Complexity and the age of languages

Umberto Ansaldo & Sebastian N. Nordhoff

Abstract

This paper addresses the issue of complexity in language creation and the time it takes for 'complex' structures to emerge in the history of a language. The presence of morphological material is often equated to a certain degree of complexity or is taken to signify a certain time-depth in the history of a language (e.g. Dahl 2004; McWhorter 2005). Though this assumption may be seen as trivial in the absence of a theoretically-based definition of complexity (Muysken 1988), or even misleading (Aboh and Ansaldo 2007; Farquharson 2007), we here put it to a test by looking at morphology in a relatively 'young' language, namely Sri Lanka Malay (SLM). SLM is a mixed language which shows considerably more morphological material and other signs of old age than 'prototypical' creoles. We explain this by arguing (a) that structural output in language genesis is closely motivated by the typology of the input languages and (b) that our understanding of rate of change needs to be revised to take into account ecological matters.

1. Foreword[1]

In this paper, we engage with the alleged relationship between a certain type of structural complexity and the time it takes in the history of a language for such structure to evolve. At least one of the above-mentioned authors has argued against an objective notion of linguistic complexity and so have many others (for a comprehensive discussion see *Linguistic Typology* 2001 vol. 5.2-3); in general, in order to be measurable, complexity needs to be defined on the basis of a specific theoretical module (Muysken 1988). It seems that, for the moment, there is no agreement amongst linguists as to what such a definition should be. For example, it is a clear result of a Eurocentric linguistic theory to view tone languages as phonologically complex; as we will see below, precisely such an assumption is part of one of our current definitions of complexity (e.g. McWhorter 2005).[2] It is legitimate to doubt how phonologically complex tone languages actually are to, say, a Thai child.

Conversely, to claim that case morphology is 'complex' may suggest that it is more difficult to acquire a case paradigm than a set of adpositions to indicate grammatical relations. More importantly, it is unclear how complexity within different modules of language interact with one another, i.e. whether we can talk of a morphologically simple but semantically complex language vs. a morphologically complex but semantically simple one and which wins in the overall assessment. Finally, it may be the case that complexity of production leads to simplicity of parsing and vice versa, which makes it difficult to talk about complex and simple languages as a whole.

In what follows, we do not address the issue of how complexity should (or should not) be understood; in section 2, we assume the validity of, in particular, one view of complexity, namely the one put forward in Dahl (2004), that allows us to at least entertain the possibility that certain types of structural features may require time in order to emerge. In section 3 we look for signs of age, i.e. 'mature' features in Dahl's sense, in Sri Lanka Malay (SLM), a language that, as we show, is historically relatively 'young'. The concepts of young and old are related to the issue of rate of change in section 4. In the light of all the observations gathered, section 5 reviews the notion of complexity that emerges from McWhorter's 'creole prototype' idea, and evaluates its relevance for our understanding of the relationship between age and structural properties in language (McWhorter 2005).

2. On complexity and new language formation

In *Defining Creole*, McWhorter (2005) presents the fundamental argument that links his notion of simplicity to time-depth. For McWhorter, "creoles are indisputably new languages" (2005: 10) *because* they all share the common feature of lacking signs of old age. Typical features of old grammars in McWhorter's view are:

1. Inflectional affixation intended as a morphological, as opposed to an abstract, UG feature, which arises over time if a free morpheme is reanalyzed as grammatical and becomes bound;

2. Tone systems, which emerge through tonogenesis from phonetic erosion, specifically grammatical tone systems involving contrastive monosyllables and

grammatical functions;

3. Derivational noncompositionality, arising through semantic drift.

These features are allegedly by-products of language change but they are not necessary for basic communication as *they are not inherent to UG* (McWhorter 2005). Therefore, they only grow over time and are an indication of emergence of complexity (or overspecification). What makes new languages such as creoles 'simple' (see McWhorter 2001) is the fact they they typically lack all or most of these features. Conversely, presence of these features can be used as a metric of complexity.

In a broader investigation of complexity, Dahl (2004) identifies a series of 'mature' phenomena in languages. Maturity is defined by Dahl (2004: 105) as follows:

> "*x* is a mature phenomenon iff there is some identifiable and non-universal phenomenon or a restricted set of such phenomena *y*, such that for a language L, if *x* exists in L there is some ancestor L' of L such that L' has *y* but not *x*."

The types of mature phenomena that can be identified according to this definition include (see Dahl 2004: 114-5):

1. Complex word structure (inflectional/ derivational morphology, incorporation)
2. Lexical idiosyncrasy (gender, INFL classes, case)
3. Syntactic phenomena depending on inflectional morphology (agreement, case)
4. Word order rules over and above internal ordering of sister constituents
5. Specific marking of subordinate clauses
6. Morpheme and word level features in phonology

However, Dahl also notes that, while it may be possible to reason about old age, it is quite unlikely that we may ever be able to reason about the youth of a language. In discussing pidgins and creoles, he notes that:

> in order to be defined as a creole, a language must have as its primary historical source a language which has a sufficiently simplified grammatical structure [a pidgin]. No grammatical property of a language can therefore be a counterexample to the thesis that

> creoles have the world's simplest grammars, because in order to be a creole, the language has to originate from an earlier language state which did not have that property (a pidgin), and if it has it there are only two logical possibilities: either that stage did not exist, in which case it is not a creole, or the property has been acquired later, in which case it is not a counterexample either, since it just means that the language is on its way to losing its creole character (2004: 111).[3]

In what follows, we present an example of a contact language that can be taken as 'young', being at best 350 years old, which would make it comparable to 'prototypical creoles', but which has a substantial amount of signs of old age. This is not intended as a counterexample to McWhorter specifically, for two reasons:

1. As explained by Dahl (see above), and contrary to McWhorter's own claim (e.g. 2005: 140), the Creole Prototype idea is *not* falsifiable.

2. A language which is structurally similar to SLM,[4] namely Sri Lanka Creole Portuguese, is described by McWhorter as having inherited its features of age through intimate contact with the inflected adstrates (Sinhala and Tamil).

Indeed we agree with the characterization provided by McWhorter in (b) above also in relation to SLM. What we aim to show is that this line of reasoning can be extended to all types of contact languages, and that the relation between time and form is not insensitive to, but rather dependent on, the typological ecology of the contact situation.

3. Complexity in Sri Lanka Malay[5]

SLM can be said to show signs of 'age' in at least six different domains:

1. Bound affixes from former free forms: these affixes can be stacked yielding an agglutinative structure;
2. Inflectional/ derivational morphology;
3. Signs of incipient agreement;

4. Arbitrary subcategorization;
5. Complex negation patterns;
6. Difference between main clauses and subordinate clauses.

In what follows we discuss the different manifestations of complexity in turn.

3.1 Agglutinative structure

SLM has substantial bound morphology that is arranged in an agglutinative way. Example (1) shows use of a nominalizing suffix, an infinitive marker, case marking and coordinating association, all expressed by bound morphology:

(1) *Rani=pe* *thaandak-an=na=le* *Farook-pe*
Rani=POSS dance-NMLZR=DAT=ASSOC Farook-POSS
nyaani *mə-dingar=na=le* *suuka*[6]
song INF-hear=DAT=ASSOC like
'I like Rani's dancing and to hear Farook's singing.'

(2) *paasir* *mə-kumpul-kan=nang=jo* *se* *Dubai=nang* *em-pi*
sand INF-much-CAUS=DAT=FOC 1S Dubai=DAT PAST-go
'It was to collect sand that I went to Dubai.'

If we accept the view that languages start off as simple and acquire accretion (or 'grammatical junk') over time, the expression of several concepts – up to a total of four (e.g 2) – in one phonological word is not something that we would locate in the early stages of a language's development.

3.2 Inflectional/ derivational morphology

It can be argued that example (1) shows derivational morphology, which according to McWhorter (2005) does not necessarily qualify as a sign of old age. The same holds for clitics, which are less 'morphological' than true affixes. But that still leaves us the infinitive marker *mə-*, which is clearly inflectional. The prefix *mə-* shown in example (1) is of Malay origin (either from transitivizer *meN-*, or from a volitive modal,

Slomanson 2006)[7] but has been reanalyzed to serve the infinitive function it has today. While Sinhala and Tamil both make use of infinitive constructions, it should be noted that the SLM infinitive differs in its distribution from both languages. The infinitive is used in Sinhala and Tamil as the complement of all modals, which is not the case in SLM (compare (3) with (3') and (3'')).

(3) *Se masthi (*mə-) pi* (SLM)
1S must INF go
'I must go.'

(3') *Mama yanɖa oonä* (Sinhala)

(3'') *Naan poha veɳum* (Tamil)
'1S go.INF must.'

As shown in Table 1, other inflectional affixes are also found in SLM, in particular TMA markers:

Table 1. Some Sri Lankan Malay TMA markers

Past	**Conjunctive Participle (CP)**	**Progressive**
anna-, su-	*s-*	*ara-*

Example (4) shows the use of the CP[8] and the progressive prefixes. Examples (5) and (6) show the other prefixes.

(4) *nyaakith oorang pada s-pi thaangan ara-cuuci*
sick man PL CP-go hand PROG-wash
'The patients come and wash their hands.'

(5) *kitham=pe* *aanak=pada=le* *karang* *bae=nang*
1PL=POSS child=PL=ASSOC now good=DAT
cinggala *su-blaajar*
Sinhala PAST-learn
'Now our children learn Sinhala well.'

(6) *Se ana-laaher* *inni* *ruuma=ka=jo*
1S PAST-be.born this house=LOC=FOC
'It was in this house that I was born.'

Considering that no other Malayic language shows TMA prefixes (see Slomanson 2006), these would have evolved in SLM in the last 350 years. We can still find traces of the unbound proto-forms, as we can see in Malay **ada* 'exists' > *ara-* and Malay **habis* (habitual marker) > *s-*. It should be noted that *anna-* is already an archaic form, as younger speakers tend to realize only either the first or the second syllable, the most reduced forms being *e-* and *n-*. These phenomena can be classified under (i) in Dahl's list, namely instances of complex word structure.

3.3 Agreement

An excellent example of grammatical accretion is agreement. There is no semantic need to express a referent twice to communicate propositional content, yet this is precisely what agreement does. In SLM, we can observe incipient agreement. Next to the subject NP, the subject can optionally be expressed by an additional cliticized pronoun (Corbett 2006: 99f.). This clitic can be proclitic (7) or enclitic (8):

(7) *Dr Draaman dua thawon blaakang* *incian=se-nniinggal*
Dr Draaman two year after 3S. POLITE =PAST-die
'Dr Draaman died two years later.'

(8) *spaaman* *awuliya su-jaadi=spaaman*
3S.POLITE saint PAST-become=3S. POLITE
'He became a saint.'

This type of agreement is not attested in the other Malayic languages. While agreement is still rare in the SLM of today, and only found in few speakers, it is the beginning of a phenomenon that can be classified under (iii) in Dahl's list, i.e. inflectional morphology for syntactic phenomena.

3.4 Arbitrary subcategorization

Another example of 'maturation' is the arbitrary subcategorization of lexemes. There is no communicative usefulness in, for example, certain German prepositions requiring the dative (*mit* 'with') and others the accusative (*ohne* 'without')[9]. The same is true for the SLM modal particles. Modal particles are a class of lexemes that carry modal information; in SLM their distribution is completely arbitrary: for example, *boole* 'can' and *therboole* 'cannot' pattern in a different way, as shown in Table 2.

Table 2. Subcategorization of modals. The first two columns show whether the modal assigns nominative or dative case. The next two columns indicate whether it is used preverbally or postverbally. The last columns show whether the verb appears in the infinitive or the bare form when combining with the modal.

	Assigns		**Position**		**verb form**		
	Nom	**Dat**	**Prevbl.**	**Postvbl.**	**Inf.**	**Bare**	**Engl.**
masthi	+	–	+	–	–	+	must
anthi	+	–	+	–	–	+	IRR
boole	–	+	+	+	(–)	+	can
therboole	–	+	+	+	+	–	cannot
mau	–	+	+	+	+	–	want

This behavior can of course be attributed to negation, since asymmetric negation is common in the languages of the world (Miestamo 2005), but this does not explain why *mau* 'want' patterns with the negative form and not with the positive one. Moreover, as shown in Table 2, *masthi* 'must' patterns in a third way, which is different again. In this respect, it behaves exactly like *anthi* 'IRR', but there seems to be no semantic motivation behind this morphological similarity. As Table 2 clearly shows, SLM modals have an idiosyncratic subcategorization pattern, which can be regarded as accretion. Again, it

should be noted that the modal subcategorization pattern cannot be attributed solely to Sinhala or Tamil influence, as both these languages have conflated the categories of desire and obligation into one morpheme (Sinh: *oonæ*/ Tam: *veɳum*), illustrated below in (8) and (8'):

(8) *Mama yanɖa oonä* (Sinhala)

(8') *Naan poha veɳum* (Tamil)

1S go.INF must/want

'I must go/want to go'[10]

This is not the case in SLM, where obligation is expressed by *masthi* and desire by *mau*[11]. As is clear from Table 2, these two particles show a quasi complementary behavior in their subcategorization patterns. Nearly everywhere where *masthi* has a +, *mau* has a – and vice versa. At least one of these patterns must thus represent a language-internal development. Since we can be sure that the language of the first Malays in Sri Lanka had neither case marking nor infinitive, this subcategorization pattern must have developed within the last 350 years. We can see this as an instantiation of idiosyncratic lexical selection, Dahl's second criterion.

A second instance of lexical idiosyncracy is the subcategorization of highly transitive verbs, of which some select the accusative (9) and some the dative (10):

(9) *Se Farook=yang ara-buunung*

1S Farook=ACC PROG-kill

'I kill Farook'

(10) *Se Farook=nang (*=yang) ara-puukul*

1S Farook =DAT ACC PROG-hit

'I hit Farook'.

3.5 Negation

Yet another example of mature structure can be found in the highly complex system of negation in which the parameters of tense, mood and verbal, nominal or adjectival

predication type play a role in the choice of negative morphemes.

Table 3 illustrates some negative morphemes in SLM, while (11) – (15) provide examples from natural discourse:

Table 3. Negative morphemes

	Verbal	**Nominal**	**ADJ**
Past	*thera*-V	N *bukkang*	ADJ *thraa*
Perfect	V *thraa*	N *bukkang*	ADJ *thraa*
Present	*thama*-V	N *bukkang*	ADJ *thraa*/*thera*-ADJ
Future	*thama*-V	N *thama jaadi*	*thama*-ADJ

(11) *Puaasa muusing* *thəra-duuduk=si*
Fasting season NEG-stay=INTERR
'You were not here in the fasting period, were you?'

(12) *Invitations* *daapath* *thraa*
invitation get NEG
'They had not received the invitations'

(13) *Mulbar* *thama-oomong*
Tamil NEG.NONPAST-speak
'They do not speak Tamil'

(14) *Sindbad the Sailor hatthu muslim, mlaayu bukkang*
Sinbad the sailor one Moor Malay NEG.NONV
'Sindbad the Sailor was a Moor, he was not a Malay.'

(15) *Kluumbu bissar* *thraa*
Colombo big NEG
'Colombo was not big.'

Sinhala and Sri Lankan Tamil have two different negation strategies for verbal and non-verbal predication (Gair 1998).[12] If we disregard *bukkang* and the nominal predication as also found in other Malayic varieties, and therefore not of more recent development, we still have to explain why SLM makes use of three different negative morphemes for verbal predications, where Sinhala uses only one, *nää*, and Tamil two, *ille* and *maTTeen*. Additionally, there is the suppletive form *mau/thussa*[13] 'want/don't want' and *jon=*, which marks negation in nonfinite clauses (participle, infinitival and imperative (Slomanson 2006); this adds up to six different negative morphemes. It is clear that this differentiation is a sign of maturation that typically involves a certain amount of redundancy. As for the previous cases of morphological elaboration, these negation patterns would have evolved within the last 350 years.

3.6 Subordination markers

A final example of maturity, captured in Dahl's fifth criterion, is specific marking of subordinate clauses. This clearly exists in SLM in the form *katha* given in (16):

(16) *Se=pe orang thuuwa pada anna-biilang [kitham=pada*
1S=POSS man old PL PAST-say 1PL=PL
Malaysia-dring anna-daathang katha]
Malaysia-ABL PAST-come QUOT
'My elders told me that we had come from Malaysia.'

We can see that the clause final *katha* marks the non-matrix sentence. We now move on to consider what this all means for our understanding of structural complexity in language creation.

4. Discussion

With regard to data such as those presented in section 3, McWhorter comments that "Creoles with a moderate degree of inflection…have long existed in intimate contact with inflected superstratal or adstratal languages" (2005: 317). As we mentioned in section 2, we fully agree on this point. Following up on it, we put forward a related claim

as its logical extension:
Creoles with clearly isolating morphology have long existed in intimate contact with isolating superstratal or adstratal languages.

The implication of our claim is that the 'typical' structural features of a number of creoles are first and foremost a result of the input languages in the contact situaton.[14] This claim is highly plausible in the light of gradualist models of genesis which, at this point in time, appear to be well grounded in historical and typological investigation (Arends 1989, 2001b; Chaudenson 2001; Ansaldo 2008), as opposed to abruptist models which, by now, has been seriously called into question (as shown in Roberts 1999, 2000, 2001; Siegel to appear).

In what follows, we would like to go beyond Dahl's scepticism regarding whether it is possible to talk about youth of language, and propose that it is questionable whether one can talk about age of language – in relation to structural properties – at all. In order to do so, we need to start with the crucial distinction between change in I-language, i.e mental grammars, and change in E-languages, i.e. socio-historical entities (Lightfoot 1999; DeGraff 2001). Such a distinction is crucial in order to understand rate of change in new language creation.[15]

4.1 Rate of change

Weinreich, Labov and Herzog (1968) already distinguish between the individual and the community as clearly different levels of linguistic analysis. Elaborating on this, in a biological view of language change, one can distinguish between grammar as a generative, 'mental' concept, and language as a population of idiolects and/ or utterances (Paul 1880; Keller 1990/1994; Lass 1999; Lightfoot 1999; Croft 2000). Based on this distinction, one can identify two different levels of analysis: (a) the psychological,[16] and (b) the social. As already pointed out in Lass (1999: 370), in discussing rate of change, it is imperative to clearly identify which domain is being analyzed. Though there may be a relation between what goes on in the minds of speakers and the verbal interaction in which they engage, it is difficult to imagine that the principles that underlie I-language would be identical to those underlying E-languages. The former, however defined, would pertain to the realm of the psychological, or the cognitive (biological grammars in Lightfoot 1999: 101); while the latter would be defined by social and historical contingencies.

Ansaldo (2009), following a suggestion in Lightfoot (1999), argues that change in I-language may be viewed as either abrupt or gradual, depending on the generative framework that is being upheld.[17] When we turn to changes that affect a language as a social construct on the other hand, we are really talking about *propagation* of variables (Croft 2000). Propagation is by definition a gradual phenomenon; whether it occurs more or less rapidly depends simply on (a) the size of a community, (b) the type of social network and (c) the presence or absence of normative constraints on the natural creativity of speakers (such as cultural scripts, educational institutions etc.; see also Milroy and Milroy 1985; Trudgill 2001; Ansaldo 2009). In other words, a given change can register and stabilize in a small group of close-knit individuals who do not share or uphold specific normative contraints within one generation. On the other hand, for a change to spread and be registered in a large, diffuse population, several generations will be necessary.

Rate of change is crucially linked with age of language, as time is regarded as a prerequisite of complexity from an evolutionary point of view (DeGraff 2001). Note that whether it really makes sense to talk of 'a language' in discussing E-languages has been repeatedly questioned (e.g. Lightfoot 1999; DeGraff 2001), and there is good reason for this. What we are really referring to when we talk about the age of, say, English, is extremely arbitrary as it rests on a number of historical and cultural assumptions that are neither objective nor linguistically indisputable.[18] Be that as it may, let us assume that we somehow know when a language is a language, and that we want to determine whether it is young or old. So let us compare a typically 'old' language with the 'young' SLM we have discussed above.

4.2 Old and young

Chinese is considered a very old language, having existed for at least two and a half millennia – what does this mean? It means that we have documents available written in a logographic script that is the precursor of Chinese characters; these texts have been historically ascribed to the beginning of Chinese civilization. Note that there is plenty of structural change between what has been reconstructed as Ancient Chinese and modern varieties. These changes affect phonology, morphology and syntax, to the point that, as is the case for all the languages for which we have good historical documentation, questions of external contact and doubts about genetic classification have been raised

(e.g. Sagart 1994; Ansaldo 1999, for Sinitic).[19] The cultural continuity of Chinese is likewise questionable: the earliest texts available, partly attributed to the Daoist school of thought, portray an anarchic social structure with a prominent role of divination and naturalist philosphies, while later, Confucian and Mohist texts, decribe well-organized feudal states, the rule of law and centralized authority. In short, the idea of Chinese being an old language is a historical and cultural construct; we would have the same idea about Mongolian if the Mongol dynasties that ruled China for over a millennium had stayed in power to this date, and if they had displayed a more colonial linguistic policy. The age of Chinese does not exist objectively or in abstraction. What exists is the time span between the present and the earliest records available.

Sri Lanka Malay can be considered young because no SLM diaspora existed until four hundred years ago, when people of Malay/ Indonesian origin were brought to Sri Lanka. To be careful, we suggest an age of roughly 300 years for the language. What we claim, based on socio-historical analysis, is that the alleged complexity of SLM would have taken 300 years to evolve. Claiming that it might have taken even less would only strengthen the claim that 'complexity' is not a sign of old age; in the case of SLM, however, we do not see any reasons to invoke rapidity. What does this mean? It means that SLM has been spoken in various forms, probably involving more Malayic elements first, and progressively substituting these with Lankan features, for roughly that period of time. It is the SLM diaspora (also some kind of socio-historical construct) that can be said to be roughly 350 years old. Therefore the variety that emerged as a result of the socio-cultural focussing (Le Page and Tabouret-Keller 1985) that characterizes the SLM diaspora may have roughly the same age. Does this make it a young grammar? If we view grammar as I-language, it doesn't really matter, as each grammar is created anew in the speaker, and we can therefore not compare different mental grammars in terms of age (unless we compare child to adult grammar, see Lightfoot 1999). The very notion of UG denies the possibility that it would take speakers of different languages more or less time to acquire their own variety, especially in the light of present-day relativistic stances regarding the notion of markedness (Rizzi 2001; Haspelmath 2006). If we view grammar as a social and historical construct, SLM may be young, but, as we show, this says nothing about the *grammatical properties* of such language, since similar properties exist in languages that we are perfectly comfortable in calling 'old', e.g. Malay, Dravidian and Indo-Aryan languages.

If we are tempted to still think that it seems like it all happened a bit too fast, we need to remember that propagation of new variants in a relatively small, dense and non-normative ecology does proceed more smoothly than in the written histories of national languages (Nettle 1999; Croft 2001). As noted in Ansaldo (2009), in the study of language histories, abruptness may also be a reflection of the lack of gradual evidence, i.e., as Darwin suggested, a measure of our ignorance. In other words, 'E-change' must always be more or less gradual,[20] as processes of grammaticalization and/ or reanalysis proceed in steps, and as each new feature needs to pass from user to user until all or most have accepted it; this will be quick if few and close speakers are involved (say a family), slow if we track the propagation of an innovative feature across generations and countries (say a language family).[21] Age, we suggest, should always be related to the external history of a language, and never attributed to grammatical features of the language, since age of 'grammar' is but an epiphenomenon of the length of documented histories (in the sense of E-language), or identical for all languages (in the sense of I-language).

5. Final remarks

5.1 On type of grammar and typology

Let us recall Dahl's (2004) definition of 'mature' phenomena given in section 2, which may be loosely interpreted as the accumulation of (redundant?) material (lexical and or morphosyntactic) in a grammar G' that was not present in the grammar of G. This accumulation can – in a way in which age and structure are relatable – be explained in various ways, including the increase of accuracy of a certain utterance, the weight that the utterance may signal to the hearer, etc. (see Dahl 2004: 5-17). It may also be historical accident, e.g. the layering of variables coming from regional varieties, social contingencies that lead to stable ditaxia etc. (Ansaldo 1999). Surely this type of maturation is possible in an environment where grammars of similar typological make-up have been, because of the socio-political histories of their speakers, connected to one another horizontally as well as vertically. This is however not *necessarily* so: for example, the history of Chinese does not show many of the signs of maturation discussed in Dahl (2004). Why would this be the case? The list of mature features presented above

is repeated here for convenience:

1. Complex word structure (inflectional/ derivation morphology, incorporation)
2. Lexical idiosyncrasy (gender, INFL classes, case)
3. Syntactic phenomena depending on inflectional morphology (agreement, case)
4. Word order rules over and above internal ordering of sister constituents
5. Specific marking of subordinate clauses
6. Morpheme and word level features in phonology

As we would expect, features (1) – (3) are largely absent in Chinese, as the language is strictly isolating and, rather than morphology developing over time, there is significant evidence that there is morphological loss in the (pre)history of Sinitic (Sagart 1999). There is however an attested history of development of stricter word-order rules as well as development of word level features in phonology. What this shows is that, though it may be possible that certain features of a grammar are signs of accumulation of historical layers, whatever accumulates depends on the basic typology from which we depart; in other words, the types of features that accumulate are, at least in part, language-type-specific. Let us see what this means for languages with a younger history.

5.2 On age and creoles

We have seen that SLM shows considerable 'complexity', intended as accumulation of mature features according to Dahl (2004), in several domains of its grammar. It seems unwarranted to call SLM a simple language, independent of any definition of simplicity, even allowing for the possibility of such a definition to be valid objectively. While certainly being far from the morphological intricacies of Georgian or Tuscarora, we must not shy away from a comparison with some older languages like other Malayic varieties (or English for that matter) as far as complexity is concerned.

We know that the language the first Malay immigrants to Sri Lanka used for in-group communication did not show the features that are discussed in this paper. Of course, early SLM migrants had command of their own native language, which might have been complex or not, but one of the codes that was available to the early migrants as a means of inter-ethnic communication, and would feed into SLM, was a reduced one, Trade Malay. Trade Malay had the main characteristics of a pidgin, i.e. a reduced

function, namely communication for commercial purposes, and morphological and stylistic poverty. In today's SLM, we find lexical traces from a number of places in the Indonesian archipelago (Adelaar 1991; Paauw 2004), but grammatical traces are very rare. There is no classifier system, no reduplication, and a complete absence of diathesis as a grammatical category. If we count from the time when the Dutch conquered Ceylon in 1656, when we know for certain that the SLM ancestors started arriving in Sri Lanka, this gives us the time span of 350 years for today's morphology to emerge.

Let us now investigate the sample[22] used by McWhorter in developing the notion of Creole Prototype (McWhorter 1998); note that, crucially, the sample consists mainly of languages *older* than 300 years (with the exception of Tok Pisin). Out of 8 languages investigated, all have a Indo-European lexifier (4 Germanic, 4 Romance), 7 have a West-African substrate, only Tok Pisin has no West African substrate, and Mauritian has Bantu and Malagasy influence besides West African. The nature of this sample is of course due to the fact that little is known about 'young' languages in other parts of the world, but this justification does not add to the validity of the sample. If we took Chinese, Lao and Thai to show that languages older than 1,000 years do not have inflection, this would obviously be seen as an opportunistic sample. The sample would have been chosen in order to satisfy the hypothesis, and furthermore it would have a strong areal and genetic bias. If a serious claim about the correlation between age and simplicity has to be undertaken, then a sample has to be employed that is genetically and areally more stratified than this one (see Dryer 1989, 2000, 2003).

From a typological perspective, McWhorter's sampling method is clearly biased: West African languages are not known for their rich inflectional morphology, neither are English or Dutch.[23] A language that develops in contact with those languages should not be expected to be morphologically hypertrophic. The claim that the languages he investigated are simple according to his criteria is likely to be an artefact of the sample to begin with. As we have shown in this paper, and as is predicted in section 4, had the sample included young languages in a morphologically rich environment, like South Asia, we might have never related 'simplicity' to young languages. And indeed, the one language that developed in an ecology rich in morphology, namely Tok Pisin, shows more morphology at the age of only roughly 100 years than the other languages after 300 years. Ecology has a far greater influence than the time elapsed since the birth of the language (Mufwene 2001) for the structural input-output relation.

What this study has shown is that an objective notion of simplicity has not been

put forward so far. Conversely, it is questionable whether we have an appropriate definition of complexity. What we do see is that typology matters; in an evolutionary framework for language change (Croft 2000; Mufwene 2001), the notion of Feature Pool supports the idea that in contact environments speakers derive their new grammars by a process of competition and selection of existing features. Therefore, in a morphologically rich environment, morphology will emerge. In a typologically isolating environment, this would obviously not be the case.[24] Just as simple does not imply 'creolization' (Ansaldo and Matthews 2001; Sampson 2006), 'complex' does not seem to imply old age. We have also supported claims against abrupt language creation, showing how both generative and sociohistorical frameworks of language change converge in disallowing the possibility of essentially differential speed in the development of grammar; speed correlates with size and type of community, but not with abstract structural processes.

References

Aboh, Enoch O. and Umberto Ansaldo. 2007. The role of typology in language creation: A descriptive take. In Umberto Ansaldo, Stephen Matthews and Lisa Lim (eds.). *Deconstructing creole*. Typological Studies in Language 73. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 39-66.

Adelaar, K. Alexander 1991. Some notes on the origin of Sri Lankan Malay. In Hein Steinhauer (ed.). *Papers in Austronesian linguistics No. 1.* Pacific Linguistics A-81. Canberra: The Australian National University. 23-37.

Ansaldo, Umberto 1999. *Comparative constructions in Sinitic: Areal typology and patterns of grammaticalization*. Ph.D. dissertation. Stockholm University. Stockholm: Allduplo Tryckeri AB.

Ansaldo, Umberto 2007. Review of John McWhorter 2005, Defining Creole. *Journal of Pidgin and Creole Languages* 22.1. 170-176.

Ansaldo, Umberto 2008. Sri Lanka Malay revisited: Genesis and classification. In K. David Harrison, David S, Rood and Adrienne Dwyer (eds). *Lessons from documented endangered languages: A world of many voice*s. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 13-42.

Ansaldo, Umberto 2009. *Contact languages. Ecology and evolution*. Cambridge: Cambridge Unversity Press.

Ansaldo, Umberto and Stephe Matthews. 2001. Typical creoles and simple languages. The case of Sinitic. *Linguistic Typology* 5.2/3: 311-326.

Ansaldo, Umberto, Stephen Matthews and Lisa Lim (eds.). 2007. *Deconstructing creole*. Typological Studies in Language 73. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Arends, Jacques 1989. Syntactic developments in Sranan. Ph.D. dissertation. University of Nijmegen.

Arends, Jacques 2001a. Simple grammars, complex languages. *Linguistic Typology* 5.2/3: 180-181.

Arends, Jacques 2001b. Social stratification and network relations in the formation of Sranan. In Norval Smith and Tonjes Veenstra (eds.). *Creolization and contact*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 291-308.

Chaudenson, Robert 2001. *Creolization of language and culture*. London: Routledge.

Corbett, Greville 2003. *Agreement*. Cambridge: Cambridge University Press.

Croft, William 2000. *Explaining language change: An evolutionary approach*. Longman.

DeGraff, Michel 1999. *Language creation and language change: Creolization, diachrony and development*. Cambridge MA: MIT Press.

DeGraff, Michel 2001. On the origin of creoles. *Linguistic Typology* 5.2/3: 213-210.

Dahl, Ősten 2004. *The growth and maintenance of linguistic complexity*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13: 257-292.

Dryer, Matthew S. 2000. Counting genera vs. counting languages: A reply to Maslova. *Linguistic Typology* 4: 334-350.

Dryer, Matthew S. 2003. Significant and non-significant implicational universals. *Linguistic Typology* 7: 108-128.

Farquharson, Joseph 2007. Creole morphology revisited. In Umberto Ansaldo, Stephen Matthews and Lisa Lim (eds.). *Deconstructing creole*. Typological Studies in Language 73. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 21.38.

Foley, William 2006. Universal constraints and local conditions in Pidginization. Case studies from New Guinea. *Journal of Pidgin and Creole Languages* 21.1: 1-44.

Gair, James and Barbara C. Lust (eds.). 1998. *Studies in South Asian linguistics: Sinhala and other South Asian languages.* New York/Oxford: Oxford University Press.

Gil, David 2001. Creoles, complexity and Riau Indonesian. *Linguistic Typology* 5.2/3: 325-370.

Grant, Anthony 2007. Admixture, structural transmission, superficial and substantive creolization. In Umberto Ansaldo, Stephen Matthews and Lisa Lim (eds.). *Deconstructing creole*. Typological Studies in Language 73. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 109-139.

Haspelmath, Martin 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42.1: 25-70.

Keller, Rudi 1990/1994. *On language change: The invisible hand in language.* London: Routledge. (Translation and expansion of *Sprachwandel: von der unsichtbaren Hand in der Sprache*. Tübingen: Francke).

Lightfoot, David 1999. *The development of language: Acquisition, change and evolution*. Oxford: Basil Blackwell.

Le Page, Robert and Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.

McWhorter, John 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5.2/3: 125-166.

McWhorter, John 2005. *Defining creole*. Oxford: Oxford University Press.

Miestamo, Matti 2005. *Standard Negation: The negation of declarative verbal main clauses in a typological perspective.* Berlin/New York: Mouton de Gruyter.

Milroy, James and Leslie Milroy 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21: 339-384.

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Muysken, Pieter 1988. Are Creoles a special type of language? In Newmeyer, Fredrick (ed.). *Linguistics. The Cambridge survey. Vol. 2. Linguistic theory: Extensions and implications*. Cambridge: Cambridge University Press. 285–301.

Nettle, Daniel 1999. *Linguistic diversity*. Oxford: Oxford University Press.

Newmeyer, Fredrick 1988. On the supposed 'counterfactuality' of Universal Grammar. Some evolutionary implications. In James R. Hurford, Michael Studdert-Kennedy and Chris Knight (eds.). *Approaches to the evolution of language*. Cambridge: Cambridge University Press. 305-319.

Paauw, Scott 2004. A historical analysis of the lexical sources of Sri Lanka Malay. MA thesis, York University.

Rizzi, Luigi 2001. Broadening the empirical basis of universal grammar models: A commentary. In Michel DeGraff 1999. *Language creation and language change: Creolization, diachrony and development*. Cambridge MA: MIT Press. 453-472.

Roberts, Sarah J. 1998. The role of diffusion in the genesis of Hawaiian creole. *Language* 74: 1–39.

Roberts, Sarah J. 1999. The TMA system of Hawaiian Creole and diffusion. In J. Rickford and S. Romaine (eds.). *Creole genesis, attitudes and discourse: Studies celebrating Charlene J. Sato*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 45–70.

Roberts, Sarah J. 2000. Nativization and genesis of Hawaiian Creole. In John McWhorter (ed.). *Language change and language contact in pidgins and creoles*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 257–300.

Sagart, Laurant 1990. Old Chinese and proto-Austronesian evidence for Sino-Austronesian. *Oceanic Linguistics* 33.2: 261-193.

Sagart, Laurant 1999. *The roots of Old Chinese*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Sampson, Geoffrey 2006. Does simply imply creole? In Mickael Suominen (ed.). *A man of measure: Festschrift in honor of Fred Karlsson on his 60th birthday*. Turku: Linguistic Association of Finland. 362-374.

Siegel, Jeff 2000. Introduction: The processes of language contact. In Jeff Siegel (ed.). *Processes in language contact: Studies from Australia and the South Pacific.* Sant-Laure Quebec: Fides. 1–11.

Siegel, Jeff 2007. Recent evidence against the Language Bioprogram Hypothesis. The pivotal case of Hawai‘i Creole. *Studies in Language* 31.1. 51-88.

Slomanson, Peter 2006. Sri Lankan Malay morphosyntax: Lankan or Malay? In Ana Deumert and Stephanie Durrleman (eds.). *Structure and variation in language contact*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 135–158.

Smith, Ian and Scott Paauw 2006. Sri Lanka Malay : Creole or convert? In Ana Deumert**,** Ana and Stephanie Durrleman (eds.). *Structure and variation in language contact*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 159-183.

Trudgill, Peter 2001. Linguistic and social typology. In Chambers, J., P. Trudgill and N. Schilling-Estes (eds). *Handbook of linguistic variation and change*. Oxford: Blackwell. pp. 707-728.

Thomason, Sarah and Terence Kaufman 1988. *Language contact, creolization and*

genetic linguistics. Berkeley: University of California Press.

---

[1] We would like to express our gratitude to the following people/ bodies: the Kandy and Upcountry Malays for their support and collaboration in our fieldwork; the Volkswagen Stiftung's initiative for the Documentation of Endangered Languages (DoBeS) for funding our research on the project 'The documentation of Sri Lanka Malay'; Lisa Lim and Anthony Grant for feedback to the paper.

[2] We refer here to McWhorter (2005) as a collated exposition of his previous work on the Creole Prototype.

[3] The fact that McWhorter's assumption equating lack of specific features to youth would be hard to test has been a recurrent criticism of the model (see e.g. Arends 2001a; Ansaldo 2007), and counterexamples of non-young languages showing the same 'simplicity' (e.g. Gil 2001; Grant 2007) as well as creoles with inflectional morphology (DeGraff 1999, 2001; Farquharson 2007) have been put forward.

[4] But crucially not easily comparable in terms of genesis (Ansaldo 2009).

[5] The data in this paper come from the Upcountry variety of SLM.

[6] The SLM examples use a practical orthography, which is based on the orthography of Indonesian, with the following additional conventions: <th> and <dh> mark dental articulation (not aspiration), <d> and <t> mark retroflex articulation, <v> or <w> mark a labiodental approximant. When digraphs are geminated, only the first letter is doubled (<tth>, not <thth>, <nny> not <nyny>).

[7] Smith and Paauw (2006: 173) also propose the debitive *mesti* as a possible origin.

[8] Conjunctive participles are an areal South Asian feature. This form expresses that the action denoted by the verb in the participle clause took place before the action denoted by the verbs of the following clauses. A rough equivalent is English *'having done X…'*.

[9] The distinction between ACC and DAT is more transparent with local prepositions in German, where DAT denotes Essive and ACC Allative.

[10] Some grammars suggest that the use of a Dative pronoun in Sinhala (maTa) conveys desire while the Nominative conveys obligation, but this has not been confirmed by Sinhala/ Tamil informants. Obligation is marked by a focus clitic on the verb instead.

[11] It should be noted that some younger speakers permit the use of *masthi* for desire. However, *mau* is never possible for obligation.

[12] The Tamil variety Gair discusses is spoken by Hindus in Sri Lanka. Some pilot research (2007) into Muslim Tamil varieties points to the conflation in Muslim Tamil of the two markers Gair found in the Hindu varieties. Standard Indian Colloquial Tamil shows this conflation as well.

[13] Note that *thussa* is also likely of Malay origin [to'sa] = 'don't need/ have to'

[14] As pointed by Anthony Grant (p.c. September 2007), there are still contact phenomena that appear to fall out of this generalization and therefore require alternative explanations. For example, Chinook Jargon owes most of its basic vocabulary to Chinook, which is highly inflected, but none of Chinook's inflections are used productively in Chinook Jargon.

[15] Whether E- and I-language can actually be taken as viable conceptual notions or not is not at stake here; the distinction is introduced in order to clarify the potential (if at all) locus of rate of change and

its nature.

[16] Or cognitive level (we do not commit to a definition of this as we are not aware of any concrete/ accepted one).

[17] For example, Lightfoot (1999) argues for abrupt change as a matter of parameter-resetting, while Newmeyer (1998) suggests an adaptationist view that implies gradual evolution of grammar.

[18] A typical example of this may be seen in the controversy regarding New English varieties, where some are considered as legitimate modern English varieties, e.g. Australian and New Zealand English, while others are still regarded as 'external' to the family, e.g. Indian English. This is clearly not a matter of linguistic similarities between each new variety and, say, RP, but rather an expression of cultural and racial differences.

[19] Similar issues regarding the evolution of English and possible 'creolization' are discussed in Thomason and Kaufman (1988).

[20] The more or less may depend on the relative markedness of a specific variable within the community.

[21] See also Arends (2001b).

[22] Ndjuka, Tok Pisin, Saramaccan, Haitian, St. Lucian, Mauritian, Fa D'Ambu and Negerhollands.

[23] The Romance languages have somewhat more morphology in their written form, although this is not necessarily so in the varieties spoken by the colonizers (Chaudenson 2001).

[24] This does not necessarily imply that specific simplification processes do not occur in certain contact situations: Siegel (2000) for example provides a good overview of processes involving also reduction of morphological inflection commonly observed in a number of pidgins. However, even in pidgin formation simplification is locally conditioned, i.e. it is a matter of typological input, as shown in Foley (2006) for the case of New Guinean pidgins.