) theoretical principles motivate the postulation that the Creole Prototype in (11) lacks “tone distinguishing monosyllabic lexical items or encoding morphosyntactic distinctions” (as in (11b))? “Monosyllabic lexical items” and “morphosyntactic distinctions” do not look like a natural class to me. The ad hoc disjunction in (11b) could well be replaced with the equally ad hoc statement “lack of tone distinguishing words with exactly two vowels or encoding distinctions in contrastive stress”. While many of the pidgin structures above can, in principle, be related in a natural fashion to well-known (psycho)linguistic facts of language acquisition/creation and language contact (e.g., language transfer, restructuring, grammaticalization, regularization, simplification, innovation, etc.), the list in (18) hardly forms a theoretically justifiable natural class for computing complexity (see Section 5 below). Compare, say, the counterexamples and counterobservations referenced in the main text with the arbitrary list of scattered features in (18). The complexity metric defined (negatively) through (18) is based, not on a theoretically-cogent “general tendency”, but on features that seem (relatively) rare crosslinguistically – rare, at least, among the “old” languages that Western linguists are most familiar with (including the lexifiers of Caribbean creoles). WSG even quotes the quite telling remark that some of its non-creole “test” languages (Kabardian and other languages of the Caucasus) are “extraordinarily complex by any linguistic standard” – if so, the “test” languages in WSG will surely make many other languages (creole or not) look extraordinarily simple by the very metric in WSG.

In any case, given the very complexity of Language and the vast space for potential distinctions at all levels of grammar (phonology, morphology, syntax, semantics, pragmatics, discourse, etc.), the ad hoc and scattered list of “incidental” features in (18) is truly “scattered”; such a list cannot reliably estimate crosslinguistic rankings of overall complexity (see Section 6.4 below).
demonstrate that origin theories based solely on evidence from well-known, well-documented mainstream pidgins and creoles are inadequate to the extent that they fail to predict the kinds of features we have illustrated.

The last sentence in (25) applies straightforwardly to the definition of pidgins in (20) and to much else in the dogma that “the world’s simplest grammars are creole grammars”. And, alongside Muysken & Law’s *mots justes* in (13), there is another caveat, from Thomason (1997: 6–7), that demystifies this dogma:

(26) [T]he structural descriptions [of the “exotic” pidgins and creoles in Thomason (ed.) (1997)] provide a strong antidote to the still common view that all pidgins and creoles have similar and simple structures. Features like systematic OSV and SOV word order patterns of Hiri Motu, the noun class system of Kitùba, and the /kp/ and /gb/ phonemes of Sango will surely help to eradicate the idea that pidgins and creoles have maximally simple and more or less identical grammatical structures.

WSG systematically skirts all empirically “strong antidote” against the view that creole grammars are simplest. WSG offers no reference whatsoever to the “exotic” pidgins and the crosslinguistically rare distinctions that have been documented in Chinook Jargon, Chinese Pidgin English, Chinese Pidgin Russian, Fanagalo Pidgin, Hiri Motu, Kitùba, Mobilian Jargon, Ndyuka-Trio Pidgin, Nubi Arabic, Pidgin Delaware, Russenorsk, Sango, Taimyr Peninsula Russian-based Pidgin, etc. Yet there exist valuable treatises on many of these “exotic” pidgins. Mobilian Jargon is one case in point: it is closely examined by Drechsel (1993, 1997) who observes that “pidgins need not reflect universal patterns as thought earlier […] but may actually exhibit highly marked features of syntax” (1993: 344).

Pidgin structures should then count as extremely diverse, with structures that often have no counterpart in many “old” languages. There is no space here to illustrate the complete range of such diversity: the possibilities may not be endless, but they do seem to go beyond what can reasonably be imagined as “basic communication”/“functionally central” requirements. This is not surprising given the history of many pidgins in conditions of intensive language contact. I encourage the reader to consult the relevant references in order to appreciate the extent to which the definition in (20) is empirically untenable; also see DeGraff (1999b: 479–508) for additional observations and references on the ontology of pidgins.31

---

31. Surprisingly, WSG’s definitions in (20) and the creole genesis scenario therein contradict em-
With Schuchardt’s and others’ detailed evidence of admixtures, restructurings, and innovations in creole genesis, any sui-genres definition of “creole” as “language” [that] ‘begins anew’ amidst pidginization” is a fallacy. The available evidence thus far does not fare well for the teleological proposal in (20) that pidgins are “communication vehicles deliberately designed to eschew all but the functionally central”. The semantics of “functionally central” should not be up for grabs: the “functionally central” in basic human communication – if definable – should be universal and should not be left to be determined on a case-by-case basis by some ecological roll of dice, lest we run into empirical and theoretical incoherence.

4.4. “Simplification”: Terminus a quo and terminus ad quem

We have now tasted our “strong antidote to the still common view that all pidgins and creoles have similar and simple structures”. Pidgin structures are typologically diverse and they are definitely not a subset of “old” language structures – many pidgin structures are nowhere to be found in functional “old” languages such as English or French.

This said, one must reckon that the products of (large-scale) language contact do give the impression that they are, to a certain degree and in certain domains, simpler than their corresponding source languages. For example, overt morphological paradigms (e.g., phonetically-realized inflectional affixes on nouns and verbs) tend to decrease in size, morphological irregularities tend to be filtered out, various sorts of semantic transparency tend to increase, etc.
But it must be stressed that such simplification is not absolute. As documented through much of the creolistics literature, what we are dealing with is gradient simplification with respect to the languages in contact and their respective complexity in particular domains of grammar. For each specific \textit{terminus a quo} whose composition is determined by contingent sociohistorical factors, simplification leads to a necessarily distinct \textit{terminus ad quem}. Given incidental variations in the particular linguistic ecology and in the relevant sociopsychological and demographic factors, the \textit{terminus ad quem} in certain cases will be more complex in certain grammatical domains than the \textit{terminus a quo} in other cases (for case studies, see references in Note 26 and also Romaine 1992: 217 for a similar observation and Muysken & Law’s 2001 important caveat in (13)).

What are the sources of simplification in language contact as in the creole-genesis cases? This is a complex question. DeGraff (1999b: 491–499, 517–518) gives one, admittedly incomplete, answer, which is rooted in a Cartesian perspective that views pidginization and creolization as reducible to individual-level mental processes that are shared across the species (cf. (3)). There I argue that simplification stems from the cognitive limitations of adult language learners and the concomitant mechanics of second-language acquisition “under duress, in the initial stages of language acquisition in the context of language contact – contact that may be massive and abrupt, and that may involve considerable social and psychological distance between speakers in different language groups” (1999b: 491). This Cartesian (i.e., mentalist) view of simplification is Uniformitarian: the underlying (psycholinguistic) causes of simplification in creole genesis are not, and could not be, exclusively “creole” (see Section 4.2). Simplification patterns are due to the idiolect-formation mechanisms that are necessarily employed in the creation of both creole and non-creole languages. In fact, we find similar simplification patterns in well-studied cases of language change via language contact and, at the individual level, in the creation of interlanguages, and in child-language acquisition (see, e.g., contributions to DeGraff (ed.) 1999 and references therein). 32

With this in mind, let us re-examine the conceptual basis of the frequent claim that simplification in pidginization, unlike simplification elsewhere, creates linguistic neonates that, as a class, start life from virtually “ground zero”

32. WSG’s view of (second) language acquisition and its effects in creole genesis appears incoherent to me. On the one hand, WSG seems to accept the sensible view that (degrees of) pidginization can be reduced to adult acquisition, as in the history of the Riau dialect of Indonesian whose “unspecified nature is almost certainly due to a degree of pidginization in its life cycle, due to extensive acquisition by adults, having ‘shaved away’ a large degree of accreted complexity” (WSG: Section 4.4; also see WSG: Note 3). At the same time, WSG argues against the view that “creoles are born via the gradual ‘streamlining’ of a lexifier language via succeeding waves of second-language acquisition” (WSG: Note 3).
Commentary on McWhorter: Michel DeGraff

of complexity (see (23)). Let’s put aside the grotesque claim that the *terminus a quo* of pidginization (i.e., the pidginizers’ native languages) is invariably at “ground zero” complexity (cf. Note 27). How can it, then, be guaranteed that simplification of, and transfer-cum-restructuration from, the source languages, alongside structural innovations, uniformly creates pidgins at ground zero of complexity?

Let’s consider the four relatively uncontroversial propositions in (27):

(27) a. “[Effects of] pidginization [can be] due to extensive acquisition by adults” (WSG: Section 4.4); “creolization is a cline phenomenon” (WSG: Note 3).


c. The structural results of second-language acquisition are characterized by various degrees of transfer from native languages, alongside various degrees of restructuration, simplification, and innovations – all based on species-uniform cognitive processes and the necessarily contingent nature of the particular linguistic ecology, its functional demands and socio-psychological profile; see Siegel (1999), DeGraff (1996b, 1999d), De Graff (ed.) (1999), Mufwene (1990, 2001) for overviews and bibliographies.

d. The source languages in creole genesis cannot (all) be at “ground zero” of complexity. This is as expected if these source languages are themselves “older language[s] that retain at all times a degree of complexity alongside […] simplifications” (WSG: Section 4.4) and if “[o]ne would find a great many of [communicatively/functionally non-basic] features in the lexifier and substrate languages that were spoken by the creators of [the simplest] creoles” (WSG: Section 6).

If all four of these propositions hold, then the claim that “the complexity emerging in a creole is arising essentially from ground zero” in (23) is a non sequitur. The conjunction of the propositions in (27) entails, a priori, that the *terminus ad quem* in creole genesis vs. language change cannot be segregated at opposing poles of some non-arbitrary global complexity metric that takes into account the entire grammars of these languages. This is because, alongside simplification, restructuration, and innovation, language contact also entails language transfer through second-language acquisition, which will inevitably carry over into the emerging contact language some of the complexity from the
languages in contact (this is sketchily illustrated and discussed in (24) and surrounding comments). Thus, language contact cannot induce a natural class that represents “the world’s simplest grammars”. The latter, it seems to me, fall in what Foucault would call a “pure and simple linguistic monstrosity” – an unavoidable result of “the quest for primitive language [, which quest induces] a world of chimera and reverie” (see (4)).

All the aforementioned facts and observations about transfer, innovation, simplification, and typology in creole genesis and beyond should wake us up from any “chimera and reverie” whereby pidgins become systems for “basic communication” that eschew “all but the functionally central” (cf. (20)).

4.5. “Basic communication”: What are the basics?

Popular scenarios for the emergence of “the world’s simplest grammars” are fraught with epistemological problems: their theoretical foundations have long been undermined by notions left critically ill-defined – including “creoles”, “pidgins”, “young” vs. “old” languages, “basic communication”, “functionally central” features. Regarding the latter two notions, in the absence of any independently justified theory for “functionally central” properties and their crosslinguistic realization, we still lack a coherent, non ad hoc notion of “basic communication” as a linguistic-theoretical concept. For now, “basic communication” remains vague and elusive.

What would be needed to adequately define “basic communication” is a theoretical framework (some universals of “basic communication”, say) that would independently motivate the “functionally central” ingredients of “basic communication” and spell out how they are minimally realized crosslinguistically at all levels of grammar. Such a framework would, for example, predict the exclusion of the features in (18) from basic communication and explain why these features alongside other “incidental” features go beyond the requirements of “basic communication” (but see Labov’s caveat about “naive or teleological design” in (16)). In contradistinction, the list of features that are “incidental to basic communication” (see (18)) is scattered through the space of typological variation and is constructed outside any independent theory of “basic communication”.

33. Creolists who firmly believe that there exist synchronic structural diagnostics that exclusively define pidgins and creoles may well decide that the “exotic” pidgins mentioned here are not really “prototypical” pidgins. If so, it will be incumbent upon these pro-prototype creolists to provide explicit and operational structural definitions of “pidgins” and “creoles” (and of “young” vs. “old”, and “simpler(st)” vs. “(most) complex”). In order to reach a minimal level of scientific adequacy, such definitions must be neither ad hoc, nor circular, nor vague. Only then will they have theoretical and empirical bite (see Sections 5 and 6).
At this point, some pro-prototype creolist may optimistically respond that it is Universal Grammar (UG) itself, or some version of the Language Biopro-
gram à la Bickerton, that tells us about (the structure of) “basic communica-
tion”. Or perhaps “basic communication” in creole formation is a community-
wide fossilized instantiation of the “basic variety” seen in Section 4.2 (cf. (22)).
And recall that the “basic variety” itself is postulated as the product of some “minimal/core UG”. So we should ask: Does UG define the “functionally cen-
tral” requirements of “basic communication”? Before answering this question, I first need to spell out some working assumptions about UG.

It is usually assumed that UG, by its very nature, does underspecify all id-
iolects, whether creole or non-creole. As far as I can tell, there is no sense (yet?) in which UG defines a scale whereby languages can be ranked as being more or less overspecified across all domains of grammar simultaneously – or more or less removed from some innate system for basic communication. Un-
derspecification is the very essence of UG qua biological template for Human Language. UG only defines the set of PERMISSIBLE languages; no ACTUAL language is defined by UG alone; see, e.g., Chomsky (1986: 145–152, 1995: 6). This is akin to the way in which genotypes underspecify phenotypes. Lan-
guage (with capital “L”, in the singular) is innate, but languages (with small “l”, in the plural) are not. That is, humans are hardwired for Language whereas the individual expression of this capacity as idiolects – (I-)languages with their particular phonetics, lexicon, morphology, syntax, semantics, etc. – is not bio-
logically programmed, even though it is biologically-constrained. 34

Generativists aim at one abstract implementation of this underspecified tem-
plate for all human languages; see, e.g., Chomsky (1986: 145–152, 1995: 6–7, etc.). In this research program, this abstract Human Language template (aka UG) consists of “principles” and “parameters”. Principles are presumably uni-
versal, ultimately hardwired in human biology. They exist alongside an array of underspecified parameter settings and/or an array of open slots for a language-particular lexicon with its concomitant phonemic inventory, morphology, lexi-
cal semantics, etc. The parameter-settings and lexical slots become (over)spe-
cified only after exposure to Primary Linguistic Data on an idiolect-specific basis (compare with the emergence of phenotypes via the interaction of innate genotypes with incidental environmental variables). In other words, UG speci-
ifies no actual parameter settings and no actual lexicon: it is the inevitably con-

34. This assumption is orthogonal to specific issues about how Language is represented in the
mind/brain (e.g., questions of modularity – with respect to possible interactions between lin-
guistic and non-linguistic cognitive processes). For my argument to go through here, it is only
necessary to assume (uncontroversially, I suppose) that whatever properties ultimately enable
language acquisition (i.e., the creation of human idiolects) are inscribed in human biology
one way or another.
tangent linguistic experience that fills in – that “specifies” – the idiolect-specific information. This framework makes it axiomatic that every actually-occurring idiolect (including creole idiolects) will be “overspecified” with respect to UG. The latter only defines the space of – the boundary conditions on – possible human languages; it does not specify any one particular language or any one particular class of languages (pace Bickerton 1988; see (28)). Neither does UG specify a global hierarchy for classifying languages in terms of overspecification at all levels of grammar taken simultaneously.

Let’s contrast this view with the proposition that creole languages, because of their alleged youth, represent the minimal – “simplest” – instantiation of some universal set of structural requirements as dictated by UG. This is the essence of Bickerton’s Language Bioprogram hypothesis. 35 Here UG is taken as a sort of lowest-common-denominator grammar with respect to which specific languages are more or less overspecified. Such overspecification is claimed to go hand-in-hand with complexity: “old” languages are the most overspecified and they are the most complex, see the following quote from Bickerton (1988: 274):

(28) The present viewpoint accounts very naturally for this “simplicity” [of creole grammars]. In older languages, the universally shared set of syntactic principles is added to, and complicated by, a wide range of lexical and morphological properties as a result of millennia of diachronic change.

It thus seems that it is overspecifications vis-à-vis UG that, over millennia, take “old” languages away from the structural ideal of “basic communication” and away from “ground zero” complexity. Witness the following Bickertonian passages in WSG:

35. In 1872 already, a somewhat similar proposal was advanced by Saint-Quentin (see (2a)). A century or so later, we hear from Seuren & Wekker (1986: 64) that, as the main factor in creole genesis, “[Semantic Transparency] enables listeners to carry out semantic interpretation with the least possible machinery and with the least possible requirements on language learning” and from Seuren (1998: 292) that “[creole grammars] lack the more sophisticated features of languages backed by a rich and extended cultural past and a large, well-organized literate society”. The later statement seems to imply that “sophisticated” morphosyntax must be backed by old and literate (“sophisticated”?!) culture and that culturally un-“sophisticated” people (whatever that means) speak morphosyntactically un-“sophisticated” languages. But must “primitive” people speak “primitive” languages and “sophisticated” people “sophisticated” languages? The answer is no, given Sapir’s well-founded observations in (7) on the “rubbish”-ness of morphology–culture correlations and my discussion in Section 6.4 of complexity–age correlations. (Also see Note 27 and DeGraff 2001b for recurring parallels between Saint-Quentin’s and later proposals on creoles as “ab ovo” creations.)
(29) a. [In older grammars, millennia of grammaticalization and reanalysis have given overt expression to often quite arbitrary slices of semantic space, the result being a great deal of baroque accretion which, while compatible with [UG], is incidental to it, as well as to even nuanced human expression. In having not existed for long enough a time for drift to encrust them in this manner to any great extent, creoles are unique in reflecting the innate component of the human language capacity more closely than older languages do. (WSG: 126)

b. The observations I have made are couched in a view of older natural language grammars as vastly OVERSPECIFIED systems in comparison to the requirements of [UG]. (WSG: 160)

c. [Are] older grammars’ structures […] completely, or even mostly, specified by, as opposed to merely compatible with, [UG?] [A]s I have argued, this is not the case. (WSG: 132)

d. [Are] creoles […] closer to [UG] than other languages ?] The present paper is an attempt to explore and support [the] provocative idea [that they indeed are]. (WSG: Note 5)

e. [Creoles represent a fundamental layer of natural language, unobscured by the results of millennia of phonological, syntactic, and semantic drift which make [UG] such a challenge to glean in older languages […]. (WSG: 155)

f. In the realm of syntax, the hypothesis that creoles are closer to an ontogenetic foundation than many other languages appears promising. (WSG: 157)

The quotes in (28)–(29) are incompatible with the conception of UG that I sketched above; see Marantz (1983) for a related critique in the context of Bickerton’s Language Bioprogram Hypothesis.

As I have already pointed out, UG is not a “basic communication”, or lowest-common-denominator, grammar in the sense of a minimal set of “functionally central” requirements. This is perhaps made clearer by taking the lexicon as example. The lexicon is yet another area of grammar where one language can “[give] overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions than another” (cf. WSG: Section 136). Individual lexica and the distinctions therein (e.g., vis-à-vis semantic distinctions as for, say, kinship terms – along with distinctions in phonemic inventory, contents of functional heads, affixal inventory, etc.) become fully specified only upon exposure to contingent Primary Linguistic Data in sociohistorically-determined environments. Independently of (say) phonemic inventory and complexity thereof, the Primary Linguistic Data give rise to arbitrary signifiant–signifié semantic oppositions of arbitrary complexity, including word-level (“opaque
On the origin of creoles

263

lexicalizations”) and phrase-level non-compositional semantics (e.g., idiom chunks) – much of this goes back to Saussure. The point here is that the lexica of natural languages are in no way fully specified sensu stricto by UG, no matter the eventual number of distinctions therein. Furthermore, lexical(alized) distinctions and the arbitrary semantic partitions they establish are, a priori, orthogonal to, say, phonemic inventories and/or their complexity: there is no reason to expect complexity qua number of distinctions (as in WSG) to increase in lockstep across all levels of grammar.

In any case, the actual (phonetic and semantic) make-up of lexical items is constrained by universal laws of phonetics, by universal constraints on semantic interpretation, constraints on argument structure, and its linking to surface representations, etc. Take Haitian Creole as an example. At the phonemic level, HC may look simpler than French or English. But, at the lexical and morphosyntactic levels, HC uses operations like reduplication and predicate-clefting for semantic stress. The precise morphological, syntactic, and semantic details of these operations are not necessarily “simpler” (i.e., with fewer overt distinctions) than the reduplication and cleft patterns that exist in “old” families like Romance and Germanic. Yet reduplication and predicate-clefting in HC must also obey UG strictures on a par with, say, affixation and verb-movement in Romance and Germanic.

As I have already mentioned, it is axiomatic that any given idiolect, though biologically bound by UG, will be overspecified with respect to UG’s initial parametric and lexical slots and so on. These slots are necessarily left underspecified (i.e., open to parametric choices) in the initial state defined by UG. The crux of the matter is that, notwithstanding apparent crosslinguistic quirks, all such quirks and myriad others (including those found in creole languages) will, by assumption, fall within the boundaries defined by UG and indeed will help us discover the make-up of UG. No amount of complexity-building via diachronic drift can take languages “beyond the bounds of the genetic specification for language” (contra WSG: Section 6.3; see (14) above). Indeed, it is tautological that our “genetic specification for language” (i.e., the genetic encoding of UG) enables us to learn ANY “overspecification” – ANY “ornamental” elaboration – in ANY human language; that is, the mind/brain is genetically pre-wired to acquire, store, produce, and process any and all the “overspecifications” that exist across the world’s languages (see Note 17). Any linguistic feature that could not be so acquired, stored, processed, etc., would just not exist in any natural language, assuming with Descartes, Humboldt, Os-thoff & Brugman, Paul, Chomsky, and others that natural languages are mental properties of *homo sapiens*. Thus, the necessity for crosslinguistic research: it is linguistic diversity that will help us elucidate the boundary conditions imposed by UG. It is by apprehending the diversity of specific languages that we will elucidate the unity of Human Language.
Commentary on McWhorter: Michel DeGraff

This Cartesian methodology puts an ironic epistemological twist on the (neo-)Schleicherian claims in (29). Is it really “such a challenge to glean [UG] in older languages” while “creoles represent a fundamental layer of natural language [that is] unobscured”? If creoles were really “the world’s simplest grammars” with the fewest distinctions possible, then this would actually make it HARDER for pro-prototype creolists to “glean UG” and the diversity it affords. Prototypical Creoles make the “prototypical” creolist’s job most trivial: “the world’s simplest grammars” require no more than the world’s simplest analyses. In this perspective, Prototypical Creoles, as defined in WSG, would have little, if anything, to contribute to theoretical progress in linguistics (be it in phonology, morphology, syntax, semantics, etc.). As Chomsky (1986: 149) writes, “[q]uite often, the study of exotic phenomena that are difficult to discover and identify is much more revealing, as is true in the sciences generally”.

Let’s imagine a would-be analogue of (29e) in the hard sciences. Imagine, say, some hypothetical claim to the effect that chemists have their best shot at discovering the molecular make-up of nature by examining the world’s simplest molecules. If the latter were the only data that chemists had to experiment with, then they certainly would have no clue about the diverse complexity of nature’s “Universal Chemistry”. In particular, chemists working with “prototypically simplest” molecules (chemistry’s equivalent of “Prototypical Creoles”) would have no opportunity to glean the complex structures of proteins and DNA – the very molecules that make our existence possible. Similarly, if our field notes and intuitions were exclusively about Prototypical Creoles as defined in WSG, then we would have no clue about inflection, tone, Saussurean lexicalizations of root-affix combinations, ergativity, grammaticalized evidential marking, inalienable possessive marking, etc. The list in (18) hints at other features that UG makes available, but that linguists could never glean from the hypothetical “simplest grammars” defined by that list (cf. Note 36). In DeGraff (2001b: 76–78, 86–88), I argue that the Creole Prototype defined by (11) – by (11c) in particular – even lacks some of the basic Saussurean properties that are usually associated with natural languages and their lexicon (see Section 2).

If UG both defines the language learner’s innate initial state and imposes boundaries on the outcome of acquisition and, thus, on each and every I-language, then it cannot be the case that UG is “a challenge to glean in older languages” because of their “incidental” features. To the contrary, UG is best studied through our exploration of the diversity of languages, and this exploration is best carried out when guided by our theoretical results about the unity of Human Language. In this view, Prototypical Creoles as theoretical constructs with simplest, unexpected, and ad hoc properties do constitute a typological-ontological (and epistemological) challenge to (the study of) our faculté de langage and the crosslinguistic structural possibilities it affords.
From the perspective of UG as sketched in this section, Prototypical Creoles become “linguistic monstrosities” (cf. (4)).

Be that as it may, we still lack any operational criteria for “basic communication” and its “functionally central” properties. Yet at this point we must raise the central question: How is “complexity” defined?

5. A most simplistic “complexity” metric

First, some words of caution: Any descriptively and scientifically adequate complexity metric requires an independent theory of complexity (that explains what is to be counted, why, and how) along with exhaustive descriptions for the languages to be compared (so we can list all that is to be counted), lest our complexity metric have exclusive scope on arbitrary bits of grammar with no consequence whatsoever for linguistic theory and global complexity across languages. No general claim about crosslinguistic levels of complexity is reliable if it focuses solely on a small set of disparate superficial patterns that are not unified in any kind of linguistic theory or psycholinguistics. More generally, simplistic and highly selective measures of complexity whose benchmarks focus on arbitrary and isolated aspects of surface strings in some handpicked sample of languages seem largely orthogonal to the theoretical and/or psychological foundations, and to the descriptive goals, of linguistic typology and theoretical linguistics (see Notes 37 and 40).

I thus agree with Muysken (1988: 288) that “the idea that creole languages are not grammatically complex in general only makes sense if one has a theory of grammatical complexity to fall back on”. Chaudenson (1994) makes similar points, as he notes the absence of any coherent evaluation metric in past and current allegations of extraordinary creole simplicity. Without any independent theory and formal criterion for complexity, we cannot even begin to determine how particular properties (or absence thereof) contribute to global complexity. With this in mind, let’s proceed to evaluate the most recent complexity metric in neo-Schleicherian creolistics.

5.1. Defining complexity via description length (= number of information bits)

The complexity metric in (30) is simply a count of “overt distinctions and/or rules” in (30a) (WSG: Section 2.4.3), which in turn is related to “length [of] descriptions” in (30b) (WSG: Section 2.4.2).

(30) a. The guiding intuition is that an area of grammar is more complex than the same area in another grammar to the extent that it encompasses more overt distinctions and/or rules than another grammar.
Our object of inquiry is differentials between grammars in degree of overspecification (as we will see, all grammars including creoles can be argued to be overspecified to some degree), to the extent that some grammars might be seen to require lengthier descriptions in order to characterize even the basics of their grammar than others.

Let’s call this view of complexity BIT-COMPLEXITY.

Bit-complexity immediately faces a number of unresolved methodological and theoretical problems that render it scientifically unusable at best and tendentious at worst.

Let’s paraphrase (30) in a transparent information-theoretic way. As sine qua non for a rigorous and objective application of the sort of complexity metric sketched in (30), we must at the very least get straightforward answers to the following questions:

(31) a. The sort of complexity that is hinted at in (30) is proportional to “degree of overspecification”, thus to “length [of] descriptions”. For any given language \( L \), what is the number \( n \) of information bits needed to describe the (entire set of) “overt distinctions and/or rules” therein? \( L \)'s complexity increases with \( n \) which is proportional to the length of \( L \)'s description.

b. The count \( n \) in (31a) presupposes a theory of grammar that would enable – or provide an algorithm for – the identification, then the counting, of language-specific “overt distinctions and/or rules”. What is the (implicit) theory of grammar that identifies the items to be counted by \( n \)?

In Sections 5.2 and 5.3, I address these two questions in turn.

5.2. Not all “bits” of grammar have theoretical bite

In order to adequately answer the “\( n \)” question in (31a) for any given language, no less is needed than (an approximation of) the description of this language at all linguistic levels (i.e., for all “area[s] of grammar [with] overt distinctions and rules”, including phonology, lexicon, morphology, syntax, semantics, discourse, etc.; cf. (30a)). Much progress has been made in typological linguistics, yet such exhaustive descriptions are not generally available for all of the world’s languages. In the absence of such descriptions, no \( n \) can be reliably estimated toward an unbiased global comparison of all natural languages (e.g.,
in order to discover what “the world’s simplest grammars” are).\(^{36,37}\)

Pending such exhaustive descriptions for all areas of creole and non-creole grammars, what one should expect from an analysis that partitions natural language into simplest and most complex classes is a general theory whereby one can safely extrapolate from the (apparent) complexity of isolated and arbitrary linguistic properties (see, e.g., (18)) to global complexity. No such theory is hinted at: the very features in (18) from a very small and selective set of (apparently) “far out” languages seem to have been picked exactly so that the few creoles chosen as “test cases” show less bit-complexity in the corresponding domains than the few non-creoles chosen as “control cases” (see Note 37).

What we have had thus far in many searches for the world’s simplest languages are formulas for “rigged” experiments – experiments that are designed to guarantee the desired “results”.\(^{38}\)

---

36. Exhaustive and reliable grammars are notoriously lacking for creole languages, which still lack strong communities of native-speaker linguists (compare, for example, the study of Sara-maccan syntax with that of Dutch syntax). And, to this day, the empirical basis of creole studies is weakened by the long-standing prejudices of linguists like Seuren who consider creoles to “lack the more sophisticated features of languages backed by a rich and extended cultural past and a large, well-organized literate society” (1998: 292) and by the too-common practice in creolistics to base one’s arguments solely on skewed and sparse samplings of unanalyzed utterances as if creoles wore their entire grammars on few superficial strings (see Appendix B). It thus seems that any arbitrarily stipulated metric that ranks creoles as “the world’s simplest grammars” is a self-fulfilling prophecy (see Note 38).

37. WSG hunts for the most unfamiliar non-creole languages with the most “exotic” linguistic properties (e.g., Tsez, Maori, Lahu, etc. with a combined total of less than one million speakers). Yet WSG systematically avoids the anti-Schleicherian “antidote” of less-familiar contact languages and their less-familiar – thus, certainly un-“basic” – features (cf. Sections 2, 4, and 6, especially the quotes in (13), (25), and (26) and the sort of data illustrated in (24) and Section 6.4). In effect, WSG: Section 1 ultimately pays only lip service to the stated “intention [for] a sustained investigation of creoles from the perspective of crosslinguistic configurational possibilities, beyond the Western European lexifier languages that have served as the primary focus of creolists’ attempts to define the term creole”. Such investigation also requires the thorough comparison, at all levels of grammar, of “old” Western European languages (e.g., English and French) with “born again” contact languages whose ancestors exclude European languages and include, say, “fearsomely elaborated” languages such as Tsez, Lahu, Maori – the non-creole benchmarks for “old language” complexity in WSG. Such systematic comparison is sorely missing, which makes hypothesis-“testing” in WSG look like a rigged experiment (see Section 6.4).

38. The coarse and a-theoretical complexity metric in WSG: Section 2.4.2 is justified as follows: “I believe that the difference in degree of complexity between older grammars and a subset of creole grammars is distinct enough that a complexity metric so fine-grained as to, for example, allow us to rank Romanian, Hausa, and Korean in terms of some general complexity quotient would be unnecessary to our project.” This a priori belief that “the world’s simplest grammars are creole grammars” becomes self-justification for replacing the theoretical necessity of a “general complexity quotient” in favor of a stipulated “metric” based on a scattered list of features based on a skewed sample of languages (see (18)). The metric itself is devised with the expressed goal of separating creoles from non-creoles; it ignores a large body of relevant
5.3. How many “rules”? The looks of languages vs. the essence of grammar

For the sake of the argument, I will now abstract away from the methodological issues in Section 5.2. Instead let’s ask this: Is complexity only, or primarily, a matter of counting information bits – “overt distinctions and/or rules” as in (30a) – no matter the source of these bits (i.e., no matter the theory underlying these distinctions and rules)? If a complexity metric involves counting, then we better make sure we know what we are counting. Linguistic distinctions and rules are not pretheoretical objects that we can gather and count without prior analysis.39

In fact, bit-complexity bears no relation to any theory where grammatical phenomena are independently identified and analyzed. WSG (Sections 2.4.1–2.4.3 and Note 6) explicitly cuts off its complexity metric from the better understood areas of (psycho)linguistics, including grammatical theory, acquisition, production, and processing. In other words, bit-complexity may well have no basis in (what we know about) Language in the mind/brain – our faculé de langage. Bit-complexity, as defined in (30)–(31) is strictly a-theoretical: this is literally bit counting with no concern for psychological-plausibility and theoretical insights.40

39. In biology, where discussion of complexity is grounded in more solid empirical and theoretical results, it is not at all clear that naive counting (e.g., the counting of genes) could lead to any scientifically satisfying notion of complexity. For example, Szathmáry, Jordán, & Pál (2001) write: “Is the number of genes in an organism’s genome an appropriate measure of biological complexity? […] The recent flurry of completed genome sequences, including our own, suggests that this is not necessarily the case […] Rather surprisingly, it turns out that the worm Caenorhabditis Elegans has 18,424 genes in its genome, the fruit fly Drosophila Melanogaster 13,601, the plant Arabidopsis about 25,498, and humans about 35,000. This suggests that there must be other, more sensible genomic measures of complexity than the mere number of genes.” Szathmáry et al.’s proposal is to use “networks of transcription factors and the genes they regulate, rather than […] simply counting the number of genes or the number of interactions among genes”. Thus, biologists go beyond simple counting of overt items and they enlist inter alia abstract computational theories about “the connectivity of gene-regulation networks”. These theories are at the core of our understanding of how genes work (e.g., the mechanics whereby certain genes are switched on and off in order to represent information and compute over these representations). No need to say that these theories go beyond my competence. Yet such a development strikes me as normal for the sciences. No matter the field, complexity measures – if they are to be scientifically constructive – must be related to broad empirical and theoretical concerns (cf. Darwin 1871: Chapter 2, 61–62 for relevant remarks on “how easily [our biological and linguistic complexity metrics] may err”).

40. The apparent divorce in WSG between bit-complexity and complexity as understood by (psycho)linguists (e.g., language acquisition and language processing researchers) is all the more unexpected given the often discussed relations between creole genesis and language acqui-
At the onset, let’s note one of the (bizarre) logical consequences of bit-complexity: the languages with the biggest lexica would be the most complex – independently of, say, their phonology and their syntax. Indeed, each new lexical item further partitions the speaker’s semantic space (recall Saus- sure’s view of the lexicon as a system of oppositions). For any given degree $d$ of complexity, a large enough lexicon automatically carries enough “overt distinctions” to make the corresponding (I-)languages complex to the $d$ degree. In this view, the proposed complexity metric (in particular, (30a)) applies even within the “same” (E-)language of a given speech community: within that community, the (I-)languages with the biggest lexica will unavoidably require the “length[est] descriptions” (cf. (30b)) and will thus be the most complex languages ceteris paribus. This strikes me as a rather naive view of language complexity. Number of (superficial) “overt distinctions and/or rules” without regard for linguistic theory (assuming for the sake of argumentation that such sets can be made available) seems, to me at least, a rather crude and uninteresting way to approach linguistic typology.

A simplistic bit-complexity creolist could well try to save the “overt distinction” metric and argue that the lexicon altogether lies outside the scope of his metric. So let’s now move from bit-complexity in the lexicon to bit-complexity in the syntax. For the syntax too, bit-complexity simplistically implies counting – here, counting of “rules” (WSG: 136):

(32) A syntax is more complex than another to the extent that it requires the processing of more rules, such as asymmetries between matrix and subordinate clauses (e.g., Germanic verb-second rules), or containing
two kinds of alignment rather than one (i.e., ergative/absolutive and nominative/accusative) [...].

But (how) do we know when the “syntax [...] requires the processing of more rules” without a theory of syntax and without a theory of processing? Here, one cannot simply rely on the “looks” of language; instead one must evoke the “essence” of grammar – one must enlist a theory of grammar lest any complexity metric becomes fatally ill-defined. Deciding whether and where a particular syntax “process[es] more rules” presupposes an independent theory of grammar which the syntactician can use to discover the language-particular “rules” to be counted. Syntax is not directly readable from strings; see, e.g., Hawkins (1988) for similar points within a comparison of generative vs. typological approaches to grammar. Even creole languages, which in neo-Schleicherian creolistics “represent a fundamental layer of natural language [that is] unobscured” (see (28) and (29e–f)), do not bear their syntax on their strings. Actually, theoretical creolistics, like much else in linguistic theory, is the theater of vivid debates about the nature of creole “rules”; see, e.g., DeGraff (1999c, 2001b) and Y. Dejean (1999a); also see Section 6.4 and Appendix B for some anti-Prototype samples. In any case, to-date we have no exhaustive list of syntactic rules for the relevant languages, and comprehensive descriptions are even more sorely lacking for “exotic” contact languages, specially those created outside the Caribbean basin (see Section 4.2 and Note 36). Be that as it may, there is an inherent methodological cum conceptual fallacy in a complexity metric that relies on the counting of (language-specific) syntactic rules without an explicit theory of syntax for identifying and classifying said rules.

One example will straightforwardly illustrate the extent of this fallacy. Let’s consider the statement that “asymmetries between matrix and subordinate clauses (e.g., Germanic verb-second rules)” entail an increase in complexity via the “processing of more rules” (see (32)). This is presumably because such asymmetries involve distinct “rules” for root vs. embedded clauses, thus an increase in the number of “overt distinctions and/or rules” (cf. (31a)). No analysis is presented for the (added) rules that underlie this added complexity – the latter is taken for granted, and mistakenly so.

Since at least den Besten (1981) it has been argued, and it now seems quite likely, that there isn’t any “asymmetric [...] Germanic V2 rule” per se. Typically, V2 in German(ic) results from the application of X0- and XP-movement rules, both of which are made universally available by UG. The finite verb moves quite high outside of VP (e.g., to the C(omplementizer) position) while a maximal projection (e.g., some topic or operator) moves to the left of the verb (e.g., in Spec(CP)). The root–embedded “asymmetry” itself is understood by many Germanicists to be epiphenomenal, emerging as a surface side-effect of the interaction between abstract syntactic ingredients (e.g., head and phrasal
movement plus the features and contents of V, C, Spec(CP), etc., and their associated functional projections in root vs. embedded contexts).

Let’s consider the German(ic) V2 asymmetry as an example – grossly oversimplifying the available analyses. German matrix clauses have a C(omplementizer) head that is usually empty, thus available as a possible landing site for V(erb)-movement, with the finite verb surfacing right-adjacent to a moved XP in Spec(CP), thus matrix V2. In German, the embedded C head is usually filled by an overt complementizer, which blocks V-to-C movement; the embedded finite V is thus stuck in the IP, in clause-final position; thus, the root–embedded asymmetry with respect to V2. Yet, in embedded clauses that allow an EMPTY C head, the “V2 asymmetry” disappears: V-to-C and XP-to-Spec(CP) take place, giving rise to an embedded V2 pattern that is “symmetric” with the root V2 pattern. Similar symmetry is robustly displayed in Germanic languages such as Icelandic and Yiddish. V2 in these languages can be argued to result from movement of the finite verb, not to C, but to a head lower than C, thus the lack of asymmetry since V2 does not depend on the (lack of) contents of C. Crosslinguistically, observed V2 (a)symmetries (as in German, Dutch, Yiddish, etc.) reduce to the complex interaction between X0 and XP movement (and the triggers and/or semantics thereof) and language-specific properties of clause structure and functional heads, etc. (See Vikner 1995 for an overview.)

Notwithstanding current debates about the exact mechanics of V2 in Germanic and beyond, the lesson from syntax about the alleged complexity of “asymmetric” rules is clear. The above (simplified) analyses for V2 teach us that “V2 asymmetry” does not necessarily entail “the processing of more rules” than (say) “V2 symmetry”. Even “symmetric” languages such as Yiddish and Icelandic instantiate similar sorts of movements, albeit within a different clausal topology. At the right level of analysis, the so called Germanic “movement rule asymmetries between matrix and subordinate clauses” become a rather superficial side-effect of a single uniform operation – head-movement of V into C and XP-movement to Spec(CP) – which applies whenever possible. The blocking of V-to-C due to overt C arises via independent morphosyntactic requirements (e.g., selectional requirements and the morphology and semantics of the CP layer). In this view, root-vs.-embedded (non-)V2 patterns in Germanic (and elsewhere) are not the result of distinct (root vs. embedded) transformations; instead such patterns result from deeperuniversal principles of the computational system of our faculté de language (e.g., structure building, selectional requirements, movement transformations, etc.) interacting with language-specific morphology and/of functional heads.

From this theoretical perspective, root–embedded asymmetries do not necessarily increase complexity in ways that are alien to creole languages. In fact, creole languages too manifest (superficial) root–embedded asymmetries,
contra the erroneous generalization in WSG: Section 6. Haitian Creole, for one, has such asymmetries with respect to WH-phrases in direct questions: the moved WH-phrase, no matter its underlying position, overtly moves to the root Spec(CP). So in HC direct WH-questions, the WH-phrase is pronounced in the root clause, never in an embedded clause; thus the appearance of a root–embedded asymmetry. This is unsurprising in any theory where (direct) questions need to be typed as such at the root level and/or where semantic operators need to take scope over their quantification domains. (Also see Syea 1997 for another possible instance of creole root–embedded asymmetry, with respect to copula distributions and V-to-C movement; cf. the Mauritian Creole data in Appendix B.) In creoles (as in non-creoles), apparent asymmetries result from deeper universals of syntax.41

More generally, syntactic theory in the generative framework has witnessed a fundamental move away from lists of language-specific and/or construction-specific “rules”. Current generative syntacticians have adopted the Principles-and-Parameters/Minimalist hunch that “constructions” arise via the complex interaction between, on the one hand, operations and constraints that are universal and, on the other hand, language-specific properties that reside mostly in the lexicon – in particular, in (the morphology of) functional heads. The number of operations made available by UG may well be few (e.g., Merge, Move, Agree, etc.). Yet these few universal operations interact in complex ways with numerous language-particular properties, thus the vast and intricate array of superficially distinct crosslinguistic phenomena (e.g., the “V2 asymmetry” in German, the WH-movement asymmetries in Haitian Creole, the copula asymmetries in Mauritian Creole, the Chinese in-situ WH-phrases, etc.). In such a framework, complexity does not reside in the number of different “overt distinctions and/or rules” (e.g., distinct/asymmetric rules for root vs. embedded

41. With respect to WH-movement, it is instructive to test the complexity claim in (30a) by a comparison of Haitian Creole (a “born again” language) and Chinese (an “old” language). HC, but not Chinese, has overt WH-movement. Thus Chinese shows fewer “overt distinctions and/or rules”: both WH-questions and their declarative counterpart manifest the same surface word order. There is no root–embedded asymmetry in the surface distribution of Chinese WH-phrases. This is unlike the overt asymmetry in the positioning of WH-phrases vs. non-WH XPs in HC. Yet, at some abstract level (“Logical Form”), it can be argued that Chinese – a WH-in-situ language – does have WH-movement of the sort found in HC. Indeed the distribution of Chinese WH-phrases and their structural relations to abstract scope positions obey some of the same constraints that regulate the distribution of overtly moved WH-phrases as in HC; see, e.g., Huang (1982) for the basic Chinese facts, Chomsky (1986: 152–155) provides a handy summary. Here too, as with the V2 case, the view in (30) is theoretically naive; it is too tied to the superficial “looks” of languages to offer any deep insights into the evolution of grammar and grammatical complexity. Indeed the metric in (30) ignores much of what syntacticians have taught us about the abstract “essence” of grammar. (See Appendix B for similar flaws in McWhorter’s 2000b Afrogenesis Hypothesis.)
German clauses). Such construction-particular and language-particular “rules” may not even exist, although their labels are often retained for taxonomic descriptive purposes (Chomsky 1995: 170).

The moral of the story is clear: there can’t be any “counting” in syntax without an explicit theory of syntax that independently tells us what needs to be counted, and how. Any comparative approach that gives the “looks” of languages priority over the “essence” of grammar runs the risk of becoming a most simplistic and misleading linguistic measure.

5.4. Complexity is no simple matter

Bit-complexity in (30) is quite ambitious: it is meant to rank the entire grammars of the entire set of creole languages against the entire grammars of the entire set of non-creole languages, with perhaps the few exceptions noted in WSG: Section 6. Recall the central claim that creole languages and non-creole languages tend to fall at opposite ends of the bit-complexity cline. This claim is not only about Lahu vs. Saramaccan, or Tsez vs. Saramaccan, or Maori vs. Saramaccan. Nor is this claim to be evaluated with respect to only a handful of linguistic features such as those in (18). Bit-complexity has universal scope: it is a claim about all languages across all areas of grammar (phonology, lexicon, syntax, semantics, pragmatics, etc.).

With this in mind and given the arguments above, it is worth stressing again that bit-complexity as “tested” in WSG enlists only an arbitrary set of linguistic properties (such as those in (18)). These properties are picked from a handful of exotic non-creole languages (e.g., those mentioned in paragraph above) without recourse to any independently-motivated theory of grammar, processing and/or acquisition. Thus, this metric has no principled implications for Language in the mind/brain (but see Note 40). Given our current state of knowledge and the complex nature of Language, we can’t yet afford a global complexity metric with global crosslinguistic scope. In the meantime, the handpicking of languages and linguistic features in implementing and testing the metric in WSG belies the purpose stated therein to elaborate “a direct comparison of certain creole grammars with older language grammars, with a view towards making more precise my grounds for the claim that creole grammars constitute a synchronically identifiable class” (WSG: Section 1). If the few “test” languages and the few “test” properties are both prejudicially chosen without regard to any independently-established criteria, then whatever we may learn from this comparison is not enough to equate creoles to the (natural?) class with the label “world’s simplest grammars”.

Empirically it has already been argued above that a slightly larger sample of crosslinguistic data and typological/diachronic observations undermine neo-Schleicherian creolistics. The rest of this critique brings additional method-
ological and empirical observations that further undermine the bases of the new Schleicherian linguistics.

6. “Learning by debunking”: The empirical (non-)basis of age–complexity correlations

Here I focus on some of the specific empirical problems that undermine the “testing” of age–complexity correlations. In a nutshell, what we are dealing with is a set of “rigged” experiments where the “test” cases and “control” cases seem carefully handpicked to provide support for a neo-Schleicherian creolistics hypothesis. But this empirical support will be shown to be illusory: the empirical claims in neo-Schleicherian creolistics ultimately lead to theoretical incoherence, specially vis-à-vis purported complexity–age correlations.

6.1. Age before complexity? Complexity before age?

Schleicherian circularity

Pre- and neo-Darwinian linguistics from, say, Schleicher (1863) to WSG rests on the following premise: Complexity increases with age – as a language gets older, it gets more complex (but see Notes 5 and 24). It is further postulated that, after tens of millennia, old languages “all come to rest at a certain ‘surplus complexity quotient’” – an evolutionary plateau of maximum complexity that excludes creole languages (WSG: Section 2.3), even though a few old languages like the Riau dialect of Indonesian and Southeast Asian languages may have slipped from the maximal-complexity plateau and acquired “pidgin-level syntax” due to “extensive adult acquisition” (WSG: Section 4.4). Temporarily putting aside the theoretical elusiveness of the notions “language birth” and “language age” qua linguistic constructs sensu stricto (see Sections 2 and 3 above), I will argue that (neo-)Schleicherian complexity–age correlations are robustly disconfirmed by the available diachronic and typological data.

Consider inflection, for example, which is taken as a marker of complexity-cum-age: “[I]nflexional morphology renders a grammar more complex than another one in most cases” (WSG: 137); “inflexion almost always complexifies a grammar” (WSG: 138); “this language [Riau Indonesian] reveals its age in having [inter alia] three inflections […]” (WSG: 155). How many inflections does a language need to “reveal its age”? As is noted in, e.g., DeGraff (2001b: 71–76) and Muysken & Law (2001) (see (13)), it is not true that creole languages lack inflectional morphology. In fact, even a(n early) pidgin can show inflectional morphology given the “right” language-contact ecology (as in, say, the case of overt transitive marking in Melanesian Pidgin in the mid-1880s described in Keesing 1991: 318–320; also see some of the contribu-
On the origin of creoles

42 If three inflections plus “some opaque derivation-root combinations, and optional numeral classifiers” are enough to “reveal [old] age”, then creoles would have to be considered “old” languages. Considering their affixal inventories and their non-pidgin syntax, some creoles (e.g., HC, Mauritian Creole, Capeverdean) must be at least as old as the Riau dialect of Indonesian per the very description and criterion in WSG: Section 4.4. This is not expected within neo-Schleicherian proposals for age-morphology correlations.

Be that as it may, does inflectional morphology always increase with age? It has long been observed in the grammaticalization and historical-linguistics literature (see, e.g., Meillet 1912) that, from a diachronic perspective, grammatical systems (e.g., Case and Tense/Mood/Aspect – TMA – marking) often evolve along analysis–synthesis cycles whereby overt markers go through the ebbs and flows of syntax (analysis/periphrasis) and morphology (synthesis/word-level processes). Free-standing auxiliaries can become verbal inflectional affixes, pre-/postpositions can become nominal case affixes, and both verbal and nominal affixes can fuse to their stems and erode over time. Given such morphology–syntax cycles, hope springs eternal for any affixless language with affix envy: “Weep not, my children, for today’s syntax is tomorrow’s morphology” (Givón 1971: 413: 1; also see Hodge 1970, Giacalone Ramat & Hopper (eds.) 1998, Heath 1998, Haspelmath 2000, and Janda 2001 for some discussion of (de)grammaticalization phenomena and their theoretical bases; also see Note 46).

Schleicher himself was well aware of the erosion of morphology in “old” languages and clever enough to try and incorporate inflectional decay in his evolutionist scenario, distinguishing between “evolution” and “history” while strenuously holding on to his complexity–age correlations (see Note 5). Classic instantiations of the rise-then-decline of morphological marking are common in the history of Romance and Germanic and, more generally, throughout Indo-European and elsewhere. Compare, say, verbal inflection and nominal case in Old English vs. Modern English; ditto regarding, say, the evolution of nominal case and verbal inflection from Latin to Romance.

Here, we get the exact opposite of the Schleicherian complexity–age correlation. Vis-à-vis case and verbal morphology in the relevant stages of Germanic and Romance diachrony, “older” implies “simpler”, assuming bit-complexity as in (30). Modern English, for example, has substantially fewer overt distinctions in, and fewer combinations of, verbal inflection than Old English. Ditto in the domain of overt case morphology on non-pronominal noun phrases: Mod-
Commentary on McWhorter: Michel DeGraff

ern English has none of the overt case affixes that were once productive in Old English. WSG: 138 does note that “English expression of case is simpler overall than Latin’s”, but it is not noted that the Modern English expression of case is also much simpler than that of Old English, and so is the Modern French expression of case simpler than that of its ancestors – in terms of bit-complexity.

The English and French cases are counterexamples to the claim that “diachronic drift […] encrusts older grammars [with complexities]”: diachronic drift has reduced the overt inflectional paradigms of both English and French, thus decreasing their inflectional bit-complexity. Therefore, it is not at all clear that inflectional paradigms can be used as reliable indicators of language “age” (also see the case of Riau Indonesian above). In fact, within single language families we do find languages with drastically different degrees of overt inflectional morphology. Compare, e.g., Malayalam to other Dravidian languages, English to Icelandic, French to Italian, or Western European languages to Balto-Slavic languages as in (8) (also see Notes 5 and 24 and the comparative data in Hodge 1970). It has long been established that much morphological variation is expected within any single linguistic phylum. Assuming that all languages within a particular family “go back” to a single ancestor proto-language (abstracting away from the theoretical difficulties in language “dating”), then such variation within single families goes against the notion that language age can be correlated with bit-complexity. Of course, this is reminiscent of Edward Sapir’s memorable quote in (7); also see (8).

It must also be noted that acquisition processes (either in ordinary situations or in situations of abrupt language contact) exert an inevitable pressure toward regularization and/or morphological “simplification” in certain domains, even if this pressure is counteracted elsewhere in the grammar by other factors such as certain types of language transfer, grammaticalization, innovation, and the like. It is not only in the diachrony of Riau Indonesian that we find “a [certain] degree of pidginization […] due to extensive acquisition by adults, having ‘shaved away’ a large degree of accreted complexity” (cf. WSG: Section 4.4). Acquisition by adults with its potential for morphological erosion is a widespread phenomenon in language contact and language change (see Notes 5 and 24). Furthermore, the standard fare in historical and

43. This observation also undermines the claim that “all natural languages [except creoles] would be equally complex by virtue of having all come to rest at a certain ‘surplus complexity quotient’”. It is hard to conceive of an individual mental algorithm (as part of language acquisition, say) that would ensure that bit-complexity remains constant across “all of the grammars [that] trace back tens of millennia” (cf. WSG: Section 2.3). This would entail that speakers of (say) individual Indo-European languages must be able to check each other’s bit-complexities in, for instance, inflectional morphology (cf. (8)) and compensate any discrepancies therein by adjusting bit-complexity in some other parts of their respective grammars. Such an algorithm is as improbable as the one critiqued in WSG: Section 2.2.
contact linguistics teaches us, for example, that: (i) sound change often leads to loss/assimilation of phonemic distinctions, (ii) morphological change often leads to regularization and analogical leveling (i.e., to loss of morphological distinctions), (iii) syntactic reanalysis often leads to structural simplification (or structural transparency), etc.; see Campbell (1999) for an overview.44

In fact, functionalists have often argued that ease of articulation, regularization, and rule extension are at the roots of language change, which is driven by functional factors – principles of least effort, economy, optimization; see Labov (1994: 547–568, 2001: 16–28) for a critical overview in the context of “maladaptive” language change. Some of these principles of economy entail a reduction of bit-complexity: as Labov notices, there is a long series of arguments that language change – sound change, in particular – may be “dysfunctional”, and in the 19th century some of these arguments were advanced within language-as-evolving-organism approaches (but see Note 5 for a sample of concurring and diverging opinions). Labov (1994: 586–599, 2001: 10–14) surveys a number of areas where sound change does reduce overt distinctions, thus bit-complexity. Labov (2001: 10) remarks: “The almost universal view of linguists is […] that the major agent of linguistic change – sound change – is actually maladaptive, in that it leads to the loss of the information that the original forms were designed to carry. Though there is a wide range of divergent opinions on the nature of sound change […] there is general agreement on the negative character of this fundamental process”.

In effect, this means that the “older” a language, the more opportunity these complexity-reducing “negative” processes would have had to reduce its bit-complexity, thus producing “loss of information” (in Labov’s terminology). Here again, we get the equation “older = simpler” at least in certain domains, with “old age” entailing less, not more, overt distinctions (as with English and French case and verbal morphology). It is also expected that processes like grammaticalization (e.g., of free morphemes into bound morphemes) may ultimately offset some of the results of regularization, leveling, morphological erosion, etc. (cf. Givón’s aforementioned quip “yesterday’s syntax is today’s morphology”, which goes back to, e.g., Bopp’s, Humboldt’s, and Meillet’s

44. Regarding syntax, the following is presented in WSG as an example of increased bit-complexity: an Irish pattern taken from Henry & Tangney (1999) which exemplifies “two kinds of alignment rather than one (i.e. ergative/absolutive and nominative/accusative)”. However Henry & Tangney’s discussion of these Case patterns suggests that in syntax too bit-complexity can decrease simply as a result of language acquisition sans creolization, as in the L2 acquisition of Irish by English-speaking children in immersion programs with non-native models. These children replace the “two kinds of alignment” in copula constructions with a uniform nominative/accusative pattern. Furthermore, Henry & Tangney (1999: 251) insist that the simplification they document “is not unique to acquisition in these circumstances, merely an acceleration of a normal process”.

44
insights – Humboldt (1836 [1988: 205]; see Note 5) talks about “the wearing-down of inflection [as] undeniable fact”; also see Note 24).

6.2. Cycles of complexity in diachrony

What these cycles suggest is a picture in which absence of (overt) distinctions in one domain of grammar (e.g., morphology with its affixes and other word-internal processes) can well be compensated by distinctions in another domain (e.g., syntax with its periphrases, word-order distinctions, selectional requirements, etc.).

For example, loss of overt inflectional morphology has often been correlated with increased rigidity in word-order. Vis-à-vis case morphology and word-order, it has long been noted, starting perhaps with Meillet 1912 [1926: 147–148], that languages seem to balance out overt case marking with grammatical marking in the syntax (via, e.g., pre-/postposition and word order). The exact correlation is hard to pin down, but the general, if overly simplistic, impression is that richness of Case morphology tends to be proportional with the scrambling of noun phrases away from their base positions (to wit: the history of English and of Romance). If so, measuring only overt morphological distinctions at the expense of abstract syntactic information leads to an incomplete and misleading metric. Similar caveats apply to the interface between any two (abstract) linguistic modules and the trade-offs therein in terms of grammatical information encoding. It seems to me that computing global complexity requires a theory of how grammatical information is encoded within and across the various linguistic modules (cf. Note 39).

45. In this respect, WSG: 144 misinterprets my views on the relationship between inflectional morphology and the syntactic differences between a creole and its sources: “DeGraff […] (1997, 1999b) argues that the differences between a creole grammar and that of its source languages are due to certain syntactic results following from loss of inflection during second language acquisition (such as lack of verb movement to I), with subsidiary results due to the filtering out of low-frequency features, and the ellipsis of certain functional categories, with the qualification that the effect of the latter two was no more marked than that upon other languages with heavy contact in their histories (DeGraff 2000).” Although I myself have not (yet) explored the fascinating and probably enlightening history of Yiddish, what I did argue is that, like in other cases of language creation, certain (morpho)syntactic properties seem correlated to properties of (overt) inflectional morphology. This is a rather commonplace, if difficult to formalize, guiding intuition in much theoretical and historical work (see, e.g., van Kemenade & Vincent (eds.) 1997 and van Kemenade (ed.) 1999 for two recent anthologies on this topic). For example, the abstract linking of (degrees of) verbal inflection and (degrees of) verb raising, although somewhat controversial, is standard fare in current research – see the references in DeGraff (1997, 1999b: 501–502, 518–521, 2001a, forthcoming), Roberts (1999). Chomsky (1995: 6) attributes to Jespersen the hunch that crosslinguistic syntactic variations can in large part be reduced to variations in morphology. In this particular respect, certain (erstwhile) contact languages such as Capeverdean Creole, Chinook Jargon, Haitian Creole, and Saramaccan (where certain kinds of overt inflectional
Hawkins (1986) is one such attempt toward formalizing the trade-off between different sources of “complexity”. Hawkins argues that, along the dimension of form–meaning opacity, English is more “complex” than German because of increased case syncretism in English. Independently of the merits of this proposal, the point here is that it can be logically argued that reduction of overt morphological distinctions (i.e., reduced bit-complexity in morphology) increases some other kind of complexity, namely form–meaning opacity which in turn increases semantic ambiguity. Hawkins places English and German morphology at opposite ends of the semantic-opacity continuum: German morphology allows for “a ‘tighter fit’ between surface form and semantic representation” (1986: 122). Then again, English’s increased “form–meaning opacity” due to case syncretism is somewhat compensated by its word-order which is more rigid than in German. Hawkins (1986: 216) speculates that:

(33) [T]here is an inherent tension [...] between the rules generating linguistic forms on the one hand, and those assigning meanings to these forms on the other. Simplicity for the one means complexity for the
Commentary on McWhorter: Michel DeGraff

other, and vice-versa, and the successful resolution of this tension defines a continuum along which the actually attested languages occur, from the least to the most complex set of formal rules, and correspondingly from the most to the least complex set of mappings between form and meaning. Over time languages can drift from one part of this continuum to another, in response to independent (e.g., phonological) changes which have clear consequences for the mapping between form and meaning.

No matter the fate of Hawkins’ arguments, his speculations shed further doubts on the theoretical validity of the simplistic sort of (bit-)complexity metric advocated in (30): it can indeed be argued on empirical and theoretical grounds that, at least in some cases, ABSENCE of “overt distinctions” (compare, e.g., case affixes in English vs. German) can itself be a distinction that should enter into some global complexity metric. This is reminiscent of Meillet’s (1912) argument on how rigid (grammaticalized) word-order emerges to replace case morphology. Here again, we see why bit-complexity is of little use in the absence of a theory of grammar that motivates the items to be counted and relates them to larger linguistic concerns – with respect to mental representations, language acquisition, language processing, language change, etc. 

6.3. Historical linguists’ “endless cycles” vs. pro-prototype creolists’ “ground zero”

Keeping these opposing trends in the picture, it must then be concluded that, in creole genesis as in other cases of language contact, distinct grammatical domains in the languages in contact will belong to distinct points on, e.g., their respective morphology–syntax cycles (à la Meillet/Hodge/Givón) and semantic-opacity cycles (à la Hawkins). With this in mind, one is forced to conclude that creolization etc. (on a par with language change via language contact) will start, not “essentially from ground zero” (contra WSG), but from a contingent (i.e., sociohistorically determined) array of non-“ground zero” structural termini ad quo. The latter fit distinct points on diachronic cycles (e.g., with respect to

46. If “morphology” and “syntax” all belong to syntax – that is, if the structure-building operation “Merge” applies at both the word level and the phrase level merge, as in, e.g., Marantz, Miyashita, and O’Neil (2000: 3–4), – then it may not make much of a difference in terms of computational complexity whether Merge takes place within words or across words; see Haspelmath (2000) and DeGraff (2001b: 72–73: Note 19) for further comments. In other words, periphrasis may be cognitively as “cheap” or “expensive” as lexical storage or derivation-root combination (see Chafe 1970: 36–37). But these are all empirical questions that deserve further study.
On the origin of creoles

their respective analysis–synthesis and semantic-opacity clines). These non-
"ground zero" structural termini ad quo determine the ecology of the Primary Linguistic Data available in the formation of creole idiolects. In other words, various components of the (pre-)creole grammar will be extrapolated (via language transfer cum restructuration, innovation, grammaticalization, simplification, etc.) from patterns that are located across the analysis–synthesis and morphological-syncretism continua, as a synchronic reflection of prior cycles in the respective diachrony of the “old” languages in contact. In a similar manner, the diachronic termini ad quem will unavoidably be scattered across the relevant cycles in various grammatical domains. Ultimately, the trace of this scattering (i.e., the resulting I-language grammar(s) whose structures are necessarily bounded by UG) will depend both on the Primary Linguistic Data and the sociodemographic conditions – all of which result from a very complex set of historical contingent factors.

So, here too, the creolist-cum-complexity-theorist must employ much caution in dealing with the competing pressures on language complexity that arise from the sociohistorical cum typological specifics of the contact situation. As Hymes (1971b: 70) reminds us, we need “to recognize pidginization as a complex process, comprising the concurrence of several component process”. At this stage, we are far away from the Schleicherian “older = more complex” perspective on language evolution. We are also quite far away from the ab ovo genesis of “born again” languages from “ground zero” complexity via “a radical reduction of [the] source languages into makeshift jargon” (cf. WSG: 149).

The empirically and theoretically responsible scenario is much more complex, thus much more fascinating, even if it removes creole languages from the category of contemporary fossils of Language evolution, and even if it deprives creolists of a most simplistic account for some of the most complex cases of diachronic development.

6.4. The empirical test: To be “old” AND “born again”?

How many non-creole languages could pass the WSG-type (neo-)Schleicherian structural litmus test for old-cum-complex languages?

WSG “carefully” picks its benchmarks for “old language” complexity, not from, e.g., Romance and Germanic (the well-documented source languages – the terminus a quo – of its Prototypical Creoles), but from Caucasian, Tibeto-Burman, and Polynesian languages (which have nothing to do with the Prototypical Creoles being surveyed). The complexity “control” cases – Tsez, Lahu, and Maori – would strike many linguists, not just creolists and creole speakers, as “fearsomely elaborated” or “extraordinarily complex by any linguistic standard” (cf. WSG: Section 3.2; also see Trudgill 1989: 237, who
can imagine “mainstream Europeans and North Americans find[ing Caucasian languages] ‘exorbitant’ or ‘incredible’”). Plus these three “control” languages are all “esoteric” languages, languages spoken in small communities in relatively low-contact situations where, for social-network reasons, speakers are most likely to maintain (and perhaps even promote) features that may APPEAR complex and/or linguistically unusual (cf. Hymes 1971b: 73, Trudgill 1989: 236). Trudgill specifically cites the languages of the Caucasus, to which Tsez belongs, as prime examples of “low contact” languages. According to Grimes (ed.) (1996), Tsez has approximately 7,000 speakers, Lahu (Shi) 600,000, and Maori 110,000 (these are approximate totals across the relevant dialects). This sociolinguistic factor alone is a delimiting confound in these complexity experiments comparing Saramaccan with Tsez, Lahu, and Maori. What should be asked is whether a hypothetical creole derived from contact among, say, Caucasian languages of the Tsez-type (or Tibeto-Burman languages of the Lahu-type or Polynesian languages of the Maori-type) would end up looking like Saramaccan. (See Notes 30 and 37 above for related methodological remarks.)

As it turns out, the neo-Schleicherian “deck-stacking” methodology, to the extent that I understand how it can be applied without bias, unsurprisingly ranks the complexity of English alongside that of Saramaccan. Yet Saramaccan, but not English, is usually regarded as a most “radical” creole. In fact, WSG considers English to be an “old” language – much “older” than “born again” Saramaccan.

This is how we can try and check whether English is “born again” or “old” by applying the criteria and methodology exemplified in WSG:

   English, like a language with “pidgin ancestry”, lacks many of the traits that are claimed to be “incidental to basic communication”. Indeed English lacks ergativity, grammaticalized evidential marking, inalienable possessive marking, switch-reference marking, inverse marking, obviative marking, noun class or grammatical gender marking, lexically contrastive or morphosyntactic tone, etc. (cf. (18)).

   English, like (the radical creole) Saramaccan, is also much less “complex” than Tsez. Indeed, given the criteria in WSG, “Tsez’s grammar is indisputably a more complex one” than English’s. In fact, the complexity of Tsez, as measured in WSG, seems much higher than that of French, Spanish, Chinese, and many other “old” languages. Thus, shouldn’t we also ask whether English, French, Chinese, etc., like Saramaccan, “have not existed for long enough a time for there to have arisen the sheer weight and depth of such features as in older languages like Tsez” (cf. WSG: Section 3.2)? Indeed, English and many other non-creole languages do resemble Saramaccan in lacking the following Tsez properties (cf. WSG: Sections 3.1–3.2): “pharyngealized uvulars”; “stops and affricates [with] phonemic ejective alternants”; “[noun] classes determined by the final segment of the stem”; “nouns [with] alternate forms for differ-
On the origin of creoles

ent suffixes”; “markers for evaluative names”; in-situ adjunct wh-phrases vs. fronting of argument wh-phrases; “grammaticalized evidential markers”; “a lative marker”; “overt delineation of experiencer verbs”; etc.

English and many other non-creole languages, like (the radical creole) Saramaccan, is also much less “complex” than Maori, given the observations in WSG: Sections 3.2 and 4.3. Indeed, English and many other non-creole languages resemble Saramaccan in lacking the following Maori properties (taken from WSG: Section 4.3): “several different interrogative constructions according to the grammatical status of the constituent questioned” (furthermore one can also say of many Germanic and Romance languages, on a par with Saramaccan, that they do not exhibit “interrogation strategies [that] vary to anything approaching this extent according to grammatical relation”); “subtle possessive distinction […] reminiscent of an alienable/inalienable distinction, but contingent basically upon dominance of possessor over possessee”; “subjects of intransitive verbs [that] are marked as possessives, while the verb itself is nominalized”.

English, like the radical creole Saramaccan, is also much less “complex” than Lahu, given the observations in WSG: Sections 3.2 and 4.1–4.4. Indeed, English (along with many other non-creole languages) resembles Saramaccan in lacking the following Lahu properties (taken from WSG: Section 4.1): “bilabial, alveolar and velar, with both aspirated and unaspirated phonemic alternants in all five places of articulation”; “seven lexically contrastive tones”; “agentive markers distinguished by sex”; “[agentive markers] distinguished by sex to denote ownership or mastership, which can also be used to nominalize clauses”; “numeral classifier, for people, animals, shapes, and more general purposes”; nominal reduplication “as a classifier”; “an accusative marker […] used with patients only to encode certain shades of emphasis”; “modal and pragmatic particles central to basic expression which are conventionalized into highly particular and idiosyncratic subdivisions of semantic and pragmatic space”; “verbs [that] occur only in concatenation with a specific other verb to convey completion”; “SOV”; “elaborate tonal system”.

In a coda to the Lahu-vs.-Saramaccan discussion, it is noted that “English is also less complex than Lahu in all but one of the features cited (derivation)” (WSG: 149). Yet, in the same paragraph, an ad hoc list of English properties is pulled out in order to argue that “English is more complex, according to our metric, than Lahu in a great many aspects”. This is yet another vacuous argument since similar ad hoc lists of non-Lahu “incidental” properties could be produced for any language. Indeed it remains possible in principle to produce a similar ad hoc list of Saramaccan features whereby Saramaccan too would look “more complex, according to [the WSG] metric, than Lahu in a great many aspects”. In fact, given the structure of UG (see Section 4.5 above) and given the contingent aspect of parameter-setting and of the lexicon, it is always possible
(in principle) for any pair of languages $X$ and $Y$ to produce an ad hoc list of grammatical distinctions that exist in $X$ but not in $Y$. Such lists exist for any pair of languages no matter whether $X$ and/or $Y$ are creole or non-creole. The handpicking of scattered “incidental” features does not constitute explanation or result. The contents of such lists are truly “incidental”, that is, language- and construction-particular in a relatively superficial way (i.e., with no explanatory power).

As a (facile) exercise, I will now overtly rig a neo-Schleicherian experiment and produce a list of “incidental” Haitian Creole features of the Lahu/Maori/Tsez type above whereby HC “is more complex, according to [the WSG] metric, than [certain varieties of French and English] in a great many aspects”:

\[(34)\]

a. Nasalized and non-nasalized vowels that, in certain contexts, occur in free variation, even when followed by a nasal consonant; nasal vowels can be analyzed as “a combination [of] oral vowel [plus] floating nasal consonant” (Cadely forthcoming).

b. “[A] regressive nasalization rule that [in certain contexts] applies within stems or underived morphemes and a progressive nasalization rule which [in certain contexts] takes place across a morpheme boundary” (Cadely forthcoming).

c. A set of personal pronouns with morphophonologically and syntactically conditioned clitic variants; the latter can behave as either proclitics or enclitics, according to complex morphophonological and syntactic conditions (DeGraff 1992a; Cadely 1994, 1995, 1997; Hilton 2000).


---

47. Cadely’s (1988) analysis of the HC syllable argues that HC has a much more extensive set of diphthongs than French: HC with its alleged “simplest” grammar has about twice as many diphthongs as its “old” French ancestor.
f. Demonstrative, definite, and plural-marking markers that are head-final in a language that’s otherwise robustly head-initial (Joseph 1988, DeGraff 1992a).

g. Head-final definite articles alongside head-initial indefinite articles (Joseph 1988, DeGraff 1992a).

h. Two predicate-clefting strategies for focus purposes: (i) one with predicate-head doubling (with verbs, adjectives, and certain bare nominals); (ii) the other with a non-verbal pro-predicate left in the in-situ position of the moved predicate XP (the latter must be [−V]). (DeGraff 1992a, 1995, 1998; also see Appendix B for further discussion with a bit of data.)

i. Certain bare nominals can undergo predicate clefting with either the predicate-doubling strategy ((i) above) or with the (non-doubling) pro-predicate strategy ((ii) above). The predicate-doubling strategy allows both a permanent/essential (individual-level) interpretation and a temporary/provisional (stage-level) interpretation of the clefted predicate while the pro-predicate strategy forces the clefted predicate to be interpreted as individual-level (Damoiseau & Saint-Louis 1986).

j. Predicate clefting-cum-doubling for the formation of various adjunct clauses (Lefebvre & Ritter 1989).

k. Three strategies for the formation of causal clauses, each of which uses some distinct CP-related position – one of these strategies also employs the predicate clefting-cum-doubling pattern (Lefebvre & Ritter 1991).


m. A rule of apocope that applies to a subset of verbs with short and long variants; this rule is morphophonologically and syntactically conditioned (DeGraff 2001b: 74–75).

The above ad-hoc “incidental” list of HC non-English properties would thus lead to the conclusion that HC is more complex than English, if one assumes the style of argumentation in WSG. In addition, HC also manifests some (quasi) English-like features that constitute “overspecification [that] goes beyond the needs of a human grammar” (cf. WSG: Section 4.2). Such (quasi) English-like “ornamental”/“incidental” properties include:

(35) a. “[T]he overt and categorical marking of definiteness” on HC nouns (Joseph 1988).

b. An “overt marker of definiteness and indefiniteness, whose occurrence is determined by referentiality as well”. In HC, we
say: Mwen wè yon fim yè swa ‘I saw a film last night’ where the marker yon marks the NP yon fim on a par with English a in a film) “despite [HC fim] being presupposed to the speaker, because it is not yet known to the hearer”. (Here, HC, like English, “goes beyond the needs of a human grammar as far as encoding definiteness is concerned”, which makes HC look older than “the thousands of grammars without such overt marking […] (e.g., Chinese and Russian)”; cf. WSG: 161.)

c. A “subtle distinction [that is] maintained” in HC between the ap future and the pral future. Note that, like English will and going to, HC ap and pral are other “feature[s] which give fine-grained and grammaticalized manifestation to a distinction lacking in Lahu as well as a great many other languages”. So in this particular respect, HC – this “born again” language – is at the very least as “complex” than two “old” languages: English and the “mega-complex” Lahu.

d. An intricate of set of (semi-)auxiliaries that can be used as (quasi) tense/mood/aspect markers (see e.g. HC pral above and its English analogue going to); also see HC konn(en) (habituality marker), fin(i) (completive marker), dwe (deontic/epistemic marker), etc.

One could argue that all the complex features of HC above – which, as in English, seem “incidental to basic communication” – have arisen “due to contact over the centuries with French” (cf. WSG: 143). But this would not work for a simple reason. Many of the features noted above have no direct counterpart in French. Furthermore, the French-like HC features (e.g., the auxiliaries ap and pral) seem to have been part of the language from very early on, resulting from the grammaticalization of French periphrastic constructions with après and après aller, respectively. Furthermore cognates of these preverbal tense/mood/aspect markers are regularly found in French varieties that are not labelled “creole” (cf. Appendix B).48

Such arbitrary and superficial comparison whereby “born again” languages may appear more complex than “old” languages given some ad hoc list of language-specific features can be extended ad libitum. Indeed, for any choice

of \( n \) (an integer), \( X \) (a so-called “born again” language) and \( Y \) (a so-called “old” language), \( X \) can be the source of an arbitrary list of \( n \) features that are absent in \( Y \). Take, say, dual marking, whose presence in any language is considered (in WSG: Section 5.3) to increase complexity/markedness in the Greenbergian sense. Dual marking, which is absent in many “old” languages, does exist in a number of “born again” languages. In Cape York Creole, as described in Crowley & Rigsby (1979), the pronominal system offers both dual and plural in the non-singular forms (for 2nd and 3rd person), plus there is a grammaticalized distinction between inclusive vs. exclusive pronouns in the 1st person. If dual is dispensable from a communicative viewpoint (which it must be given its absence in many “old” languages), then Cape York Creole is surely more complex than, say, English in that respect. A similar example is provided by Chaudenson (1994: 50) who argues that dual marking in French-lexifier creoles in the Indian Ocean (e.g., in Réunion) makes these creoles more complex than their lexifier vis-à-vis the grammaticalization of number marking. Dual marking is also found in Taimyr Peninsula Russian-based Pidgin (Wurm 1996: 83). Such “incidental”/“ornamental” systems of pronominal reference and number marking are not found in “old” languages such as many English and French varieties. Here too, we have “born again” languages with arbitrary distinctions that are not found in “old” languages (also see Section 4 for further examples of pidgin/creole features that are “incidental to basic communications”).

It is worth repeating that, given the partially contingent nature of parameter-setting and lexicon formation, any number \( n \) of arbitrary (and so called “incidental”/“ornamental”) distinctions can, in principle, be found in ANY language, whether “old” or “born again”. Indeed, the list of “incidental” features provided by creolists-cum-complexity-theorists in their creole-vs.-non-creole comparisons (see, e.g., (18)) is truly incidental.

Lastly, one could argue that the reason why English’s bit-complexity looks low when compared to Tsez, Maori, and Lahu is that English is, after all, a “creole” as can be perhaps adduced from its history. Such argument would go on to claim that English’s history of contact is the reason why so many creoles (such as HC and Cape York Creole, say) have features that are at least as “complex” as their English counterparts. Well, I suspect that many other languages besides English would also rate low given the complexity metric in WSG and its exotic benchmarks and “incidental features” list. In any case, the English-as-creole argument would drastically weaken the theoretical bite that WSG would like to assign to the term “creole”, especially in the absence of an unambiguous STRUCTURAL definition for “creole”. 49 In any case, English

49. Calling Middle English (ME) a “creole” – as, e.g., in Bailey & Maroldt (1977) – raises more questions (contradictions, really) than it resolves. Indeed, English (even as another “cre-
IS claimed as one of these multi-millenarian languages that are generally more complex than young languages (see, e.g., WSG: Section 4.2).

In fact, it can be argued that language contact in the history of English did contribute to reducing bit-complexity in restricted domains, as in, e.g., inflectional morphology. Meillet, Weinreich, Trudgill, and Chambers, among many others (see Notes 5 and 24) have argued that language learning in contact situations induces “erosion” of morphology and/or regularization of overt morphological distinctions. And it has been argued many times before that language acquisition itself can induce various degrees of “simplification” in the concomitant new idiolects. Thus, any language (whether “old” or “born again”) can undergo certain sorts of “simplification”, independently of its creole status or origins.

What the above remarks suggest is that the complexity-reducing effects of “pidginization” in certain domains (e.g., in overt inflectional morphology) are quite widespread – even when there are no recognizable pidgins(-to-be) in sight (recall that (I-)pidginization as an individual-level process is in principle distinct from the creation of stable (E-)pidgins – the latter crystallize through the focusing of norms and other group-level sociolinguistic processes). This, of course, is not a novel observation. Schleicher himself considered that language contact (e.g., in the history of English) was a degenerative factor while Humboldt took the “undeniable fact” of inflectional erosion as a sign of intellectual maturity (see Note 5)! It has been commented over and over again in the sociolinguistics and historical linguistics literature that language contact, across space and time, often entails structural simplification in various domains (see Note 24). As Hymes (1971b: 73) wrote, “simplification may prove to be, not an isolated phenomenon, but one pole of a continuum applicable to all languages” (see Note 50). Others have proposed that one key factor driving structural simplification in various domains is vehicularization, i.e., the socio-historical process by which certain languages become lingua francas, as in the case of English; see Mufwene (2000b, 2001). At the individual level, the sim-

ole”) still resembles a “genetic” Germanic language according to standard Stammbaumtheorie methodology (see, e.g., Campbell 1999 for an introduction). Furthermore, the sociohistorical specifics of English diachrony do NOT resemble those of (“radical”) creole genesis as in the emergence of HC. For example, it is widely accepted that (varieties of) English remained the native idiolects for the majority of speakers in England throughout the Norman Conquest. Thus, I somewhat agree with Domingue (1977: 90) that: “[I]f ME […] can be described as another creole, there must be many more creoles around than we think. On the contrary, if ME does not fit in the creole category, it will be because stringent characteristics of that category will have been established, a task yet to be done”. Vis-à-vis “stringent [structural] characteristics of [the creole] category”, the “task yet to be done” remains elusive still and seems to me impossible – thus, this paper.

50. See, e.g., Henry & Tangney (1999), Newport (1999), Roberts (1999), and references therein.
plification processes often noted in language contact and vehicularization must be ultimately rooted in the cognitive capacities and constraints that underlie second language acquisition (see, e.g., DeGraff 1996b, 1999b, c, d for discussion and references).

7. Envoi: The descent of the creole speaker

The various lists elaborated by (neo-)Schleicherian linguists in the past three centuries in order to isolate “the world’s simplest grammars” still seem irreducible to any fundamental principle(s) of linguistic theory. Thus far, these lists cannot count as scientific explanation for any robust set of linguistic phenomena. At best, they identify scattered “historical accidents” in scattered domains of grammar in scattered samples of languages.

In any UG-based framework (along the lines sketched in Section 4.5), language-specific “incidental overspecifications” (such as pro-predicate morphemes, dummy verbs, dual marking, ergativity, grammaticalized evidential marking, pharyngealized uvulars, clicks, etc.) ARE “historical accidents” – the sociohistorically contingent choices made by particular idiolects cum societal conventions within the boundaries set by the biological necessities of our faculté de langage. Given the rich linguistic ecology and the (socio)linguistics of language contact, it is no surprise that “historical accidents” of various sorts also happen in creole genesis, as documented above.

One key question facing modern linguists is: What is the structure of UG such that, across the species, language learners faced with incidental and relatively shallow Primary Linguistic Data (PLD) unfailingly (re-)create idiolects with all sorts of abstract complex properties that are not evident from the PLD (cf. Chomsky’s 1986 “Plato’s Problem”).

The central assumption here is that every idiolect is somewhat created anew at every instance of acquisition. A related, but distinct, assumption is also found in the grammaticalization camp. There too, notions like “old” and “new” languages seem to make little theoretical sense: “Students of grammaticalization realize that worrying about where one grammar ends and the next grammar begins is a totally meaningless and futile pursuit. For the ‘new grammar is constantly being created on top of the wiling and yielding ruins of the old’ […]” (Matisoff 1991: 447)

From the Cartesian-Uniformitarian perspective espoused here, it can be reasonably argued that ALL (I-)languages evolve via an initial “break in transmission”: grammars are not inherited, but (re-)created (Paul 1890; Meillet 1929; Halle 1962; Chomsky 1981, 1986, 1995; Lightfoot 1999; etc.). In each of its individual instantiations, language acquisition sensu stricto is not language transmission, but UG-guided language (re-)creation with contingent, sparse, and heterogeneous PLD drawn from idiolects (i.e., speakers) in contact (Chom-
The exact and particular nature of this “contact” is sociohistorically determined and so are the tempo, amplitude, and group-level effects of the individual “breaks”. Yet, whether in “creole genesis” or through “language change”, the PLD always underdetermine the attained grammar: there always exist structural “breaks” (however subtle) between “old” and “new” idiolects. Thus, the ineluctability of language change/creation. Whatever general tendencies may exist across instances of creole genesis partake of the same mental processes that underlie all other cases of (I-)language creation.

As of “(I-)pidginization”, to the extent that it reduces to (aspects of) second-language learning (L2A) in adulthood, the “pidgin-to-creole life-cycle” will have congeners far and wide, in all cases of language contact where non-native utterances contribute to the PLD that are used in the formation of native idiolects (L1A). In a Cartesian-Uniformitarian perspective, the “pidgin-to-creole cycle” is naturally mirrored by an L2A–L1A cycle – or L2A–L1A “cascade relationship” in the terminology of DeGraff (1999b: 497, 504). In this vein, the discrepancies between “old” and “new” idiolects may seem more dramatic in the creole cases than in the non-creole cases, but it can reasonably be argued that the difference (if any) is a matter of degree, rate of spread, and/or subjective perception, not of quality.

If so, the sui generis structural category “world’s simplest grammars” with “born again” genealogical status is only “chimera and reverie” and, worse yet, “linguistic monstrosity”, as hinted at by Foucault in (4). Thus, the fallacy of dualist Neo-Darwinian scenarios for the origin of creoles. Given UG and given the sociolinguistics and ubiquity of language contact, there is not, and there could not be, a constant and exclusive set of creole structures that are FUNDAMENTALLY special, across time and across space, independently of the specific linguistic ecology (see references in Note 12). I thus agree with, inter alios, Mufwene and Muysken, as per the following quotations (compare with (10)):

(36) a. The very notion of a “creole” language from the linguistic point of view tends to disappear if one looks closely; what we have is just a language. (Muysken 1988: 300)

  b. [N]o language-development processes were involved [in creolization] that were unique to [creole languages,] just the same ones usually assumed in historical linguistics except for the emphasis on language contact. (Mufwene 1996: 107)

What we have everywhere seems to be simple evolution of languages from one state to another in different ecological conditions. (Mufwene 1998: 324)

In my native Haiti and elsewhere in the Caribbean and wherever else we find creole speakers, arbitrary (pseudo-)linguistic measures are still employed
to “classify humanity” and de-humanize (monolingual) creole speakers (see references in Note 9). This “linguistic apartheid” is undermined (theoretically, at least) by observations like those in (36); also see the quote in (10) and the references in Note 12. As a creolophone creolist, I find that there is grandeur in this (Cartesian-Uniformitarian) view of Language: Creoles, on a par with ALL other languages (irrespective of genealogy), are reflections of our (species-uniform and species-specific) human biology, which is among the “most beautiful and most wonderful [forms that] have been, and are being, evolved” (cf. Darwin 1859 [1979: 459–460], Chomsky 1995: 1–4, 2001: 2). 51

Appendix A: Were Haitian affixes “borrowed late”?  
As we saw throughout the main text (see, e.g., Sections 2.1, 3.3, 6.1, 6.4), there is robust, inescapable evidence that HC – the creole formerly known as “most creole of creoles”, “‘pure’ [Creole Prototype] case”, and “basilectal creole” (McWhorter 1998: 809, 812; 2000b: 206) – is far removed from a structurally “simplest” Creole Prototype with ancestry in a structurally “simplest” affixless pidgin. For instance, HC affixes (with cognates in French affixes) straightforwardly disconfirm the catastrophic pidgin-to-creole scenario whereby HC affixes would have emerged via grammaticalization of erstwhile free morphemes. How would a neo-Schleicherian creolist reconcile the postulation of a pidgin-to-creole catastrophic cycle (and its radical morphological bottleneck) with the well-documented French-based affixes of HC qua “most creole of creoles”, “basilectal creole”, etc.?

51. Reading this last sentence, some linguists may ask, with due caution, whether its contents jeopardizes scientific objectivity. Since this paper is ultimately about the mismeasure of creole speakers, my response is appropriately taken from Gould’s The Mismeasure of Man (1996: 36–37):

Scholars are often wary of citing […] commitments [to social justice], for, in the stereotype, an ice-cold impartiality acts as the sine qua non of proper and dispassionate objectivity. I regard this argument as one of the most fallacious, even harmful, claims commonly made in my profession. Impartiality (even if desirable) is unattainable by human beings with inevitable backgrounds, needs, beliefs, and desires. It is dangerous for a scholar even to imagine that he might attain complete neutrality, for then one stops being vigilant about personal preferences and their influences – and then one truly falls victim to the dictates of prejudice. Objectivity must be operationally defined as fair treatment of data, not absence of preference. Moreover, one needs to understand and acknowledge inevitable preferences in order to know their influence – so that fair treatment of data and arguments can be attained! No conceit could be worse than a belief in one’s own intrinsic objectivity, no prescription more suited to the exposure of fools. […] The best form of objectivity lies in explicitly identifying preferences so that their influence can be recognized and countermanded.

Gould then proceeds to debunk a number of rankings of human cognition across the “races”, in, e.g., the practice of 19th-century craniometry and 20th-century psychometrics.
McWhorter (2000a: 107) claims that “[HC] fits quite neatly into the Creole Prototype model as a case [ . . . ] in which contact with the lexifier over the centuries pulled the creole away from the Prototype to which it honed at its genesis”. This claim is somewhat echoed in WSG (Section 3.2) where HC, “due to contact over the centuries with French”, is considered to “have borrowed many French lexicalized derivation–root combinations and thus does not exemplify the Creole Prototype in the purest possible form”. Such a scenario is ahistorical given established socio-demographic facts of Haitian history as sketched in 2.1 above (also see Note 15 and Appendix B).

On the empirical front, McWhorter (2000a: 106) mistakenly relies on what he calls “[Goyette’s (2000) demonstration] through painstaking historical analysis that the derivation markers in modern [Haitian Creole] [ . . . ] cannot have been incorporated into the creole at its birth, and in fact were borrowed from French in later periods.”

Not only is Goyette’s scenario as ahistorical as McWhorter’s (see the sociohistorical sketch in Section 2.1), Goyette’s empirical generalizations about (the diachrony of) HC and French are flawed, and so are his deductions.

Take, for example, Goyette’s discussion of the timecourse of the HC prefix re- /rekile/ (cf. French re-). From Fattier (1998), he quotes the lone HC form ekile /rekile/ ‘to move back’ (a variant of HC rekile /rekile/). Goyette claims ekile as a cognate of some (dialectal) metathesized French variant erculer (cf. Standard French reculer /rekyle/). In Goyette’s scenario, erculer’s counterpart in HC is ekile and not erkile since most HC dialects forbid syllable-final /l/. Why is the single HC form ekile, taken from Fattier’s 6-volume dissertation (1998), of such significance? Goyette’s argument is based on the premise that “[i]n 17th-century French, [the prefix re-] was consistently metathesized” while “where /re/ is a prefix [in HC] such metathesis is wholly unknown”. Thus, Goyette argues, the first-syllable /e/ in HC ekile is the un-productive remnant of the metathesized French prefix er-; crucially the productive prefix re- in contemporary HC was not part of proto-HC morphology. Goyette’s conclusion: re- in contemporary HC was borrowed late, not inherited early.

To the extent that I can understand Goyette’s argument, it seems to contain at least one empirical flaw, even if one abstracts away from Goyette’s overly simplistic claims about the phonology, distribution, and diachrony of the er-re-alternation across French dialects (the complex diachrony of French dialects sheds doubt on the categorical claim that “[i]n 17th-century French, [the prefix re-] was consistently metathesized”). Given (inter alia) that ekile and rekile have the same initial vowel /e/, Fattier (1998) reasonably takes ekile as a case of apheresis, not metathesis. Fattier also documents a third variant, which she writes ’ekile, where the first segment (the superscript /l/) is a phonetically weakened variant of the /l/ in rekile (see Fattier 1998: Volume 1, 230–232, Volume 2, 449 for details). Furthermore, throughout her 6-volume thesis, Fattier documents robust cases of apheresis in a variety of environments. Similar apheresis is documented in Ducourjoly’s (1802) creole-teaching manual. The latter also documents robust affixation in early HC, virtually all of it derived from French affixes (including apparently un-metathesized re-). Thus vanishes Goyette’s single data point in arguing for the late borrowing of HC re-.

The other HC suffix discussed by Goyette (2000) is agentive -è /as in mantè /mãtÈ/. For Goyette, “[HC -è] is the normal reflex of final French -eur [pronounced /œr/] [ . . . ]
On the origin of creoles

In the seventeenth-century, [this French] suffix lacked a final /r/ (...) (menteux instead of menteur ‘liar’). (...) We should expect the modern [HC] form of the suffix to be [/e/ since French /ø/ is mapped into HC /e/]" (emphasis added). Here too, Goyette’s scenario is overly simplistic and empirically inaccurate. Although exact dates are uncertain, it seems that the reduction and dropping of final /r/ in French -eur (as part of a larger pattern of final-consonant reduction) was most robust in the Middle French period up until the 16th century. The opposing trend (pronunciation of final /r/) gained strength in the second half of the 17th century, becoming re-established in many 18th-century dialects and their descendants up to the present.

So could it be true that “[i]n the seventeenth-century, [the -eur] suffix lacked a final /r/?” The crucial observation here is that, even at its apex, final-/r/ reduction in French was not categorical across all dialects; see, e.g., Fouché (1966: 669) and Morin (1986: 173–175). Instead many dialects manifested (phonologically, morphologically, semantically, and sociolinguistically conditioned) variation between dropping and retention of final /r/. In some cases, final /œ/ would enter into external sandhi phenomena, being pronounced before a vowel or a pause. In other cases, final /r/ was pronounced for emphasis or semantic nuance or for ill-understood sociolinguistic reasons (e.g., hypercorrection).

A very telling case is the 1547 example Ajoustes si tu veux les Perfumeux, les Balleurs ‘Add if you want the Perfumeux and the Balleurs’ where -eur and -eur alternate in the very same sentence (Brunot 1906: 290); also see Brunot (1913: 211–212, 1924: 671) for other instances of variation with semantic and/or sociolinguistic nuances and for further remarks on the diachronic course of final-/r/ reduction.

In addition to the variable rule of final-/r/ reduction, there are other factors affecting -eur in French diachrony. Discussing the passage from /œ/ to /œ/ or /ø/ through Old and Middle French (cf. flor > fleur, dolor > douleur, etc.), Nyrop (1899: 163) mentions 15th- and 16th-century dialects where œ and ø do not exist and where sœur and sur enter into “near rhymes that are said to be either ‘Provencal’ or ‘Gascon’, either ‘Normand’ or ‘from Chartres’ “. Nyrop adds that such near rhymes are widespread in the 15th and 16th centuries. Also relevant here is the following synchronic alternation: chœur/choral, docteur/doctoral, fleur/floral, mœurs/moral, pasteur/pastoral, etc. (I thank Dominique Fattier for bringing this alternation to my attention.)

Pending further research, one can reasonably speculate that the above alternations in French diachrony and synchrony are related to another case of variation in HC’s synchronic morphophonology. As Fattier (1998) repeatedly notes (also see Freeman & Laguère’s 1998 dictionary), è/ò alternation in HC word-final syllables is quite widespread, although not generalized: flatè/ò ‘flatter’ (cf. flat ‘to flatter’), gadègadò ‘watchman’ (cf. gade ‘to watch’), vòlè/vòlò ‘thief’ (cf. vôle ‘to steal’), etc. (also see fyèl/fyòl ‘godchild’, lè/lò ‘time’, sè/sò ‘sister’, etc.); see Fattier (1998: Volume 1, 132, 269, Volume 2, 501, 740, 755). This è/ò alternation does not seem to have been borrowed late from French as spoken in Haiti: as far as I can tell, contemporary Haitian French manifests no analogous alternation or remnants thereof; see Pomplius (1961) for a sketch of Haitian French. (Although Goyette mentions HC mantè/manòt from Fattier’s thesis, he fails to mention the variant mantò/manòt and the corresponding alternation vis-à-vis the -e suffix.)

On the conceptual front, it must be noted that, unlike, say, Québécois French (per
Commentary on McWhorter: Michel DeGraff

Goyette), many contemporary (non-creole) French dialects (e.g., Standard French and Haitian French), on a par with HC, do not productively use er- for re- and neither do they productively drop their final /r/ in -eur /œr/ (contrast with Québec French éruler and ménèus as cited by Goyette). Yet the absence of such alternations surely does not constitute evidence that the contemporary affixes re- and -eur in these French varieties sans er- and sans final /r/ dropping “were borrowed […] in later periods”. Where would these “later period” French speakers “borrow” the (UN-metathesized) re- prefix and the l-/œr/ suffix (WITH phonetic realization of final /r/) if ALL prior dialects consistently lacked such affixes?

Now, let’s assume (purely for the sake of argument) that (some? most?) French dialects in 17th-century Haiti did metathesize re- into er- and did drop final /l/ in, e.g., -eur. Then, whatever (socio)linguistic processes led to contemporary absence of er- and the contemporary retention of final /l/ in (the “later periods” of) Standard French, Haitian French, etc., could, in principle, also account for the corresponding facts in HC without evoking an (unattested) affixless-pidgin stage. In any case, we must reckon with the complex French and HC variations noted in this appendix. Pending detailed work on the (socio)linguistics of Caribbean French colonists and their entourage, it is quite unlikely that all the relevant French varieties would “consistently metathesize [-er]” and that all “lacked a final /l/ in their pronunciation of -eur].” (See, e.g., Chaudenson & Mufwene 2001: 145–153 for preliminary remarks and caveats on the intricate mix of Romance lects – standard, “patois”, non-native varieties, koines, etc. – in the colonial French Caribbean and elsewhere in the New World.)

To sum up, the data and observations that disconfirm Goyette’s (and McWhorter’s) late-borrowing scenario are taken from Goyette’s own bibliographical source (i.e., Fattier 1998) and from well-known facts about French diachrony as can be found in, e.g., Nyrop (1899, 1903), Brunot (1906, 1924), Zink (1986), Morin (1986), Pierret (1994). Lastly, Goyette’s (2000) categorical claim about the postulated (but undocumented) existence of a totally affixless proto-HC is based both on faulty logic and on erroneous empirical generalizations about the diachrony of (the diachrony of) only two HC affixes – not a representative sample, by any measure.

Appendix B: On the “afrogenesis” of Haitian (and Mauritian) Creole

All available evidence, coupled with theoretical considerations, suggests that the ancestors of HC as spoken in colonial Haiti (then known as Saint-Domingue) were never affixless, contra the claims of the “classic” pidgin-to-creole scenario (see the data, argumentation and references in Sections 2.1, 3.3, and Appendix A). But what if the “real” ancestor of HC goes back even further, to some UNDOCUMENTED French-based proto-pidgin spoken somewhere in West Africa, say in Senegal, around a slave fort? Is it, then, this hypothetical pidgin qua HC’s proto-ancestor that would have most closely honed to the (“simplest”) Creole Prototype?

McWhorter’s (2000b) “Afrogenesis Hypothesis” (hereafter AH) pushes the origin of Caribbean and Indian Ocean creoles back to a small number of hypothetical pidgins that would have been created around West African slave castles (cf. Goodman 1964: 129–132). The AH considers Caribbean and Indian Ocean French-lexicon creoles as expan-
On the origin of creoles

sions of a SINGLE French-based pidgin ancestor that was created around a slave-trading fort on the Senegalese coast in mid-17th century. The AH is empirically, methodologically, and theoretically flawed. Only a small sample of these flaws can be discussed in this appendix, abstracting away from the lack of historical evidence (see, e.g., Bickerton 1998).

Empirically, the AH’s linguistic evidence is based on a misleading and skewed survey based on superficial comparisons. The latter are string-based with no attempt whatsoever at structural, distributional, and semantic analyses. Independently of robust and documented morphosyntactic and interpretive differences, any superficial similarity between, say, tense/mood/aspect markers across Caribbean and Indian Ocean French-lexicon creoles is judged “too close to be attributed to chance” and taken to “suggest common ancestry” in the hypothesized SINGLE Senegalese French-based pidgin (McWhorter 2000b: 149). Given the granularity of such comparison, a host of fundamental differences among French-lexicon creoles (vis-à-vis their syntax and semantics) are either ignored altogether or dismissed as “minor paradigmatic variation [that] is not counterevidence to a common [Senegalese pidgin] parent”.

Take tense/mood/aspect markers. On one hand, robust TMA-related differences across French-lexicon creoles are amply documented. On the other hand, non-creole French varieties (e.g., Québec French, Missouri French, Cajun French, and 17th-/18th-century French) exhibit cognates of the same preverbal TMA FORMS that are enlisted from HC and Mauritian Creole as support for AH. (The references in Note 48 document a variety of (dis)similarities in TMA and clause structure across creole and non-creole French-related varieties.) Do the “creole-like” preverbal TMA markers in regional and diachronic French varieties trace back to a single pidgin spoken somewhere in Africa? As many have noted, the origin of these markers is, in all likelihood, not from a Senegalese French-based pidgin, but from (the grammaticalization and, in some cases L1-influenced, restructuration of) verbal periphrases in earlier French varieties.

More generally, do loose and superficial similarities in the phonetics, distribution, and interpretation of preverbal TMA markers, IN ADDITION TO SYSTEMATIC MORPHOLOGICAL AND LEXICAL CORRESPONDENCES, constitute “conclusive evidence of a common origin” (cf. McWhorter 2000b: 148–151, 178–179, etc.)? If so, then HC and Mauritian Creole readily join Québec French, Cajun French, Missouri French, etc., as bona fide co-descendants with common origins in FULL-FLEDERED varieties of French. At this point, the discussion of (non)common PIDGIN origins for French-lexicon creoles becomes moot. (Also note that Baker’s (1995: 14) survey of pidgin and creole characteristics tentatively takes the combination of preverbal markers as one of the few “potential candidates for linguistic features which might distinguish creoles from pidgins”. If Baker is right, then the “rudimentary” would-be pidgin spoken around 17th-century West African slave castles could not have provided any stable structural model for the complex structures of HC and Mauritian Creole’s TMA systems, contra the premises of the AH.)

Related methodological remarks apply to McWhorter’s (2000b: 151–155) use of HC’s ye and Mauritian Creole’s ete as evidence for the AH. (See, e.g., HC ye in Se yon lengwis Bouki ye ‘It’s a linguist that Bouki is’ and Mauritian ete in En volver Malis ete ‘A thief Malis is.’) HC ye and Mauritian ete are classified, without structural analysis,
Commentary on McWhorter: Michel DeGraff

One problem here is that McWhorter skirts around central empirical and theoretical details of HC and vs Mauritian predication patterns and their (language-particular and universal/theoretical) implications. Here are some tidbits of ongoing research and debate on the nature of HC ye vs. Mauritian Creole ete, just enough to illustrate the fragility of the facile comparisons that underlie the AH.

On the Haitian side, I myself have analyzed ye as a morpheme that spells out certain traces of non-verbal (i.e., [−V]) predicates when the latter move outside the clause (i.e., outside IP into some operator position) for contrastive stress or WH-formation; see DeGraff (1992a, b, 1995, 1998, 1999c); cf. (34h)–(34j) above. In Se yon lengwis Bouki ye, ye would spell out the trace of the displaced nominal predicate yon lengwis ‘a linguist’, even when the (phonetically realized) trace is not in sentence-final position as in Se yon lengwis Bouki ye vre ‘It’s a linguist that he really is’ (like HC ye, Mauritian ete does surface even when not “exposed” in sentence-final position; Syea 1997: 34). This disconfirms McWhorter’s (2000b: 152) generalization that “copular overtness is sharply restricted to sentence-final position”; “exposed copula” is somewhat a misnomer. In reality, spell-out of [−V] predicate traces by ye is subject to subtle syntactic and semantic constraints, having to do with the licensing of movement and traces and with the syntax-semantics of quantification. These constraints are rooted in UG, even if they result in distributional and interpretative details that appear specific to HC.

On the Mauritian side, Baker & Corne (1982: 46, 103) argue explicitly against the copula status of ete while Baker & Syea (1991: 172) take ete to result from Case-assignment requirements. As of Syea’s (1997) analysis, it takes ete to result from the need to strongly head-govern a predicate trace; in root questions such as Kot Malis (ete)?, strong head-government of kote’s trace is ensured either via ete or via (the co-indexed trace of) a null copula that has moved from V to C and agrees abstractly with the moved predicate in Spec(CP). Though elegant, this analysis does not account for the (apparent?) restriction of ete to WH-, nominal, and prepositional predicates (do we get ete with movement of verbal and adjectival ([+V]) projections?). Neither does it account for the facts noted by Baker & Syea (1991: 167–170) whereby ete in contemporary Mauritian Creole is OBLIGATORY in matrix non-negated present-tense questions with ki, ki kote, and other WH-phrases distinct from kot (but see Syea 1997: 28–29 for crucially different – dialectal? – judgements). Lastly, if the null-vs.-overt alternation is driven by economy considerations vis-à-vis ECP satisfaction (“null except when it can’t be”; Syea 1997: 52), then the null-vs.-overt alternation in Kot Malis (ete)? is incorrectly ruled out. Of course, these are all delicate theoretical problems that McWhorter’s superficial “description” has nothing to say about.

Most relevant to the discussion here, four observations are in order on the cross-creole/crosslinguistic syntax of (so called) “exposed copulas”: (i) HC ye and Mauritian ete seem to have evolved via distinct diachronic routes, judging from their respective etymology and from Baker & Syea’s (1991) historical analysis; (ii) predication and predicate-movement patterns are not isomorphic across HC and Mauritian Creole: differences obtain, e.g., in root clauses and in comparative clauses; see Baker & Corne (1981: 31–48), Syea (1997: 30, Note 11); (iii) the HC and Mauritian predicate-movement patterns find
rough analogues in English varieties, which motivates Syea (1997) to extend his analysis to John isn’t a teacher vs. I wonder what John isn’t’s (now) vs. What isn’t John; in a similar comparative mode, DeGraff (1998) analyzes certain parallels between HC and Irish predication patterns and extends his analysis from HC to Irish; (iv) some of the Haitian and Mauritian non-verbal predication patterns find parallels in Hebrew, Arabic, Russian, etc. (DeGraff 1992a, b, 1998; Syea 1997).

Observations (i) and (ii) disconfirm McWhorter’s (2000b: 152) claim that HC and Mauritian show “the same occurrence pattern”, which would originate from a single common pidgin. Here, again, detailed distributional and structural analysis is of the essence: linguistics, after all, is about structure, not strings, since what you see is often NOT what the structure gets (see Section 5.3 above for related remarks). Observations (iii) and (iv) illustrate the methodological perils of scant and skewed comparisons as “linguistic evidence” in phylogenetic speculations: the structural resemblances in (iii) and (iv) surely do not suggest that HC, Mauritian Creole, English, Irish, Hebrew, Arabic, and Russian all descend from a Senegalese French-based pidgin.

In any event, as analyzed thus far, the HC predicate-movement strategies (see (34h)–(34j)) and their Mauritian analogues are built on the intricate interaction of delicate morphosyntactic and/or semantic constraints, and so are their TMA systems. As such, they could hardly qualify as pidgin features, specially given the definition of pidgins in WSG (Section 2.3) as youngest/simplest “rudimentary codes” that eschew all but the “functionally central” (also see Baker’s 1995: 14 comment, cited above, on pidgins’ apparent lack of TMA combinations). The predicate-movement and TMA strategies in HC and Mauritian do not exist in many (functional) “old” languages. Therefore, the syntax and semantics of TMA and predication in HC and Mauritian – alleged telltale of a common pidgin ancestor – could not have been part of any pidgin that was created on the Guinea Coast as “rudimentary code [that is unlike] full language”.

Another conceptual flaw in the AH argumentation concerns its idiosyncratic use of the comparative method. Take, say, Romance languages and the uncontroversial fact that they, like French-lexicon creoles, exhibit structural correspondences aplenty at various levels of grammar (including across their Latin-derived lexica). Pan-Romance correspondences are more reliably documented and more numerous than McWhorter’s few HC–Mauritian correspondences (the latter number a dozen or so). The logic of the AH comparison and argumentation, when applied to Romance, would have us erroneously conclude that all Romance languages originated in a single locale via a single contact language created in a single encounter. However pan-Romance similarities are not due to a single encounter with Latin; they are due to historically and geographically separate encounters between related varieties of Latin and diverse “substrate” languages. The Romance case teaches us that it is an error to claim monogenesis in a single pidgin in order to explain grammatical correspondences among certain French-lexicon creoles (e.g., HC and Mauritian Creole), lest we throw away our usual comparative-historical heuristics when dealing with creole genesis.

A logical flaw in AH concerns the (non-)evidence for a common French-based PIDGIN ancestor vs. related evidence for FULL-FL E DGED French ancestors. If pidgins are structurally reduced lowest-common-denominator compromises among the source languages – “rudimentary codes not fulfilling the needs of full language […] that eschew
Commentary on McWhorter: Michel DeGraff

all but the functionally central” (WSG: Section 2.3; also see discussion in Section 4.2 above) – then the few bits of superficial comparative data in the AH also count, in principle, as evidence for common ancestry in (native and non-native) FULL-FLEDGED French varieties, as spoken in the relevant contact situations. Given assumptions in the AH (and in WSG), full-fledged French varieties (or, rather, simplifications thereof) contributed to the reduced patterns in the hypothesized Senegalese (proto-)pidgin – recall that pidginization “‘shaves away’ a large degree of accreted complexity” from the source languages (WSG: Section 4.4; also see McWhorter 2000b: 4). If so, there logically is no way to demonstrate single ancestry for, e.g., HC and Mauritian Creole in a common French-based PIDGIN while excluding separate ancestry in independent situations of contact between the corresponding FULL-FLEDGED varieties of French and (some of) the corresponding FULL-FLEDGED substrate(s), unless it can be documented that speakers of the relevant varieties entered into contact once and only once, in one and only one locale.

As den Besten, Muysken, & Smith (1995: 88–89) write, any sort of theory of monogenesis from a single pidgin is “fundamentally flawed” and “completely irrational” because “[a] UNIQUE example of any TYPE of phenomenon connected with human conceptual and cultural activity is just inconceivable – anything that can happen once can also happen more frequently”. If (some) adult plantation slaves in Haiti, Jamaica, Barbados, Cuba, Columbia, Mauritius, Seychelles, Mauritius, Réunion, New Caledonia, etc., did approximate some (ANY) variety of some European language (cf. (ii) and (iv) in Note 15), then they, like language learners everywhere, could not have acquired that variety overnight. This is specially so in the psycho-social context of colonial plantations from the perspective of the African-born who were taken to the colony as adults (see the contemporary reports in Pelleprat 1655, Girod-Chantrans 1785, Moreau de Saint-Méry 1797, Descourtilz 1809, etc.). These African-born slaves (the bossales) must have passed through a “pidgin(ized)” (QUA EARLY INTERLANGUAGE) stage with some structural features similar to those of the corresponding “pidgins” created by their in-situ compatriots who dwelt around African slave forts. It is not accidental that the speech of the Bossales – the numerical majority on 18th-century Haitian plantations – was often ranked as markedly “inferior” and “unintelligible” as compared to the speech of the locally-born (“Creole”) slaves. Sociolinguistic factors, some of which remain to be elucidated, would determine the eventual fate of these early pidgins/interlanguages in the Old and New Worlds. However Cartesian-Uniformitarian assumptions about language acquisition/creation (see Sections 4.2 and 4.4) guarantee the existence of these “pidgin(ized)” varieties, at least as transitory individual-level lects, across all instances of language contact in the Old and New Worlds and beyond.

As for full-fledged and stable French-lexicon creoles, systematic correspondences between them and the corresponding FULL-FLEDGED varieties of French as spoken in, say, the then-colonized Caribbean and Indian Ocean islands (e.g., systematic correspondences at the level of morphology and lexicon; see Section 3.3) are many times more robust and numerous than the few (about a dozen) superficial correspondences claimed by the AH as genetic tracers linking Caribbean and Indian Ocean French-lexicon creoles. Yet these systematic lexical-morphological correspondences across creoles and their respective lexifiers, while they belie exclusive ancestry in a structurally reduced
pidgin, are not taken into account by the AH. In effect, the AH enlists scant and skewed correspondences among Caribbean and Indian-Ocean French-lexicon creoles in order to argue for the un-broken transmission of some hypothetical French-based Senegalese pidgin while it discards massive and robust correspondences between these creoles and their French lexifier in order to argue for a radical break in the transmission of French. This methodological paradox is unlike standard practices in comparative-historical linguistics.

Keeping the latter in mind, we can conclude that certain commonalities among, say, Mauritian and HC need not be due to “the same encounter with French” and need not “trace back to the same pidgin” (contra McWhorter 2000b: 147, 150). Like in the better understood Romance case, commonalities can independently arise from separate encounters (in the plural) among overlapping sets of languages and from universal strategies of language acquisition/creation. Common ancestry and common patterns do not necessarily entail common birthplace. In this perspective, differences among French-lexicon creoles (somewhat on a par with differences across Romance) are due to, inter alia, ecological variations across contact situations – variations that are now being documented (which varieties were spoken and learnt where, how, by whom, to what ends, by how many, for how long, etc.?).

On a structural UG-related note, let’s ask a question that goes beyond the specific concerns of the AH (this question is related to the general methodology in WSG; see Section 5.3 above): With respect to modern historical-comparative syntax, what is the status and import of string-based comparisons of few isolated and superficial patterns?

With clever handpicking, any degree of ad hoc superficial (dis)similarity can be established between any pair of languages, whether they are historically related or not (see Note 20). Yet twelve or so superficial similarities seem enough for the AH to consider HC and Mauritian Creole “too alike not to have had a common [pidgin] ancestor”, in spite of the lack of sociohistorical evidence for such a pidgin. One can’t help but notice that there do exist, across a wide range of creoles and non-creoles across distinct genetic phyla, deep structural similarities in clausal structure, including TMA structure (see, e.g., Cinque 1999). In AH’s parlance, these similarities become “too close to be attributed to chance”, “suggest[ing] common ancestry”. Thus, the widespread structural similarities in Cinque’s large-scale and theoretically grounded crosslinguistic comparisons would erroneously suggest a monogenesis scenario for all languages – perhaps from a single (affixless?) Ur-creole spoken by Eve in prehistorical Africa! (See Note 18.) Humor put aside, Cinque’s own, and more reasonable, conclusion is that certain crosslinguistically common patterns, even if they superficially look idiosyncratic, can – actually, MUST – arise independently across languages, simply due to UG. In this view, common patterns do not necessarily entail common ancestry in one single E-language. Common patterns are often due to the common biological ancestry of the species Homo sapiens and universal constraints on I-languages. (See Marantz 1983: 16 for similar arguments, as part of a critique of Bickerton’s Bioprogram Hypothesis.)
Correspondence address: Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, U.S.A.; e-mail: degraff@mit.edu

Acknowledgements: For helping me (re-)write this paper with more depth (and more pleasure) than I ever thought I could and for kindness beyond expectations, my heartfelt thanks go to Yves Dejean, Dominique Fattier, Tometo Hopkins, Alec Marantz, Heliana Mello, Salikoko Mufwene, Marilene Phipps, and Adrienne Talamas. I am specially grateful to Yves (Papa Iv) for constant guidance, to Dominique for her seminal work on the Haitian Creole lexicon, to Salikoko for his always “heretical” inspiration, and to Marilene for endless brainstorming.

This paper and much else in my thinking about “creoles” and “creolistics” benefited from many, many discussions with students and colleagues – e.g., in my Topics in Linguistic Theory class at MIT (Spring 2001), at the meeting of the Society for Pidgin & Creole Linguistics in Coimbra, Portugal (June 2001), and on the Internet List CreoLIST. I’ll single out Marlyse Baptista, Melissa Cho, Hans den Besten, Yolanda Rivera-Castillo, Maria Carlota Rosa, Michael Tsai, and Henri “Nou la!” Wittmann.

I would also like to take this opportunity to express my gratitude to various institutions that, over the past couple of years, have kindly invited me for colloquia or courses where I could share, and improve on, some of the ideas developed in this paper and related work: Bates College, Boston University, Boston’s Haitian Scientific Society, Cornell University, Florida International University, University of Maryland (College Park), Ohio State University, UMass Boston, Universidade Federal de Minas Gerais, Universidade Federal do Rio de Janeiro, University of Chicago, University of Rochester, University of Tulsa, and the University of the West Indies (Barbados) – host of the June 2001 conference (Re)Thinking Caribbean Culture. Much of my “(re)thinking” is due to feedback I received at these venues.

Abbreviations: AH Afrogenesis Hypothesis (see Appendix B), BV Basic Variety (Klein & Perdue 1997), HC Haitian Creole, PLD Primary Linguistic Data, TMA tense/mood/aspect, UG Universal Grammar, WSG McWhorter’s target article in this issue of Linguistic Typology.

References

Bacon, Francis (1620 [1994]). Novum Organum. [Translated and edited by Peter Urbach & John Gibson, Chicago: Open Court.]


On the origin of creoles

— (1993). *Alerte, on tue!*. Montréal: CIDIHCA.


On the origin of creoles

Commentary on McWhorter: Michel DeGraff


Meillet, Antoine (1906 [1926]). L’état actuel des études de linguistique générale. Opening lecture for the course on comparative grammar at the Collège de France. [Reprinted in Meillet 1926: 1–18.]


On the origin of creoles


Commentary on McWhorter: Michel DeGraff

On the origin of creoles


Commentary on McWhorter: Michel DeGraff


1. Three questions, all to be answered in the negative

Our paper was prompted by three questions. First, the creole prototype (CP) question: Can creoles be defined as a synchronically viable structural class as suggested in McWhorter (1998)? Second, the break in transmission question: Can we really distinguish between normal transmission and abnormal transmission in the formation of a language (along the lines proposed in Thomason & Kaufman 1988)? Third, the typology of Sinitic question: Chinese varieties exhibit morphological simplifications of the same kinds as McWhorter sees in creole grammars; are Chinese varieties therefore creoles?

The answers, we suggest, are all negative. The latest debates within creolistics are a sign of a new maturity of the field which is slowly but steadily finding its way back to linguistic theory rather than defining itself as some special branch. Moreover, insights emerging from the debate can potentially redefine concepts of language change that have been taken for granted for too long.

In this paper we show three things. First, there is no structurally definable class of languages called creoles on typological grounds; in other words, creole languages cannot be treated as synchronically different from any other language. This will be shown by applying the CP to the typology of Sinitic; in this experiment we will see that Chinese varieties could well qualify as fairly typical creoles. We will argue that a more constructive way to view the structural features described by McWhorter can be found within a broader perspective of heavy contact situations involving typologically distant varieties (see Section 3).

Second, as already suggested in several studies (Croft 2000, DeGraff 2001), there is no dichotomy between broken and continuous transmission as often assumed in historical linguistics as well as creolistics. The question is not whether creoles arise out of broken transmission or not but rather that there is no such thing as normal transmission since all languages undergo some kind of restructuring at some point of their evolution. What there is can be defined in terms of slower or faster degree of restructuring as well as higher or lower amount of convergence in the evolution of any language. In other words, there is no such thing as creolization in diachronic terms but only change.

Third, the structure of Chinese varieties puts them in a category of creoloids if this is used as a general term for languages that have undergone significant restructuring and shifted away considerably from the original parent-languages.
Since this category would include a broad range of natural languages the dichotomy between creole and non-creole languages can be dissolved.

These claims have important consequences for general linguistic theory as they, on the one hand, reabsorb creole studies into mainstream linguistics, as already argued in Mufwene (1986), Corne (1995), etc.; on the other hand, they offer a fresh view of language change and language contact. Hybridity, we will argue, is to different degrees the norm, rather than the exception, in language change. We are therefore questioning the validity of the distinction between genetic and non-genetic evolution as drawn by Thomason & Kaufman (1988). Having mixed parents, we argue, is more the norm rather than an exception, though a long-held view of historical linguistics has kept this notion frozen and propagated internal reconstruction and the family tree model of genetic relationship as the main trend. This view has weakened considerably in recent years: among others, Dixon (1997) questions the validity of the family tree model when applied to languages that have been in prolonged contact with languages of different genetic affiliation. DeGraff (2001) argues against the concept of creolization as a unique process within language change and in Ansaldo (1999) as well as Croft (2000) we find the suggestion that hybridity may be more common than actually assumed so far. By recognizing that even “normal” languages often have more than one ancestor we ultimately dissolve the special status of creole as well. That is why the concepts of creolization and genetic evolution are intimately related.

We will give evidence for the above by conducting an experiment based on the notion of CP applied to a “normal” language family. In particular, we re-evaluate the CP showing that it does not convincingly distinguish what are assumed to be prototypical creoles from languages believed to have developed through normal transmission, namely Chinese varieties. And we look at the history of a particular language family, namely Sinitic, which, though assumed to have evolved “naturally”, shows different degrees of restructuring and a serious amount of genetic distance in between varieties.

Ironically, the CP, by failing to capture the supposedly essential, typological nature of creoles shows that there is no reason for distinguishing between the latter and so-called normal languages. As the only serious recent approach to define creoles in structural terms, it ultimately proves that creoles are, at best, a socio-historical concept, but not a linguistically valid one. Only by insisting on the confluence of the three features can the CP come close to defining a structural class. In that sense, however, the CP has to be taken as a prototype

1. Please note that any mention of “creoleness” and/or “normal” transmission in this paper should be taken as presented here, in quotes.
Typical creoles and simple languages

313

concept and as such it allows for semi-creoles and creoloids (as McWhorter acknowledges, (1998) and personal communication). 2

Even considering the latest addition to the prototype manifesto, McWhorter’s target article in this issue of LT, which introduces a concept of relative structural complexity in natural languages, the basic weaknesses of the prototype idea remain. Though we find the idea that natural languages offer different degrees of complexity rather appealing we want to point out that one fundamental question needs to be answered: complex to whom? This question is of particular relevance for the prototype idea especially when we deal with the implication that tonal languages are more complex to acquire than non-tonal varieties (McWhorter 1998). As we point out below, there is no empirical evidence for this in studies of acquisition. Nor is there empirical evidence for the fact that tonal systems undergo simplification in all contact situations. In fact, we will argue, there is evidence against this in Sinitic varieties. On a different note, the complexity idea does suggest that, basically, isolating languages would be younger than say agglutinative or polysynthetic languages (cf. McWhorter 2000). Again, though we find this yet another very appealing idea, empirical evidence seems to be lacking. 3

Sinitic languages, we argue, appear as creoloids, as do many “sinospheric” languages such as Zhuang and Hmong. Non-tonal Austro-Asiatic languages such as Chrau, by meeting the tonal criterion, may be even closer to the CP. Here we shall focus on the Sinitic cases since we are particularly familiar with them and since previous discussions about the creole nature of Chinese varieties are still unresolved (cf. Branner 2000). Moreover, even accepted creoles may fail one or more of the criteria as DeGraff (2001) has argued with respect to Haitian derivational morphology, showing that the “conjunction” cannot be an absolutely necessary condition. In a typological diachronic approach the class of creoles dissolves: creoles are a subset of isolating languages and each of their properties are all attested in “non-creoles” as well. Parallel to this, the process that leads to the formation of a creole is not intrinsically different from language evolution processes such as restructuring and convergence although it may differ at times in the rapidity and drasticness of those changes. It becomes

2. Note also that in McWhorter’s presentation the three features are on the same level, not related implicationally.
3. There are other problems with the notion of complexity, though to address them all would require a paper in its own right. For example, though it is plausible that an older language, developing sophisticated registers also develops many redundant grammatical features, “younger languages” such as creoles often entail a serious amount of variation and the ability to negotiate codes within a continuum scenario. So, while grammatically they may be said to appear less complex, socio linguistically they offer rather complex situations. The Tok Pisin speaker in other words may deal with just as much complexity as say the speaker of Swahili, though in different domains. Again, the question remains: complex to whom?
obvious then that a dichotomy in synchronic and diachronic terms between creoles and non-creoles is unnecessary. Dissolving this dichotomy leads us to conclude, first, that creolization is simply one subset of linguistic changes that are commonly found in languages of the world (just like, e.g., grammaticalization – which does not have the status of a theory, as emphasized in Newmeyer 1998, nor would we want to identify a specific set of languages as the “grammaticalized languages”); second, that genetic affiliation among languages is something that always needs to be negotiated between internal reconstruction and language contact phenomena, with multiple parenthood being the norm rather than the exception; third, that transmission is always “normal” though the rapidity of acquisition of the grammar and the degree of divergence from the adults’ generation can vary significantly.

These assumptions will be systematized in Section 4 where we will present what we call the Hybridity Cline Hypothesis.

2. The creole prototype: Creoles as a structural class

The extent to which creoles, or more generally contact languages, are structurally unique has become a major point of controversy. On the one hand, Bickerton’s Language Bioprogram Hypothesis (Bickerton 1981, 1984) strongly implied that they are: “these languages show similarities which go far beyond the possibility of coincidental resemblance and which are not explicable in terms of conventional transmission processes such as diffusion or substratum influence” (Bickerton 1981: 132).

On the other hand, Thomason (1996), for example, argues that the class of contact languages does not correspond to any structural typology (1996: 3):

This definition [of contact languages] is fundamentally historical; it is based on diversity in the sources of linguistic structures, rather than on (say) typological characteristics of the language. The reason for insisting on a historical definition is that synchronic definitions don’t work; there is, for instance, no such thing as a master list of linguistic features that are universally shared by, and exclusive to, contact languages, or even pidgins and creoles as a set.

Against this view McWhorter (1998) argues that creoles do constitute “a synchronically definable typological class of languages”, characterized by the conjunction of three traits: (i) little or no inflectional affixation; (ii) little or no use of tone (a) to distinguish monosyllables, or (b) to encode syntactic distinctions; (iii) semantically regular derivational affixation.

Crucially, it is the simultaneous occurrence of all three properties which defines the CP. McWhorter argues that they are all the result of imperfect acquisition by “emergency learners”; inflection, irregular derivation and extensive use of tone are opaque to learners in such situations, and appear only as a result of protracted development over time.
2.1. Inflectional morphology

McWhorter argues that some creoles have at most one or two inflectional affixes. For example, Tok Pisin has two: the transitive marker -im and the adjectival marker -pela, exemplified in (1) from Mühlhäusler (1982: 195–196):

(1) a. Em i kam pain-im Jisas.
   he PM come find-TR Jesus
   ‘He came to find Jesus.’

b. Dis-pela man i save slip long ples matmat.
   this-ADJ man PM HAB sleep in cemetery
   ‘This man slept in a cemetery.’

Also, inflectional morphemes are seldom obligatory in creoles as the following examples from Negerholland Dutch Creole show:

(2) a. Ham a jak ši kabrita sini a sabān
   he PAST drive his goat PL to savanna
   ‘He drove his goats to the savanna.’ (Stolz 1986: 122)

b. Anānsi a ko, so los alma si hunda Ø abit
   Anansi PAST come so let all his chicken out
   abini di yard.
   in the yard
   ‘Anansi came and let all his chicken out into the yard.’ (Stolz 1986: 122)

2.2. The tonal criterion

The tonal criterion, which has the effect of excluding Sinitic and “sinospheric” languages, appears problematic to the extent that some creoles do have lexical tone. In Papiamentu, Römer (1977) noted some 250 minimal pairs such as pápà ‘porridge’ vs. pàpá ‘father’ and mâtâ ‘plant, tree’ vs. màtâ ‘kill’. For Principense, Devonish (1986) cites pairs such as fuúta ‘breadfruit’ (< Portuguese fruta) vs. fúuta ‘steal’ (< Port. furta), peétu ‘black’ (< Port. pretu) vs. péetu ‘near’ (Port. perto). High tones typically correspond to the stressed syllable in Portuguese, while low tones (unmarked) correspond to unstressed syllables in Portuguese (Günther 1973), e.g., tôsi < Port. tosse (stress on first syllable) vs. tosi < Port. tossir (stress on second syllable). Rising tone is a development from Portuguese words that lost a final stressed syllable in the course of creolization, e.g., kwē < Port. correr [kunér].

Ndjuka and Saramaccan, often considered “prototypical” creoles, are real tone-languages. Saramaccan has three underlying tones: H, L, and neutral. These tones have grammatical as well as lexical functions, e.g., unspecified tones become H between adjoining words in a syntactic relationship carrying a H tone. A limited number of minimal pairs involve monosyllables, e.g.,
Commentary on McWhorter: U. Ansaldo and S. J. Matthews

dá 'to give' vs. da 'to be' (Voorhoeve 1961). Perhaps most significantly, as McWhorter acknowledges, Saramaccan exhibits tone sandhi, e.g., mí ‘my’ + tata ‘father’ → mí tata ‘my father’. The operation of sandhi appears comparable in complexity to that in many Sinitic varieties: indeed, Ham (1999:73) makes explicit comparisons with the Min dialect of Xiamen regarding the syntactic domain in which tone sandhi operates.

In Guyanese Creole English (Carter 1987), tone also has syntactic functions: iteration is expressed by downstep on the second high tone, e.g., tall! tall! while in reduplication the first tone is incorporated in the second, e.g., tall! tall!

The CP can also be used to classify various languages as semi-creoles, e.g., Bantu-based contact languages, e.g., Kituba and Lingala (cf. McWhorter 1998:811). In this respect, it is interesting to observe that Sinitic languages share with these the contrastive use of tone and are actually closer to the CP with respect to the criteria of inflection and derivation. For example, Kituba has a plural prefix for nouns and Lingala has subject agreement on the verb while Sinitic varieties are much more impoverished in this respect. So even if we accept this criterion, without much critical insight it appears obvious that Sinitic varieties should be considered at least as creoloids. How can the CP account for this state of affairs?

The rationale behind the tonal criterion is McWhorter’s assumption that tone is dropped in rapid acquisition situations because of the necessity of simplifying the system (basically the same claim accounts for the reduced inflectional morphology and transparent derivational morphology). Of the three features making up the CP, the tonal criterion seems the most problematic because of lack of empirical evidence to substantiate this claim. It seems to be the case that tone is typically lost in the course of creolization but this is largely true with regard to situations where only one substrate language exhibits tone. Where both substrate and superstrate are tonal, as in the African varieties, tone is frequently retained in the language (e.g., Kituba, Lingala). Moreover, the studied cases involve African tonal systems; tone systems such as those in Southeast Asian languages are very different and could behave differently. Phonological theory has already suffered from the mistaken assumption that tone-systems of Southwest Asia could be treated along the lines of typical African systems, as can be seen, for example, in the application of Autosegmental Theory to Chinese tone-systems. Indeed there are cases in Southwest China where the number of tones increases under language contact, see, e.g., the 15 tones of Dong.

---

4. In Tay Boy, a Vietnamese-French creole, the five tones of Vietnamese were preserved, cf. Phillips (1975:258).
5. E.g., the assumption that contour tones should be represented as a combination of underlying level tones.
Kam-Sui language. McWhorter’s account assumes that the acquisition of tone is somehow a complex task for learners, a point that needs to be demonstrated. Again, complex to whom? For example, acquisition of tone does not seem to be particularly problematic in Sinitic languages even for bidialectal children acquiring two separate tone systems. The tonal criterion appears therefore at least weak on theoretical basis and does not completely exclude Sinitic varieties from the CP.

2.3. Derivational morphology

As an example of the transparency of derivational affixes let us look at Tok Pisin (McWhorter 1998, Mühlhäusler 1985). In this language we find the item pasin (from English fashion) for deriving abstract nouns, e.g., gutpasin ‘virtue’ (from gut ‘good’), isipasin ‘slowness’ (from isi ‘slow’), proutpasin ‘pride’ (from prout ‘proud’), and paitpasin ‘warfare’ (from pait ‘fight’).

Note, however, that DeGraff (2001) shows that Haitian derivational morphology is more extensive and less clearly transparent than the CP would imply. The “inversive” prefix de- for example, does not exhibit consistently transparent semantics any more than its cognates in other Romance languages.

3. Sinitic and the creole prototype

Apart from the doubts raised about the empirical validity of the three criteria, we suspect that in general the CP as a set of defining characteristics may have been misunderstood. If the empirical situation is that the languages known as creoles tend towards the three characteristics mentioned above, rather than being taken as diagnostics of creolehood or lack thereof, those features could be seen as a scale on which to map actual languages and their approximation to the “ideal” creole character.

Perhaps a more constructive way to see the “prototypical creole” traits is that languages which have been subject to intensive contact involving several typologically distant varieties will tend to show some combination (or subset) of these features. We shall proceed to show that the idea of the CP can be productively applied to languages in general. Looking at varieties of Chinese in the light of the CP, strong similarities between Sinitic grammar and the prototype features emerge.

Two of the three features identified as part of the CP are easily found in varieties of Sinitic. The tonal criterion would indeed rule out Sinitic varieties from the CP. This criterion is however dubious: Does Sinitic fall short of creole status on this criterion or is its validity somewhat overestimated?
3.1. **Tone**

As shown in Section 2.2, rather typical creoles exist with subtle tonal differentiations. It is not clear that these systems are any less complex than, for example, some northern Mandarin varieties with three lexical tones and minimal tone sandhi. But even if we were to concede that Mandarin ranks higher on the tonal complexity scale than some typical creoles, in the next two sections we show that it ranks lower than typical creoles in respect to the other two features of the CP.

McWhorter relies on the tonal criterion to exclude Chinese and other Southeast Asian languages. Virtually all varieties of Chinese are tonal, which might seem to mark the typology of Sinitic languages as clearly distinct from that of creoles. However, as we have seen, a number of creoles do make limited use of tone. To make the tonal criterion more precise, McWhorter argues that creoles do not use tone (a) to distinguish monosyllables, or (b) for syntactic purposes. He argues that those creoles which show tone make almost no grammatical use of it.

On the first of these sub-criteria, monosyllables, we should note that the monosyllabic character of Chinese has been exaggerated (DeFrancis 1984 even speaks of the Monosyllabic Myth). Concerning the second sub-criterion, grammatical tone, this may represent a striking contrast with West African tonal languages, but a less strong contrast with Chinese that (as far as we know) makes little grammatical use of tone. One of the few clear cases involves the perfective derived by tone change in some Yue dialects:

\[(3) \quad \text{a. } sik^6 \text{ faan}^6 \quad \text{b. } sik^2 \text{ faan}^6 \]

\[
\text{eat rice eat-PFV rice} \\
\text{‘eats’ ‘has eaten’}
\]

This rising tone alternates with, and can be derived from, the perfective suffix \( \text{zo}^2 \).

3.2. **Inflection**

Turning to McWhorter’s inflectional criterion, the match between creole and Sinitic typology appears close: both have extremely limited inflection. McWhorter acknowledges that some creoles have at most one or two inflectional affixes. A similar situation holds for varieties of Chinese. Most analyses of Chinese morphology recognize two inflectional morphemes, i.e., the aspectual markers -le and -he in Mandarin and -zo\(^2\) and -gan\(^2\) in Cantonese. They appear as suffixes to the verb (see (4) for Mandarin and (6) for Cantonese) and with coverbs as in (5):

\[\]

Brought to you by | Radboud University Nijmegen
Authenticated | 131.174.248.224
Download Date | 1/28/13 11:16 AM
(4) a. Wǒ zài fāngzi lǐ zuò-zhē
   I at house in sit-DUR
   ‘I am sitting in the house.’
b. Wǒ mǎi-le hěn duō shū
   I buy-PFV very many book
   ‘I bought a lot of books.’

(5) Tāmen wàng-zhe chuán shàng fāng qiāng
    they toward-DUR boat on fire gun
   ‘They fired at the boat.’ (Li & Thompson 1981: 361)

(6) a. Keoi⁵ hai² dou⁶ gōng² guān² dīnwāa²
     (s)he at talk-PROG phone
   ‘(S)he is talking on the phone.’
b. Ngo⁵ ji² ging¹ saan¹ zo² mun⁴
     I already close-PFV door
   ‘I have already closed the door.’ (Matthews & Yip 1994: 206)

3.3. Derivational morphology

Again we can find very similar aspects in Chinese; the examples from three dialects given below show the transparency and regularity of the few derivational affixes found.

In Mandarin kē- derives adjectives: kē-ài ‘lovable’ (< ài ‘to love’); kē-pà ‘frightful’ (< pà ‘to fear’); kē-xiào ‘laughable’ (< xiào ‘to laugh’); kē-kào ‘reliable’ (< kào ‘to lean on’).

In Cantonese -dei² with reduplication derives attenuated adjectives: fei⁴-fei² -dei² ‘rather fat’ (< fei⁴ ‘fat’); laan⁴-laan² -dei² ‘a bit broken’ (< laan⁴ ‘broken’); sau²-sau² -dei² ‘thinnish’ (< sau¹ ‘thin’).

In Chaozhou sio³³ - derives reciprocal verbs: sio³³ -p hak² ‘hit each other’; sio³³ -me¹¹ ‘scold each other’; sio³³ -tai²⁵ ‘make war/fight’.

Where (derivational) morphology is found, it always occurs locally, that is, in forms unrelated to the morphology found in other dialects. For this reason, we cannot justify reconstructing morphology based solely on attested spoken varieties of Chinese (cf. Branner 2000: 160).

Considering the cases of opaque derivational forms noted by DeGraff (1999), the derivations in the Sinitic languages we have looked at appear rather more regular and transparent than those found in “prototypical creoles” such as Haitian.

These parallels could mean any of the following. First, the CP in fact suggests the existence of a class of languages with isolating typology, showing reduction with respect to their ancestors. The CP would then not be unique to creoles: there would be no “creole prototype” defined as a unique class of languages. Second, the CP suggests that Sinitic languages should be included in
Commentary on McWhorter: U. Ansaldo and S. J. Matthews

the structural class defined by it. This would revive questions about the evolution of Chinese in terms of diachronic syntax as well as sociolinguistic issues. Chinese varieties would be regarded as creoloid. Third, there is no CP and a language family always undergoes a certain amount of restructuring due to contact but the output depends on the typological nature of the input.

3.4. A sociohistorical sketch of the Sinitic family

In his recent book on the history of Chinese dialects, Branner (2000) suggests a thought experiment in which English vanishes and we are left with English based creoles. So the languages available for inspection are Sranan, Hawaiian Creole, and Tok Pisin. The lexicon will point to a common ancestor but there won’t be much shared morphology. This is very much the situation that we are faced with when we look at Chinese varieties such as Cantonese, Mandarin, and Min.

Indeed a long-standing problem of historical Chinese linguistics has been the typological chasm between the Sinitic and Tibeto-Burman branches of Sino-Tibetan. This chasm is well illustrated by DeLancey’s typological description of the Tibeto-Burman family: “With the exception of Karen, all the Tibeto-Burman languages are postpositional SOV languages with predominantly agglutinative morphology […] and this must also have been true of Proto-Sino-Tibetan. […] Proto-Tibeto-Burman is now reconstructed with a split-ergative case marking and verb agreement system” (DeLancey 1987: 806).

Given the clear genetic affinity, the Sinitic languages must have developed from such an ancestor. However, Branner (2000: 164) notes that while early Chinese appears to be typologically a form of Tibeto-Burman, modern Chinese varieties belong to a typologically distinct family. The gap between early Chinese and modern varieties is indeed puzzling and different explanations have been suggested. The fact that the transition from early to modern Chinese shows a great deal of restructuring, not unlike what is usually labelled as creolization, can be explained by the long history of contact between Sinitic and typologically distant varieties, in particular Altaic and Austro-Tai (Ansaldo 1999). In fact, the evolution of Chinese varieties has been described as a gradual shift from Austro-Tai to Altaic typology in the Altaicization theory (Hashimoto 1984). This is related to the prolonged and intense history of contact with neighbouring peoples and the vast population movements recorded throughout the history of China. In particular two phenomena appear to have been crucial for the evolution of Sinitic varieties: contact with Altaic population for roughly a millennium and the deep South of China as a linguistic area in which Tai-Kadai, Austronesian, and Sino-Tibetan varieties coexisted.6 Despite

Typical creoles and simple languages

321

a tendency towards acquiring Mandarin syntax, modern varieties of Sinitic still show clear traces of a north–south divide in various aspects of grammar. We do not intend to discuss this issue at length as it is only tangential to the central claims of this paper and as it would need to be dealt with in its own right. What we want to point out is the following: if natural languages such as Sinitic show creole-like structure as well as a history of drastic restructuring then language shift of the type observed in creolization cannot be unique to creole languages.

4. The Hybridity Cline Hypothesis

So what then are creoles? In this section we would like to present an idea that concludes this study by at the same time opening up a whole new way of thinking about language change. The view introduced in Ansaldo (1999) on hybridization in language is paralleled in Croft (2000) and takes a “contextualized typological” view (cf. Ansaldo 1999) of the field of language contact and change in general. Croft points out that pidgins and creoles, though arising in a context of extreme sociolinguistic variation, ultimately resemble other mixed languages that arise in much more stable situations:

Essentially, the linguistic result is the same as in shift and semi-shift: most of the vocabulary is drawn from a single parent language. But unlike shift or semi-shift, the grammatical elements are not drawn from the same parent language, and in fact have disputed lineages. (Croft 2000: 221)

In other words, all languages are hybrids, albeit to different extents. Where the typological and sociolinguistic variation is overwhelming, we suggest, new grammar is created heavily and rapidly because there is not enough common material available (unless a conscious effort is made to maintain social distance; cf. Mühlhäusler 1985). The restructuring is heavier in these situations as mutual accommodation is difficult. This depends presumably on the degree of typological divergence of the varieties involved (cf. Thomason & Kaufman 1988). Under this hypothesis, here called the Hybridity Cline Hypothesis, the following assumptions are postulated: first, most societies have some degree of bi/multi-lectalism/lingualism; second, all multilingual/multilectal speech communities show language contact to different extents; third, language contact leads to different degrees of restructuring in a community, i.e., all languages are hybrids; fourth, the degree of restructuring is determined by the structural (typological) affinity of the varieties in contact; fifth, the speed of restructuring is determined by the level of multilingualism of the speech community.

This view is based on the fundamental assumption that language contact necessarily leads to some kind of restructuring, be it borrowing or shift. The abruptness and drasticness of this restructuring is inversely proportional to the structural affinity between the varieties involved and to the degree of multilingual competence of the speakers concerned. Therefore, in a scenario in which
typologically divergent languages are involved and multilingualism is rare, the restructuring will be heavier. Speakers in this situation will have fewer prescriptive limits and more urgency to reach a common medium of communication than speakers in a more stable scenario. In this sense heavy contact and high diversity would only allow for minimal specification according to the requirements of Universal Grammar. Where typological divergence is low and multilingualism widespread, there will be a more concrete target variety and less urgency to restructure the system. In such a contact situation, where the languages involved are very similar, we would expect less restructuring as the pool of common structure available to speakers is rather large. Also, a certain amount of mutual intelligibility can be expected and therefore the urgency of the restructuring is low. This situation would be more conducive to overspecification. However, where languages differ typologically and where competence is low, the restructuring has to be drastic to arrive at a common medium of expression, and mutual unintelligibility would make this restructuring rather fast. Within the Hybridity Cline Hypothesis the evolution of a language can be represented as shown in Figure 1.

So-called creole varieties would be placed higher up on the restructuring axis than languages arising from a typologically more homogenous context in which multilingualism is more widespread. Sinitic varieties would have to be placed quite high up on this axis as opposed to, say, European varieties as in the latter case multilingualism was more widespread and typologically affinity is higher (cf. Branner 2000, Ansaldo 1999).
5. Final remarks

The current debate on the typological status of creole languages has been of enormous importance for the field of language change. It has uncovered the importance of language contact for change and has highlighted several typical processes of restructuring. It has brought about the insight that languages like Chinese may actually have undergone a similar process, and has therefore ultimately led to the understanding that language contact is a significant force in all types of change. It would, however, be counterproductive at this point to insist on claiming that there is something like a creole class of languages, a process of creolization as distinguished from other processes of change, or a creole-genesis of grammar. The heated debate that sprung off the CP idea and the more recent debate on creolization/convergence have had the important effect of showing once and for all that creolistics may indeed be a discipline facing a precipice, to paraphrase McWhorter’s words (2000). It shows that, ultimately, socio-cultural rather than structural approaches such as those advocated in Mufwene (1986) are the only viable ways to describe creolists as some kind of sub-group of linguists, just like we have typologists, phonologists, etc.

The linguistic mechanisms of change involved in the formation of what we regard as typical creoles as well as the sociohistorical processes accompanying them must have happened in the history of humanity many times before and may happen again. In the case of what we call creoles the difference is that we were almost able to observe them growing, at least in part. This, and only this, is what is really special about them.

Received: 12 May 2001
Revised 16 June 2001

Correspondence addresses: (Ansaldo) Department of English Language and Literature, National University of Singapore, Block As5, 7 Arts Link, Singapore 117570, Singapore; e-mail: ansaldo@nus.edu.sg; (Matthews) Department of Linguistics, Pokfulam Road, University of Hong Kong, Hong Kong, S.A.R., China; e-mail: matthews@hkucc.hku.hk.

Abbreviations: ADJ adjectival marker, DUR durative, HAB habitual, PFV perfective, PL plural, PM predicate marker, PROG progressive, TR transitive marker.

References


Commentary on McWhorter: U. Ansaldo and S. J. Matthews


Can one tell that a language is a creole just by looking at it, that is to say, by examination of its synchronic grammatical properties, without reference to its history? Given that the definition of a creole language is couched exclusively in historical terms, this question is a substantive one, on which opinions may legitimately differ. Indeed, although there is a widespread view that creoles are synchronically just like other languages, my own personal experience suggests that many linguists actually believe otherwise. On numerous occasions, when presenting the results of my investigations into the Riau dialect of Indonesian, linguists have asked me “Is it a creole?” – thereby revealing that they harbour certain convictions about the ways in which creoles differ synchronically from other languages. To the extent that I have been able to ascertain the motivation for such questions, what seems to be involved is the notion of complexity, and the belief that creoles are generally less complex than other languages. Which leads to a second question:

Can one characterize some languages as less complex than others, not just in particular subsystems within their grammars, but rather with respect to their overall structures? In previous eras, there was a widespread belief that, in comparison with their European counterparts, the languages of Africa, Asia, and the Americas were simpler, or more primitive, or plain inferior; in many cases these assumptions about the languages were coupled to other assumptions about their speakers which today would be judged as morally reprehensible. With the advent of modern linguistics and greater familiarity with the world’s languages, such beliefs were duly discarded; however, their place was taken not by serious empirical investigation of the issues involved, but rather by another dogma, to the effect that all languages are of roughly equal overall complexity. Why this dogma has taken root so pervasively is a matter for
Commentary on McWhorter: David Gil

speculation that need not concern us here. To some extent it is presumably due to political correctness and an irrelevant and misguided introduction of ethical considerations, however noble in and of themselves, into the realm of empirical investigation: some people seem to think that if one language were shown to be more complex than another, then it would follow that the latter language is in some sense inferior, which in turn would entail that the speakers of that language are inferior, and from here we’re only one short step to ethnic cleansing. But there were probably other, less extraneous reasons for the spread of this dogma: linguistics over the course of the last century has simply chosen to concern itself with a different range of issues, and besides, perhaps most importantly, complexity of linguistic structure is a notion that is extremely difficult to formalize in an explicit and quantitative manner. None of the above, however, should be reasons not to try and address the issue of complexity – as indeed is suggested in a number of recent articles, including Comrie (1992), Romaine (1992), and McWhorter (1998, 2000, 2001, this volume).

In his article in this volume, McWhorter proposes explicit and interrelated answers to the two questions above, both in the affirmative. He argues that languages can and do differ with respect to their overall degree of complexity, and that this variation provides a synchronic diagnostic for the identification of creole languages, as encapsulated in his title: “the world’s simplest grammars are creole grammars”. His essential idea is that languages typically accumulate complexity over time, and that creoles, by definition, have not had enough time to acquire such complexity. Thus, although both creole and other languages may exhibit variation with respect to their overall degree of complexity, all creoles (or at least all prototypical creoles) have simpler grammars than all older languages. Although McWhorter considers his own conclusions to be controversial, they are hardly out in left field: witness the many occasions, mentioned above, when I have been asked, on the basis of its apparently very simple grammar, whether Riau Indonesian is a creole.

In his paper in this volume McWhorter mentions Riau Indonesian briefly, characterizing it as the older language “which comes the closest to exhibiting the degree of complexity of a typical creole”, and therefore the closest to providing a counterexample to his major thesis. This paper picks up where McWhorter left off, offering a more detailed examination of Riau Indonesian and the ways in which it sheds light on the two questions above. First, it is argued that, on the basis of available historical evidence, Riau Indonesian is not a creole language (Section 2). Next, the structure of Riau Indonesian is contrasted with that of McWhorter’s stock example of a creole language, Saramaccan, showing that indeed Riau Indonesian is of lesser overall complexity than Saramaccan (Section 3). Finally, it is suggested that, on the basis of the Riau Indonesian facts, McWhorter’s proposed correlation between creoles and complexity needs to be weakened from a bi-directional to a uni-directional im-
plication: whereas all creole languages are indeed simple, older languages may run the gamut from simple to complex (Section 4).

2. Riau Indonesian in social and historical perspective

To the best of my knowledge there is no written documentation of the history of Riau Indonesian. Of course, this is hardly surprising, given that the existence of such a dialect remained virtually unacknowledged until I started working on it and writing about it in the early 90s. Thus, in the absence of such attestations, there is no alternative but to start with the present and work one’s way back in time through reasonable processes of inference.

The Indonesian province of Riau occupies a sizeable chunk of the east-central part of the large island of Sumatra, plus about 3,200 smaller islands in the straits of Malacca and the South China Sea; its population is over 3,300,000, of which 89% are Muslim (according to a 1990 census, cited in Cribb 2000: 51). Riau Indonesian is the name given to the variety or varieties of colloquial Indonesian spoken throughout the province. In actual fact, there is no reason to believe that the arbitrary and relatively recent political boundaries of Riau province correspond to any linguistic reality; future research may point to different dialectal boundaries which will in turn entail finding a more felicitous name for what is currently referred to as Riau Indonesian. But one has to start somewhere.

Riau Indonesian is not alone in the province: the linguistic landscape of Riau is one of considerable intricacy. As in many other parts of Indonesia and Southeast Asia, multilingualism and multiglossia are the norm rather than the exception. A large proportion of the population of Riau, probably over half, are migrants from other provinces or their first- and second-generation descendants. The most significant immigrant language is Minangkabau, a Malayic language from neighbouring West Sumatra province, which has itself acquired the status of a lingua franca throughout much of Riau. Other immigrant languages include Javanese, Toba Batak, Bugis, and many other non-Malayic Austronesian languages from throughout the archipelago. Finally, there is a significant population of Chinese, most of whom speak one of the southern Min Chinese languages.

All of the indigenous language varieties of Riau belong to the Malayic subgroup of Austronesian, which has an estimated time-depth of perhaps 2,000 years. A multidimensional classification of these isolects is provided in Figure 1.1

---

1. In this paper, the terms “language variety” and “isolect” are used interchangeably, as neutral cover terms for “language” and “regional/ethnic/social dialect”.

---
In accordance with Figure 1, the indigenous isolects of Riau may be classified in accordance with three partially independent dimensions: geography, ethnicity/religion, and register. Geography is obvious: in Riau, like elsewhere, people in different places speak differently. But that is just the beginning of the story.

Within each individual location, language varies as a function of ethnicity, which, in the local context, involves also cultural and religious factors. As suggested by the central horizontal axis in Figure 1, the indigenous language varieties of Riau may be divided into four major types: (a) non-mainstream isolects, spoken by small, remote, predominantly non-Muslim, indigenous “tribal” communities such as the Sakai, Orang Akit, Orang Asli, and Orang Laut; (b) mainstream isolects, spoken by the large majority of the indigenous population, who practice Islam, and who generally identify themselves as Malay, except in some western parts of the province, where various place-names function as ethnonyms; (c) Indonesian isolects, used for interethnic and increasingly also intraethnic communication by most inhabitants of the province, indigenous and non-indigenous; and finally (d) Outsider isolects, spoken by and to the Chinese inhabitants of Riau. The big “X” in the middle of the diagram stands for “Cartesian product”. What this means is that in
order to specify an isolect in Riau, it is necessary both to situate it geographically and to classify it in accordance with one of the above four types: for example, the non-mainstream isolect of the island of Pulau Padang, spoken by the local Orang Asli, or the mainstream isolect of the Siak river basin, spoken by the Malays of that region.

As indicated in Figure 1, these four types form a linear hierarchy, whose import is both geographic and sociolinguistic. Geographically, the amount of variation tends to decrease as one moves along the hierarchy from left to right. Thus, non-mainstream isolects differ greatly from place to place, mainstream isolects perhaps somewhat less, Indonesian isolects considerably less again, and Outsider isolects little if at all. Sociolinguistically, if a speaker of an ethnicity associated with an isolect further to the left encounters a speaker of an ethnicity associated with an isolect further to the right, the preferred language of communication will be the latter, or rightmost isolect. For example, an Orang Sakai and a Malay will typically communicate in a mainstream isolect of Malay; a Malay and a Javanese will generally communicate in a variety of Indonesian; and a Javanese and a Chinese will most often communicate in an Outsider isolect.

Isolects belonging to these four types are used in a variety of everyday contexts, fulfilling a wide range of functions. Nevertheless, they are, for the most part, basilectal language varieties. If and when speakers find themselves in a more formal situation, requiring recourse to a more acrolectal register, they speak Standard Indonesian to the best of their abilities. As suggested by the vertical arrow in Figure 1, there exists a continuum of registers, extending from the local varieties of basilectal Indonesian, all the way up to Standard Indonesian.

It is to these local isolects of basilectal Indonesian, indicated in Figure 1 in boldface, that the term Riau Indonesian applies. Like any other language name, Riau Indonesian conveys a convenient but unavoidable abstraction. As suggested in Figure 1, Riau Indonesian differs from place to place, it is in constant contact with other language varieties, and it lies at the lower end of a cline connecting it to the official standard language. Nevertheless, it is a real language variety, one of the native languages of a few million people, and the most commonly used variety throughout Riau province.

This, then, is Riau Indonesian in the present. But where did it come from, and what was it like in the past? Of course, in a trivial sense, Riau Indonesian is a “new” language, since Indonesian itself is the national language of a country that gained its independence only in 1945. But this is mere nomenclature: in fact, there is good reason to believe that Riau Indonesian is a relatively direct descendant from various contact varieties of Malay which, one may safely presume, were spoken in east-central Sumatra, like elsewhere in the archipelago, over the course of the preceding millennium.
As mentioned above, many of the inhabitants of Riau province are migrants from other parts of Indonesia. However, the available evidence suggests that this is not a recent phenomenon; in fact, east-central Sumatra was probably the scene for migration and ensuing language contact for the last several centuries if not longer.

In a small-scale informal survey, I interviewed a sample of inhabitants of a small rural town in Riau province, Sungai Pakning, in order to obtain information with regard to their ethnicity. This survey was linked to a broader project involving the collection of DNA samples for genetic analysis; the location was originally chosen as one likely to produce a homogeneous sample of ethnic Malays, though the results of the survey proved this to be a false assumption. Excluding those who were Chinese, the sample consisted of 97 unrelated males, mostly in their teens and twenties. Of these 97 subjects, 48, or about 50%, reported that both their parents were Malay, while the remaining 49 reported that one or both of their parents were non-Malay. Projecting back in time, the subjects were also asked about the birthplace and ethnicity of their grandparents. Of the 194 grandparent couples (2 for each of the 97 subjects), 149, or 77%, gave birth to the subject’s parent in Riau province. And of these 149 couples, 112, or 75%, were both Malay, while the remaining 37, or 25%, contained at least one non-Malay member. Thus, the latter figure, 25%, represents the percentage of households in the Riau of some 50 years ago in which at least one of the core members was an immigrant, and in which it is likely that some contact language resembling Riau Indonesian was spoken. Recall, now, that Sungai Pakning was actually chosen in order to obtain an ethnically homogeneous population of Malays; one can safely assume that the figures for non-Malays in other locations would be considerably higher. What these figures suggest, then, is that Riau province was the target of substantial immigration for the last two generations at least, and that some contact language, presumably the immediate ancestor of Riau Indonesian, would have had to have gained wide currency.

Such patterns can be projected back further in time, using other methods of inference. Nowadays, the most common path of migration is that of the Minangkabau, from their ancestral home in West Sumatra province, north-eastwards to Riau province; many of these migrants then go on to cross the straits of Malacca, to the north-east, to enter Malaysia, either legally, or in many cases as illegal immigrants. Thus, several neighbourhoods of Kuala Lumpur, the capital city of Malaysia, contain significant populations of Minangkabau, many of whom were born in West Sumatra before passing through Riau on their way to Malaysia; often, these Minangkabau have relatives not only in their ancestral villages in West Sumatra but also in Riau. However, this path of migration is hardly new. Clear cut evidence for its history is provided by the population of another state of Malaysia, namely, Negeri Sembilan. Ac-
According to historical records, Negeri Sembilan was settled by Minangkabau migrants beginning in the fifteenth century, in the era of the Malacca sultanate; the influence of these settlers can still be observed in various cultural artefacts, as well as in the local dialect of Malay, which exhibits considerable influence from Minangkabau. So the Minangkabau have been migrating from West Sumatra to peninsular Malaysia for centuries (see Cribb 2000: 52). Which raises the question: what language did these migrants speak, in previous centuries, while passing through what is now Riau province in the course of their migrations? Although we have no written records, we can make an educated guess. Since Minangkabau and the Malay dialects of Riau are closely related, a Minangkabau and a Malay would not have had to choose one of the two languages to the exclusion of the other in order to communicate, nor would they have had to resort to a third, completely different language, such as that of one of the European colonial powers. Rather, they could easily have met in the middle, making use of contact varieties of Malay in order to communicate. Presumably, such contact varieties would have been the precursors of present day Riau Indonesian.

Strategically situated on the shores of the straits, directly across from the historical sultanate of Malacca, it is safe to assume that contact varieties of Malay were spoken in east-central Sumatra for many centuries into the past. Historical records are replete with stories of Bugis sailors, Chinese emissaries, European colonialists, and merchants from almost everywhere passing through the straits (Collins 1996, Cribb 2000, and references therein). Although the focus of activity was in Malacca, some of it presumably spilled over to the Sumatran coastline, and besides, Malacca would also have attracted a significant amount of local travel across the straits. So the existence of some contact variety of Malay would have become almost inevitable.

However, even without Malacca across the straits, some variety of contact Malay would have had to have developed along the coasts and in the hinterland of east-central Sumatra in the course of the preceding centuries: so much can be concluded from the contemporary linguistic evidence alone. Although occupying a relatively small part of the Malayic-speaking region, the isolects spoken in Riau do not belong to one small subgroup within Malayic; on the contrary, they exhibit a considerable degree of structural diversity, spanning much if not all of the diversity evident within Malayic as a whole. Some illustrations of this diversity, from phonology, morphosyntax and the lexicon respectively, are presented in (1)–(3) below:

(1) Phonology
   a. word-final Proto-Malayic *[a], as in *[mata] ‘eye’:
      (i) within Riau province
Commentary on McWhorter: David Gil

(a) [mata] Pulau Padang Orang Asli
   [mato] Siak Malay
   [matɔ] insular Riau Malay

(ii) elsewhere
   [mata] Irian Indonesian
   [mato] Kelantan Malay
   [matɔ] Urban Peninsular Malay

b. word-final Proto-Malayic nasal stop, as in *[makan] ‘eat’
   (i) within Riau province
      nasal [makan] Siak Malay
      pre-oralized [makatn] Pulau Padang Orang Asli
      oral [makat] Sakai
   (ii) elsewhere
      nasal [makan] Jakarta Indonesian
      pre-oralized [makatn] Selako (Hudson 1970)
      oral [makat] Urak Lawoi (Hogan 1988)

(2) Morphosyntax
a. possessives
   (i) within Riau province
      postnominal bare Siak Malay
      prenominal with reduced punya Outsider Malay
   (ii) elsewhere
      postnominal bare Jakarta Indonesian
      prenominal with reduced punya Irian Indonesian

b. causatives
   (i) within Riau province
      periphrastic Outsider Malay
      with *-kan Siak Malay
      with *-Vn Bangkinang
   (ii) elsewhere
      periphrastic Irian Indonesian
      with *-kan Cocos Malay (Adelaar 1996)
      with *-Vn Jakarta Indonesian

(3) Lexicon
a. ‘give’
   (i) within Riau province
      beri Siak Malay
      agiah Bangkinang
      kasi Outsider Malay
Examples (1)–(3) make reference to the following Malayic isolects of Riau province. Pulau Padang Orang Asli is the non-mainstream isolect of the Orang Asli, literally “original people”, of Pulau Padang, an island just off the east coast of Sumatra, opposite the mouth of the Siak river; these people are also occasionally referred to as the Orang Hutan, literally “forest people”, and are possibly akin to the Orang Hutan described by Kähler (1960) on other nearby islands. Siak Malay is the mainstream isolect of the Malays living along the banks of the Siak river, where a small but influential sultanate was based, and in neighbouring regions, including the small town of Sungai Pakning mentioned previously; to the best of my knowledge there are no previous descriptions of this dialect. Insular Riau Malay is the mainstream isolect of the Malays living in the Riau archipelago, a cluster of islands off the tip of the Malay peninsula, which, during the previous century, constituted the cultural capital of the Malay world: although this dialect is renowned as having formed the basis for the standardization of the official languages of Malaysia and Indonesia, the available grammatical descriptions are mostly obscure (H. Idrus et al. 1993, Kailani 1994, Kailani et al. 1985, Kailani et al. 1983, Saidat et al. 1991, and Saidat et al. 1986). Sakai is the non-mainstream isolect of the Orang Sakai, who, until very recently, were nomadic forest dwellers in the region between Pekanbaru, the provincial capital, and the port city of Dumai; the only linguistic material that I am familiar with is a dictionary which does not reflect many of the distinctive properties of the language (A. Gani et al. 1985). Outsider Malay is a generic term for the isolects spoken by and to Chinese inhabitants throughout the province; impressionistically, these varieties seem to differ relatively little from place to place, though they have hardly been studied in any detail. Finally, Bangkinang is a mainstream isolect used by residents of the town bearing the same name, situated in the western part of the province, on the main road leading from Riau to West Sumatra; again, I am aware of no previous descriptions of this variety in the linguistic literature.
Commentary on McWhorter: David Gil

Example (1a) illustrates what is perhaps the “flagship” feature of phonological variation in Malay/Indonesian dialects, the different realizations of word-final *-a in Proto-Malayic; as shown above, the patterns of variation within Riau mimic those of other regions of the archipelago. Example (1b) involves a substratum phenomenon characteristic of many Austronesian languages of Borneo and also Mon-Khmer languages of peninsular Malaysia, namely the partial or complete oralization of word-final nasal stops (Blust 1997); within Riau, this process occurs among the tribal varieties of Malay. Example (2a) shows the two most common strategies for the expression of attributive possession, postnominal bare, as in Siak Malay buku Rudy ‘Rudy’s book’, or prenominal with a reduced form of the possessive marker punya, as in Outsider Malay Rudy mia buku, with the same meaning; similar constructions are attested throughout the region. Example (2b) illustrates the three strategies that are available for expressing causativity: periphrasis as in Outsider Malay kasi mati ‘give die’, or ‘kill’, ‘extinguish’; affixation with -kan as in Siak Malay matikan; and affixation with -Vn as in Bangkinang motiin, both with similar meaning – again, these three strategies are also available in other parts of the Malay/Indonesian speaking region. And examples (3a) and (3b) provide two instances of lexical variation within Riau that are mirrored by the occurrence of cognate forms with the same meanings in other areas. Many more such examples, from all realms of grammar, can be easily adduced. Admittedly, some of the linguistic variation within Riau may be due to recent borrowing from other dialects spoken outside the province. And some other patterns may be the product of parallel independent innovations (see, for example, Tadmor 2001, on Proto-Malayic *-a). However, not all of the linguistic diversity within Riau can be accounted for in this way; much of it probably reflects the time-depth of the Malayic language family, and the results of ancient migrations into (and presumably also out of) Riau province. Thus, it may be safely concluded that present-day Riau province was the venue of substantial language contact over much of the last 2,000 years. Accordingly, various contact varieties of Malayic must have arisen during this lengthy period. And such contact varieties constitute plausible ancestors for what is now Riau Indonesian.

This, then, is the best guess that can be hazarded for where Riau Indonesian comes from: on the basis of the arguments presented above, it is probably most appropriately viewed as the descendant of a long and uninterrupted line of contact languages, spoken in the region throughout the last thousand years and maybe even longer.

On the other hand, there would seem to be little or no reason to characterize Riau Indonesian as a creole language. The logic is basically that of Occam’s Razor. Creoles, by definition, are the product of abrupt linguistic restructuring, such as the native-language acquisition of a pidgin. Such rapid linguistic changes result from very specific kinds of language contact, involv-
ing languages with little or no mutual intelligibility, emerging from specific configurations of political, economic, and social circumstances, prototypically exemplified by sailors on a boat or workers in a colonial plantation. There is, quite simply, no positive evidence whatsoever for any of this in the history of Riau Indonesian. Accordingly, in the absence of such evidence, there is no justification for characterizing Riau Indonesian as a creole.

One objection that might reasonably be raised at this point is that the history of Riau Indonesian need not be restricted to the region in which it is currently spoken: conceivably, Riau Indonesian could be descended from a creole language which came to east-central Sumatra from some other part of the archipelago where it originated in some local pidgin. This scenario gains in plausibility when it is recalled that a variety of colloquial Malay, sometimes referred to as “Bazaar Malay”, played the role of a trade language throughout the pre-colonial era – see, for example, Reid (1988), Adelaar & Prentice (1996), Collins (1996), and many others. If this trade variety of Malay were a pidgin or a creole, and if Riau Indonesian were descended from it, then Riau Indonesian would also qualify to be labelled as a creole. In fact, however, the first if is unclear, and the second one even more unlikely.

A number of scholars have argued that certain varieties of Malay, descended from this trade language, deserve to be characterized as creoles. Notably, however, most of these arguments seem to depend more on intrinsic synchronic features than on extrinsic historical circumstances. For example, with reference to the language of the Straits-born Chinese of Malacca and Singapore, Ansaldo & Matthews (1999) observe that “[Baba Malay] has many of the typological properties expected of a language which has undergone creolization. It exhibits the gross typological characteristics of creoles such as isolating morphology, SVO word order and preverbal TMA particles”. Of course these characteristics are also exhibited by the local Malay dialects as well as by the Minnan Chinese languages which contributed to the lexicon of Baba Malay – none of which are presumably creoles.

Perhaps the most extensive proposal along these lines is that of Adelaar & Prentice (1996), who coin the term “Pidgin Malay Derived” (or “PMD”) dialect, and apply it to a wide range of Malay/Indonesian isolects, including the “Bazaar Malay” of Malaysia and Singapore, various forms of Java Malay, and East Indonesian varieties such as those of Ambon, the North Moluccas, Manado, Bacan, Irian, and Kupang. According to them, these isolects are all derived from the pre-colonial trade variety of Malay; however, they do not provide any explicit arguments in support of their characterization of the original trade language, or “Bazaar Malay”, as a pidgin. In the absence of any such arguments, it is hard to evaluate the appropriateness of the term “Pidgin Malay Derived” dialect and of the resulting characterization of such dialects as creoles.
Table 1. Adelaar & Prentice’s “Pidgin Malay Derived” features and Riau Indonesian

<table>
<thead>
<tr>
<th>characteristic “Pidgin Malay Derived” features (Adelaar &amp; Prentice 1996)</th>
<th>Riau Indonesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 possessive constructions consisting of possessor + *punya + possessed item</td>
<td>questionably</td>
</tr>
<tr>
<td>2 plural pronouns derived from singular pronouns + *orang (‘human being’)</td>
<td>marginal</td>
</tr>
<tr>
<td>3 retention of *ter- and *ber- as the only productive original Malayic affixes</td>
<td>no</td>
</tr>
<tr>
<td>4 *ada, the Malay existential marker, indicating progressive aspect</td>
<td>no</td>
</tr>
<tr>
<td>5 reduced forms of the demonstratives *ini and *itu preceding a noun and functioning as determiners</td>
<td>no</td>
</tr>
<tr>
<td>6 the use of a reduced form of *pergi ‘to go’ as a verb as well as a preposition meaning ‘towards’</td>
<td>no</td>
</tr>
<tr>
<td>7 causative constructions consisting of the auxiliaries *kasi/*beri (‘to give’) or *bikin/*buat (‘to make’) + the head verb</td>
<td>occasionally</td>
</tr>
<tr>
<td>8 the use of *sama or another word as a multifunctional preposition (also for direct and indirect objects)</td>
<td>yes</td>
</tr>
</tbody>
</table>

Instead, they provide a list of eight structural features which they claim are characteristic of PMD dialects. In their words, these features “are not diagnostic in themselves for being PMD-derived, but large configurations of them are” (p. 675). These eight features are reproduced in Table 1.

Regardless of whether Adelaar & Prentice are justified in associating these eight features with PMD dialects, it is instructive to examine the extent to which these features are applicable to Riau Indonesian. As suggested by the rightmost column in Table 1, they are not, on the whole, very applicable. Let us briefly examine each of these features in turn.

[1] In Riau Indonesian, it is indeed possible to form expressions such as Rudy punya buku, meaning ‘Rudy’s book’. However, unlike in other isolects, e.g., Outsider Malay, exemplified in the discussion of (2a) above, this is not a dedicated possessive construction. Rather, it is completely parallel to an open set of other expressions, such as Rudy beli buku ‘book that Rudy bought’, Rudy baca buku ‘book that Rudy read’, and so forth. Thus, a more perspicuous gloss of Rudy punya buku would be ‘book that Rudy owns’; it is thus an instance of an apparent right-headed or internally-headed relative clause of the type discussed in detail in Gil (1994, 2000a). What clinches the argument that it is not a grammaticalized possessive construction is the fact that, unlike in Outsider Malay and most or all of Adelaar & Prentice’s PMDs, punya does not undergo
phonological contraction – it is just a regular lexical item, meaning ‘have’ or ‘own’.

[2] In general, this does not occur in Riau Indonesian. The one exception that I have encountered is in the variety spoken in the city of Tanjung Pinang, where *dia orang*, literally ‘3 person’, is occasionally used as the 3rd person plural pronoun; interestingly, this construction does not undergo reduction, as it typically does in eastern Indonesian dialects, thereby suggesting that it is a relatively recent innovation or borrowing.

[3] In Riau Indonesian, *ter-* and *ber-* are in fact considerably less productive than the other two prefixes in the paradigm, *N-* and *di-*; see Gil (1999, to appear d) for detailed discussion and analysis.

[4] In Riau Indonesian *ada* can occur in sentences associated with any aspect whatsoever.

[5] In Riau Indonesian, the demonstratives *ini* and *itu* occasionally do surface as *ni* and *tu*, however they only very rarely occur as prenominal determiners.

[6] In Riau Indonesian, *pergi* is never reduced, and it never means ‘towards’.

[7] The most common causative construction in Riau Indonesian involves the affixation of *-kan*, just as in the Siak Malay example in the discussion of (2b) above; however, considerably less frequently, *kasi* ‘give’ is used to form a periphrastic causative, as in the Outsider Malay example also given above.

[8] In Riau Indonesian, *sama* is indeed associated with a wide variety of functions, which include the marking of comitatives, various other obliques, agents, but not patients, as seems to be suggested by Adelaar & Prentice for PMD dialects. In addition, *sama* often occurs as the translational equivalent of the conjunction ‘and’ and the adjective ‘same’.

Summing up, of the eight features suggested by Adelaar & Prentice to be characteristic of PMD dialects, only one is incontestably characteristic of Riau Indonesian. One other feature is of partial applicability, two others are dubious, and the remaining four are completely inapplicable. Thus, Riau Indonesian does not fit Adelaar & Prentice’s profile of a PMD dialect. Accordingly, to the extent that these eight features are indicative of a common origin in some trade variety of Malay, or “Bazaar Malay”, then Riau Indonesian does not seem to be derived from this putative common language. In particular, even if Adelaar & Prentice are right in characterizing this trade variety of Malay as a pidgin, which, as suggested above, is not at all clear, they would still not be able to invoke these features in order to characterize Riau Indonesian as a creole language.\(^2\)

---

\(^2\) In fact, preliminary work suggests that of the Malay/Indonesian isoelects spoken in Riau Indonesian, it is Outsider Malay that most closely fits Adelaar & Prentice’s profile of a PMD dialect.
Commentary on McWhorter: David Gil

What we have seen in this section, then, is that although there is scant direct evidence pertaining to the linguistic history of east-central Sumatra, certain plausible inferences can be made with regard to the ancestry of Riau Indonesian – and what these inferences suggest is that there is little or no reason to characterize Riau Indonesian as having the history of a creole language. Thus, Riau Indonesian constitutes a bona fide example of an old language, which may accordingly be contrasted with a creole language, in order to put to the test McWhorter’s hypotheses regarding the correlation between creoles and complexity.

3. Riau Indonesian and Saramaccan: A contrastive analysis

McWhorter proposes, in Section 2.4.3, four general “diagnostics of grammatical complexity” which, in conjunction, provide a characterization of the overall complexity of a language. These four measures are reproduced in slightly paraphrased form in (4):

(4) McWhorter’s four diagnostics of grammatical complexity
   a. Phonology: A language is more complex than another to the extent that its phonology has more marked members.
   b. Syntax: A language is more complex than another to the extent that its syntax requires the processing of more rules.
   c. Grammaticalization: A language is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions.
   d. Morphology: A language is more complex than another to the extent that it has inflectional morphology.

McWhorter describes these four general diagnostics as “chosen to arouse the least possible controversy from as wide a spectrum as possible of linguists” (Section 2.4.3). He then puts them to immediate practical use, in contrastive analyses of Saramaccan, his stock example of a creole language, and two selected non-creole languages: Tsez, a Daghestanian language of the Caucasus, exhibiting many of the grammatical intricacies characteristic of languages from that part of the world, and Lahu, a Tibeto-Burman language of Southeast Asia, a typical exemplar of the allegedly simpler isolating language type. These contrastive analyses show, rather convincingly, that both Tsez and Lahu are of greater overall grammatical complexity than Saramaccan. On the basis of these analyses, McWhorter concludes that, in general, older languages are of greater overall complexity than creole languages.

However, if Saramaccan is contrasted not with Tsez or Lahu but rather with Riau Indonesian, a very different picture emerges. As we shall see below, McWhorter’s diagnostics characterize Riau Indonesian as being of lesser over-
all complexity than Saramaccan. Somewhat more specifically, Riau Indonesian emerges as significantly less complex than Saramaccan with respect to the first two diagnostics, phonology and syntax, but of roughly the same degree of complexity as Saramaccan with respect to the latter two diagnostics, grammaticalization and morphology. In conjunction, then, McWhorter’s four diagnostics may be said to characterize Riau Indonesian as being of lesser overall complexity than Saramaccan.

Before embarking on the comparison, however, a few words about methodology are in order. One problem, acknowledged and discussed in detail by McWhorter, is that each of these four diagnostics constitutes in effect an open-ended family of more specific diagnostics, involving individual linguistic features. This raises the possibility that the analyst may, if not consciously than inadvertently, “stack the deck”, choosing those features which support his or her hypothesis, while ignoring those that go against it. For example, McWhorter could have focussed in on those features with respect to which Saramaccan is simpler than Tsez or Lahu, while downplaying other features which might have contributed to a characterization of Saramaccan as more complex than those two languages. Similarly, in the comparison that follows, I had to be wary not to seek out those features with respect to which Riau Indonesian is simpler than Saramaccan, while ignoring other features which might have painted a different picture. I can think of no hard and fast way to avoid this methodological trap other than to simply be as honest as one can with respect to the available facts.

A second problem is that in many cases, the analyst will be much more familiar with one of the two languages than with the other, and must rely on secondary sources to obtain information with regard to the less familiar language. Thus, McWhorter is more familiar with Saramaccan than with Tsez or Lahu, while I know much more about Riau Indonesian, which I’ve worked on for many years, than about Saramaccan. In the contrastive analysis that follows, for Saramaccan, I have made use of the materials that were available to me, Bakker, Smith, & Veenstra (1995), Veenstra (1996), and of course McWhorter (this volume), supplemented with further helpful information kindly provided by McWhorter. Nevertheless, the comparison still suffers from my considerably lesser familiarity with Saramaccan. Though, if one may try to make a virtue out of a vice, the limitations on the Saramaccan materials did provide me with a partial solution for the “stacking-the-deck” problem mentioned in the previous paragraph: with less available for Saramaccan, I came close to simply taking everything I had.

A third problem, emerging out of the second, is that the languages being contrasted are sometimes described within different descriptive traditions, significantly skewing various measures of complexity. For example, as mentioned above, one of my few sources for Saramaccan was Veenstra (1996), who de-
scribes its syntax within the minimalist framework, which, with respect to its proposed inventories of syntactic categories, is a misnomer if ever there was one. Thus, if I were about to contrast the inventory of syntactic categories of Riau Indonesian and Saramaccan, I could easily say, well, Riau Indonesian, as per Gil (1994, 2000b), has just one open category and one closed one, while Saramaccan has categories galore – NP, VPs, functional projections and whatnot – in accordance with Veenstra (1996). But this is quite obviously a difference between descriptive frameworks, not between languages: if I were to describe Saramaccan and Veenstra to turn his attention to Riau Indonesian, the picture would be very different, perhaps almost the mirror image. Thus, in order to conduct the comparison fairly, it is necessary to abstract away from such differences, in order to try to get at the languages behind the descriptions.

Nevertheless, in spite of the above problems, the above four diagnostics provide a practical tool which, as McWhorter shows, makes it possible to arrive at substantive and, one might also add, intuitively appealing, characterizations of the relative overall complexity of the grammatical systems of different languages. So let us now turn to the task at hand, and apply the four diagnostics to a contrastive analysis of Riau Indonesian and Saramaccan.

The first diagnostic, in (4a), suggests that a language is more complex than another to the extent that its phonology has more marked members, where markedness is defined not in articulatory but rather in implicational or distributional terms – marked items are those whose presence entails the presence of other, less marked items, but not vice versa. The results of comparing Riau Indonesian and Saramaccan with respect to their inventories of marked phonological items are summarized in Table 2.\footnote{In Table 2 and subsequent tables, the margin with which one of the two languages emerges as more complex than the other with respect to each property is indicated, iconically, with "plus" symbols: one plus for "somewhat more complex"; two plusses for "substantially more complex".}

The phonemic inventory of Riau Indonesian is presented in Table 3. Segments in parentheses are those which are marginal in one or more of the following ways: (a) their status as independent phonemes, as opposed to allophones, is questionable; (b) they occur only in some varieties of Riau Indonesian; (c) they occur only in words that have been borrowed from other languages.

As evident in Table 3, the phonemic inventory of Riau Indonesian is very simple. The only sounds which may be characterized on universal grounds as marked are some of the ones that occur in parentheses: these are the prenasalized stops and the high-mid vowels.

The first set of marked segments in Riau Indonesian are the prenasalized stops, [mʰp], [mʰb], [nʰt], [nʰd], [nʰc], [nʰj], [nʰk], and [nʰg]. The possible occurrence
Table 2. Riau Indonesian and Saramaccan: Phonology

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Riau Indonesian</th>
<th>Saramaccan</th>
<th>More Complex Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>prenasalized stops</td>
<td>possibly</td>
<td>yes</td>
<td>Saramaccan</td>
</tr>
<tr>
<td>vowel height distinctions</td>
<td>three, plus</td>
<td>four</td>
<td>Saramaccan</td>
</tr>
<tr>
<td></td>
<td>possible fourth</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>vowel nasality distinctions</td>
<td>none</td>
<td>yes</td>
<td>Saramaccan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>vowel length distinctions</td>
<td>none</td>
<td>three lengths</td>
<td>Saramaccan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>lexical tone</td>
<td>none</td>
<td>two or three</td>
<td>Saramaccan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>++</td>
<td>tones;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tone sandhi</td>
</tr>
</tbody>
</table>

Table 3. Phonemic inventory of Riau Indonesian

<table>
<thead>
<tr>
<th>p</th>
<th>t</th>
<th>c</th>
<th>k</th>
<th>(?,)</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>d</td>
<td>j</td>
<td>g</td>
<td>i</td>
</tr>
<tr>
<td>(m)p</td>
<td>(n)</td>
<td>(p)c</td>
<td>(p)k</td>
<td>(p)d</td>
</tr>
<tr>
<td>(m)b</td>
<td>(n)</td>
<td>(p)j</td>
<td>(p)g</td>
<td>(p)</td>
</tr>
<tr>
<td>(f)</td>
<td>s</td>
<td>(x)</td>
<td>h</td>
<td>(a)</td>
</tr>
<tr>
<td>m</td>
<td>n</td>
<td>m</td>
<td>ŋ</td>
<td>t</td>
</tr>
<tr>
<td>r</td>
<td>l</td>
<td></td>
<td>ɻ</td>
<td>a</td>
</tr>
<tr>
<td>w</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

of prenasalized stops divides Riau Indonesian into two subdialects, which we shall label arbitrarily as A and B. (In general, it would appear to be the case that dialect A is associated more with ethnic Malay speakers, or areas in which Malay dialects are spoken, while dialect B is associated more with ethnic Minangkabau speakers, or regions in which Minangkabau is widespread as a lingua franca.) Whereas dialect A provides no evidence for the existence of prenasalized stops, there is some reason to believe that prenasalized stops do occur in the B dialect. This evidence derives from the ways in which words containing the relevant sequences of sounds are syllabified. Consider a word containing the sequence [... VNCV ...], where V stands for vowel, N for nasal stop, and C for oral stop. Whereas in dialect A the syllable boundary, denoted by $, falls between the nasal and the oral stops, [... VN $ CV ...], in dialect B the syl-
lable boundary falls before the nasal stop, [... V $ NCV ...]. 4 One way, though not the only way, to account for this difference between the two dialects, is to analyze the sequence [...NC...] as constituting a single complex segment, that is to say a prenasalized stop, in dialect B but not dialect A – its unitary nature blocking the insertion of a syllable boundary. Thus, it is possible that more highly marked prenasalized stops occur in one subdialect of Riau Indonesian.

Turning now to Saramaccan, we find a series of four prenasalized stops, [mb], [nd], [ndj], and [ng] (Bakker, Smith, & Veenstra 1995: 170). Although these are actually two less in number than in the B dialect of Riau Indonesian, there is no suggestion in the available sources that these segments are marginal in any way in Saramaccan. Thus, with respect to prenasalized stops, it may be concluded that there is some reason, though not overwhelming, to characterize Saramaccan as more complex than Riau Indonesian.

The second set of marked segments in Riau Indonesian are the high-mid vowels [e] and [o], but here the facts are less clear cut, with lots of variation between dialects, between speakers, and even within individual speakers. In some cases, [e] and [o] appear to be in free variation with the corresponding high vowels [i] and [u]. In other cases, though, variation seems to be governed by the phonological environment, suggesting that [e] and [o] may be allophonic variants of [i] and [u]. For example, for some speakers, in a [CVCVC] environment, there appears to be a rule of vowel harmony, in which the quality of the second vowel depends on that of the first. Specifically, if the first vowel is high, the second vowel will also be high, e.g., [sikit] ‘a little’, [busu?] ‘stink’, whereas if the first vowel is [a], the second vowel will be high-mid, e.g., [saket] ‘hurt’, [maso?] ‘enter’. However, even for such speakers, there exists a small residue of forms in which the quality of the vowel seems not to be predictable by the phonological environment, thereby suggesting that [e] and [o] are on their way to acquiring the status of individual phonemes. But the judgements are often shaky, and I have not yet been able to come up with robust minimal pairs. Moving on to Saramaccan, here there are high-mid vowels [e] and [o] which contrast phonemically with high vowels [i] and [u] and also with low-mid vowels [ɛ] and [ɔ] (Bakker, Smith, & Veenstra 1995: 170). Nevertheless, there are “very few minimal pairs” for high-mid and low-mid vowels, and the low-mid vowels occur very infrequently (McWhorter, personal communica-

4. Evidence for these different syllabifications comes from three independent sources, which can be illustrated with respect to the proper noun [anton]. (a) When asked to “break the word into two halves”, speakers of dialect A produce [an-ton] while speakers of dialect B produce [a-nton]. (b) When applying a rule of truncation, creating a monosyllabic vocative form from names, titles and kinship terms by omitting the “weak” syllable(s) and retaining the “strong” one, speakers of dialect A produce [ton], while speakers of dialect B produce [nton]. (c) When speaking in a ludling which reverses the order of syllables within a word, speakers of dialect A change [anton] to [tonan], while speakers of dialect B change [anton] to [ntona].
tion). Thus, once again, it may be concluded that Saramaccan is somewhat, though not massively, more complex than Riau Indonesian.

Whereas each of the two would-be marked sets of segments in Riau Indonesian is matched by a similar and more robust set of segments in Saramaccan, Saramaccan also boasts three additional kinds of marked items which have no counterpart whatsoever in Riau Indonesian (Bakker, Smith, & Veenstra 1995: 170–171). The first of these is vowel nasality: while Riau Indonesian has no distinction between oral and nasal vowels, in Saramaccan each of the seven oral vowels contrasts with a phonemically distinct nasalized counterpart. The second is vowel length: whereas Riau Indonesian has no vowel length distinctions, Saramaccan exhibits a crosslinguistically rare distinction between short, long, and extra-long vowels. And the third is lexical tone: while Riau Indonesian is not a tonal language, Saramaccan has two phonemically distinct lexical tones, high and low, which are argued to result from an underlying three-way distinction between high, low, and unmarked, to which rules of tone sandhi apply (Bakker, Smith, & Veenstra 1995: 104, 170–171). Thus, the comparisons of marked segment inventories, as summarized in Table 2, point clearly towards the conclusion that Riau Indonesian is of lesser overall phonological complexity than Saramaccan.

Moving on to other structural domains, the second diagnostic, in (4b), states that a language is more complex than another to the extent that its syntax requires the processing of more rules. A comparison of the relative syntactic complexity of Riau Indonesian and Saramaccan is provided in Table 4.

In Riau Indonesian, expressions denoting activities, such as pergi ‘go’, kejar ‘chase’, kasi ‘give’, and so forth, can stand on their own as complete non-elliptical sentences; there is no requirement that the participants in the activity be overtly expressed. (In current parlance, Riau Indonesian is a “pro-drop” language, though elsewhere, see Gil 2000b, I have argued that in such constructions there are actually no structural positions occupied by phonologically-null elements.) In contrast, in Saramaccan, the participants must be overtly expressed, with either a pronoun or a complete NP (McWhorter, personal communication). By dint of this requirement, Saramaccan is more complex, as it has a specific constraint that is absent from Riau Indonesian.

5. With regard to vowel length, McWhorter (personal communication) argues that the would-be long and extra-long vowels are more appropriately analyzed as sequences of distinct vowels. To the extent that this is true, this would weaken the claim that Saramaccan is more complex than Riau Indonesian with respect to the property in question, but not negate it entirely: Saramaccan would still be more complex by dint of the existence of certain particular sequences of identical vowels which are absent from Riau Indonesian, as well as from many other languages – the latter suggesting that such sequences should be viewed as highly marked.
Word order in Riau Indonesian is free: in particular, in a sentence containing an activity and two participants, any of the six logically possible orders of the three expressions constitutes a grammatical sentence (Gil 1994, 1999, to appear d). In contrast, Saramaccan has rigid SVO word order (Bakker, Smith, & Veenstra 1995: 175, Veenstra 1996: 12). Thus, with respect to word order,
Saramaccan is also more complex, as it has a specific constraint which Riau Indonesian lacks.

Question words in Riau Indonesian, such as *apa* ‘what’, *mana* ‘which, where’, may occur in any position that the corresponding non-interrogative expression might also occur. (To use generative jargon, WH words may occur either in situ or fronted – though “fronted” in this case is a misnomer, since sentence-initial WH words are also in situ.) In contrast, in Saramaccan, question words are obligatorily fronted (McWhorter, this volume). Once again, Saramaccan proves to be more complex, by dint of an ordering rule that is lacking in Riau Indonesian.

Riau Indonesian has no grammatical elements bearing any kind of resemblance to the category of copula. In contrast, Saramaccan has not one but two copulas, *de* and *da*, with quite different grammatical properties (Bakker, Smith, & Veenstra 1995: 175–177). Whereas *de* has verbal status, can take TMA markings, and can occur with NP-, PP-, and AP-complements, *da* is of pronominal nature, cannot take TMA markings, and can occur only with NP-complements, not with PP- or AP-complements. Thus, in this aspect of syntactic organization, Saramaccan is, once more, of greater complexity than Riau Indonesian.

A similar picture is presented by complementizers, elements that – to use theory-neutral terminology – serve to introduce themes of expressions referring to acts of cognition, perception, and speech. Whereas Riau Indonesian has no such item, Saramaccan has the complementizer *táa*, a grammaticalized form of the verb ‘say’ (Veenstra 1996: 154–157). Yet again, Saramaccan is more syntactically complex than Riau Indonesian.

Often, NP-internal syntactic structure is more articulated than its clausal counterpart, so here would be a good place to go into the quest for at least some syntactic complexity in Riau Indonesian. One parameter along which languages may vary is whether a property-denoting expression can stand by itself as a referring expression, or whether it needs to be “supported” by an additional grammatical marker, a proform referring to the attributive theme, before it can assume that function. In Riau Indonesian, expressions denoting properties may indeed stand on their own as referring expressions; for example *bagus* ‘good’ can mean either ‘good’ or ‘good one’. In contrast, in Saramaccan, adjectives cannot head NPs on their own; in order to do so, an additional marker is required, in fact the same *one* as in English, the proform or attributive theme marker *wan*, for example *bun wan* ‘good one’ (McWhorter, personal communication). Thus, whereas in Riau Indonesian property-denoting expressions such as *bagus* can function in a variety of different ways, in Saramaccan, the corresponding forms are more constrained in their syntactic distribution, thereby providing another reflection of the greater syntactic complexity of Saramaccan.
Another measure of NP-internal complexity is provided by the extent to which languages distinguish between various kinds of nominal attribution, corresponding to English possessives, adjectives, and relatives. In Riau Indonesian such differentiation is almost completely absent. In a construction of the form HEAD ATTRIBUTE, with no overt grammatical construction marker of any kind, the attribute may denote a possessor, as in *buku Rudy ‘Rudy’s book’*; a property, as in *buku bagus ‘good book’*; or an event, as in *buku Rudy beli ‘book that Rudy bought’*. Whereas the English translations of these expressions instantiate three distinct construction types in English, namely possessive, adjectival, and relative clause respectively, the three Riau Indonesian expressions belong to one and the same construction type: Riau Indonesian has no dedicated constructions for possessive, adjectival, or relative clause attribution. In contrast, in Saramaccan, there is at least some measure of differentiation in this functional domain. In Saramaccan, there is a prenominal ATTRIBUTE HEAD construction in which the attribute may denote either a possessor or a property – but not an event. For this latter function, a different postnominal construction is required, of the form HEAD MARKER ATTRIBUTE, in other words, a dedicated relative clause, formed with a relative pronoun. Thus, nominal attribution is more highly differentiated in Saramaccan than in Riau Indonesian, thereby contributing even further to the greater syntactic complexity of Saramaccan.

A somewhat different strategy for contrasting the syntax of Riau Indonesian and Saramaccan is to compare various items, or sets of items, which, although superficially analogous to each other, may be seen under closer inspection to behave in very distinct ways which are revealing of the different underlying ground plans of the two languages. In such cases, the consistent pattern is that in Riau Indonesian the items’ functions are purely pragmatic or semantic whereas in Saramaccan they are at least partially formal, that is to say syntactic. Such items thereby provide yet additional support for the greater syntactic complexity of Saramaccan.6

One obvious set of items are the pronouns. In Riau Indonesian, the choice of pronouns is governed by an array of politeness considerations. For example, for 2nd person reference, the speaker can choose between *kau* or its variant

---

6. The reader may wonder whether, in cases like these, the greater syntactic complexity of Saramaccan is not being compensated for by a greater pragmatic or semantic complexity in Riau Indonesian. The answer is negative, for the following reason. Given that the essential nature of language is a system of correspondences between sounds and meanings, the more direct the correspondence between the two, the simpler the overall structure of the language. Syntax, by definition, is one of several arbitrary way stations on the road from sound to meaning, and as such contributes inherently to complexity. Thus, a generalization that makes reference to syntactic elements is more complex than an alternative generalization requiring recourse to elements of meaning, since meanings are there anyway, by the very nature of language, and thus come at no additional cost.
enkau (familiar, or impolite and condescending), kamu (neutral and rather impersonal), anda (very formal), and several other forms borrowed from other Malayic isolects. Or the speaker can use a kinship term, reflecting the relative ages of the speaker and addressee as well as the addressee’s gender. Or the speaker can use the name of the addressee. Or – as is often the case – the speaker can give up in despair at having to commit to such an explicit formulation of interpersonal relationships, and make use of the very convenient additional option of leaving the addressee completely unexpressed. The use of pronouns is thus perhaps one of the most complex features of Riau Indonesian; however, such complexity clearly lies outside of what may reasonably be subsumed under syntax.7 Not so Saramaccan. In Saramaccan the speaker also faces certain choices, but here the choices are a matter not of politeness but rather of grammar proper. Following Veenstra (1996: 30–40), there are two pronoun series, weak and strong. Whereas weak pronouns only occur in subject position, strong pronouns may occur in object position and as possessive pronouns. In addition, when strong pronouns occur in subject position, they are interpreted as emphatic, and can be associated with high tone, or reduplicated. When pronouns are followed by the negative marker, their form changes, and the weak/strong distinction is neutralized. After devoting ten pages to a description of the pronominal system, he concludes, justifiably, that “[t]he pronominal system of Saramaccan is rather elaborated and quite complex”. And in fact, there is even more complexity than that: as pointed out by McWhorter (1997: 98–99), Saramaccan also has “quirky case-marking”, with the subject pronoun of a non-verbal identificational predicate obligatorily taking the strong form. So both Riau Indonesian and Saramaccan have complicated pronoun systems; but what is crucial to us here is the nature of the complexity. Whereas for Riau Indonesian it is pragmatic, for Saramaccan it is clearly hard-core syntactic, making reference to grammatical relations such as subject and object, grammatical constructions such as non-verbal predication, as well as the negation operator and the relationship of identification. Thus, a comparison of the pronouns contributes further to the characterization of Saramaccan as having more complex syntax.

7. In addition to the above, it should be acknowledged that a few speakers of Riau Indonesian make infrequent use of an alternative series of pronouns whose distribution is governed by more hard-core grammatical factors; these are the agentive proclitic forms ku- for 1st person and kau- for 2nd person. However, these forms are clearly marginal in Riau Indonesian, and are probably more appropriately considered as reflecting “seepage” from Standard Indonesian, where they occur much more frequently.
A second grammatical item highlighting the differences between the two languages is reduplication. In both languages, reduplication is used frequently and productively; however, the functions of reduplication in the two languages differ in a revealing way. In Riau Indonesian, reduplication, either partial or, most frequently, complete, is associated with a wide variety of functions, including the expression of distributivity, iterativity, large number, amount, or extent, universal quantification, negative polarity, concessivity, atelicity and more. For example, from *sapu* ‘sweep’, ‘broom’, one can derive *sa-sapu* or *sapu-sapu*, some of whose many meanings include ‘sweep repeatedly’, ‘sweep lots of places separately’, ‘sweep casually, without actually cleaning’, ‘lots of brooms’, ‘every broom’, ‘brooms all over the place’, and so forth. Crucially, all of these functions are semantic in nature: reduplication has meanings, in fact lots of different ones, interrelated in complex ways, but it does not impinge in any respect on the syntax. Turning to Saramaccan, here reduplication also has some semantic functions, including the expression of approximation, repetition, and distributivity (McWhorter, personal communication). However, in addition, it has two well-defined syntactic functions: converting transitive verbs to attributive adjectives and resultatives, and converting verbs into nouns. For example, from the verb *síbi* ‘sweep’ one may derive the noun *sísíbi* ‘broom’ (McWhorter, this volume, citing Bakker 1987: 21). Thus, the different functions of reduplication in the two languages follow the same pattern, reflecting yet again the greater syntactic complexity of Saramaccan.

A similar contrast is displayed by alternative forms available for the expression of negation in the two languages. In Riau Indonesian there is a choice between a set of four free variants, *tak, ndak, nggak*, and *tidak*, and a distinct fifth form with different usage, *bukan*. Given a structure of the form (X) NEG-Y, the choice of negative marker depends on the semantic relationship between X and Y. If Y denotes an activity and X is one of its participants, or if Y denotes a property and X is its theme, then one of the four variants, *tak, ndak, nggak*, or *tidak* will be chosen. Conversely, if the relation between X and Y is equational, then *bukan* will be selected. As a result, *tak, ndak, nggak*, or *tidak* typically negate expressions whose translations into English involve verbs or adjectives, while *bukan* typically negates expressions which correspond to English nouns. Typically, but not always, as shown by the following examples from a corpus of spontaneous speech specimens:

(5)  

```
Ndak Royal do?
NEG Royal NEG.POL
[Speaker asks me where I’m staying; I say ‘City Hotel’; speaker asks about the bigger hotel down the road]
‘Aren’t you staying at the Royal?’
```
Creoles, complexity, and Riau Indonesian

(6) Bukan buka
NEG open

[Speaker rummaging through interlocutor’s belongings, picks up a bag; interlocutor tells him not to open it; speaker responds]

‘What I’m doing isn’t opening it’

In the context of (5), Royal, the name of a hotel, actually denotes an activity associated with the hotel, namely, staying at it, and the (covert) addressee is a participant in the activity; hence one of the four free-variant negative forms, ndak, is chosen. Conversely, in (6), the speaker had a choice of two different perspectives within which to cast his message. If he had chosen one of the four variants, tak, ndak, nggak, or tidak, he would have been explicitly negating his own participation in the activity; a more felicitous translation would have then been ‘I’m not opening it’. However, in the given context, the speaker chose to downplay his own involvement in the activity, and to focus, instead on the nature of the activity itself, in order to deny that it constitutes an opening; this results in an equational assertion, which entails the choice of bukan. Thus, examples (5) and (6) show that the choice between the four free variants, tak, ndak, nggak, and tidak, and the fifth form, bukan, is dependent on the semantic relationship that holds between X and Y in the scheme (X) NEG-Y.

A seemingly similar yet crucially different contrast is presented by the two negative forms of Saramaccan, á and ná. More often than not, á will correspond to one of the four variants, tak, ndak, nggak and tidak, while ná will correspond to bukan. However, as argued by McWhorter (this volume), the choice between á and ná is in fact dependent on a syntactic distinction: á is used to negate verbal predicates, ná is used elsewhere. Thus, whereas the choice of negative form in Riau Indonesian is semantic, in Saramaccan it is syntactic, thereby providing yet another reflection of the greater syntactic complexity of Saramaccan.

Other specific forms in the two languages replicate the same pattern. Riau Indonesian has a prefix peN- which is used, non-productively, as a word-formation device, in derivations such as minum ‘drink’ > peminum ‘drunkard’, curi ‘steal’ > pencuri ‘thief’, takut ‘fear’ > penakut ‘coward’. This prefix seems to bear a close resemblance to the Saramaccan suffix -ma, which is used, also non-productively, as a nominalizer, in derivations such as bebe-daan ‘drink rum’ > bebe-daan-ma ‘drunkard’ (Bakker, Smith, & Veenstra 1995: 173–174). However, the similar English translations for peminum and bebe-daan-ma as ‘drunkard’ obscure a deeper difference between the two forms (which has nothing to do with the rum).

The Riau Indonesian prefix pen- actually consists of two distinct but closely fused prefixes, pe- and N-, each of which has a function which may be described in purely semantic terms. The prefix pe- has quantificational force; its
meaning is ‘frequently’ or ‘typically’. The prefix $N_-$, one of the more common prefixes in the language, marks the item to which it is attached as being associated with a participant bearing the thematic role of actor (see Gil 1999, to appear d for detailed discussion and analysis). Combining the two, a construction of the form $peN-X$ has the meaning ‘$X$ associated with an actor, and with great frequency’. A crucial feature of Riau Indonesian semantics is the indeterminacy of most expressions with regard to ontological category: given the right context, the same expression can refer to a thing, an activity, a time, a place, and so forth. For example, $minum$ prototypically refers to an activity of drinking, but it can also refer to a participant (actor, patient, or whatever) associated with a drinking, or to a time or place in which a drinking occurs, and so on. Putting this all together now, we see that whereas $peminum$ indeed often denotes a thing, namely an ‘actor frequently engaged in drinking’, or simply a ‘drunkard’, in other contexts it may refer to an activity, namely a ‘frequent drinking associated with an actor’. In some cases, admittedly, it is hard to tease the two possibilities apart: if somebody says $Rudy peminum$, it is hard to tell whether the intended meaning is ‘$Rudy$ is a drunkard’ or ‘$Rudy$ drinks frequently’, and the construction is therefore most appropriately analyzed as being vague between the two. But in other contexts $peminum$ clearly refers to an activity. Consider the following example, from the beginning of a folk tale about the traditional protagonist Yong Dyolah:

(7) Ha, Yong Dyolah ini ‘kan, dia suka
EXCL Yong Dyolah DEM.PROX NEG 3 like

peminum
FREQ.AG.drink

[Beginning folk tale, introducing main character]
‘Ah, Yong Dyolah, right, he used to like to drink a lot.’

In the above example, the word $suka$ ‘like’, directly preceding $peminum$, narrows the semantic range of $peminum$, forcing it to be interpreted as an activity:

---

8. The representation of this prefix as $N_-$, following traditional practice in Austronesian linguistics, reflects its nature as an underspecified nasal, whose form is dependent on the first segment of the stem to which it attaches. When this segment is itself a nasal, as for example in $minum$, $N_-$ is “invisible”; in other cases it materializes as a homorganic nasal stop, either before the first consonant, as in $penjaga$, or in place of it, as in $penakut$. It should be noted that in some cases, such as when the first segment of the stem is a voiced obstruent, the form of $N_-$ differs, depending on whether it occurs by itself or in conjunction with $pe$-. Thus, for example, whereas $N$-$jaga$ surfaces as $nyaga$, $pe$-$N$-$jaga$ is realized as $penjaga$. In cases such as these, the realization of $N_-$ when in conjunction with $pe$- is the same as Standard Indonesian. This suggests that forms containing $peN$- may be “borrowings” into Riau Indonesian from the standard language. However, as shown below, the syntactic and semantic behaviour of such forms is quite different from that of their Standard Indonesian counterparts.
as the continuation of the story makes amply clear, Yong Dyolah did not like drunkards, he was a drunkard, or, as the sentence says, he liked to drink frequently. Thus, the above example shows that the basic meaning of peminum is not ‘drunkard’ but something more like ‘drinking associated with an actor, and with great frequency’, which may refer, according to context, to a person, an activity, or various other kinds of entities.

Turning now to the Saramaccan form bebe-daan-ma, a very different picture emerges. Whereas peminum is vague with respect to ontological category, bebe-daan-ma refers unequivocally to a person, specifically a drunkard. Thus, constructions corresponding to (7) can only mean ‘He likes drunkards’, not ‘He likes to drink a lot’ (McWhorter, personal communication). Moreover, this semantic difference is paralleled by a syntactic one. While peminum has the same distributional privileges and other syntactic properties as minum, bebe-daan-ma exhibits very different behaviour from bebe-daan. Specifically, whereas bebe-daan is a verbal expression, bebe-daan-ma exhibits all the characteristic properties of a nominal expression. (The relationship between bebe-daan and bebe-daan-ma is thus similar to that between the English drink and drunkard.) Accordingly, whereas the Riau Indonesian prefix peN- has an exclusively semantic function, the Saramaccan suffix -ma has a fundamentally syntactic function, that of nominalizer, the difference between the two affixes thereby providing yet another reflection of the greater syntactic complexity of Saramaccan.

Summarizing, then, the contrastive analyses of pronouns, reduplication, negation, and the peN- and -ma affixes reflect the different ground plans of the two languages, with the semantic/pragmatic orientation of Riau Indonesian contrasting with the greater syntactic elaboration of Saramaccan. These more specific items thus join forces with the other more general syntactic features, discussed previously, in contributing to the greater syntactic complexity of Saramaccan. Taken together, the weight of the evidence summarized in Table 4 leads inexorably to the conclusion that the overall level of syntactic complexity of Riau Indonesian is substantially, indeed one might go so far as to say overwhelmingly, lesser than that of Saramaccan.

Although superficially unrelated, the specific syntactic properties of Riau Indonesian represented in Table 4 may be argued to fall out from a single more general property, namely an impoverished inventory of syntactic categories. In Gil (1994, 2000b, 2001) it is argued that in Riau Indonesian, almost all words and larger expressions belong to a single open syntactic category, which is essentially that of the Sentence. Among the complete sentences of Riau Indonesian are names, such as Rudy, words denoting things, such as kucing ‘cat’, words denoting activities, such as kejar ‘chase’, and a wide variety of other items such as aku ‘1SG’, tiga ‘three’, sudah ‘PFCT’, and all combinations of such items. Thus, Riau Indonesian does not distinguish between
nouns, adjectives, verbs, and prepositions, nor between lexical and phrasal categories. (What Riau Indonesian does have, though, is a small finite class of a few dozen items with different syntactic properties, which may accordingly be assigned to a closed syntactic class distinct from the Sentence.) The properties of Riau Indonesian listed in Table 4 are all straightforward consequences of this simple inventory of syntactic categories. Expressions denoting activities may stand alone as complete, non-elliptical sentences because they are in fact S(entence)s. Relative order of subject, verb, and object is free because there are no subjects, verbs, or objects, a consequence of the absence of any relationship of government between a verb and its nominal arguments — the relevant items are all Ss. There are no rules specifying the position of interrogative expressions, such as “WH-movement”, because there is no distinct grammatical category of interrogative expressions to which such a rule could refer: all interrogative expressions are also Ss. There are no copulas because the function of copulas is to distinguish between nominal, adjectival, and verbal predication, and there are no such distinct predicate types, since they all belong to the category S. There are no complementizers because the function of complementizers is to convert Ss into NPs, but there is no distinction between Ss and NPs, since they are all Ss. Expressions referring to properties may occur in the same positions as expressions referring to things because they belong to the same syntactic category, S. Expressions referring to possessors, properties, and events may occur in the same attributive constructions because they belong to the same syntactic category, S. And finally, pronouns, reduplication, negation, and various suffixes such as peN- cannot exhibit the kind of syntactic behaviour characteristic of the corresponding items in other languages because, in Riau Indonesian, there are no distinct syntactic categories to which such behaviour may make reference. Thus, all of the properties of Riau Indonesian illustrated in Table 4 are consequences of a single, more fundamental property, with respect to which Riau Indonesian is simpler than Saramaccan, namely its inventory of syntactic categories.

---

9. Readers who wish to maintain a sceptical attitude with respect to the analysis outlined in the previous paragraph may be reassured that the main claim of this paper, namely that Riau Indonesian is of lesser overall complexity than Saramaccan, is not dependent on whether indeed Riau Indonesian has but a single open syntactic category, from which all else follows. If, contrary to the above analysis, the properties listed in Table 4 are taken to be logically independent of each other, then the lesser syntactic complexity of Riau Indonesian would be equally in evidence, and arguably even more remarkable. Perhaps the only interpretation of these facts which might lead to a different conclusion would be one which posited complex underlying syntactic structures common to Riau Indonesian, Saramaccan, and presumably all languages, possibly in accordance with some variants of the generative framework. Given such universal structures, Riau Indonesian would emerge as more syntactically complex, by dint of the heavier machinery which would need to be invoked in order to obliterate all evidence for such structures, via flattening syntactic trees, neutralizing categorial distinctions,
Table 5. Riau Indonesian and Saramaccan: Grammaticalization

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Riau Indonesian</th>
<th>Saramaccan</th>
<th>More Complex Language</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>tense and aspect</td>
<td>very weakly</td>
<td>weakly grammaticalized</td>
<td>Saramaccan</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>grammaticalized</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number</td>
<td>1st and 2nd</td>
<td>all pronouns,</td>
<td>Saramaccan</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>person pronouns,</td>
<td>definite article,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>optionally</td>
<td>obligatorily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>classification</td>
<td>numeral classifiers,</td>
<td>none</td>
<td>Riau Indonesian</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>optionally</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Whereas the first two diagnostics in (4) characterize Riau Indonesian as substantially less complex than Saramaccan, the latter two diagnostics paint a somewhat different picture. The third diagnostic, in (4c), suggests that a language is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions. A comparison of Riau Indonesian and Saramaccan with respect to some basic semantic categories is provided in Table 5.

Tense and aspect marking is only weakly grammaticalized in both languages, though perhaps more weakly in Riau Indonesian. In Riau Indonesian, all of the forms whose meanings pertain to tense and aspect are independent words which may occur in any syntactic environment, including on their own, as complete non-elliptical sentences: like most other expressions in the language, they thus belong to the category of Sentence. For the expression of time, there are words such as nanti for proximate future, tadi for proximate past, and dulu for distal past; however, these forms are no more grammaticalized than any of an open set of other temporal expressions such as selasa pagi jam lapan ‘Tuesday morning eight o’clock’, and therefore hardly merit being classified as tense markers. The only candidate for tense marker is the form mau, whose basic meaning is ‘want’: in many cases it occurs in contexts which are clearly non-volitional, and with a meaning that appears to be approaching that of a simple future, thereby following a crosslinguistically frequent path of grammaticalization (Bybee & Dahl 1989, Bybee, Perkins, & Pagliuca 1994). As for aspect, the expression of progressivity is relatively infrequent, and is restricted to particular subvarieties of Riau Indonesian. One way of expressing the progressive and assigning phonologically-null realizations to various positions. But I would consider such an analysis to be rather perverse.
Commentary on McWhorter: David Gil

is with the form lagi, a conjunctive operator with a wide range of meanings which include ‘also’, ‘more’, ‘still’, and ‘again’; this usage of lagi occurs more often in urban varieties of Riau Indonesian, suggesting that it is may be a recent borrowing from the more prestigious Jakarta Indonesian, where its use is much more frequent. A second way of expressing the progressive is with the form sedang, whose basic meaning is ‘moderate’ or ‘intermediate’; whereas for most speakers this form is associated exclusively with Standard Indonesian, some speakers occasionally make use of it also in basilectal Riau Indonesian. In both cases, the progressive usage may be viewed as little if anything more than a particular case of a more general meaning associated with the form in question. The expression of perfectivity is considerably more widespread. In many cases, perfectivity is expressed through forms which are associated with more concrete basic meanings: habis ‘finish’, ‘consume’, and siap ‘prepare’, ‘ready’. More often, however, perfectivity is expressed by means of the dedicated perfective marker sudah. Like all of the other forms considered in this paragraph, sudah can stand alone as a complete non-elliptical sentence; in fact, it does so with great frequency, with meanings, depending on context, such as ‘That’s enough’, ‘I’m done’, or ‘Are you through yet’? Nevertheless, sudah is probably as close as Riau Indonesian comes to a grammaticalized tense-aspect marker. One characteristic that it shares with grammaticalized forms in other languages is its highly abstract meaning. A second characteristic, and one that distinguishes it from all the other forms considered in this paragraph, is that it often occurs in a reduced form, udah or even dah.10 Thus, sudah may be characterized as an aspectual marker exhibiting a certain degree of grammaticalization. But not really that much, when compared with the highly grammaticalized tense-and-aspect systems of European and other languages. All in all, then, tense and aspect in Riau Indonesian may thus be characterized as, at best, very weakly grammaticalized.

Turning, now, to Saramaccan, the available evidence suggests that tense and aspect markers are no less grammaticalized than in Riau Indonesian, and perhaps somewhat more so. Admittedly, Veenstra (1996: 12–25) argues explicitly that forms such as o-, denoting ‘intention, irrealis, future, and the like’, are unbound, functioning mostly like main verbs. However, Bakker, Smith, & Veenstra (1995: 167) treat forms such as o- as prefixes, thereby implying a higher degree of grammaticalization. Moreover, McWhorter (personal communication) points out that the markers ta ‘continuous’ and bi ‘past’, even though they can be stranded, “have no verbal equivalents and do not cleft”. Thus, it may be tentatively concluded that Saramaccan exhibits a somewhat

---

10. The alternation between word-initial s- and Ø- as in sudah ~ udah is characteristic of a handful of high-frequency forms, occurring also in saja ~ aja ‘only’, ‘just’ and si ~ i PERS.
greater degree of complexity with respect to the grammaticalization of tense and aspect.

Number marking is of very limited distribution in both languages. In Riau Indonesian, number marking is limited to 1st and 2nd person pronouns, following the pattern shown in Figure 2.

As evident in Figure 2, Riau Indonesian has dedicated forms for 1st person singular and 2nd person plural, but not for 1st person plural (exclusive or inclusive) or 2nd person singular. However, even for the expression of 1st person singular and 2nd person plural, other forms are available which are unspecified for number.  

11. It should be noted that some of the forms that are unspecified for number do have preferred interpretations in which number is specified: *kami* is preferentially 1st person plural exclusive, while *kita* is preferentially 1st person plural inclusive. This suggests that such forms may be analyzed as basically plural, with their singular interpretations derived by the application of further rules of a pragmatic nature, perhaps not unlike the English “royal ‘we’”. However, until now, I have not been able to find any arguments adjudicating between such alternative analyses. But in any case, even if *kami* and *kita* were analyzed as basically plural, there would remain other pronominal forms for which there seems to be no preference for either singular or plural interpretations, and which, accordingly, would justify the characterization of number marking as optional for 1st and 2nd person pronouns. (In a fascinating discussion of 1st person plural pronouns in other Malay/Indonesian dialects, Donohue & Smith 1998 trace various diachronic developments in the forms and meanings of these pronouns, but do not address the issue of distinguishing between basic and derived meanings. Also, the 1st person singular usage of *kami* in Riau Indonesian provides a counterexample to their claim, on p. 71, that “it is not, however, in general use as a first-person singular pronoun in any variety of Malay/Indonesian”.)
Commentary on McWhorter: David Gil

1st and 2nd person pronouns, and elsewhere it is unavailable. In contrast, in Saramaccan, number marking is obligatory for 1st, 2nd, and 3rd person pronouns. In addition, it is also obligatory for the definite determiner: *dì wòmi* ‘the man’, *déé wòmi* ‘the men’ (McWhorter, personal communication). Thus, number marking is more highly grammaticalized in Saramaccan than Riau Indonesian, though by a margin that is hardly overwhelming when viewed from the perspective of an average European language.

Classification of words denoting things provides the first, and as it turns out the only, criterion which respect to which Riau Indonesian appears to be of greater complexity than Saramaccan. Neither Riau Indonesian nor Saramaccan have gender marking of any kind. However, Riau Indonesian may be argued to possess a rudimentary system of classification, by dint of the occurrence of numeral classifiers, in expressions such as *tiga ekor beruk* ‘three macaques’, where *ekor*, literally ‘tail’, connecting *tiga* ‘three’ and *beruk* ‘macaque’, functions as a numeral classifier, associated with the class of animals (including those such as macaques which are almost completely lacking in tails). Nevertheless, this classificatory system plays a minor role in the grammar. Unlike in many other languages, classifiers in Riau Indonesian occur only with numerals, not with non-numeral quantifiers, demonstratives, or other kinds of attributive expressions. Moreover, even with numerals, the use of numeral classifiers is optional and relatively infrequent: in a sample corpus of mixed conversation and narrative containing 75 numerals, only 10 of these, or 13%, occurred together with a classifier. Thus, although Riau Indonesian and Saramaccan are both generally lacking in nominal classification, the numeral classifiers present in Riau Indonesian – but not in Saramaccan – contribute towards the characterization of Riau Indonesian as somewhat more complex with respect to the particular feature.

Summing up then, as reflected in Table 5 above, Riau Indonesian and Saramaccan do not seem to differ systematically from each other with respect to the degree to which various semantic distinctions are overtly grammaticalized. Although the points of comparison were few in number, they encompass some of the core grammatical domains in language, and besides, the overall “feel” of

---

12. Elsewhere in this case includes kinship terms and titles with 1st and 2nd person reference, pronouns with 3rd person reference, and all other expressions referring to things. One limited exception to this generalization is the 3rd person plural form *dia orang*, in the Tanjung Pinang variety of Riau Indonesian, mentioned in the previous section. Another possible exception is the reduplicated form *anak-anak*, from *anak* ‘child’. In general, in Riau Indonesian, when an expression denoting a thing is reduplicated, the reduplication tends to express distributivity (‘things distributed over space, time, or some other dimension’), quantification (‘lots of things’), or negative polarity. However, in the single case of *anak-anak*, a simple plural interpretation (‘children’) also seems to be available.
Table 6. Riau Indonesian and Saramaccan: Morphology

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Riau Indonesian</th>
<th>Saramaccan</th>
<th>More Complex Language</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>morphophonemics and suppletion</td>
<td>very little</td>
<td>very little</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>declensional and arbitrary allomorphy</td>
<td>none</td>
<td>none</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>morphological agreement</td>
<td>none</td>
<td>none</td>
<td>=</td>
<td>=</td>
</tr>
</tbody>
</table>

The two languages, to the extent that one is entitled to speak of such, suggests that any other choice of features would have yielded similar results.

The fourth and final diagnostic, in (4d), suggests that a language is more complex than another to the extent that it has more inflectional morphology. Simply put, neither Riau Indonesian nor Saramaccan have any inflectional morphology whatsoever. A slightly more fine-tuned comparison of the two languages making reference to some of the consequences of their common isolating typology is provided in Table 6.

As suggested in Table 6, Riau Indonesian and Saramaccan have very little by way of morphophonemics and suppletion. For Riau Indonesian, the complexity in this domain is well-nigh exhausted by the behaviour of the prefix N-, mentioned earlier (as part of the discussion of the complex prefix peN-). As noted, the realization of N- varies in accordance with the first segment of the stem to which it attaches. A further complicating factor is that the distribution of the prefix N- varies in accordance with the choice of stem in ways that are partly phonologically governed and partly arbitrary. Phonologically, the prefix N- occurs very infrequently in front of stems beginning with a voiced obstruent. But a large residue of arbitrariness remains: for example, N- occurs frequently before kopi ‘coffee’ to yield ngopi, but only extremely rarely in front of teh ‘tea’ to yield neh (see Gil 1999, to appear d for detailed discussion). But that’s just about it, as far as morphophonemics and suppletion in Riau Indonesian are concerned. Turning to Saramaccan, a roughly equivalent picture emerges, with just two instances of suppletion mentioned by McWhorter (this volume). Thus, there would seem to be no clear justification for characterizing either one of the two languages as more complex than the other with respect to morphophonemics and suppletion.

As for the remaining two properties in Table 6, neither language has any declensional and arbitrary allomorphy, or any morphological agreement phenomena – again, with respect to these two features, the two languages emerge
as equivalent. Thus, with respect to the fourth diagnostic, inflectional morphology, Riau Indonesian and Saramaccan are of roughly equal complexity, as two typical exemplars of the isolating type.

Summarizing the results of this section, the contrastive analysis of Riau Indonesian and Saramaccan points clearly toward the conclusion that Riau Indonesian is of lesser overall complexity than Saramaccan. Whereas with regard to diagnostics (4c) and (4d), grammaticalization and morphology, the two languages are roughly on a par in terms of complexity, with regard to diagnostics (4a) and (4b), phonology and syntax, a clear and consistent pattern emerges, with Riau Indonesian exhibiting significantly lesser complexity than Saramaccan. As McWhorter emphasizes, the criteria are not such that one can award points and then add up for a total score. However, in the case at hand, the overall pattern is quite consistent – every bit as consistent as those patterns which lead McWhorter to conclude that Saramaccan is less complex than Lahu and Tsez. Thus, using McWhorter’s own criteria, there is no choice but to conclude that Riau Indonesian is a more simple language than Saramaccan.

4. How did Riau Indonesian get to be this way?

This paper began with the question: Can one tell that a language is a creole just by looking at it? McWhorter’s answer is “yes”: he proposes a bi-directional correlation between simplicity and age – simple languages are young, complex languages are old. However, the results of this paper suggest that McWhorter is only half right: Riau Indonesian is a simple language that is not young. In other words, one cannot tell that a language is a creole just by looking at it. In particular, those scholars who, on the basis of synchronic descriptions of Riau Indonesian, have wondered whether it is a creole, have been entertaining an invalid presupposition.

Nevertheless, in some cases, one can tell that a language is not a creole just by looking at it. Thus, while simple languages, contra McWhorter, can be either young or old, complex languages can only be old – and for precisely the reasons that McWhorter spells out in some detail. Only through the course of time can a language accumulate all of the features, phonological, syntactic, grammaticalized, and inflectional, which together contribute to its characterization as more highly complex. Looking at things from the other direction, creoles are simple for precisely the reasons that McWhorter suggests, namely that they have not had enough time to accumulate complexity.

However, the results of this paper suggest that an additional scenario needs to be acknowledged, namely that of an old language, in the course of time, shedding its complexity and becoming as simple as, or even simpler than, a typical creole language. Given what we know about the history of the Austronesian family, this is almost certainly what must have happened somewhere
in the history of Riau Indonesian. Thus, just as McWhorter addresses the issue of why creoles are simple, we might ponder the question what mechanisms might be responsible for an older language becoming simpler in the way Riau Indonesian most probably did.

It is not immediately obvious that this question has a general, principled answer. The structure of a language consists of a large number of features which, although partially interrelated (“tout se tient”), are also partially independent (“autonomous”): nasalization, complementizers, number marking, and suppletion, to name just a few that were discussed in the preceding section. Each of these features exhibits a degree of crosslinguistic variation within the bounds imposed by universal constraints on language; in particular, such variation encompasses a range from simple (e.g., no nasalization, complementizers, number marking, suppletion) to complex (e.g., lots of nasalization, complementizers, number marking, suppletion). Moreover, as argued convincingly by McWhorter (this volume), it is hard to imagine a holistic mechanism whose effect would be to balance the relative degrees of complexity exhibited by each of these features, coordinating them in such a way that complexity here is cancelled out by simplicity there. Now pick one of the world’s languages at random. This language will exhibit an arbitrary degree of complexity with respect to each of the relevant features. Given the laws of probability, the chosen language is likely to be simple with respect to some features, mid-range with respect to others, complex with respect to yet others. Now keep on picking languages, and, after examining them with respect to their individual features, assess their overall levels of complexity. The result will be of the sort representable by a bell-shaped curve, with lots of languages in the middle, exhibiting a near-average overall level of complexity, but also a few outliers, with either a very high or a very low overall level of complexity. And indeed, the latter case suggests one reasonable way to view Riau Indonesian, namely as the coincidental result of numerous distinct features each in its own way exhibiting a very low level of complexity.

Of course, nothing is really coincidental; randomness is a euphemism for our inability to formulate an accurate and all-encompassing account of the immensely complicated and diverse factors which contribute to the shaping of a language in time. However, this is not to say that we are totally ignorant. While at one level there may be no definitive answer to the question how Riau Indonesian got to be the way it is, we do know quite a bit, enough perhaps to enable us to say that it is not really too surprising that the language that exhibits such an exceptional degree of overall simplicity is Riau Indonesian, and not, say, the standard literary variety of Sorbian. Following, in (8) below, are three possible reasons, or determinants, for why Riau Indonesian is as simple as it is:
Table 7. Determinants of the structure of Riau Indonesian: Assessment through pair-wise language comparisons

<table>
<thead>
<tr>
<th></th>
<th>Southeast Asia</th>
<th>other regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>basilectal</td>
<td>acrolectal</td>
</tr>
<tr>
<td>contact</td>
<td>Riau Indonesian</td>
<td>Standard Indonesian</td>
</tr>
<tr>
<td>non-contact</td>
<td>Siak Malay</td>
<td>Standard Minangkabau</td>
</tr>
</tbody>
</table>

(8) Determinants of the structure of Riau Indonesian

a. ETHNICITY
   Riau Indonesian is so simple because it is a contact language.

b. REGISTER
   Riau Indonesian is so simple because it is basilectal.

c. GEOGRAPHY
   Riau Indonesian is so simple because it is spoken in Southeast Asia.

The above three factors are not mutually exclusive; on the contrary, all three are significant, though to different degrees. In (8) above, the three factors are arranged in order of increasing importance, with ethnicity less important than register, and register less important than geography.

In order to evaluate the relative contribution of these three factors to the structure of Riau Indonesian, it is necessary to isolate each factor in turn and evaluate its contribution, while holding the remaining two factors constant. Since the three factors are logically independent of each other, they define a three-dimensional space. Simplifying somewhat, each factor may be assigned two values, resulting in a $2 \times 2 \times 2$ matrix into which different languages may be placed, as indicated in Table 7.\(^\text{13}\) The relative weights of the three factors may now be assessed by briefly contrasting Riau Indonesian with each of the other three languages indicated in Table 7 in boldface: these are the languages which differ from Riau Indonesian with respect to just one of the three determinants.

---

\(^{13}\) Unfortunately, since this journal only provides two-dimensional pages, the three-dimensional $2 \times 2 \times 2$ matrix is represented flattened out, in the form of two two-dimensional $2 \times 2$ matrices side-by-side: the reader should try to envisage the right half of Table 7 superimposed on the left half.
The idea that Riau Indonesian owes its exceptional structural simplicity to its function as a contact language is intuitively appealing. Thus, McWhorter (this volume) writes that “creolization is a cline phenomenon, and many natural languages were born via a language contact process that resulted in simplification of a degree less radical than pidginization”. More specifically, with regard to Riau Indonesian, he notes that “this language is a koine, traditionally used as an interethnic lingua franca, meaning that its particularly unspecified nature is almost certainly due to a degree of pidginization in its life-cycle, due to extensive acquisition by adults, having ‘shaved away’ a large degree of accreted complexity”. This hypothesis may be tested by comparing Riau Indonesian with neighbouring basilectal varieties which do not have the function of contact languages, such as for example Siak Malay. However, such a comparison reveals that the function of Riau Indonesian as a contact language is not the main reason for its simple structure. Broadly speaking, Siak Malay is very similar in its structure to Riau Indonesian; it is only some of the finer differences between the two isolects which might, perhaps, be attributed to the different functions of the two language varieties.

Phonologically, the two isolects have largely similar systems. Unlike one variety of Riau Indonesian, Siak Malay has no prenasalized consonants. However, its vocalic system is somewhat richer, with a clearly phonemic [ə], and a possibly more robust four vowel-height distinction – though here it is the low mid vowels whose distribution appears to be limited, often to recent loans from Indonesian and other languages. In addition, it has a parallel series of nasalized vowels, though their occurrence is infrequent and phonologically constrained: limited to word-final syllables, in a pattern tantalizingly similar to that of Aslian languages such as Semelai (Kruspe 1999) and Jahai (Burenhult, personal communication), and only after a glottal consonant, exemplifying rhinoglottophilia (a term coined by Matisoff 1975) – it is thus unclear whether they are deserving of the status of independent phonemes. So in balance, the phonology of Siak Malay may be a little more complex than that of Riau Indonesian, but not by much.

Similar observations hold with respect to morphosyntax. Most grammatical structures in Riau Indonesian have precise calques in Siak Malay: one has to look hard to find systematic differences between the two isolects – though when such differences are observed, they do tend to point in the direction of greater complexity associated with Siak Malay. Among such differences are the following. The prefix N- occurs with noticeably greater frequency in Siak Malay (though its form and function are the same as in Riau Indonesian). Several high-frequency bisyllabic function words sometimes undergo reduction to a monosyllabic form; these include *dengan* > *ngan* ‘with’; *dekat* > *kat* ‘in’, ‘at’; *lagi* > *gi* ‘also’, ‘more’, ‘still’, ‘again’; and *jugo* > *go* ‘also’. When the high-frequency form *nak* ‘want’ is negated, the resulting collocation *tak nak* ‘not
want’ may be replaced with the suppletive form tendak. The aspect marker sudah, mentioned in the discussion of Table 5 in the preceding section, has undergone further grammaticalization, and may occur in two different positions with distinct interpretations: preceding the activity expression it expresses the perfective, while following it expresses the perfect – indeed, it may cooccur in both positions within the same construction, thereby expressing both notions simultaneously. (The behaviour of sudah in Siak Malay thus presents a remarkable parallel to that of the Mandarin Chinese aspect marker le, cf. Li & Thompson 1981, though there is no reason to believe that this parallelism is not the result of independent developments in both languages.) Thus, what differences are in evidence between the grammatical structures of Riau Indonesian and Siak Malay seem to point towards the characterization of Siak Malay as more complex, but by a rather unimpressive margin. Indeed, if I had chosen to use Siak Malay rather than Riau Indonesian in the comparison with Saramacan, the results would not have been very different: Siak Malay would have emerged as the simpler of the two languages in its overall structure.

In conclusion, then, it is plausible to suggest that, as a contact language variety, Riau Indonesian may have chosen to streamline itself and “shave away” some of the idiosyncratic complexities that it encountered in its neighbouring languages, such as some of the features described above for Siak Malay. In this sense, then, Riau Indonesian may be said to owe some of its structure to its function as a contact language. But not that much. Given that the various non-contact isolects of Riau province, such as Siak Malay, also exhibit very simple grammatical structures, an explanation must clearly be sought elsewhere for the remarkably simple structure characteristic of Riau Indonesian.

A more significant determinant for the simplicity of Riau Indonesian is provided by register, namely its character as a colloquial, informal, and unwritten language variety. It has often been observed that basilectal varieties differ systematically from their acrolectal counterparts in a number of respects which may contribute to their characterization as of lesser overall complexity. Thus, for example, Givón (1979) distinguishes between “pragmatic” and “syntactic” modes of organization, arguing that these modes are characteristic of spoken and written language respectively; a similar theme recurs in the work of many other scholars. In particular, Benjamin (1993) suggests that colloquial and formal Malay are distinguished by “condensed” versus “articulated” modes of expression. To examine the effect of register on the structure of Riau Indonesian, it may be compared with its very own acrolectal counterpart, which also happens to be a contact language, namely Standard Indonesian. Here the picture is quite clear cut: Riau Indonesian is indeed significantly simpler than Standard Indonesian.

Where this is least obvious is in the phonology: as a common acrolectal variety used across the archipelago, Standard Indonesian is generally spoken in a
variety of local accents reflecting the provenances and linguistic backgrounds of its speakers. As a result, descriptions of its phonology tend to differ, depending on the sources that are used. Thus, for example, whereas most descriptions (e.g., Alieva et al. 1972, Mintz 1994, Hasan et al. 1998) posit a six-vowel system, others (e.g., Oetomo et al. 1980) posit an eight-vowel system, with four distinctive vowel heights. In morphology, too, Standard Indonesian is not significantly more complex than Riau Indonesian – at least not according to the criteria under consideration here. True, even the most cursory textual comparison will suggest that the average number of morphemes per word is much higher in Standard Indonesian than in Riau Indonesian. However, all of the morphology in Standard Indonesian is derivational and (again, with the exception of the prefix N-) formally transparent. Thus, Standard Indonesian resembles Riau Indonesian with respect to all three properties listed in Table 6, with very little morphophonemics and suppletion, no declensional and arbitrary allomorphy, and no morphological agreement. Moving on to grammaticalization, a difference begins to emerge. Of the three grammaticalization properties listed in Table 5, Standard Indonesian resembles Riau Indonesian in two: tense and aspect is weakly grammaticalized, and there is an optional use of numeral classifiers. However, with respect to number, Standard Indonesian is more complex: number marking is obligatory for all pronouns, and optional for all nouns – where it constitutes one of the semantic functions of reduplication.

However, as soon as one turns to syntax the difference between the two language varieties becomes overwhelming. Of the eleven syntactic properties listed in Table 4, Standard Indonesian resembles Riau Indonesian in just three: overt expression of participants is optional, the position of the question word is free, and the choice between alternative negative markers is semantically rather than syntactically based. With respect to the remaining eight syntactic properties, Standard Indonesian exhibits greater syntactic complexity than Riau Indonesian, in fact bearing a close resemblance to Saramaccan. Basic word order is relatively rigid. There is a copula, adalah, which occurs in predicate nominal constructions. There is a complementizer, bahwa, used to mark sentential complements. Adjectives such as bagus ‘good’ cannot head NPs on their own; in order to do so, they must be preceded by a relative marker; for example yang bagus ‘good one’. In a construction of the form HEAD ATTRIBUTE, the attribute may denote a possessor or a property but not an event; in the latter

14. With regard to the latter property, there is in fact a widespread belief that in Standard Indonesian, tidak and its occasional variants negate verbs and adjectives while bukan negates nouns; see, for example Kwee (1965: 16) and Hasan et al. (1998: 378–384). However, other scholars, such as Mintz (1994) and Sneddon (1996), have observed that even in Standard Indonesian there are exceptions to this generalization, and have attempted to account for them in semantic terms.
Commentary on McWhorter: David Gil

case, the same relative marker is required: *buku yang/*Ø Rudy beli ‘book that Rudy bought’. The choice of pronouns is governed not only by politeness but also by grammatical factors: in addition to the common independent pronouns, there is a series of agentive proclitic pronouns (mentioned in Footnote 8 above) and also a series of possessive enclitic pronouns. Reduplication shares many of the semantic functions that it has in Riau Indonesian; however it also has at least one syntactic function, converting adjectives into adverbs, for example *pelan ‘slow’ > pelan-pelan ‘slowly’. Finally, the prefix *pen- is indeed a nominalizer, forming expressions which may only denote things: in Standard Indonesian, sentence (7) could only mean ‘He likes drunkards’. Thus, in terms of its syntax, Standard Indonesian is indeed significantly more complex than Riau Indonesian.

The lesser syntactic complexity of Riau Indonesian is undoubtedly due to its role as a basilectal language variety. The prototypical use of Riau Indonesian is to talk about the here and now; it thus contrasts with Standard Indonesian, which must also be able to convey, in speech or writing, information about matters that are remote in space and time – as in a politician’s speech or a newspaper article. Accordingly, Riau Indonesian can permit itself to leave more aspects of meaning without formal encoding. Context will usually disambiguate, and in those rare cases when it does not, a simple ‘What do you mean?’ will elicit the necessary clarification. Thus, register is clearly an important determinant for the overall simplicity characteristic of Riau Indonesian. But once again, it is a long way from being the whole story. As suggested by the above discussion, even acrolectal Standard Indonesian exhibits an overall degree of complexity comparable to that of Saramaccan, or, in other words, significantly less than the most basilectal varieties of many older languages from other parts of the world.

To fully appreciate how Riau Indonesian got to be the way it is, it is necessary to take cognizance of the third and most important factor: geography. In a nutshell, Riau Indonesian is so simple primarily because it is located in a neighbourhood where simplicity of structure is a widespread characteristic feature. To see this, it suffices to compare Riau Indonesian with a similar language from some other part of the world: a basilectal variety of a major language, used as a contact variety in a region in which many other languages are also spoken. An example of such a language is the variety of Russian spoken in Dagestan, a constituent republic of the Russian Federation, occupying a relatively small mountainous area in the northern Caucasus where approximately 30 distinct languages are spoken in close proximity. (In the absence of any available written descriptions of Dagestani Russian, the facts described below are based on discussions with linguists who have conducted field work in the region: Konstantin Kazenin, Andrej Kibrik, and Sergei Tatevosov.) Broadly speaking, Dagestani Russian is of comparable complexity to Standard Rus-
sian, and hence of much greater complexity than Riau Indonesian. Even the most superficial glance at Daghestani Russian with its palatalized consonants, morphological cases, aspectual verbal prefixes, morphosyntactic strategies for subordination, suppletive verbal forms, declensional allomorphy, and various kinds of agreement, should dispel any doubt that this basilectal contact language is of significantly greater overall complexity than Riau Indonesian, or, for that matter, Siak Malay and Standard Indonesian. Admittedly, in certain areas, Daghestani Russian has undergone simplification, which might be attributed to its function as a basilectal contact language; some examples include the loss of morphological comparative forms of adjectives, the neutralization of gender agreement in attributive adjective constructions, and the use of nominative case with toponyms expressing direction or location. On the other hand, it has also acquired from neighbouring Caucasian languages certain features whose effect is to increase its overall level of complexity, such as ejective consonants and echo reduplication (a particular type of reduplication in which the first segment of the second reduplicand is replaced with a constant consonant, in the case at hand [m]). Thus, one might say that Daghestani Russian looks just like what one would expect from a Slavic language spoken in that hotbed of structural complexity, the Caucasus. Accordingly, the contrast between Daghestani Russian and Riau Indonesian suggests that however important contact function and register may be as determinants of structural complexity, they are of lesser significance than the locations in which the languages are spoken.

Although at an extreme with respect to its overall level of simplicity, Riau Indonesian is actually quite a typical Southeast Asian language; for some perspectives on the characteristic simplicity of Southeast Asian languages see Riddle & Stahlke (1992), Huang (1994), Bisang (1996), and Gil (to appear a). More specifically, the properties of Riau Indonesian are just what one would expect from a language that is spoken in central Sumatra. Much of its structure can be viewed as reflecting its fortuitous location in the intersection of two regional sprachbunds, associated, respectively, with mainland Southeast Asia and the Pacific. The first of these is the well-known sprachbund of isolating languages, which encompasses, among others, major languages such as Chinese, Vietnamese, Khmer, and Thai. Riau Indonesian differs from the prototypical isolating language in its preference for bisyllabic rather than monosyllabic words. However, in terms of its characteristic morpheme-to-word ratio, it is, if anything, an even purer exemplar of the isolating type, in that – unlike many other languages of Southeast Asia – it does not make productive use of compounding. The second is a sprachbund whose defining feature is the underspecification of syntactic categories, and which encompasses a large swathe of the Pacific and adjacent regions. Within Austronesian, the underspecification of syntactic categories has been observed for a variety of languages, including Tagalog (Shkarban 1992, 1995, Gil 1993a, 1993b, 1995), Tongan (Tchekhoff
1984, Broschart 1997), and others. Outside of Austronesian, similar observations have been made for Eskimo-Aleut, Salish and Wakashan (the most renowned of these being Swadesh 1939 for Nootka); and in the other direction also for Munda (Bhat 1997). Geographically, Riau Indonesian is located right where these two sprachbunds meet, namely, off the southern tip of mainland Southeast Asia, at the beginning of the vast archipelago. What could be more natural, then, than for Riau Indonesian to display characteristics of both sprachbunds, with its isolating typology and underdifferentiation of syntactic categories.

Some of the more specific effects of geography on the structure of Riau Indonesian can be appreciated through a perusal of some of the maps currently in preparation for the *World Atlas of Language Structures*. To cite just a few examples, in Gil (to appear b), Riau province is located within a region characterized by the low differentiation of attributive constructions. In several maps, east-central Sumatra falls in an area characterized by the absence of various inflectional categories; some of these include Corbett (to appear) for gender, Iggesen (to appear) for case, Dahl & Velupillai (to appear) for future tense, and Siewierska & Bakker (to appear) for person agreement on verbs. And in Veselinova (to appear), Riau province is right in the middle of a region characterized by the absence of suppletion in tense and aspect categories. Thus, as these and many other maps show, the structure of Riau Indonesian is to a very large extent determined by its location. Of the three factors responsible for the overall level of simplicity of Riau Indonesian, the most significant is clearly geography.15

Of course, Riau Indonesian did not absorb its structural characteristics from the tropical air or the muddy soil. Unlike the first two determinants, contact function and basilectal register, whose causal effect on simplicity is principled and well-motivated, there is no intrinsic relationship between location and structure, obliging, say, any language spoken in east-central Sumatra to be simple. Rather, to claim that Riau Indonesian owes its overall level of simplicity to geography is to suggest that it acquired the relevant structural characteristics through the mechanisms which are generally assumed to be responsible for the creation of sprachbunds, all of which involve one form or another of language contact. In other words, the simplicity of Riau Indonesian is something that it picked up from its neighbours. But how did its neighbours get to be so simple? At this point, it seems, there is no choice but to return to the notions of coincidence and randomness, albeit in a somewhat more geographically-

15. The significance of geography is also evident in those fewer cases where Riau Indonesian owes an element of complexity to its location. The most obvious example of this is the presence of numeral classifiers, which is a well-known characteristic feature of Southeast Asian languages (Nichols 1992, Gil to appear c).
informed fashion. Imagine, as before, a large number of partially interrelated but also partially independent properties; however, this time, instead of assigning them to individual languages, associate them with a geographical region. Thus, instead of languages varying randomly within the bounds of universal constraints, we may think of geographical regions as being the venues for variation. (Given that languages themselves are idealizations, this actually represents a mere quantitative shift in perspective rather than a qualitative ontological leap.) Why Southeast Asia and the nearby islands should be associated with simplicity of structure is a question which cannot be provided with an insightful and principled answer. However, given that this happens to be the case, this coalescence of areal features constitutes what is by far the most significant reason for the overall structural simplicity of Riau Indonesian.

Thus, Riau Indonesian owes its simplicity of structure to a limited extent to its function as a contact language, to a much greater extent to its basilectal nature, but more than anything else to its geographical location. In its overall simplicity, it stakes out an extreme position in the typological space within which languages are free to vary. As a language that fulfils its communicative functions as well as the next one, it underscores the degree to which much of linguistic structure is not strictly necessary for communication, thereby lending further support to the view that much of the elaborate structures exhibited by most languages are incidental to universal grammar, the products of historical processes of accretion. However, as a descendant from Proto-Austronesian without any known history of abrupt, radical restructuring, Riau Indonesian shows that the accumulation of structure is not an inexorable unidirectional process: in some circumstances, the direction may be reversed, and a language may choose to shed much of its structure, moving instead from complexity towards simplicity.

Received: 23 April 2001    Max-Planck-Institut für evolutionäre Anthropologie
Revised: 14 May 2001

Correspondence address: Max-Planck-Institut für evolutionäre Anthropologie, Inselstraße 22,
04103 Leipzig, Germany; e-mail: gil@eva.mpg.de

Acknowledgments: I would like to thank Frans Plank for providing the original impetus for writing this paper. I am particularly grateful to John McWhorter for much-needed assistance with the Saramaccan data plus lots of helpful discussion, without which I would not have been in a position to pass critical judgement on his ideas. I am also indebted to John McWhorter and Claire Lefebvre for their helpful comments on an earlier draft. This paper would not have been possible without the assistance of numerous speakers of Malay/Indonesian, who, intentionally and otherwise, contributed to my understanding of their diverse language varieties, and showed me how simple they could actually be; thanks, in particular, to Doni, Pai, and Bowo for uttering the sentences cited in (5)–(7). Except where references are given, all of the Malay/Indonesian data cited in this paper are from my own field work.
Abbreviations etc.: 1 1st person, 3 3rd person, AG agentive, DEM demonstrative, EXCL exclamation, FREQ frequentative, NEG negation, NEGPOL negative polarity, PERS personal, PFCT perfect, PROX proximate, SG singular. Whereas in the context of phonological discussions, forms are cited in square brackets, in other contexts, forms are cited in italics, in a standardized orthography which may obscure a few phonemic distinctions.

References


Creoles, complexity, and Riau Indonesian


Contact and simplification: Historical baggage and directionality in linguistic change

by Peter Trudgill

Nearly twenty years ago I wrote (Trudgill 1983: 106):

It is usual for laymen to claim that some languages are easier to learn than others. Linguists have tended to play down this suggestion, and to point out that it depends on what your point of departure is: Spanish is easier for an English speaker to learn than Chinese, but for a speaker of Thai it might be the other way round. However, I think it is legitimate to suggest that some languages actually are easier for adults to learn, in an absolute sense, than others. If one were given a month in which to learn a language of one’s choice, I think one would select Norwegian rather than Faroese, Spanish rather than Latin, and Sranan rather than English.

Crucially, I then added that the preferred languages for rapid learning were all “languages which […] have undergone more contact” (see also Trudgill 1989). My thinking was, and is, that “linguistic complexity”, although this, as McWhorter says, is very hard to define or quantify, equates with “difficulty of learning for adults”. I therefore entirely agree with him, as the above quotation shows, that some languages are more complex than others.

I also agree entirely that, other things being equal (see below), the older a language is the more complexity it will have. As I wrote in Trudgill (1999: 148) of the complexity resulting from grammatical gender:
Commentary on McWhorter: Peter Trudgill

Grammatical gender marking in languages such as European languages which have only two or three genders seems to be almost totally non-functional. [...] We are used to the idea that human languages contain and indeed need redundancy to aid with processing. But do not these particular forms of gender marking represent redundancy on a somewhat nonfunctional scale? The only way we can explain these phenomena satisfactorily would appear to be historically. We know that languages drag along with them a certain amount of, as it were, unnecessary historical baggage. This is most obvious in the case of grammatical irregularities which all languages appear to be able to tolerate up to a point. If the plural of foot in English is feet rather than *foots, native learners can cope with this, and linguists can explain why it is so on historical grounds. But it may well be that in languages, or at least in some languages, there is much more of this afunctional historical baggage than has sometimes been thought. For example, the presence of different declensions for nominal forms and different conjugations for verbal forms in inflecting languages would appear to provide good evidence that languages can demonstrate large amounts of complex and non-functional differentiation which provide afunctionally large amounts of redundancy and whose presence in such languages can again, presumably, only be explained satisfactorily in historical terms.

The implication is that the longer a language exists, the more “historical baggage” it acquires. For some reasons why this might be, see Trudgill (1995).

It is also clear how this state of affairs can endure (Trudgill 1999: 149):

Gender marking occurs with a very high degree of frequency indeed in those languages which have it, and is thus a feature with a very high degree of entrenchment in the sense of Langacker (1987: 59). It is thus very readily maintained in the speech of individuals; and because of the amazing language learning abilities of the human infant, languages readily maintain this type of complex historical baggage from one generation to another even though it represents a complication and/or an excess of redundancy, and even though it may have no particular or very important function.

But it also seems to me crucial to understand that “other things” are very rarely equal and that old languages can lose complexity as well as acquire it. Note that, above, I not only preferred a new language, Sranan, to English; I also preferred Spanish to Latin and Norwegian to Faroese. I entirely agree that creole grammars are the simplest grammars, but I believe they are simply at one end of a continuum of complexity and simplicity. Just as complexity increases through time, and survives as the result of the amazing language-learning abilities of the human child, so complexity disappears as a result of the lousy language-learning abilities of the human adult. Adult language contact means adult language learning; and adult language learning means simplification, most obviously manifested in a loss of redundancy and irregularity and an increase in transparency. This can indeed be seen at its most extreme in pidgins
and hence in creoles (Trudgill 1996a). But it is not confined to these types of language.

I accept that “creole genesis entailed a transformation of source language structures which far bypassed the relatively non-disruptive processes which German dialects underwent in becoming Yiddish” (Section 4) but I am still inclined to think that this is a difference of degree rather than kind. It is not an accident that Faroese, as a low-contact language not subject to adult language-learning, has maintained a degree of inflectional complexity which Norwegian has lost. As Braunmüller (2000: 291) argues of Icelandic, Faroese, and North Frisian: “Diese kleine Sprachen werden übrigens typischerweise kaum einmal von Fremden gelernt und können DEshalb hohe Allomorphik und geringe morphologische Transparenz bewahren” (my emphasis; PT; typically these little languages are rarely learnt by foreigners, and they can THEREFORE retain high allomorphy and low morphological transparency).

An important consequence for typologists follows from this. In Trudgill (1992), I argued that it was interesting to consider the title of Labov’s influential paper “On the use of the present to explain the past” (1975). I suggested that the present is in fact going to be increasingly unlike the past in demographic and social network terms, and that this might well lead to differences in the direction of linguistic change and in the distribution of structures over the world’s languages. I suggested that increasing language and dialect contact means that creoles, creoloids, and koines (Trudgill 1986) will be on the increase, and that languages spoken in small, isolated communities with tightly-knit social networks (Trudgill 1996b) – which I hypothesised were the types of language most likely to produce historical baggage in the form of complexity and redundancy and to transmit them successfully to descending generations (for fuller argumentation, see also Trudgill 1998, 2001, forthcoming) – were becoming less and less common. It is therefore not unlikely that languages with large numbers of afunctional or nonessential grammatical devices will become less numerous, and indeed it is not entirely impossible that complexity of the type so ably described by McWhorter will one day disappear completely from the languages of the world. If this is so, we should do as much as we can, as quickly as we can, to investigate languages with a high degree of complexity before it is too late.
Commentary on McWhorter: Östen Dahl

References


— (forthcoming). Language in Contact and Isolation.

Complexification, erosion, and baroqueness by Östen Dahl

McWhorter’s paper touches upon many issues that are central to our understanding of how human languages and their grammars work, and I am in basic sympathy with his general approach. Like McWhorter, I think that it makes sense to compare grammars with regard to complexity, and that the thesis that all languages are equally complex is not well founded. Furthermore, I think that there is a clear sense in which grammars tend to become more complex over time. (This formulation is actually stronger than any one I have found in McWhorter’s paper, but I think it is consonant with what he is saying.) The diachronic processes that condition this tendency are varied but an essential
role is played by the phenomena studied under the heading of grammaticalization. The tendency is counteracted, however, by processes that are largely due to language contact and “abnormal transmission” of languages, that is, cases where a language is transmitted from one generation to another wholly or partly through the mediation of non-native speakers. I would myself tend to see creolization as the limiting case of such a situation; I do not know if McWhorter would agree with this or if he thinks that the intervenience of pidgins somehow essentially changes things – in any case, no evidence for such a position is presented in the paper.

Now, for the objections.

What is the nature of the processes that lead to increased complexity? McWhorter, throughout the paper, uses phrases such as “baroque accretion”, “random complexities”, and “millennia of usage and drift”. Judging from the context, the term “drift” here is used rather in the sense found in biology (changes due to random factors) than in Sapir’s sense of a directed trend. But grammaticalization processes are not random. Rather, they proceed along a set of crosslinguistically definable paths, as has been amply demonstrated in typological research over recent decades. This does not mean that the result of these processes is not complex. There are several sources of this complexity. One is that the introduction and spread of new grammatical patterns may leave parts of the vocabulary untouched, entailing the creation of lexical idiosyncrasies. Another is through phonological change, to be commented upon below. Finally, even if there is a limited number of paths of grammaticalization, they have enough possible branchings and parameters of variation to be far from predictable in their actual output.

As McWhorter points out, the semantic distinctions involved in grammatical markings often seem incidental to communication. I do not think, however, that the quote from Trudgill that they serve no purpose is necessarily true. It appears rather strange that such a large proportion of human languages should be equipped with complex systems that serve no purpose whatsoever. Although there is certainly a great deal of inertia in cultural systems, it appears improbable to me that something like the Semitic verb morphology could be preserved over several millennia, as we know it has been, if it were a complete useless heap of baroque complexities. What kind of function it has is not directly obvious but I think that saying that it contributes to an efficient transmission of information by creating a suitable kind of redundancy is not too far from the truth.

The term “erosion” is frequently used as a metaphor for what happens in language change. McWhorter links it with simplification, which “is an ongoing process in older languages, as phonetic erosion and analogy exert their effects over time”. I assume that by “erosion” McWhorter is thinking of reductive processes in phonology. McWhorter says that simplification is “comple-
mented by emerging complexifications”. I would make this statement stronger, claiming that it is the “erosion” process itself that leads to complexification. McWhorter’s example from Lahu is actually an example of this, but there are much clearer and more familiar ones. Compare the relatively straightforward relationship between the masculine and feminine genders in Italian adjectives such as bello : bella, caldo : calda, bianco : bianca (‘beautiful’, ‘hot’, and ‘white’, respectively) with the corresponding forms in French: [bo] : [bɛl], [jo] : [jɔd], [blɔ] : [blɒ]. The “erosion” of final consonants and schwas in French has led to a significant increase in the complexity of the system. Such a process seems rather different from one where a certain inflectional marker is just dropped – something that I think happens less often in “normal” language change. In general, I am skeptical about the “erosion” metaphor. Erosion, as we know it from the decay of physical objects, such as marble statues, means that the object loses the molecules it consists of in a stochastic fashion. But what we see, for instance, in the French case does not resemble this very much. Segments are lost, but the loss is compensated for in other places.

Bickerton identified the creole prototype with the pure output of the linguistic bioprogram. McWhorter seems to entertain a view of language that comes fairly close to this idea. In Section 5.2, two possible properties of an Ursprache are discussed: that of total lack of inflectional morphology (and also, as far as I understand, of all other “complexities” discussed in the paper), and that of being the basis of Universal Grammar. It is not quite clear what the status of these properties is in McWhorter’s argumentation. However, the assumption that both hold of the first human language leads to strange consequences, in my opinion. It means that our genetic endowment prepares us for the maximally simple language but for no other. At the same time, we know that precisely the “complexities” listed by McWhorter are those properties of human languages that most clearly distinguish first and second language learners, a fact that is also consonant with the tendency for these properties to be filtered out in language contact situations. The assumption that Universal Grammar does not encompass “complexities” in any sense would rather lead us to predict that they would be filtered out fairly quickly also in “normal” (monolingual) language acquisition. But as I have already mentioned, complex grammatical sub-systems such as Semitic verb morphology exhibit an astonishing stability over time, once they have entered a language.

On the other hand, there is really no reason to assume that human language has passed the kind of zero point that McWhorter’s scenario presupposes. A genetic predisposition for language may well have co-evolved with language itself. That is, just because grammatical gender, inflectional tense and aspect, lexical tone, and other “baroque” structures develop by historical change, we need not assume that we are not genetically predisposed to acquire them. But
the kind of predisposition that is involved here may be quite different from the common view of universal grammar as a closed set of parameters with finite sets of values. Such a view entails that languages move within a restricted domain of possible synchronic states. But in a structure-building process like the one we can observe in grammaticalization, decision trees typically branch out at later stages, when combinations of possibilities multiply.

Received: 27 April 2001

Correspondence address: Institutionen för lingvistik, Stockholms universitet, 106 91 Stockholm, Sweden; e-mail: oesten@ling.su.se

Creoles, complexity, and linguistic change
by Wolfgang Ullrich Wurzel

1. The questions raised by McWhorter about complexity, markedness, and functionality are not only relevant for creole studies, but also bear on key theoretical issues in comparative and historical linguistics. What makes his own suggestions especially valuable is that he transcends the kind of pure speculation and mere plausibility reasoning that is all too common in theorizing about complexity and markedness.

2. I fully agree with McWhorter on his two basic assumptions. First, languages need not all be equally complex in their grammars – contrary to a traditional view that is, however, not supported by sufficient evidence, but in line with recent work like that of Ross (1998). Second, creole languages in general have simpler grammars than other languages – as has been argued time and again in creole studies, though not uncontroversially. While these two assumptions can, in my opinion, be safely made irrespective of one’s definition of complexity, the particular definition of complexity chosen does influence one’s evaluation of individual structural domains of different languages or languages as a whole.

3. There is no single definition of grammatical complexity that is generally accepted; there are several, more or less pre-theoretical notions of complexity, and it is not obvious which of them is the most adequate, if indeed there is
Commentary on McWhorter: Wolfgang Ullrich Warzel

a single most adequate one. Aware of this problem, McWhorter proposes an evaluation metric based on four main criteria, relating to phonology, syntax, grammatical (morphological and syntactic) categories and distinctions, and the extent of inflectional morphology. While these criteria as such are not implausible, what remains rather unmotivated is why it is these four, rather than any conceivable others, that are attributed such significance.

4. It seems to me an important question here how to conceive of the relationship between complexity and markedness, two evaluation measures which are no doubt related but at least partly competing. Are complexity and markedness equivalent? Should complexity include markedness or perhaps vice versa? Do they intersect or are they independent?

Other creolists emphasize the importance of markedness theory for determining complexity: “the idea that the creole languages are not grammatically complex in general only makes sense if one has a theory of grammatical complexity to fall back on, and this brings in markedness theories” (Arends, Muysken, & Smith 1995: 12). Considering only generative conceptions of markedness (as espoused among others by Bickerton 1999, Roberts 1999, or Lightfoot 1999), this position is rejected by McWhorter: “what distinguishes grammars in terms of complexity according to my definition is largely independent of the syntactician’s conceptions of markedness or optimality” (Section 5.2). For McWhorter, markedness thus has no role to play in clarifying what “the world’s simplest grammars” are, at least in syntax. Given that the original reason for elaborating the notion of markedness was precisely that it would help define grammatical complexity, this is a rather surprising conclusion. However one sees the relationship between complexity and markedness in specific instances, they should have something to do with one another in the sense that markedness also manifests itself as complexity. Otherwise, what should markedness theory be about and what should it be good for?

Arguably, McWhorter’s pessimism about the relationship between syntactic complexity and markedness stems from the limitations of his own notion of syntactic complexity and of the inadequacy of markedness as practiced in generative syntax. Taking markedness in a more resolutely Greenbergian (and, one should add, Jakobsonian) sense, and applying it in syntax as for example in Haiman (1985), there would surely be grounds for optimism.

5. Turning to McWhorter’s criteria for complexity, phonological complexity is defined in terms of the complexity of phonemic and tonal inventories. Here, unlike in syntax, markedness does play a defining role: “A phonemic inventory is more complex to the extent that it has more marked members” (Section 2.4.3). As McWhorter says, while there is nothing complex about marked sounds as such, an inventory with marked sounds is more complex than one
with unmarked sounds because the former comprises marked in addition to unmarked sounds. Two questions arise here.

First: What markedness concept does McWhorter have in mind when he speaks of languages with only unmarked sounds? There are no such languages. Polynesian languages, invoked by McWhorter, do not just have the one vowel /a/, which is the only (fully) unmarked vowel; they also have spirants, which are marked relative to stops in manner of articulation, etc. (Besnier 1992). Rather, McWhorter is probably referring to the distinction between unmarked or weakly marked vs. strongly marked sound classes, as his examples of what is supposedly marked/unmarked seem to indicate (“ejectives, clicks, and labialized consonants vs. stops, rounded back vowels, and glides”). It should have been made clearer that markedness is a gradual rather than a binary property of linguistic units.

Second: If (adult!) native speakers can deal with them without difficulty, does it really follow that the quality of individual sounds is irrelevant for the simplicity or complexity of a grammar? Or can, for example, their manners of articulation differ in complexity? Articulatory phonetics and language acquisition research both unambiguously suggest that the latter is the case.

Of greater overall significance, however, is that McWhorter sees phonological complexity only in terms of phonemic inventories, i.e., paradigmatically. For him, the combinations of segments in syllables and words are completely irrelevant. But clearly, these syntagmatic relationships ought to play a role for phonological complexity: languages would seem simpler when they are limited to CV syllables than when they permit CCCVCCC syllables. Also worth bearing in mind is the trade off between large segment inventories and simple morpheme and word structures on the one hand and small inventories and complex combinatorics on the other. Here McWhorter might have found additional arguments for his position, given that creoles typically simplify the consonant clusters of their lexifier languages. More attention should also be paid to whether or not, and to what extent, individual phonemes are adapted to their phonological environments in syllables and words – for example, whether only voiceless or both voiced and voiceless obstruents, being in a clear markedness relation, can occur before syllable boundary (cf. German vs. English). Ignoring such relationships has far-reaching consequences, as will be shown presently.

6. Unlike the universalist phonological criterion, the syntactic one is couched in language-particular terms: “A syntax is more complex than another to the extent that it requires the processing of more rules” (Section 2.4.3). It is perhaps not implausible to assume, as McWhorter does, that different basic orders for main and subordinate clauses or split ergative/accusative alignments increase complexity beyond that obtaining in corresponding uniform syntactic systems; but more detailed justification would have been appreciated here too.
7. While it is equally reasonable to assume that “[a] grammar is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions than another”, there is the question of grammatical distinctions which are semantically arbitrary, at least partially. As cited by McWhorter himself, verb government in languages like German, English, and French is an example. Inflection classes, potentially devoid of any semantics, are another; but they fall under McWhorter’s next complexity diagnostic.

8. Stating that “[f]ormal morphology renders a grammar more complex than another one in most cases” (Section 2.4.3), McWhorter rightly notes, though, that a language which expresses grammatical relations through inflection is not per se more complex than one doing the same with free morphemes. Rather, morphological complexity is a “secondary” effect of such typical ingredients of inflection as morphophonemics (of the stem), allomorphy (of inflections – not only caused by phonological factors but also due to inflection classes), and suppletion, which are especially characteristic for flexional (or fusional) languages.

This raises the question of whether this is all there is to morphological complexity. Is it really appropriate to assume that morphological symbolization is not inherently more complex than analytic constructions, given that it may entail complex word forms, such as verbs in Turkish as illustrated in (1)?

(1) dol-dur-ma-yabil-ir-di-m
   fill(itself)-CAUS-NEG-IMPOSS-AOR-PRET-1SG
   ‘I could have refrained from filling (it/something) in’

Conceivably, extreme morphological complexity of word forms as such, as found with agglutination and especially incorporation, might contribute to the complexity of the language as a whole, and syntactic symbolization is really inherently simpler in such cases and universally preferred.

Elsewhere, McWhorter (1995: 796–797) had assigned greater importance to word formation for defining the “Creole Prototype”, where derivation is purportedly more restricted and semantically more transparent than in non-creoles. In the present paper, its contribution to complexity is more marginal.

9. With complexity in general being characterized as “overspecification”, it is striking that McWhorter’s four criteria for complexity are essentially independent of each other: a change of complexity in one respect has no consequences for complexity in other respects. However, it is well known that morphological complexity (in precisely the sense of McWhorter) again and again comes about through simplifications of the phonological structure of words, and that the reduction of morphological complexity can, on the other hand, lead to
phonologically more complex word structures. There is a certain balance between phonological and morphological complexity, which eludes McWhorter because he sees phonological complexity exclusively in terms of phonemic inventory. Thus, relevant phonological changes, difficult to motivate independently, are evaluated as leading to an increase in systemic complexity even though they may decrease morphological complexity. In markedness or also optimality theory such interdependencies are accounted for by assuming conflicting markedness principles or constraints.

10. In order to explain the differences in complexity between creole and non-creole languages, as he sees them, McWhorter (with reference to Lightfoot 1999: 250) suggests the following considerations: “One might stipulate that after countless millennia of usage and drift, we might expect that by a certain point all grammars had, by the sheer dictates of change, developed various random complexities in parts of their grammars. This might follow from the mounting evidence of the inherent tendency of natural systems to complexify with the passage of time according to apparently universal principles of self-organization” (Section 2.3).

While this hypothesis appears to be compatible with McWhorter’s own comparisons of creoles with non-creoles, it is hardly consistent with what is known about language change in general, and for a good reason: unlike for instance biological organisms, human language is not a natural system in the true sense. The principles of self-organization, which ensure ever better adaptation to the environment and hence necessarily lead to an ever more complex internal structure, do not apply to it. Human language is, in Keller’s terminology (1990: 83), a phenomenon of the “third kind”. It is characteristic of such phenomena that, like natural phenomena, they are unplanned, while at the same time, like artifacts, they are the result of human action. Strictly speaking, we cannot say that a language, of its own, DEVELOPS. It is formed by its speakers through change, not according to a plan but not lacking direction either (Wurzel 1997). Grammatical change is always “improvement” in respect of a given parameter – which also motivates change in the first place (Vennemann 1989).

As a consequence, “improvement” on one parameter very often, though not always, entails “deterioration” on another, introducing new complexity. Nonetheless, the evidence from attested language histories suggests that such newly acquired complexities themselves can be gotten rid of again. And why indeed should languages accumulate complexity, especially when it is dysfunctional? More than 100 years ago, Hermann Paul cogently described in his Principien der Sprachgeschichte how morphological and phonological complexity can not just enter the system but eventually are also removed again, compensated for by increases and decreases in complexity elsewhere in the language system.
McWhorter is not in fact unaware of such interdependencies, for he notes the “secondary” effects of flexional morphology on the complexity of grammar. Evidently, the development from Proto-Indo-European to the modern Indo-European languages, or from Latin to the modern Romance languages, led to massive decreases of overall complexity in these respects. The diversity of synthetic case forms and declensions classes were replaced by uniform and none-too-numerous prepositions, the number category of the dual was abandoned, etc. It is conceivable that such simplifications entail compensating complications elsewhere, but this is not a logical necessity.

Rather obviously, such developments from synthesis to analysis, by no means unique to Indo-European, are at odds with McWhorter’s hypothesis of an inevitable gradual increase in the complexity of languages in the course of their history. I believe that this is indeed the general picture whenever we look at long-term language histories or when we compare individual consecutive stages in the recorded or reconstructed histories of a language. If overall complexity does not perforce increase in “normal” language development, but rather on the contrary, this also throws doubt on McWhorter’s structural identification of creoles as being less complex (having not had sufficient time to develop).

11. This does not mean, however, that aging processes are of no significance for the structure of languages and their grammatical complexity. But matters are far subtler than McWhorter’s formulation suggests (cf. Wurzel 1998). Indisputably, non-creole languages have much longer histories in comparison with creoles, and are in this sense “old”. All the same, even old languages may contain young grammatical material, for instance forms ensuing from recent grammaticalizations. Also, aging primarily affects individual forms and structures rather than languages as a whole. Like in modern typology, it is therefore advisable in diachrony too to focus on individual variables first, and only then draw conclusions about systemic co-variation or co-development.

A few typical examples from inflectional morphology must suffice to show how individual grammatical phenomena age and what consequences this has for complexity.

As one would expect, “young” inflectional forms and paradigms, innovated not too long ago, like those of Turkish noun inflection, are agglutinative – and thus do not add complexity in McWhorter’s sense. According to McWhorter, the factors contributing to complexity include the morphophonemics (non-uniformity) of the stem and allomorphy of inflections (presence of inflection alternations and inflection classes, disregarding “automatic” alternations such as those of vowel harmony, as in Turkish); and one might add the occurrence of fusions at morpheme boundaries (restriction of morphological transparency).
Complicating non-uniformity of stems is caused by phonetically determined phonological changes, as in the case of the Germanic umlaut, as mentioned by McWhorter. Significantly, however, history does not end here. There is a strong universal tendency for morphologically arbitrary stem alternations subsequently to be eliminated from paradigms – either by levelling or by morphological systematization. Both ways can be observed with umlaut in German (and in other Germanic languages). Already in early Old High German, the alternation in the paradigms of weak masculine nouns was completely undone:

(2) SG

<table>
<thead>
<tr>
<th></th>
<th>NOM</th>
<th>hano 'rooster'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN</td>
<td>henin &gt; hanin</td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>henin &gt; hanin</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>hanun</td>
<td></td>
</tr>
</tbody>
</table>

PL

<table>
<thead>
<tr>
<th></th>
<th>NOM</th>
<th>hanun</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN</td>
<td>hanōno</td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>hanōm</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>hanun</td>
<td></td>
</tr>
</tbody>
</table>

Also in Old High German, the paradigms of masculine i-stems are regularized so that a singular without umlaut contrasts with a plural with umlaut, with plural thus uniformly marked through umlaut:

(3) SG

<table>
<thead>
<tr>
<th></th>
<th>NOM</th>
<th>gast</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN</td>
<td>gastes</td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>gaste</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>gast</td>
<td></td>
</tr>
<tr>
<td>INS</td>
<td>gestiu &gt; gastiu</td>
<td></td>
</tr>
</tbody>
</table>

PL

<table>
<thead>
<tr>
<th></th>
<th>NOM</th>
<th>gesti</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN</td>
<td>gestio</td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>gestim</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>gesti</td>
<td></td>
</tr>
</tbody>
</table>

Both ways, morphological complexity that had been occasioned by phonology is eliminated or reduced.

Inflection classes can also be brought about by phonological changes. Thus, problematic though the reconstruction is (cf. Szemerényi 1996: 160), it is assumed for Indo-European that at the proto stage the same set of case-number inflections was used with all nouns – with no diversification of inflection classes and hence no morphological complexity of this kind. This uniformity of inflection was eliminated, and complexity correspondingly increased, through the working of phonological changes that caused reductions and fusions of inflections proper with stem-forming elements. (Similar developments of inflection classes can be observed, for instance, in Finno-Ugric.) Again, however, history
continues. In all languages with inflection classes, there are trends towards uniformity of inflectional forms (“one meaning – one form”). In English, for example, the *s*-plurals and *s*-genitives, originally restricted to certain declensions, were generalized, so that with the exception of a few remnants declensional variety no longer exists. Similarly in the continental Scandinavian languages, the *s*-genitive occurs today (as possessive) in all nouns not only in the singular but also in the plural (Wurzel 1984: 140):

\[
\begin{array}{llll}
\text{Old Swedish} & \text{Modern Swedish} \\
\text{SG NOM} & \text{kvinn-a} & \text{kvinn-a} & \text{‘woman’} \\
\text{GEN} & \text{kvinn-u} & \text{kvinn-a-s} \\
\text{PL GEN} & \text{kvinn-u} & \text{kvinn-or-s}
\end{array}
\]

Such levelling renders inflection more uniform, hence reduces morphological complexity.

Limitations of morphological transparency due to fusions at morpheme boundaries also come about through phonological processes, occurring again and again in particular languages. Fusions as illustrated in (5) were very frequent, for example, in Middle High German verb forms:

\[
\begin{array}{llll}
\text{Old High German} & \text{Middle High German} \\
\text{leite-n, preterite} & \text{ich leite-ta} & \text{leit-en, ich leite} & \text{‘I guided’} \\
\text{ahtō-n, preterite} & \text{ich ahtō-ta} & \text{aht-en, ich ahte} & \text{‘I valued’}
\end{array}
\]

Such morphologically “bad” (complex) forms, with morpheme boundaries no longer identifiable, were subsequently improved again – cf. Modern High German *leit-en, ich leit-ete; acht-en, ich acht-ete*.

What these cases have in common is that morphological complexity first emerges and is then reduced again. More specifically, phonological markedness is first reduced at the expense of morphological markedness, and then vice versa (cf. Wurzel 1994: 90–93). That is, morphological complexity in McWhorter’s sense does not continually increase in normal language history, but effectively returns to the original level. Constantly undergoing “rejuvenation”, forms with a long history thus end up having a “young”, non-complex structure (Wurzel 1998: 141). This applies to particular forms in an inflectional system; the entire system does not necessarily have to undergo wholesale and fast reduction of morphological complexity.

12. Now, how can McWhorter’s assumption that creoles display relatively little complexity in comparison to older languages, which I share, be reconciled with my conclusion that the assumption of constantly increasing complexity in language history is untenable?
For complexity in inflectional morphology, the argument might go as follows. Due to their specific genesis, creole languages have little or no inflectional morphology. Any incipient inflection that might occur, attesting to the general tendency of isolating structures to become morphologically bound in the first phase of what has been called the “typological spiral” (Gabelentz 1901: 255–258), will as a rule be young, hence agglutinating. A case in point is the morphologized tense-aspect-mood marking in the Portuguese-based “late creoles” of West Africa (Thiele 1989). Creoles will thus have no or very little morphological complexity. As agglutinating languages with young inflectional morphology, which initially shows little morphological complexity (few morphophonemic stem alternations, virtually no suppletion, little marker allomorphy except through vowel harmony – as in Turkish, remaining at this stage perhaps longer than usual), are developing further, inflection will become more and more fusional, with complexity thereby massively increasing. (This developmental phase is exemplified by Finno-Ugric, with Estonian and Livonian having progressed furthest on this route; cf. Korhonen 1996: 208). Continuing on the typological spiral, fusional inflection will tend back towards isolation, and morphological non-complexity, eventually eliminating inflectional exponents, with secondary agglutinative structures (as seen above for Swedish noun inflection (4)) as possible intermediate stages.

Advancing on the typological spiral takes time. Morphological complexity is bound to emerge sooner or later; but once it is there, it will be hard and even more time-consuming to get rid of again. For example, in the 5000 or so years since Proto-Indo-European, little of what has been accumulated in complexity has been reduced in such conservative languages as Icelandic or Russian; and even languages which have lost much of their former inflection, such as English or French, still show more complexity than creoles especially in what has remained of verbal inflection. A comparison of the history of Indo-European with that of Finno-Ugric suggests that reducing inflectional complexity takes much more time than building it up. Thus, between two and three thousand years ago Proto-Saamic was purely agglutinative, while its modern descendants are essentially fusional (cf. Anttila 1989: 301, Korhonen 1996). A possible reason for such differences in the rate of opposite developments is that phonological change, bringing about such morphological complexity as manifests itself in inflection classes and stem alternations, usually affects entire classes of words, whereas complexity-reducing morphological change, such as cutting down on inflection classes through changes in class membership and levelling out stem alternations, usually affect only individual words or small groups of words.

If reducing morphological complexity takes longer than building it up, it follows that creoles relatively rich in inflection (like the Portuguese-based ones of West Africa) will still be less complex than many old languages, where
Commentary on McWhorter: Wolfgang Ullrich Wurzel

complexity has had more time to come and go.

It remains to be seen whether complexity increase and decrease in domains other than inflection can be accounted for analogously.

13. In my comments I have only addressed issues which I consider problematic from the point of view of a historical linguist of the markedness persuasion. There is much else in the target article that creolists, historical linguists, and typologists will find worth reading – and perhaps disagreeing with.

Received: 4 May 2001 Zentrum für Allgemeine Sprachwissenschaft
Revised: 27 June 2001

The proofs came too late for the author to approve or disapprove of my occasional rephrasings of his and his consultant’s English. On the phone in hospital, he had given me card blanche; that was to be our last conversation.

Wolfgang Wurzel died on 4 August 2001, just 61.

As was easy to see for discerning phonologists and morphologists wherever, he wrote the best that was to be had in phonology and morphology in Germany, following in Hermann Paul’s footsteps. That will now be missed greatly, as will he.

References


Creoles, complexity, and linguistic change


