

## 1 Introduction

---

You are sitting in the tropical garden of a café nursing your thirst with a cold beer when you hear the following conversation:

- *Eh41 yu got watch PCK las nait or not?*
- *Ya lor55. Wa lau he kena sen English class! His English so lau ya one meh55?*
- *Ya wat21! Gamen say mas spik gud English ma21. Don play-play! Oderwise people ting yu dam chingchong la21.*<sup>1</sup>

You understand some of the words, but not all of them, because, although they sound English, some are pronounced in ‘strange’ ways. More importantly, you cannot quite make sense of all that is being said. So was it really English you heard or did you just imagine it? If it was, what was going on with the grammar? Because, surely, there was something wrong somewhere. And what about those little exclamations at the end of each sentence with such strange intonation? You might think that you have just heard some rather ungrammatical English being spoken. Or was it Chinese with some English words in it? According to some linguists, it was neither. What you have just heard is a language variety known as Singlish, or Colloquial Singapore English in academic circles. Singlish is a native variety of Singapore whose lexicon and grammar in part derive from English, Chinese and Malay. It is spoken by a majority of the young generation of Singaporeans, who relish it as their native tongue. It is officially labelled not ‘good English’ by the government and active steps have been taken in order to discourage its use. Educators and politicians in Singapore speak negatively of it, yet use it in the classroom and in political rallies. Ask five different language experts (or linguists) what they make of such a language and you are likely to get four different answers. This is the world of contact language formation, i.e. the evolution of new languages in multilingual environments. This book deals with the many issues – social, historical, and grammatical – that such languages raise.

### 1.1 Introduction to contact language formation (CLF)

Why do languages change over time? What leads a community of speakers to adopt a new language or, even more interestingly, to develop a new one?

2 Contact Languages

Language change – in particular contact-induced change – has developed into one of the most fascinating areas of linguistics since Weinreich's (1953) seminal work. There are several reasons to justify the appeal of the field: perhaps the most striking is its multidimensionality, spanning as it does traditional sub-disciplines of linguistics, namely sociolinguistics, historical linguistics, structural analysis, and the ideology of language. In addition, understanding CLF, the topic of this book, holds the key to some of linguistics' most fascinating questions: How does a new grammar (and a new language) emerge? What are the socio-historical conditions that trigger it and what factors determine the grammatical outcome? What is it that speakers do in situations of language contact and why do they do it? Are there constraints on the creativity of language or is CLF completely unpredictable?

In many parts of the world, children and adults function in multilingual environments where several languages are spoken, perhaps in different domains, such as one language for the home (used predominantly in the family or one's ethnic group), one or more language(s) for the market place, and yet another language for official interethnic communication. Often these languages are mastered very early on and learnt in an informal environment, that is, without the support of school or other institutions. The strict sense of grammatical norm that speakers of Western cultures have become accustomed to in the past few centuries is not felt to the same degree in these communities, where variation between speakers along social class, ethnolinguistic group, religious sphere or generation may be very pronounced. In these multilingual environments switching between and mixing of different languages is the norm. Even in Western societies, mixing of codes can be readily observed in bilingual environments, such as within immigrant communities (think of *Gastarbeiterdeutsch* 'foreign workers' German', the variety spoken by predominantly Turkish immigrants) or in border areas (e.g. Italy's Northwest, where the Italian-German variety known as *Alto-Adige* or *Südtirol* is spoken).

In certain times and places, practices of code-mixing have eventually led to the formation of a new language. Perhaps the most familiar evidence comes from Creole languages, typical products of the displaced communities that emerged out of European colonial exploitation of, in particular, West African, Caribbean, and Pacific regions. But population movements, forced and free migrations, and population admixtures have in fact recurred in human history long before the recent European colonial era (fifteenth to twentieth centuries) and are still ubiquitous today. Indeed we can detect them in all regions of the globe for which we have good historical documentation.

The first time I experienced the fascination of a mixed language and culture was in Melaka, Malaysia. As a second-year student of Mandarin Chinese, I had already experienced the marvel of 'variation' when I encountered Cantonese in Hong Kong and later Minnan (or Hokkien) in Penang. I had expected to find

other Sinitic languages to be somewhat similar to the Mandarin I had learnt but I was instead struck by completely different grammatical systems; the Sinitic varieties were not mutually intelligible, which rendered the application of the label ‘dialects’ completely useless in their case. In Melaka I discovered another dimension of marvel in the Peranakan<sup>2</sup> culture, with its intriguing fusion of Hokkien and Malay elements, in food, rituals, language, and other cultural domains. When and how had this happened? Why had the Chinese *Babas* and the Malay-Indonesian *Nyonyas* developed this new language known as Baba Malay?<sup>3</sup> And how exactly could one explain the grammatical outcome? I carried these questions with me during my graduate years in Stockholm and Hong Kong, and found other, fascinating phenomena, such as Makista (or Macanese), the variety of Portuguese infused with Sinitic elements in Macau, which was, until recently, a well-preserved example of a city-port in which East and West met. The encounters that produced these mixed language varieties also produced hybrid architectural styles and innovative cultures that offer a fascinating insight into cultural and linguistic admixture.

My first academic appointment was at the National University of Singapore, where I had the opportunity to study in depth the incredible multilingual ecology of Singapore, in which Singlish and Bazaar Malay have emerged. Also, thanks to a university research grant, I was able to conduct fieldwork on Cocos Malay and Sri Lanka Malay, two interesting contact Malay varieties of the Indian Ocean. In more recent years I have been working with the Sri Lanka Malay speech community as part of Volkswagen Stiftung’s initiative for the documentation of endangered languages (DoBeS). The present book reflects the knowledge that I have gained to date from this rich linguistic experience, knowledge based on case studies that I conducted on salient aspects of CLF in these Asian contexts. Some of the restructuring is still in process and was investigated first-hand in fieldwork I conducted over the years. Other varieties are no longer spoken; thus the contact scenarios had to be reconstructed through historical and typological analysis. I discuss all of them in the following pages.

## 1.2 Research questions

This book aims at providing a conceptual framework, as well as a number of case studies, for exploring one underlying theme: the interaction between social factors and grammatical features in the process of CLF. In order to further our understanding of how new languages emerge, it is important to study the division of labour between ‘external’ and ‘internal’ factors in the evolution of new grammars (Mufwene 2001). This means that we need to bring together observations of a socio-historical nature with grammatical (here: functional-typological) analysis, laying emphasis on the former as the main force behind

4 Contact Languages

CLF, as advocated by Thomason and Kaufman (1988). Specifically, throughout this book, I attempt to offer some answers to the following questions:

1. What kinds of ecologies trigger the emergence of new languages?
2. What is the relationship between input languages and the new grammar?
3. How do speakers use languages in multilingual environments?
4. Why do new languages evolve?
5. Which are the theories of language that can be reconciled with the lessons learned from CLF?

Following in the footsteps of Thomason and Kaufman (1988), I seriously consider the claim that languages change because speakers change them, by looking at both speakers' agency and grammatical systems in accounting for CLF in Asian contexts. Through the study of cultural and linguistic contact in Asian contexts, I further explore the relationship between the formation of new identities and the emergence of new languages, already established by Le Page and Tabouret-Keller (1985) in the case of the West Indies. Moreover, the case studies in this book show that a unified approach to the processes of language change and language maintenance clarifies and improves our current understanding of language change, in line with an evolutionary approach to language change as exposed in Croft (2000) and Mufwene (2001). In evolutionary terms, language change and the lack of it are seen as processes that can be explained by the same cognitive mechanisms of selection and replication of utterances. Finally, this book stresses the importance of ecology, in the sense of the environment in which CLF takes place ('external ecology'), as well as the systemic interaction of the linguistic codes in contact ('internal ecology'), as already pioneered by Mufwene (2001, 2008). While at the University of Amsterdam, I have enjoyed the benefit of exchanging ideas with colleagues whose expertise lies in language contact in different regions of the world, in particular West Africa, the Caribbean, and South Africa. It is in the hard work of these (and other) colleagues before me that I find inspiration to present what I hope will be novel, useful materials to advance our understanding of the creative linguistic behaviour typical of usually multilingual diasporas around the globe.

In the approach taken in this book (see chapters 4 and 5), the underlying assumption is that CLF, including the development of pidgins and Creoles, or mixed languages in general, is a natural outcome of multilingual environments in which languages are learnt in informal settings (the home, the work place, the market), as opposed to the school. This is particularly so in the case of displaced communities – diasporas that can be regarded as 'minorities' or peripheral within the geopolitical contexts in which they are situated. I will show that in polylectal speech communities, external factors such as (a) size of the community, (b) type of intra- and intersocietal networks, and (c) multilingual practices actively influence and, for the most part, determine the outcomes of

contact situations. ‘Speech community’ is intended here in the sense of a web of social networks (Milroy 1987; Milroy and Milroy 1992) with group-internal relationships as well as intercommunal relationships that, to various degrees, determine the types of linguistic attitudes that define the community.

Cutting across these parameters is the typological dimension, i.e. the similarity or difference between the grammatical systems in contact (‘congruence’ in Mufwene 2001). This dimension is of the utmost importance as I assume that it is only by relating grammatical analysis to points (a) to (c) above that a complete picture of CLF can be achieved. In this sense, I aim to explain patterns of CLF as essentially not divorced from observations derived from historical linguistics, language variation and change, areal typology, and socio-linguistics (e.g. Siegel 2004; Heine and Kuteva 2005; Ansaldo, Matthews and Lim 2007).

The case studies presented in this book offer substantial evidence for the fact that CLF is the natural – though not necessary – outcome of population admixture in shifting socio-historical contexts. The findings present compelling evidence in support of the following claims:

- language use is inherently creative;
- multilingualism and/or casual transmission support innovative use of language;
- the typological kinship of the input languages determines to a high degree the grammatical outcome;
- CLF goes in parallel with new identity formation;
- CLF is not the consequence of a negative process (i.e. failure to acquire a target language) but rather of a constructive process involving speakers’ agency; and
- CLF happens most often in minority groups, displaced groups, or otherwise geopolitically peripheral communities, who are not necessarily disempowered but often in a position of socio-cultural brokers.

### 1.3 The role of ecology in Asian contexts

If we acknowledge the role that ecology plays in the evolution of new varieties, it is extremely rewarding to step away from the traditional contexts in which the dynamics of CLF have been heavily researched so far. It is obvious that the lessons learned from these encounters, no matter how valuable, have limitations. In terms of both external and internal ecology, we have been looking at encounters between Western European languages and West African or Pacific languages in colonial environments within which the power relation is in favour of the Western party. This means that, in terms of linguistic typology as well as historical and social dynamics, we still have not yet explored the full range of possibilities in which CLF can occur.

6 Contact Languages

In this book then, I discuss CLF in East, South and Southeast Asia, focusing on the role of Malay and Chinese in various encounters before, during, and after the European colonial era (fourteenth to twentieth centuries AD). The geographical area I cover extends from the South China Sea to the eastern half of the Indian Ocean, including the Indonesian archipelago. As has often been pointed out in traditional historical literature (e.g. Reid 2000), this region is first and foremost a climatic zone, ruled by the two monsoon seasons that determine the lifecycle of its populations in all significant aspects of life; it is also known as Monsoon Asia. Malay and Chinese varieties played an important role throughout the history of the region as the languages associated with the economically most powerful and influential territories, a status made more evident by the spice trade between the Maluku Islands and China (see chapters 3 and 4). This offers an opportunity to observe internal and external ecologies that are different from the better-known settings in which CLF has been studied, i.e. the Atlantic and Pacific regions.

Take the case of China Coast Pidgin (CCP), for example, the oldest attested Chinese–English pidgin we know of: here the linguistic encounter involves Cantonese and English. The former is a strongly isolating language; that is, it makes no use of inflections for person, number, and tense. Also importantly, its sound system is radically different from that of English, especially in its use of lexical tone. Moreover, the Chinese were by no means colonial subjects in the formative period of CCP. Quite the contrary, as I will argue, they were in a position of power, determining the local conditions of trade between them and the English. What are the social dynamics of such encounters? Was CCP the result of the Chinese failing to master English or was it the outcome of attempts by Western merchants to master Chinese? Where does the agency lie in the formation of this new language and how do we make sense of its grammar?

Ecological diversity is also conspicuous in the evolution of contact Malay varieties. For example, there were contexts of relative isolation, in which varieties such as Cocos Malay emerged that were influenced by pre-existing lingua francas, just as there were settings of intimate contact, where varieties evolved that were influenced by adstrate languages, as in the case of Baba Malay, which bears Chinese influence. The consequences of such ecological variation lie in the different grammatical outputs of the language restructuring process. Moreover, there are cases where contact spanned a very long period of time, such as that of Bazaar Malay, which was attested for roughly a millennium, and these offer us a new perspective on the nature of pidgins and lingua francas. Finally, the patterns of multilingualism observed in certain Asian ecologies are quite unlike what we know of, say, the Caribbean region, again providing us with valuable comparative material. This is, for example, the case of Sri Lanka Malay, a language formed through the practice of widespread trilingualism (see chapter 6).

As suggested above, using the case studies presented in the book, it will be possible to tackle the following questions in relation to CLF in Monsoon Asia:

- Does contact language formation in the Asian context present us with novel evolutionary patterns?
- What are the dominant external factors in Asian ecologies?

In order to answer these questions properly, it is important to appreciate fully the nature of the ecological factors that define the various contexts in Monsoon Asia, which are the following:

1. Slavery and manpower: the study of slavery and manpower in different regions and historical phases helps us understand the patterns of indentured labour and population movements that led to heteroglossic networks as seen in particular in the city-ports. This relates to questions of (i) mode of transmission, i.e. how new features and varieties got integrated with the already existing ones, and (ii) notions of ‘target language’ and feature pools available to speakers.
2. Economy and politics: seasonal trade routes determined by the monsoons gave rise to a number of multicultural and multilingual enclaves. How were the numerous languages negotiated and what were the outcomes of such negotiations in the different ports? And how did the transition of power from local networks through ‘interactive’ colonization (Faraclas, Walicek *et al.* 2007), involving trade, intermarriage and other forms of mutual exchanges, to exploitation colonization impact language change in the region?
3. Social relations: mixed cultures that arose in precolonial times from quasi-equal social relationships between the parties (e.g. marriages between Chinese merchants and local women in Malaysia, or communities that emerged from the mixing of Portuguese, Malays, and Chinese) provided ideal settings for the evolution of new languages.
4. Language varieties and linguistic codes: the typology of the languages involved, Malay and Sinitic, can show remarkable congruence at a superficial level. It is crucial to distinguish between written and spoken registers in both cases because there is a great divergence between the registers, and it was predominantly features from colloquial varieties that were involved in interethnic communication. It is also important to look at the appropriate varieties in each case, as there is significant regional variation. When looking at Malay influence, for example, we have to be aware of the variety of languages spoken in the Indonesian world and their significance in the past. When looking at Sinitic influences, we usually have to focus on southern varieties, Hokkien (Min) and Cantonese, typologically very divergent from Mandarin (or Standard Chinese).

8 Contact Languages

In this context, I find it particularly enlightening to look at cases of CLF of precolonial times as a direct product of the social, cultural and political dynamics of Monsoon Asia. The development of pidgin Malay varieties such as Bazaar Malay (Adelaar and Prentice 1996) is a clear indication of the ubiquity of multilingualism and language admixture, and shows us that CLF can occur under historical circumstances that are not particularly extraordinary. Nonetheless, the question of why humans develop new languages over time, rather than speak like their neighbours or their ancestors, is a very intriguing one. As we will see, when social and historical conditions are ripe for a new community to emerge from the blending of previously existing ones, CLF is very likely to occur. The external conditions for CLF typically involve one or several of the following situations (see chapters 2 and 3):

- (a) trade between different communities, in particular where there is a stable space where the groups converge in order to trade, i.e. a market, a port, etc.;
- (b) interethnic marriage, especially when that happens in a neutral or ‘third space’ as in the case of the Hokkien Babas and Malay/Indonesian Nyonyas in Melaka (see chapters 3 and 7);
- (c) resettlement as a consequence of migration, whether forced or spontaneous (indenture, economic, political, etc.), as observed in the Malay Diaspora of Sri Lanka (see chapter 6).

In newly formed ecologies created by such conditions, a community that has achieved a critical demographic mass and stability in the new space will often start identifying itself as a discrete cultural unit. As part of this process, there can be a societal convergence (both conscious and unconscious) towards a shared set of communicative practices that results in what appears as the evolution of a new language. For example, the case of the Peranakans of Melaka shows how a new community that arose out of intermarriages between Chinese- and Malay-speaking people came to develop a new, hybrid variety of Chinese–Malay, although both Chinese and Malay were fully available and functional in the ecology in which the Peranakans lived. The new language, Baba Malay, was not a communicative necessity or a failure in mastering either Chinese or Malay but rather an expression of the new cultural identity of the Peranakans, neither Chinese nor Malay. This is an example of why, in the remainder of this book, CLF is treated as an aspect, as well as a result, of new identity formation (chapter 7). Situations where language contact was extensive, but no new identity formation occurred, typically did not result in CLF. Extensive contact without CLF can be observed in ‘linguistic areas’ where the effects of contact result in typological convergence, as discussed by Gumperz and Wilson (1971) for the Indian village of Kupwar, and Ross (1996) for the case of New Guinea.

Another contribution of the case studies in this book is that, in all cases, there is substantial historical evidence to reconstruct a plausible ecology. In the cases of varieties still spoken today, such as Cocos Malay, Sri Lanka Malay and Bazaar Malay, I have been in the fortunate position of collecting first-hand data within the communities, not only relying on data elicited from informants but also recording natural language use and variation, as well as cultural data and oral history. In this sense, these case studies offer an invaluable opportunity to sketch theories based on sufficient information, rather than absence of data, as has often been the case for CLF in the Atlantic or the Mediterranean. For the varieties no longer spoken, there is often reliable socio-historical documentation for plausible reconstruction, as is the case for Baba Malay and Makista. This is discussed further in section 1.5 as well as in chapter 3.

#### 1.4 Theory of language and CLF

In linguistics there are basically two views regarding the nature of language. In one view, our linguistic faculty is determined by innate structures that specify the range of things human language can do. In this sense, language is ‘generated’ by what has been identified as Universal Grammar, the biological endowment for language, or the language organ, which is common to all modern humans. In the other view, our linguistic abilities are determined by the functions of language; in this sense, language is the product of usage and, if there are universal aspects of language, they are to be found in general cognitive structures. Besides the generative view, developed by Noam Chomsky and his colleagues, and functionalist views, advanced in the work of Talmy Givón, among others, we also find intermediate views in which both positions are reconciled, as most clearly advocated in the work of Fredrick Newmeyer (see Newmeyer 2005).

Many linguists inspired by the first approach like to talk of language as an abstract entity with identifiable structural organizations that cross-linguistically have identical components. For example, they talk about ‘mechanisms’ of language contact and of ‘acquisition’ and the systemic transfer of features from one idiolect into another. However, in recent years, our increased understanding of linguistic diversity has proven time and again that universal features of language are hard to pin down, and it has been suggested that all aspects of syntactic structure may be language specific (Croft 2001: 364). For example, even the broadest, apparently universal features of language appear weak: Subject-Verb-Object (SVO) as the ‘default’ parameter (Kayne 1994), for instance, is severely questioned in typological literature (Dryer 1992; Hawkins 1994; see also chapter 6), and the generally assumed ‘universal’ noun–verb distinctions in parts of speech (Croft 2001) are questioned in particular in Gil (2000, 2001). Even more problematic is the obvious inherent diachronic change in language

that seems like a logical contradiction to the idea of a static, hard-wired linguistic ‘organ’ (Tomasello 2003). In fact, philosophers and psychologists, as well as evolutionary biologists, have become increasingly disenchanted by generative ideas of acquisition, and the case for innate structure is at this point at best weak, while there is an increasing amount of knowledge that argues against language-specific cognitive patterns (Tomasello 2003; Lieberman 2006).<sup>4</sup> Moreover, from the point of view of contact linguistics, despite attempts to identify structural constraints on contact-induced change, so far no such constraints have emerged (Koptjevskaja-Tamm to appear).

My work is very much part of the second trend and therefore incorporates a usage-based view of language and transmission; in this sense the context in which language functions is a primary domain of inquiry. The determining role of ecology in the process of CLF has been demonstrated in the evolutionary approach to language change developed in Mufwene (2001), in terms of ‘internal’ ecology (or the system-internal relations that can be observed when we abstract language from its social context) and ‘external ecology’, i.e. the socio-historical domain. When discussing the currents that affect the internal ecology of CLF, I focus on functional-typological features and cognitive aspects of language. Having established the primary role of external ecology, I assume that if any constraints are to be uncovered, they will be found in the socio-historical domain.

#### 1.4.1 *Evolution of grammar*

Language evolution can be seen as the cumulation of innovations occurring in individual speech acts of individual speakers (Mufwene 2001), and languages can be seen as complex systems “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).

In functionalist and cognitive theories of grammar (part of the second position presented above), 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 1979a; Langacker 1987). In both cognitive and functionalist theories, grammar 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 not been the dominant one in linguistic theory, in recent years a number of trends have been accumulating substantial evidence in favour of a usage-based theory of language, in particular in the field of grammaticalization studies, language acquisition, and functional-typological linguistics (Bybee