

# Contact language formation in evolutionary terms<sup>1</sup>

**Umberto Ansaldo**

## **Abstract**

The aim of this paper is to present a view of contact language formation in which language creation in multilingual ecologies follows the same principles as language maintenance in monolingual ecologies, i.e. selection and replication of features available to speakers in a given environment. In order to do so, I introduce the foundations underlying an evolutionary framework to contact language formation and the views they offer for our understanding of language contact and change. The view of grammar as an evolving system, I believe, can be best appreciated in a functional-typological theory of language. For this reason, I first introduce the basic functionalist, usage-based linguistic theories required for an evolutionary framework. I then synthesize a view on language contact and change in evolutionary terms based on Croft (2000, 2006a) and Mufwene (2001). Finally, I apply the views presented here to a case of contact language formation, namely the evolution of case markers in a variety of Sri Lanka Malay. These are particularly interesting as, from a classic or orthodox view, they might be seen as ‘complex’, ‘marked’ or at least ‘unexpected’ instances of contact-induced change. The evolutionary framework however can explain these as natural acts of linguistic replication in multilingual settings, thus avoiding exceptionalist explanations. Instead, an evolutionary framework offers an integration of socio-historical and functional-typological observation, something that our current approaches to language change still largely lack (Croft 2006b). Among the advantages of the framework applied here, as discussed in the concluding section, is the suggestion that overall structural complexity, however defined, does not change as a result of contact language formation: a new grammar is simply the result of a recombination of grammatical features of the input languages.

---

<sup>1</sup> I would like to thank Enoch Aboh for his useful comments on this paper.

## 1. Introduction

One of the most fascinating aspects of linguistics is the diversity that is found, albeit decreasingly, in the languages of the world. One question that is still, in my opinion, a fundamental one in linguistics is: How did all these languages come about? Different answers are possible here but it seems clear that any answer will have to be able to account for one thing: the divergence (and convergence) of different languages. In other words, my aim is to contribute to our understanding of language speciation, which is ultimately a window into human history. In doing so, I assume that in the early history of humankind, there have been many episodes of communities being newly created, because of geographical separation, encounters with other populations, etc. Therefore, the study of how new languages emerge offers an important field for understanding how and why language speciation happens.

The area under investigation in this paper is the formation of new grammatical structure salient in – but not exclusive of – contact-induced change as can be observed in the formation of contact languages. Following Thomason (2003), no significant difference is assumed in the processes that lead to the traditional ‘types’ of contact languages (pidgins, creoles and mixed bilingual languages). In this sense, and following, among others, Mufwene (2001) and Ansaldo and Matthews (2007), such labels, where used, are purely for socio-historical identification. What I argue for in this paper is simple: (a certain version of) an evolutionary framework of language, change and contact offers an illuminating take on what to look for when diagnosing causes and effects of contact language formation (henceforth CLF). It may not force us to abandon other more orthodox approaches, but it adds a new dimension to our current understanding, in particular in offering an integrated approach to both socio-historical and functional-typological analysis. The rationale behind the wish for such a model of CLF rests on three, related ideas:

- (1) Whichever theory of language and grammar we subscribe to, we view language as a set of cognitive patterns that all humans have in common.

- (2) Multilingual ecologies have always been and to some extent still are the norm in human societies. The patterns of language use and language change observed in such environments, which include CLF, must be absolutely ‘normal’ and should not require exceptionalist explanations (Ansaldo and Matthews 2007).
- (3) It is widely acknowledged that language change and contact-induced change are brought about by speakers; thus, if we want to understand grammatical outcomes of language contact, we must integrate grammatical analysis with sociolinguistic theory.<sup>2</sup>

The framework presented here can be seen as a synthesis of already existing proposals.<sup>3</sup> My intention is simply to offer a non-exceptionalist scenario of CLF, i.e. to account for contact-induced change in ‘normal’ terms (see section 5; DeGraff 2001, 2003, 2005; Ansaldo and Matthews 2007). Exceptionalism and contact linguistics are closely related as a result of two problematic assumptions common within (historical) linguistic approaches: (i) that the types of change observed in monolingual environments, with normative sociolinguistic tendencies (e.g. standardization) and institutionalized education, are ‘normal’; and (ii) that, in reconstructing CLF, whatever cannot be attributed to the lexifier or the substrates must be UG. These, I argue, are questionable assumptions: as stated above and as argued, among others, in Mufwene (2001) and Ansaldo and Matthews (2007), schooling and linguistic homogeneity are recent, modern, ideological reflections of societies from which we should only theorize with extreme care (see also Bakhtin 1981; Kroskrity 2003). In addition, in many cases of CLF, we do not have a full set of documentation available in order to reconstruct the ecology in which a new grammar evolves. What cannot be reconstructed does not necessarily indicate UG,

---

<sup>2</sup> Even ‘internal’ types of change must be ultimately be brought about by speakers, i.e. through idiolectal variation (see Mufwene 2001).

<sup>3</sup> Moreover, it is a ‘lighter’ evolutionary scenario than the ones presented so far, as I only pursue the analogies between different types of evolution to the minimally necessary requirements; for example, I do not talk of competition, exaptation, memes, etc. This is a choice derived, in part, from what are a number of negative reactions to evolutionary analogies pushed too far (e.g. Andersen 2006).

universal cognitive patterns or other abstractions, but may simply indicate a gap in our knowledge (Arends 2001; Ansaldo and Nordhoff this volume). Instead, let us derive generalizations on CLF from well-documented cases where we do have reconstructions of almost complete ecologies. These generalizations can help us establish a not-too-speculative model that will help us reconstruct the process of CLF where ecological information is not fully available

## **2. On the history of evolutionary frameworks in linguistics**

Evolutionary theory in the history of (historical) linguistics is offered a brief and critical treatment by Joseph and Janda (2003: 7-11). As the current study falls within that apparently problematic tradition, a few words are required to orientate the reader. Perhaps not surprisingly, Joseph and Janda (2003) offer a critical view of organic metaphors of language arising in 19<sup>th</sup> century Europe (von Schlegel, Bopp, Schleicher, etc.) as basically mistaking language for a natural organism. With respect to these views, they oppose the enlightened view of language as a historical entity expressed, among others, by Bonfante (1946). A little more credit is given to a later attempt found in Lass, most likely the first in modern historical linguistics to present a view of “language as a population of variants moving through time and subject to selection” (Lass 1997: 377). But even Lass attracts their criticism, based on Milroy’s (1999) paper which basically identifies the major weakness in Lass’s proposal, namely the absence of the role of speakers in language change. Joseph and Janda however fail to appreciate what is the fundamental intuition in Lass, which, far from echoing organic metaphors of the past century, tunes in to very contemporary notions of evolutionary theory. Lass (1997) expresses a belief in a theory of a ‘historically evolved system’, in which both biology and language fall. In Lass’s view (1990, 1997), evolutionary theory may help us explain language change, not because languages are seen as biological systems, nor because language is seen as a genetic feature (in the Chomskyan sense). Lass’s idea, rather, is that linguistics and biology may be seen as systems that are in some ways similar, most

notably in the fact that they evolve over time.<sup>4</sup> The most important question we should worry about, as he aptly puts it, is the non-essentialist one:

...which (kind of) facet is the best one to look at given the particular epistemic game one happens to be playing? From this point it doesn't really matter if one is a realist or an instrumentalist, if one wants to grant priority to one particular facet or another. The point always is fruitfulness, not 'truth'. We are Model Builders. We are not (or should not be arrogant enough to think that we are) in the truth Business. (Lass 1997: 384-85)

Lass was partly inspired by Dawkins (1976), who suggested that his work on evolutionary biology could be applied to the evolution of cultural systems. The idea therefore is not that language, change and contact are the same as biological evolution, but rather that they share interesting, common properties worth exploring.

In a critical take on evolution applied to linguistics, Andersen (2006) highlights a number of mismatches between analogies drawn in the fields of evolution of language (change) and biological evolution. This is a common approach that fails however to understand (a) that it is not necessary to attain a total overlap between explanatory models in order for the analogies to be useful (Croft 2000, 2006a, b), and (b) that, at least in the sense of Croft (2000), it is the field of *conceptual evolution* – and not biological evolution – that provides the necessary foundations for a linguistic framework (see Dawkins 1976; Hull 1988). Moreover, Andersen reveals a strong belief in typical products of European historical linguistic scholarship, such as the *Stammbaum* model and the possibility of clearly distinguishing between mechanisms of change, e.g. neologisms, extension, reanalysis and adaptation. It is clear by now that the *Stammbaum* model cannot easily account for the speciation that has occurred in all language families (see e.g. Dixon 1997 for Australian languages), and competing models have existed for a long time (e.g. *Wellentheorie*). Regarding mechanisms of change – which are not discussed in this paper as I believe it is premature to generalize on this aspect of the model – a note is required:

---

<sup>4</sup> Note also that Croft (2000) develops a strong usage-based theory of change, recognizing the speaker as the central element, one of Joseph and Janda's (2003) *desiderata*.

typical mechanisms such as the ones listed above, are extremely difficult to keep apart in actual analysis, as can be seen in Andersen's own treatment (2006: 68, 72, 73), which reveals a certain awareness of how blurred these notions can be. Moreover, these mechanisms may serve us to some extent in discussing the histories of languages with relatively uncomplicated, prevalently monolingual ecologies and well-documented histories. This is however clearly not the case in CLF; if we proceed to generalize a set of tools developed to 'fix' 'marked' aspects of speciation (i.e. speciation of written traditions) to fit more 'normal' cases such as CLF, we would be moving in the wrong direction according to point (2) in section 1 above. From this short historical introduction, we derive three important criteria for a framework of (contact-induced) change:

- (i) it needs to recognize the historical dimension in language;
- (ii) it needs to recognize the role of the speaker;
- (iii) it should be fruitful in offering new insights through novel domains of analysis.

### **3. Necessary theoretical asides**

My framework is inspired by Croft's (2000) Theory of Utterance Selection (see section 4.2), which builds on Hull's (1988) generalized theory of conceptual evolution. I also re-interpret the notion of Feature Pool in language contact – developed in particular in Mufwene (2001) – in a purely functional-typological perspective. The theoretical foundations are functional-typological (following Bybee 1998, 2006; Croft 2000, 2006a) and usage-based (Tomasello 2003). The latter is a choice that derives from a philosophical orientation and is not a must; in other words, it should be possible to adjust the framework to different theoretical beliefs, as addressed further below.<sup>5</sup>

---

<sup>5</sup> This study is necessarily programmatic in nature, as it summarizes a framework that is still being developed and applied to a number of contact languages of Asia (Ansaldo in prep.). The reader will have to

### 3.1. Usage-based linguistics

Language can be viewed as a complex system in which the processes that occur in individual usage events, [...], with high levels of repetition, not only lead to the establishment of a system within the individual, but also lead to the creation of grammar, its change and its maintenance within a speech community. (Bybee 2006: 730)

Not so long after Chomsky's (1959) review of Skinner, it had already become evident (to some) that the notion of competence was too limited to account for what really constitutes the knowledge a speaker needs in order to be really competent (Gumperz 1964, 1968; Hymes 1971, 1972). In being concerned with what is normally termed 'performance', these and other authors looked at how to seriously integrate variation and diversity in a theory of language. Hymes (1971) for example replaces Chomsky's notion of competence as tacit knowledge of grammatical rules with that of communicative competence, and brings back a behavioral dimension of acquisition of competence. In this tradition, the unit of analysis is the linguistic community or speech community (Gumperz 1968), within which typically a number of varieties are represented; language is therefore regarded as a heterogeneous phenomenon and it is its diversity that becomes a prominent aspect of study (Hymes 1971).

In functionalist and cognitive theories of grammar, linguists do not isolate the structure of language from language use, as grammar is intended as the cognitive organization of a speaker's experience with language (e.g. Givón 1979; Langacker 1987). Grammar in both cognitive and functionalist theories is seen as a set of cognitive representations that rely on general cognitive abilities of categorization, generalization, representation, etc., and are therefore not language-specific. While this view has until recently not been the dominant one in linguistic theory, as opposed to the orthodox, structuralist view of

---

bear with a somewhat lengthy introduction, considering the brevity of the overall paper, followed by illustrations through a case study.

grammar developed in Chomskyan linguistics, in recent years a number of trends have been accumulating substantial evidence in favor of what is clearly a usage-based theory of language, in particular the field of grammaticalization studies, language acquisition and functional-typological linguistics (Bybee 1998). As noted in Bybee (1998), a very important discovery that these approaches share relates to how living languages create new grammar. Research in these areas over the past two decades shows that grammar evolves naturally out of pre-existing lexical material in language use, as is abundantly documented in the literature on grammaticalization (Hopper and Thomson 1993). Grammar is a constantly evolving system in this view and “there is every reason to believe that all existing grammar came about in just the same way we observe in the documented cases at our disposal...” (Bybee 1998: 250). Grammar therefore can be seen as an emergent system, complex and dynamic, similar to other complex systems observed in biology (Lindblom, McNeilage and Studdert-Kennedy 1984; Hopper 1987).

Probably the most systematic and compelling arguments to date in favor of a usage-based view of language can be found in the work of Tomasello (late 1980s onwards) in which sound empirical psycholinguistic evidence and serious theoretical engagement lead to a definition of linguistic competence that recognizes three phases (Tomasello 2003): (i) phylogenetic, as the biological adaptations enabling members of a species to communicate linguistically; (ii) historical, i.e. the process through which a set of conventions emerge in a given speech community; and (iii) ontogenetic, in the sense of the development that leads an individual to acquire full competence during life. Unlike generative grammar, which recognizes a ‘periphery’, a superficial domain of language that changes over time, and a ‘core’, the linguistic competence often referred to as UG, Tomasello presents a usage-based approach that recognizes the essence of language in its symbolic dimension, i.e. the ways in which humans use linguistic symbols for interpersonal communication. In accordance with much work in cognitive-functional grammar (e.g. Bybee 1985; Langacker 1987; Givón 1995; Croft 2001), patterns of usage are ritualized (or grammaticalized) into constructions, and new generations inherit such constructions through exposure. The ability of abstracting general grammatical rules, Tomasello shows, is enabled by biological cognitive skills that are not language-specific, but fall under the categories of ‘intention-reading’ and ‘pattern-finding’. Grammar is not

a biological adaptation but a historical and ontogenetic process. Crucial to this view is the notion of dual inheritance, which emphasizes that organisms inherit both their genes and their environments (e.g. Durham 1991). This applies to humans as well, but with a twist, as humans are adapted both for pre-existing structures in their environments as well as for acquiring completely new knowledge from the socio-cultural context (Tomasello, Kruger and Ratner 1993). Therefore humans acquire language because they are biologically prepared for it and because they are exposed to their own linguistic culture.

### **3.2. On usage and the role of frequency**

The above-mentioned approaches to language all share the basic assumption that structural aspects of language need to be explained by also making reference to linguistic function, or usage. This is because, as argued above, in usage-based linguistics, grammar is seen as a dynamic system which varies within the life-span of adult speakers; in this view, phenomena of language use influence the representations of grammatical knowledge. In particular, there is wide agreement that frequency plays a dominant role (Langacker 2000; Croft 2000, 2001; Bybee 2001, 1998, 2006). Frequency of usage is responsible for the depth of entrenchment and the degree of abstract generalization of grammatical rules in the evolution of grammar (Bybee 2006; Croft 2006a). As we will see below, frequency is a determining factor in CLF, as it allows us to understand what gets replicated and how the replication happens (see sections 4.1, 4.2, 5.1, 5.2).

Bybee (2006) offers a review of frequency effects relevant to processing and storage of grammar:

1. High-frequency words and phrases undergo phonetic reduction at a faster rate than low and mid-frequency sequences, e.g. *do not* > *don't*. This can be explained by the fact that articulatory representation depends on neuromotor routines which increase in fluency when repeated. The increase in fluency can lead to a group of words being identified as a single unit (also Anderson 1993).

2. High-frequency strings become more entrenched in their morphosyntactic structure and more resistant to restructuring. Frequency strengthens the memory representation of words, phrases and syntactic structures alike making them easier to access and more conservative (Bybee and Thompson 1997). Bybee (2006: 715) offers the example of high-frequency irregular verbs in English such as *keep, kept* that maintain the irregularity while low-frequency ones such as *reap, reaped* are more prone to regularize.
3. Related to the above, high-frequency strings can lose semantic and syntactic transparency and become autonomous (e.g. words with derivational affixes that increase in frequency lose transparent relations to their base forms, Bybee 1985).

Based on exemplar theory developed in particular for representation of phonetic variation (Pierrehumbert 2001, 2001), Bybee (2006) suggests the following observations about the relation between use and cognitive organization:

- (i) Organization, or storage in memory is probabilistic, not categorical. While an adult who experiences a new string will be only minimally influenced as s/he already possesses a wealth of older, more entrenched strings, a child, for whom most items are new, will be more influenced in its organization. Repetition or ritualization increases ease of access (Haiman 1994).
- (ii) Repeated linguistic experiences, just as experiences of other kinds, leads to strong memory of such experiences and loss of non-repeated ones (Goldinger 1996).
- (iii) Linguistic memories do undergo reorganization.

Frequency effects, as argued here and below, play a significant role in the evolution of grammar, independently of the ecological environment in which such evolution takes place. In particular, we can talk about two broad types of frequency (this is elaborated in sections 3.3, 3.4 and 4 below, see also Croft 2000):

- I. Token-frequency, i.e. the frequency of a linguistic item in discourse, texts etc.
- II. Type-frequency, i.e. the frequency of a certain construction in grammar(s).

### **3.3. Functionalism, variation and language change**

From a functionalist perspective, language and grammar are historical entities; it is necessary to embrace this view if we want to seriously embrace the dimensions of variation and diversity that characterize language use, in particular in multilingual, heteroglossic environments typical of contact languages. In Croft (2006a), language change is simply seen as variation at a broader level; his model distinguishes between three different orders of variation. First-order variants are individual variants that occur in all subdomains of grammar (phonological, semantic, syntactic etc.). When first-order variants take on a sociological significance in a community, they become second-order variation. When second-order variation becomes conventionalized in a speech community and leads to divergence, we have third-order variation (i.e. variation across languages). Here, as will be shown below, the study of grammaticalization and typological congruence becomes a central part of the explanation of new grammar creation.

### **4. A generalized view of change**

To understand how we arrive at an evolutionary view of language and language change, we need to first understand the origins of a theory of evolution applied to conceptual systems in general (and not limited to biology). An evolutionary synthesis of complex, evolving systems can be found in Hull (1988); the essence of Hull's position is presented below, followed by an application of his theory to language change as proposed in Croft (2000).

#### 4.1. Selection and replication

Hull (1988) is first and foremost aimed at resolving a number of controversies in evolutionary biology, specifically regarding the objects of selection in the neo-Darwinian debate. At the same time, Hull intends to offer a view of selection that transcends biology and can be applied to conceptual systems at large. For the purpose of this study, and for a general framework of language contact in evolutionary terms, it is not necessary to grasp the actual debate which gives rise to Hull's view, but rather to highlight the product of his work and how it relates to language change and language contact.

(i) The nature of change: As an evolutionary model, Hull (1988) is concerned with change by replication; change by replication is one of two (or more) possible types of change, the other being inherent change. While inherent change refers to transformation over time, replication is basically a type of copying. However, copying is not always perfect, as change can enter the replication process in various ways, resulting in *variation* and *differentiation*. In evolutionary biology, copying is an iterative process that is cumulative in nature and that results in a varied set of replicators, due to random mutation.

(ii) The trigger for change: The most important aspect of *differential replication* is *selection*. Selection is a consequence of the adaptations that occur through the *interaction of an organism with its environment*. We can therefore generalize with Hull (1988) as follows: iterated copying produces variation; environmental interaction can lead to differential replication.

(iii) The locus of change: Hull (1988) assumes a population theory of species. In a population theory, a population is defined as *a set of interbreeding individuals who are reproductively isolated from other populations*; it is within the species that change takes place. This is a departure from an essentialist view of species where a species is identified on the basis of a set of constant structural properties. A population is spatio-temporally

bound; there are no inherent properties that define the set but rather relational properties between members of the set.

With these basic notions in mind, let us see how they can be applied to language change, following Croft (2000, 2006a, b).

#### **4.2. Croft's evolutionary model of language and change**

Based on Hull's model of change by replication, Croft (2000) proposes the following generalizations (my adaptation). As we can see, points (a) to (e) derive directly from Hull's generalizations presented in the previous sections, as applications of his notions of change, replication and population to language:

- (a) In language, as in other complex evolving systems, change is caused by replication.
- (b) In communicating, speakers replicate linguistic features they have been exposed to. This replication can be more or less faithful, resulting in 'normal' or 'altered' replication. Altered replication means change.
- (c) Change happens at a population level. A population is a set of linguistically interbreeding individuals. The locus of language change is therefore the speech community or the social network.
- (d) Specifically, a language is interpreted as a population of utterances in a speech community (functional, usage-based theory of language, see section 2).
- (e) Grammars are spatio-temporally bound; they are basically idiolects and each speaker's knowledge may include more than one grammar.

Following from the above and from Hull (1988), a population can split into 'varieties' because of geographical isolation; this allows us to conceptualize a phylogenetic classification of languages. We can see that what is appealing in a population view of

language is the fact that a close linguistic equivalent of a biological population can be found in the notion of speech community (Croft 2000: 17-20).

In communication, simply put, Croft proposes that speakers engage in the following activities: they select (features of) utterances for replication; they replicate, and in doing so they might replicate identically or innovatively (i.e. ‘altered replication’, Croft 2000: 38 and chapter 3.3);<sup>6</sup> finally, they are involved in the propagation of replications within the speech community (2000: 4 and chapter 7.3). His view of language change is summarized in the *Theory of Utterance Selection* (TUS) (2000: 39):

- TUS does not preclude other selection processes from occurring at other levels of language.
- TUS does not require a specific set of causal mechanisms for the selection of utterances.
- TUS puts linguistic convention at the centre of language change: normal replication means conforming to conventions of the speech community, altered replication means departure from those conventions, i.e. innovation.

From the point of view of TUS, utterances matter, not sentences, meaning that it is the actual occurring linguistic features (texts, dialogues, etc.) that are relevant, not the potential instantiations of grammar. This is obviously a theoretical stance on what constitutes a grammar/ language. In Croft’s (2000) view, and in accordance with the theories illustrated in this section, languages are best understood as constantly evolving systems that defy taxonomic categorization. In this sense, language can be seen as a population of linguistic features and grammar as a combination of idiolects (section 4.3): communication thus entails interbreeding of different idiolects. This means that, when speakers interact, they exchange utterances; in such an exchange, they may replicate the

---

<sup>6</sup> It can be claimed that replication is never exactly ‘identical’, in the sense that there are not two identical idiolects of a language out there (Mufwene 2001). Therefore ‘identical’ here needs to be taken in the relative sense of a replication which only minimally modifies the input.

linguistic features of their environment (their community, network, tribe etc.) identically, or they may replicate them with alterations, innovations, etc. An *identical* replication is what we might expect in an environment with a high degree of monolingualism and strong normative tendencies, e.g. a clear notion of standard, compulsory, institutionalized education, etc. Note that language change nonetheless does occur in environments where contact is not present – or salient – which means that altered replication happens in language transmission between speakers in a homogeneous environment. It should follow then that, in a highly multilingual environment with low normative tendencies, an *altered* replication is in fact quite likely (Ansaldo 2008). In other words, in a typical situation of language contact, where speakers afford multilingual competence and where negotiation among different linguistic codes is the norm, altered replication can be expected to occur with rather high frequency. It seems therefore clear that, in order to understand CLF, we need to focus our attention on the dynamics of altered replication.<sup>7</sup>

For a framework based on selection and replication, it is important to clarify the following definitions:

1. The distinction between innovation and propagation is a crucial one that is absent in much literature on language change (but see Lightfoot 1997; Croft 2000; Ansaldo 2008). The general idea is that innovation is something that an individual speaker may do, potentially instantaneously, while propagation is something that happens to innovation at the level of population. The latter is, in accordance with functional-typological theory and grammaticalization studies, gradual.
2. Selection (and propagation) is not functional adaptation. Recall that Hull (1998) is not an evolutionary theory in the strict Darwinian sense; therefore there is no notion of selection and competition as ‘survival of the fittest’. In accordance with

---

<sup>7</sup> I want to stress how this is a theoretical stance that is, as all theoretical stances are, a matter of one’s philosophy of language. As Lass reminds us, we are not into the Truth Business; therefore a different take, say, a formal one, may also be possible. As we will see below, Croft (2000) identifies the utterance as the relevant unit that undergoes change in language use. This can be easily replaced by ‘sentence’ in a different theoretical orientation, and allows for other types of generalization to be invoked.

abundant sociolinguistic literature, selection of variables is socially determined, and may be related to issues of accommodation, imitation, differentiation etc.

3. The mechanisms of innovation are predominantly functional-typological (see section 3.3).

### **4.3. CLF as innovative replication**

In an evolutionary view of CLF, speakers in a multilingual community have a very heterogeneous set of utterances from which to extract material for replication. This framework assumes that an extremely low level of multilingual competence is sufficient for different linguistic systems to interfere in a speaker's use of language (already in Weinreich 1953), and that new language creation is the natural, creative process that typifies CLF (see Heine and Kuteva 2005).<sup>8</sup>

From what we have seen so far, an evolutionary view of CLF leads us to pay attention to the mechanism of replication and the types of the replication that occur, in particular altered or innovative replication. It also highlights selection and propagation as important areas to look into in order to understand how new variables move within a population. Following Hull (1988) and Croft (2000), and echoing views found, among others, in Thomason and Kaufman (1998) as well as Joseph and Janda (2003), an evolutionary view of contact-induced change acknowledges that it is speakers who change languages, not languages that change themselves. The main area of analysis for a proper understanding of language contact is therefore the socio-historical (and political, economic, cultural dimension). However, as suggested in Croft and as argued below, there is one phase of the process that can be looked at as partially independent from socio-historical dynamics, namely innovation. For the purpose of simplification, let us visualize the main phases of CLF as follows:

---

<sup>8</sup> This is not meant to imply that every multilingual society must lead to CLF. Obviously in societies where standardized varieties are institutionally supported and normative tendencies are enforced, the linguistic ecology may be stable and not particularly prone to high degrees of innovative replication.

- I. Selection: Whether speakers select certain features over others is predominantly a matter of social dynamics. There is plenty of literature on this topic, from variational sociolinguistics to the study of ethnolinguistic vitality to show that different forms of prestige play a role in CLF. For example, in the formation of Singapore English, notwithstanding the fact that English is a powerful lexifier with economic capital and institutional support, speakers selected abundant material from Sinitic varieties, due to the fact that these represent a substantial part of the population and do carry both overt and covert prestige (Ansaldo 2004; Lim 2007). Moreover, a few features from Malay were also selected at a point in time where Malay-based contact languages were present in the ecology of Singapore English, and due to founder effects they survive to this day (Mufwene 1996; Ansaldo in prep.; also Ansaldo, Lim and Mufwene 2007).
  
- II. Innovation: How variables interact with one another depends, at least in part, on their typological and functional properties (Croft 1995). This phase can thus be – to some extent – investigated on its own terms, though it is never completely independent from socio-historical principles. As I will argue below, frequency needs to be identified as a central factor behind creation of grammar, and related effects such as salience and congruence need to be taken into account. For example, in accounting for the emergence of case in Sri Lanka Malay, we can see how token frequency, i.e. frequency of occurrence, and type frequency, i.e. frequency that emerges out of typological congruence of two adstrates, lead to case features being selected in the replication process (Aboh and Ansaldo 2007; Ansaldo in prep.; and section 5).
  
- III. Propagation: Propagation happens at the broader level of social organization, and requires that we take into account social network dynamics such as density and multiplexity (Milroy and Milroy 1992). The observations here vary, from suggesting that small, tight-knit communities are more conservative to opposing views suggesting that change can spread fast in such communities (cf. Nettle 1997; Trudgill 2002). I suggest that both views are correct, in the sense that *small, tight-*

*knit communities afford a higher control on usage*, and can therefore intervene successfully on maintenance or change, depending on socio-historical situations. More importantly, speed relates to community type, as propagation can be fast in smaller and tighter groups, but is slower in large, diffuse communities. This explains why CLF, most typical of medium to small and often tightly-knit groups, has been often described as ‘rapid’ in the literature (see also Matras and Bakker 2003; Ansaldo and Nordhoff this volume).

Remember that the three ‘processes’ listed above are not really separate phases with a hierarchical ordering, as clearly explained for evolving systems in general in Hull (1988). Nor can we treat them as referring to different domains of analysis: just as the interaction between replication and evolution is independent of the levels of organization in evolutionary biology, so can selection, propagation and innovation occur simultaneously and iteratively at any level of organization. In particular, selection and propagation are really concomitant processes, and innovation is abstracted in order to allow functional-typological analysis to take place.<sup>9</sup>

Thus far, we have addressed two important questions. The first question is the *what*-question: what are we looking for when investigating CLF? This has been answered in evolutionary terms, identifying the process of replication as the driving mechanism of change and suggesting that differential replication, i.e. innovation, constitutes one of the salient aspects of CLF. We have also specified *where* such processes can be found, i.e. in the speech community and in language use (this leaves us with a number of possible analytical domains, see the next section). The *how*-question, I have argued, needs to be answered in functional-typological terms: since the relevant domains of enquiry are usage

---

<sup>9</sup> Note that I do not explicitly bring in the notion of competition in this approach. Obviously a type of competition takes place at least in the social (and political, economic, cultural) domain, which does impact on the selection of certain variables over others. One could also claim that frequency patterns do indeed represent a form of competition, as higher frequency items seem to ‘win’ over lower frequency ones. However, for the purpose of this paper, the notion of frequency seems sufficient to capture the dynamics of CLF I have in mind. We will surely need more on competition in the future.

and diachrony, we look at functional-typological properties of language as well as grammaticalization patterns (as a holistic study of emergence of new grammar). There is a fourth question that needs to be answered, for any framework to be worth considering, and this is the *why*-question. This will be addressed in the concluding section.

#### **4.4. The typological matrix and ecology**

The notion of the typological matrix (TM) is intended here as the total set of linguistic variables available to a group of people related by shared patterns of communication (a linguistic network/ a speech community). The reader may recognize here previous proposals such as Croft's (2000) notion of Lingueme Pool and Mufwene's (2001) Feature Pool. Some minor differences between TM and these alternative notions are the following:

- (1) Croft's notion of Lingueme Pool is justified by the desire to keep the genetic analogy as tight as possible and implies the notion of 'lingueme' (2000: 29) as the linguistic replicator. In genetics, it is recognized that genes come in highly organized structures (e.g. DNA strings) but, as Croft suggests, language change may not show such an organization. This applies even more so to CLF, where attempts to identify (typological) constraints have so far largely failed (Heine and Kuteva 2005; Koptjevskaja-Tamm to appear). Though I am not suggesting that there are no organizational constraints whatsoever in the TM, I want to suggest that there are few and very flexible ones.
- (2) Mufwene's Feature Pool is a looser concept than Croft's, implying issues of markedness as significant factors in determining the outcome of this process. Moreover, there is also a dimension of discourse salience (e.g. 'perceptual salience' and 'semantic transparency' in Mufwene 1991) that plays a role in the selection process. These are partly useful notions, as I will show below, but in the TM, frequency substitutes for markedness and salience.

The notion of TM simply suggests that, in CLF where typically variables of different typological origins interact, if it is not in the social history of speakers, it is in typological analysis that we find the key to understanding *how* new grammar emerges. This is particularly relevant in the context of this volume as a TM-based approach predicts that the output of a contact ecology is largely dependent on the input (see also Mufwene 2001). In other words, according to the TM, complexity or simplicity however defined do not dramatically increase or decrease in CLF, as the output is a recombination of variables of the input. Moreover, TM provides a convenient way to identify this particular version of evolutionary frameworks of CLF.

The notion of TM offers a convenient abstraction for the purpose of function-form analysis, and allows us to distinguish between socio-historical investigation, which is required to understand selection and propagation, and actual innovation, i.e. different patterns of recombination of variables. Remember however that these phases are not entirely independent from one another and that innovation does depend on selection and successful propagation.

In analyzing TMs, the most salient factor is frequency. As we saw in section 2, from a functional-typological perspective, frequency is the dominant factor in evolution of grammar, and it comes in two types (Croft 2000, 2006a):

- (i) Token-frequency: This can be seen as discourse frequency. Linguistic items that are frequent in discourse may be those that are grammatically obligatory, semantically salient or pragmatically more relevant (Aboh and Ansaldo 2007). One such case is Experiencer marking in SLM, as will be shown in section 5.
- (ii) Type-frequency: Linguistic items that are type-frequent in CLF are those constructions that are more common ('unmarked') in a specific grammar, for example, because of typological congruence, where the occurrence of the same type in two adstrates reinforces its presence in the TM, as in the case of Dative markers in SLM, as will be seen in section 5.<sup>10</sup>

---

<sup>10</sup> This proposal was still being developed at the time of writing. See Ansaldo (2009) for a fuller account.

Note that frequency patterns, as argued in Haspelmath (2006), can usefully replace the typological notion of markedness in all its linguistically relevant uses, whether in the sense of common, complex, overtly marked or otherwise. It therefore also replaces my earlier use of ‘local markedness’ as *relative to a single linguistic system*, in CLF (Aboh and Ansaldo 2007).

## **5. Application**

In this section, I look at cases of altered replication involved in the formation of Sri Lanka Malay. By applying the notions of selection, propagation and innovation introduced above, I account for the emergence of what appear as ‘complex’ constructions as predictable instances of innovation given the ecology of the contact language and its related TM. First the (subset of the) TM of a variety of Sri Lanka Malay, namely Kirinda Java (KJ), is identified based on the principles of selection. Then the relevant domain of the TM is investigated to show its internal dynamics. As we will see, the evolutionary framework is particularly suited to unveil the interaction between the socio-historical principles and the functional-typological dynamics that characterize the three ‘phases’ introduced above.

### **5.1. Kirinda Java**

Kirinda Java (KJ) is the southernmost variety of Sri Lanka Malay (SLM) (Ansaldo 2005a, 2008). Though the differences between varieties of SLM may not be overwhelming (Nordhoff in preparation), the Kirinda community is a distinctive speech community clearly ‘isolated’ from other communities (Lim and Ansaldo 2006). As first noted in Hussainmiya (1987), the restructured Malay spoken in Sri Lanka has a fully-developed case system, a marked feature for Malayic languages which typically lack case altogether. Ansaldo (2005a, 2008) describes the Malay of Sri Lanka as a mixed language of trilingual base, in order to capture the fact that this is a heavily ‘restructured’ variety that evolves in a multilingual situation where Sinhala, Tamil and Malay mix in almost all

aspects of grammar. Lexically these varieties show heavy maintenance of Malayic words and a pronominal system of clear pidgin-Malay-derived nature (PMD, see Adelaar and Prentice 1996). These are pronominal forms originally from Hokkien (Sinitic) found in many contact varieties of Malay in the region, a strong evidence of a Founder Principle effect in this case (Mufwene 1996). From the point of view of complexity, as already argued in Aboh and Ansaldo (2007), as well as Ansaldo and Nordhoff (this volume), a case system is clearly a ‘complex’ type of development in CLF and is typologically marked within Austronesian languages (see also Dahl 2004). This study however is not a full description of the case system of KJ (see Ansaldo 2005a; Aboh and Ansaldo 2007; as well as Smith et. al 2004 for general SLM case features). I focus here only on the core cases in KJ, in order to apply an evolutionary framework to the evolution of case in CLF.

## 5.2. What goes into the matrix

As argued above, speakers and their socio-history are the salient factors behind change, which occurs through replication. The environment in which KJ evolves is a multilingual speech community without particular segregation between the Malay, Sinhalese and Tamil ethnic groups. The history of the early Malays in Sri Lanka is one of active interethnic exchange and communication (Hussainmiya 1987, 1990); in an evolutionary sense, this means free mixing as elements from the three different languages are and were present in the every day communication between speakers of Malay, Sinhala and Tamil.<sup>11</sup> Therefore, a priori, we must assume that Malay (colloquial Malay), Sinhala and Lankan Tamil features were all available for selection and replication to the (early) Malay immigrants in Sri Lanka.

However these features were not *equally* available for selection. In Sri Lanka, Sinhala was and is the language of the majority, with political and economic prestige. Because of numbers and prestige, we have to assume that Sinhala features had very high frequency in the ecology in which Sri Lanka Malay evolved; frequency was identified in

---

<sup>11</sup> Interbreeding is strictly communicative; it does not suggest actual sexual mating between the different groups. Though this is also a possibility, it is, in the case of Sri Lanka Malay, a very unlikely one based on historical documentation (see Ansaldo 2008, cf. other proposals, e.g. Smith et al. 2004).

section 3 as the dominant force behind usage and therefore replication. This means that Sinhala features were more likely to be selected. At the same time, we cannot disregard the potential vitality that the ancestral language, in this case colloquial form(s) of Malay, carries in many contact environments, i.e. the wish to maintain features of one's own ethnic variety. Therefore, so far, we may want to assume good reasons for early Malay speakers to accommodate to Sinhala, considering the basic tendency of speakers to accommodate to the stronger 'norm'; in other words, selection of Sinhala features by Malay speakers should be expected.<sup>12</sup> But we also need to factor in a potential counter-tendency: the wish to maintain one's own language, i.e. selection of Malay features. The third language, Tamil, was, just as it is today, in an inferior position to Sinhala, numerically and economically, as the language of a predominantly lower-class minority, and was therefore undoubtedly less prestigious (Ansaldo 2008). Tamil is thus also present in the ecology and, though less likely to be a candidate for selection, cannot be totally excluded.

Let us now abstract the dimension of functional-typological analysis and reflect on innovation: what can happen in the TM of KJ, where Sinhala, Malay and (to some extent) Tamil features 'interact'? Sinhala has high token-frequency (numerical dominance), which results in salient discourse presence of Sinhala features. Here Tamil starts playing a role, as Sinhala and Tamil have converged typologically over the past thousand years and share many areal features (Masica 1978). From the typological point of view, Tamil reinforces the type-frequency of Sinhala features, where the grammars correspond. For example, Sinhala is SOV, Tamil is SOV, and Malay is SVO. A TM with these word orders would clearly imply the possibility that the output be an SOV language, because one of the SOV orders is numerically dominant in discourse, and the two congruent orders 'gang-up' and are selected (as well as propagated). Indeed the basic word order of KJ (and other varieties of Sri Lanka Malay) is SOV (though variation occurs, of course). Another example is found in morphological processes: the TM has [inflectional (Sinhala)

---

<sup>12</sup> I do not mean to suggest that overt prestige is the only form of prestige that matters in CLF. Covert prestige can be just as influential in selection, but it does not seem to apply to this particular scenario (see also Ansaldo 2008).

+ inflectional (Tamil) + mildly agglutinative (Malay)] morphology; the outcome is an inflecting language, as argued for KJ in Ansaldo (2005a) and other SLM varieties in Ansaldo and Nordhoff (this volume). Where they do not show congruence, this may either undermine the frequency, by offering another alternative for selection, or simply be irrelevant. Moreover, we have to consider functional properties of specific features, i.e. whether they are obligatory in discourse or highly regular, as this increases frequency or ‘visibility’ in terms of selection (Aboh and Ansaldo 2007 and below).

### 5.3. Selection and innovation

In KJ Nominative case is used primarily for prototypical agents and is zero-marked. Dative is the most frequently expressed and functionally diverse case, marking predominantly Experiencer roles (the ‘non-nominative subject’ of South Asian languages, see Bhaskararao and Subbarao 2004). Accusative case is highly optional, mostly used for definiteness and/ or emphasis (see examples 1-3 below). As mentioned above, case is not a feature of Austronesian typology and is absent from Malay varieties. It is clear that case features enter the TM of KJ from Sinhala and Tamil. The functions of core cases in these languages is summarized in Table1:

Table 1. Case alignment in Kirinda Java

CASE	MAIN FUNCTION	BEHAVIOR
Nominative	Agent [+ control]	
Dative	Experiencer/ Patient/ Goal [ control] (Sin. and Tam.)	obligatory
Accusative	+ Animate (Sin.) + Definite (Tam.)	optional

As we can see, there is a significant functional overlap between the cases of Sinhala and Tamil and the ones of KJ. Cases are realized as post-nominal suffixes, showing at least partial structural overlap as well, as in examples (1)-(3):

- (1) *master=nang pena mau*  
 teacher=DAT pen want  
 ‘Teacher wants a/the pen’
- (2) *pohong=yang potong*  
 Tree=DOM<sup>13</sup> cut  
 ‘(I) cut the tree’
- (3) *go=dang minum mau*  
 I=DAT drink want  
 ‘I want/ would like a drink’

We can say that the grammar of case is ‘Lankan’ (Sinhala/ Tamil) but the case suffixes are of Malay origin. *-nang* (in 1) is a directional preposition in Malay; *-yang* (in 2) is most likely derived from a Malay colloquial definite marker *-nya* (Uri Tadmor p.c. 2004; for more on this see Ansaldo 2005a; Slomanson 2006); *-dang* (in 3) is a variant of *-nang* found in the pronominal system (most likely due to Animacy, see below). It is clear that case-marking in KJ evolves as innovative replication in a TM in which case becomes a very frequent option, in the sense that it reflects the linguistic usage of a majority of speakers because it is high in both token- and type-frequency. The specific frequency patterns in this TM can be explained as follows:

(i) Dative case is assigned identically in Sinhala and Tamil: we have a complete functional overlap as it covers the same semantic roles of Experiencer, Goal and (certain types of) Patient (see Ansaldo and Aboh 2007). This gives it high type-frequency and therefore high visibility. It is obligatory (or regular), which results in extremely high discourse (or token-) frequency and strengthens its dominance. Moreover it is used to encode the Agent-Experiencer opposition, which seems like a rather salient communicative function; this means visibility to the speaker.

---

<sup>13</sup> Following Ansaldo (2005a, b) what is normally seen as Accusative case is here glossed as ‘definite object marker’, with possibly emphatic function.

(ii) Accusative case shows only partial semantic overlap in the adstrates and is highly irregular; therefore, while token-frequency may be high because of a numerical dominance of Sinhala (and to a lesser degree) Tamil utterances in the environment, type-frequency is much lower than Dative. This is related to and enhanced by the fact that the functions it marks are less communicatively crucial. Animacy in Sinhala is usually only marked in high-register variants (Gair and Paolillo 1997) and definiteness is not a constantly salient feature (Accusative marking cross-linguistically is mostly optional, Blake 1994); this also reduces the token-frequency of the Accusative.

The result is that, in the grammar of KJ, Dative case is a regular, frequent and robust feature, while Accusative is a highly irregular, elusive, optional feature, as shown in examples (1) and (3) above. From these observations we derive that a framework such as the one tested here seems to allow us to reflect not only on what is realized through the innovation process but also on how specific features may be realized.

Finally, it is important to note that spatial adpositions are recognized as the most common source of case markers in grammaticalization studies; in this case, as assumed in general in this paper, there is no difference between the processes of change that occur in CLF and elsewhere in historical linguistics (Traugott and Heine 1991; Hopper and Traugott 1993). Moreover, the construction follows predictable typological motivations: the post-nominal position of case markers is coherent with the verb-final word-order.

#### **5.4. A note on propagation**

Having reflected on selection and innovation, we now briefly consider propagation. As noted above, there is another crucial dimension to propagation, namely network type. Considering the case system as a whole, it may appear remarkable that such a system may evolve in 300 years of history. Indeed, as also discussed in Ansaldo and Nordhoff (this volume), such a complex set of features is not supposed to arise ‘quickly’ in evolution of grammar (Dahl 2004), and not at all in CLF (McWhorter 2005). But, as argued in section 3, propagation is a matter of social dynamics, not a structural process

with an inherent clock. The Malay communities of Sri Lanka are, to this day, small, tightly-knit but open communities. In such communities, I suggest (and see Matras and Bakker 2003 for similar claims on ‘mixed languages’) that enforcement of usage is highly effective, whether it is innovative or conservative (cf. Nettle 1997, cf. Trudgill 2001; and see Ansaldo in prep.). This suggests that speed is to a large degree relative to type and size of speech community.<sup>14</sup>

## 5.5. Summary

The evolution of a complex feature of grammar such as case in CLF can be accounted for within an evolutionary framework based on principles of selection, innovation and propagation, with the help of functional-typological analysis of the matrix. In this way, sociohistorical dynamics and functional-typological features are integrated within the same framework. By relying on theories of grammatical evolution and sociolinguistic principles that underlie communication at large, an evolutionary framework suggests that the process of language creation is the same across societies, but that different ecologies lead to diverse outputs. A generalized approach like this is not only desirable; it is necessary if we still subscribe to the view that communication, and therefore language change, including CLF, is a basic human feature that we all share and that functions according to some, perhaps very few, universal principles. The principles invoked to account for case in KJ are in harmony with those presented in sections 1-3, namely:

1. In CLF, a salient aspect is differential replication, as we are trying to account for the emergence of new structure (and new languages).
2. We look at language as a population of utterances or speech community – in CLF this means a dynamic, heteroglossic population; its functional-typological make-up may, for convenience of analysis, can be abstracted as the Typological Matrix.

---

<sup>14</sup> It may be the case that relatively unmarked variables might move faster through a community than marked ones, but then again it may be the case that highly infrequent variables do not get selected at all, which makes this difficult to prove.

3. In CLF, selection, innovation and propagation occur iteratively and feed into one another. The most likely candidates for selection and propagation are determined based primarily on sociohistorical analysis and typological make-up, within which frequency patterns play a dominant role.

The notion of TM suggests that the output in CLF can be fully made sense of by looking at the input languages; I do not necessarily hold that other types of change, such as internal or independent changes, may not occur. For example, I have hypothesized that an earlier stage of the Dative-Accusative case system may have seen a conflated case that then split into the two cases we have today (Ansaldo 2005a, b). The reason for such an hypothesis was the fact that, in the history of languages, we have at least one attested case of such a diachronic pattern, where an adposition grammaticalizes first into a more generic marker and then specializes, namely Persian (Blake 1994). But it is a rare change and we do not have appropriate historical evidence available to prove it or disprove it, so it is ultimately a little futile to discuss it. However, I believe that ‘invisible hand’ changes should be treated very carefully, as a last resort in trying to account for CLF, and we should always be aware of the possibility that, more often than not, these hypotheses arise from imperfections in the data and/or in our knowledge and understanding of the TM. A case in point is the pronoun system of KJ that shows Animacy effects, as pronouns are marked by the suffix *-dang* (rather than *-nang*, see Ansaldo 2005, 2008). This we could label an independent innovation of KJ; at the same time, Animacy is a feature present in various domains of Sinhala grammar. In a sense, therefore, Animacy is present with the TM in which KJ evolves; calling it ‘independent’ may simply be a way to say that, for now, we cannot make sense of why it is selected and then realized as innovation within the pronominal system.

## 6. Concluding remarks

At this point let us consider the *why*-question: ‘do we really need all this in order to talk about CLF?’ The answer, I believe, is a clear ‘yes’. In contact-linguistics, in particular in

the field of language creation, there has been relatively slow progress in crucial areas of investigation. As this volume (and others, e.g. Ansaldo, Matthews and Lim 2007) shows, for example, the field has struggled with notions of complexity, notwithstanding the alarm bells rung as early as Muysken (1988) that a proper discussion on the matter of complexity would have to be based on a clear and shared theory of what is really grammatically complex and what is not. The approach presented above offers a theoretical domain in which complexity can be properly investigated in terms of input-output relations. The views of transmission proposed in the field often assuming some kind of ‘failure’ on the part of speakers in CLF, or at least some imperfect or abnormal process at play (imperfect acquisition, rapid change etc., see Ansaldo and Matthews 2007), have been criticized for ideological bias as well as theoretical shortcomings (DeGraff 2001, 2003, 2005). The framework presented above offers a neutral approach to CLF in which sociohistorical and functional-typological analysis find integration. Work on alleged universal properties of contact languages such as Creoles have made a number of controversial claims (e.g. McWhorter 2005); such claims, I believe, can be more thoroughly investigated within this framework, and promise less controversial solutions, that can be integrated into general linguistic theories of language rather than requiring exceptional treatment.

## References

- Aboh, Enoch O. and Umberto Ansaldo 2007. The role of typology in language creation: A descriptive take. In Ansaldo, Umberto, Stephen Matthews and Lisa Lim (eds). *Deconstructing Creole* [Typological Studies in Language 73], Amsterdam/Philadelphia: John Benjamins Publishing Company. 39-66.
- Adelaar, K. Alexander and J. D. Prentice 1996. Malay: Its history, role and spread. In Wurm, Stephen A., Peter Mühlhäusler and Darrel T. Tryon (eds). *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin: Mouton de Gruyter. 673-693.

- Andersen, Henning 2006. Synchrony, diachrony and evolution. In Thomsen, Ole N. (ed.) *Competing models of linguistic change: Evolution and beyond*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 59-90.
- Anderson, J. Robert 1993. *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Ansaldò, Umberto 2004. The evolution of Singapore English: Finding the matrix. In Lim, Lisa (ed.) *Singapore English: A grammatical description*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 127-149.
- Ansaldò, Umberto 2005a. Typological admixture in Sri Lanka Malay. Ms. University of Amsterdam.
- Ansaldò, Umberto 2005b. Contact-induced morphologization. Motivations for the emergence of case in Kirinda Java. *New Reflections on Grammaticalization 3*. Santiago de Compostela, Spain, 17-20 July 2005.
- Ansaldò, Umberto 2008. Revisiting Sri Lanka Malay: genesis and classification. In Harrison, K. David, David S. Rood and Arienne Dwyer (eds). *Lessons from documented endangered languages: A world of many voices*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 13-42.
- Ansaldò, Umberto. 2009. *Contact languages. Ecology and evolution*. Cambridge: Cambridge University Press.
- Ansaldò, Umberto, Lisa Lim and Salikoko Mufwene. 2007. The sociolinguistic history of the Peranakans: What it tells us about ‘creolization’. In Ansaldò, Umberto, Stephen Matthews and Lisa Lim (eds). *Deconstructing Creole* [Typological Studies in Language 73], Amsterdam/ Philadelphia: John Benjamins Publishing Company. 203-226.
- Ansaldò, Umberto and Stephen Matthews. 2007. Deconstructing creole: The rationale. In Ansaldò, Umberto, Stephen Matthews and Lisa Lim (eds). *Deconstructing Creole* [Typological Studies in Language 73], Amsterdam/ Philadelphia: John Benjamins Publishing Company. 1-18.
- Ansaldò, Umberto, Stephen Matthews and Lisa Lim (eds). 2007. *Deconstructing creole* [Typological Studies in Language 73]. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Ansaldò, Umberto and Sebastian Nordhoff. 2008. Complexity and the age of language. In

- Arends, Jacques 2001. Social stratification and network relations in the formation of Sranan. In Smith, Norval and Tonjes Veenstra (eds). *Creolization and contact*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. pp. 291-307.
- Bakhtin, Michael M. 1981. *The dialogic imagination*. [Translated and edited by C. Emerson and M. Holmquist]. Austin: University of Texas Press.
- Bhaskararao, Peri and Karumuri Venkata Subbarao (eds). 2004. *Non-nominative subjects*. (2 Vols.) Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Blake, Barry 1994. *Case*. Cambridge: Cambridge University Press.
- Bonfante, Giuliano 1946. 'Indo-Hittite' and areal linguistics. *American Journal of Philology* 67: 289-310.
- Bybee, Joan 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Bybee, Joan 1998. A functionalist approach to grammar and its evolution. *Evolution of Communication* 2: 249-278.
- Bybee, Joan 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4): 711-733.
- Bybee, Joan and Sandra A. Thompson 1997. Three frequency effects in syntax. *Berkeley Linguistic Society* 23: 378-388.
- Chomsky, Noam 1959. Review of *Verbal Behaviour* by B.F. Skinner. *Language* 35: 26-58.
- Croft, William 1990. *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William 1995. Autonomy and functionalist linguistics. *Language* 71: 490-532.
- Croft, William 2000. *Explaining language change: An evolutionary approach*. London: Longman.
- Croft, William 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William 2006a. Evolutionary models and functional-typological theories of language change. In Kamenade Ans v., and Bettelou Los (eds). *Handbook of the history of English*. Oxford: Blackwell. pp. 68-91.

- Croft, William 2006b. The relevance of an evolutionary model. In Thomsen, Ole N. (ed.), *Competing models of linguistic change: Evolution and Beyond*, 91-132. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Dahl, Östen 2004. *The growth and maintenance of linguistic complexity*. Amsterdam/ Philadelphia: John Benjamins.
- Dawkins, Richard 1976. *The selfish gene*. New York: Oxford University Press.
- DeGraff, Michel 2001. On the origins of creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5.2-3: 213-230.
- DeGraff, Michel 2003. Against Creole exceptionalism. *Language* 79: 391-410.
- DeGraff, Michel 2005. Linguists' most dangerous myth: The fallacy of creole exceptionalism. *Language in Society* 34: 533-591.
- Durham, William 1991. *Coevolution, genes, culture and human diversity*. Palo Alto: Stanford University Press.
- Gair, James and John Paolillo. 1997. *Sinhala*. München: Lincom Europa.
- Givón, Talmy 1979. *On understanding grammar*. New York: Academic Press.
- Givón, Talmy 1995. *Functionalism and grammar*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, memory and cognition* 22.1. 1166-1183.
- Gumperz, John J. 1964. Linguistic and social interaction in two communities. *American Anthropologist* 66.6: 137-153.
- Gumperz, John J. 1968. The speech community. In *International Encyclopedia of the Social Sciences*. New York: Macmillan. 381-386.
- Haspelmath, Martin 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42.1: 25-70.
- Heine, Bernd and Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Haiman, John 1994. Ritualization and the development of language. In Pagliuca William (ed.) *Perspectives on grammaticalization*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 3-28.

- Hopper, Paul J. 1987. Emergent grammar. *Berkeley Linguistic Society* 13: 139-157.
- Hopper, Paul J. and Elizabeth C. Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hull, David L. 1988. *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.
- Hussainmiya, B.A. 1987. *Lost cousins: The Malays of Sri Lanka*. Kuala Lumpur: Universiti Kebangsaan Malaysia.
- Hussainmiya, B.A. 1990. *Orang Rejimen: The Malays of the Ceylon Rifle Regiment*. Kuala Lumpur: Universiti Kebangsaan Malaysia.
- Hymes, Dell 1971. *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.
- Hymes, Dell 1972. On communicative competence. In Pride, Janet B. and John Holmes (eds). *Sociolinguistics*. Harmondsworth: Penguin. 269-293.
- Janda, Richard and Brian Joseph. On language, change and language change – or of history, linguistics and historical linguistics. In Joseph, Brian and Richard Janda (eds). *Handbook of historical linguistics*. Oxford: Blackwell. 3-180.
- Koptjevskaja-Tamm, Maria (to appear). Linguistic typology and language contact. In Song, Jae-Jung (ed.). *The Oxford Handbook of Typology*. Oxford: Oxford University Press.
- Kroskrity, Paul V. (ed). 2000. *Regimes of language*. Santa Fe, NM: School of American Research Advanced Seminar Series.
- Langacker, Ronald 1987. *Foundations of cognitive grammar*. Vol. 1. Stanford: Stanford University Press.
- Langacker, Ronald 2000. A dynamic usage-based model. In Barlow, Michael and Suzanne Kemmer (eds). *Usage-based models of language*. Stanford: Center for the Study of Language and Information. 1-63.
- Lass, Roger 1990. How to do things with junk: Exaptation in language change. *Journal of Linguistics* 26: 79-102.
- Lass, Roger 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.

- Lieberman, Paul 2006. *Toward an evolutionary biology of language*. Cambridge, MA/ London, England: Belknap Harvard University Press.
- Lightfoot, David 1999. *The development of language: Acquisition, change and evolution*. Oxford: Blackwell.
- Lim, Lisa 2007. Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes* 26(4): pp. 446-473.
- Lim, Lisa and U. Ansaldo 2006. Keeping Kirinda vital: The endangerment-empowerment dilemma in the documentation of Sri Lanka Malay. In *Amsterdam Center for Language and Communication Working Papers* 1: 51-66.
- Lindblom, Björn, Peter McNeilage and Michael Studdert-Kennedy 1984. Self-organizing process and the explanation of phonological universals. In Butterworth, Brian, Bernard Comrie and Östen Dahl (eds), *Explanations for language universals*. New York: Mouton. 181-203.
- McWhorter, John 2005. *Defining Creole*. Oxford: Oxford University Press.
- Masica, Colin 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
- Matras, Yaron and Peter Bakker (eds). 2003. *The mixed language debate: Theoretical and empirical advances*. Berlin/ New York: Mouton de Gruyter.
- Milroy, James. 1999. Roger Lass, historical linguistics and language change. *Diachronica* 16(1): 179-191.
- Milroy, Leslie and James Milroy 1992. Social network and social class: Towards an integrated sociolinguistic model. *Language in Society* 21: 1-26.
- Mufwene, Salikoko S. 1991. Pidgins, creoles, typology and markedness. In Byrne, Francis and Thom Huebner (eds), *Development and structures of creole languages: Essays in honor of Derek Bickerton*. Amsterdam/ Philadelphia: John Benjamins Publishing Company. 123-143.
- Mufwene, Salikoko S. 1996. Creole genesis: A population genetics perspective. In Christie, Pauline (ed.) *Caribbean language issues old and new*. Kingston, Jamaica: University of the West Indies Press. 168-209.
- 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, David 1997. *Linguistic diversity*. Cambridge: Cambridge University Press.
- Nordhoff, Sebastian (in preparation). A grammar of Sri Lanka Malay. Ph.D. dissertation, University of Amsterdam.
- Pierrehumbert, Janet 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, Joan and Paul J. Hopper (eds), *Frequency and the emergence of linguistic structure*, Amsterdam/ Philadelphia: John Benjamins Publishing Company. 137-158.
- Pierrehumbert, Janet 2002. Word-specific phonetics. In Gussenhoven, Carlos and Natasha Warner (eds). *Laboratory phonology 7*. Berlin/ New York: Mouton de Gruyter. 101-140.
- Ross, Malcolm 1996. Contact-induced change and the comparative method. Cases from Papua New Guinea. In Durie, Mark and Malcolm Ross (eds). *The comparative method reviewed: Regularity and irregularity in language change*. New York: Oxford University Press. 180-217.
- Ross, Malcolm 2001. Contact-induced change in Oceanic languages in northwest Melanesia. In Dixon, R.M.W. and Alexandra Aikhenvald (eds), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press. 134-166.
- 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 Deumert, Ana and Stephanie Durrleman *Structure and variation in language contact*, (eds). Amsterdam/ Philadelphia: John Benjamins Publishing Company. 135–158.
- Smith, Ian, Scott Pauuw and B.A. Hussainmiya. 2004. Sri Lanka Malay: The state of the art. In Singh, Rajendra (ed.) *Yearbook of South Asian Languages 2004*, Berlin/ New York: Mouton de Gruyter. 197-215.

- Thomason, Sarah 2003. Contact as a source of language change. In Joseph, Brian and Richard Janda (eds). *Handbook of historical linguistics*. Oxford: Blackwell. 687-712.
- Thomason, Sarah and Terence Kaufman 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Tomasello, Michael 2003. *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA/ London, England: Harvard University Press.
- Tomasello, Michael, Anne Cale Kruger and Hilary Horn Ratner. 1993. Cultural learning. *Behavioral and Brain Sciences* 16: 495-552.
- Trudgill, Peter 2001. Contact and simplification. Historical baggage and directionality in linguistic change. *Linguistic Typology* 5: 372-375.
- Trudgill, Peter 2002. Linguistic and social typology. In J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds). *The handbook of language variation and change*. Malden MA/ Oxford UK: Blackwell. 707-728.
- Weinreich, Uriel 1953 (1968 2<sup>nd</sup> ed.). *Languages in contact: Findings and problems*. The Hague: Mouton.