1 Perspectives on contact

1.1 INTRODUCTION

On 26 December 2004, an undersea earthquake occurred in the Indian Ocean with an epicentre off the west coast of Sumatra, Indonesia. At 9.0 on the Richter scale, it triggered a series of tsunamis along the coasts of most landmasses bordering the Indian Ocean, which inundated coastal communities with waves up to thirty metres, and killed more than 230,000 people in fourteen countries. Close to the epicentre, most of the Andaman and Nicobar Islands – an archipelago of 572 islands in the Bay of Bengal between Myanmar and Indonesia – were devastated. As it was feared that North Sentinel Island, one of the Andaman Islands, would also have been swamped, an Indian Coast Guard rescue helicopter was sent to check on casualties. A single naked bowman emerged from the forest and fired arrows to drive the helicopter away. This made the news, not only for the fact that the Sentinelese had survived the tsunami. What piqued the world's interest then – and continues to do so – was the existence of a 'Stone Age' tribe (as it came to be dubbed), numbering about 250, thought to be directly descended from the first human populations to emerge from Africa. The tribespeople were estimated to have lived on the Andaman Islands for up to 60,000 years, but with very little contact with the outside world for thousands of years. That their language is so different even from other Andaman Islanders suggests this lack of contact; that they continue to actively and violently reject any form of outside incursion or contact suggests they prefer it this way.

News of 'uncontacted peoples' such as the Sentinelese – that is, communities who live, or have lived, either by choice or by circumstance, without significant contact with globalised civilisation – never fails to garner the world's attention. The idea of having no contact with any other community, especially in today's global world, strikes most as incredible, in all respects. After all, contact between communities, and their languages, is such a fundamental human experience, one which

has been occurring since almost the beginning of humanity – an experience which by the very nature of human beings involves communication. It is of small wonder, then, that the study of languages in contact is of interest to many.

1.1.1 Speakers in contact

The study of language contact is an integral and important part of linguistics for a number of reasons that will become clear below. But before we even step into the realm of linguistics, it is important to appreciate that the study of language contact is really, as pre-empted above, the study of speakers in contact. In other words, it involves analysing and understanding what speakers of different languages do when they need to communicate across linguistic barriers imposed by their different languages. As such, language contact is the study of human societies, their histories, cultures and ideologies, and how these impact on human behaviour. In this sense, language is but a window on human nature, a documentable and describable domain, thanks to the analytical tools developed in linguistics, which can lead to deep insights into the way in which humans, as individuals as well as social networks, negotiate identity, power, ideology and ethnicity through linguistic acts. Language contact is therefore one of the fields of linguistics most likely to hold relevance for other traditional fields of humanities, such as history and philosophy, as well as education and the social sciences. Its importance cannot be underestimated; we live in a globalising world in which the economic tension between centre and periphery, the developmental challenges between North and South, and the scientific and technological competition between East and West exacerbate the necessity and urgency of cross-cultural communication and interethnic admixture. In this world, understanding how to manage the power of language, as a mediating tool as well as a site of struggle, is fundamental for a more enlightened and just society.

Since the publication some sixty years ago of Uriel Weinreich's renowned monograph *Languages in Contact* (1953), which, most scholars concur, marked the beginning of modern contact linguistics, work in the field has burgeoned. It has become one of the fastest growing areas of linguistics, with scholarship also drawing from diverse fields such as sociolinguistics, historical linguistics, dialectology and second language acquisition (SLA). A seminal introduction to the field is Appel and Muysken's (1987) textbook, while Thomason and Kaufman's (1988) widely cited book lays the foundation of language contact from a historical and genetic perspective. Pidgin and creole languages are given thorough treatment in the works of Holm (1988, 1989), as well as

Arends, Muysken and Smith (1994). The phenomenon of code switching is captured in studies such as Myers-Scotton (1993), Milroy and Muysken (1995), and Muysken (2000). Influential work in SLA and language contact can be found in Andersen (1983).

What do we look at when we study languages in contact? In this book, we will suggest that the most powerful approach to language contact is a sociolinguistic one. Six decades ago, Weinreich (1953: 83–110) demonstrated that the linguistic outcomes of language contact are in large part conditioned by sociocultural variables, including extent and degree of bilingualism; length of contact; geographical and demographic distribution; social factors such as religion, race, gender and age; use in different social functions such as education, government, media and literature; and political and ideological factors, including those of prestige and language loyalty. Indeed, it has thereafter been repeatedly argued, in other influential work such as Le Page and Tabouret-Keller (1985) and Thomason and Kaufman (1988), that speakers' agency is an important, possibly crucial factor behind language change. Thus, although it is at times useful to abstract the study of language away from speakers, we should not forget that it is ultimately in the speaker's actions, conscious or unaware though they may be, that the key to the processes of language contact lies.

A sociolinguistic approach is also important if we are prepared to acknowledge the limitations of structuralist approaches to language contact. Despite the intense search for constraints that might lead to a deeper understanding of universal aspects of language structure, the field of contact linguistics has time and again proven that no constraints hold when speakers of different grammars come into contact. Language admixture can be as varied and creative as the sociocultural contexts allow it to be. This, to be fair, is in line with latest developments in Chomskyan linguistics, which acknowledge that universal or hard-wired aspects of language structure might have to be confined to a few elementary operations of generic cognitive nature (Hauser, Chomsky and Fitch 2002). Acknowledging that 'everything goes' as a defining principle of language contact is not admitting defeat; quite the contrary, it is a step forward in understanding that, if constraints exist, they must be sought outside grammar, in sociolinguistic patterns of language use. This is corroborated by at least three different fields of linguistic enquiry:

i. linguistic diversity, or linguistic typology, which, with an increasing understanding of structural variation, is calling for a theory of language that aims at explaining diversity, rather than

universality, as a defining cognitive capacity of humans (Evans and Levinson 2009);

ii. usage-based acquisition and language socialisation theory, which conclude that children acquire language largely through exposure, based on cognitive routines that are not language-specific, thus highlighting the role of society and, in particular, adults in understanding patterns of language use (Tomasello 2005; Kulick and Schiefflin 2007); and

iii. language evolution, which recognises that contact between different linguistic features is a fundamental underlying force in language evolution at large (Mufwene 2001; Ansaldo 2009a).

In this book, we approach the emergence of contact languages as a product of various forms of 'choices' that speakers in multilingual ecologies make. Our approach takes the view that contact languages including pidgins and creoles arise as solutions to specific contexts, as already noted in Baker (1995; see also Ansaldo, Matthews and Lim 2007).

1.1.2 Ecology and evolution

Language contact is as old as humanity, since it defines the earliest and most vital encounters between different human groups in contexts such as markets, intellectual exchanges, rituals and conflict. As such, it must account for one of the most important forces behind language change and speciation. In an evolutionary framework, it is reasonable to assume that there are two paths to the diversity that characterises human languages: (a) mutation – i.e. change that can occur in relatively isolated populations over time – and (b) adaptation – i.e. contact-induced change. Both are ultimately motivated by the same human need: identity construction, or alignment, as argued recently in Ansaldo (2009a). In other recent work, Dediu et al. (2013) similarly identify in-group-out-group oppositions as a force in the development of different languages. Whether through detachment from an original group or through admixture with a new population, humans manifest their social identity through the creation of new cultural traits that set them apart from neighbouring groups. Language is part of this cultural profile, and the emergence of diversity can thus be captured as the manifestation of new identities in human dispersal. The study of language contact is aimed at explaining (b). Considering that contacts between human populations have existed since the dawn of human society and permeate even the smallest and relatively isolated groups, it is fair to say that language contact must be a central part of any framework of linguistic theory.

In this book, we take seriously the notion of ecology as applied to the evolution of languages in contact, where ecology is used as a metaphor from population genetics and biology, a paradigm that has been developed most recently in linguistic contact by Salikoko Mufwene (2001 and other work). Language evolution, and the emergence of contact-induced varieties, can be regarded as speakers making selections from a pool of linguistic variants available to them in a contact setting. This feature pool consists of the sum total of the individual forms and variants that each of the speakers involved, with different language backgrounds and varying linguistic experiences, brings into the contact situation. Which variants from this feature pool are chosen as stable elements of the newly emerging variety depends on the complete ecology of the contact situation: the set of relevant conditions and circumstances, both extralinguistic and intralinguistic. All these parameters enter a complex 'contact equation', which allows for some degree of chance impact. The following components are involved: the numerical (demographic) and social relationships (including mutual attitudes and power distributions) between the participants in a contact situation; the amount and type of communicative events; and the nature of the linguistic input elements, in terms of cognitive and typological congruence between the systems involved. An evolutionary approach to contact language formation (Ansaldo 2009a) is elaborated on in Chapter 4, and an examination of the outcomes of language contact using the ecological approach is presented in Chapter 5. There the interplay of both external and internal factors in the selection of linguistic features from the feature pool, and the significance of the founder population, will emerge with clarity.

1.2 CONTACT PHENOMENA

When speakers of different languages come into contact, linguistic interference may occur. It is highly unlikely that speakers of different languages exist in each other's proximity without the other languages and cultures affecting the languages and cultures that define them. Even in the most resistant of cultures, as this book will show, transfer of linguistic features occurs. Language contact can be observed in even superficially multilingual contexts; it occurs frequently in bilingual and multilingual societies, in trading environments, through technological transfer as well as colonisation and globalisation.

A number of phenomena have been traditionally identified as linguistic outcomes of languages in contact, and in what follows we provide a preliminary sketch of these outcomes and other related concepts that

define the field of contact linguistics. Note that here we use the terms conventionally used in the literature, so that the reader can, in this first chapter, make the necessary connections with what they may have encountered elsewhere.

1.2.1 Borrowing

The most common type of influence that a language exerts on another language is borrowing, which is the incorporation of foreign elements into the speaker's native language. Borrowing can include words, grammatical structures, phonology and the replication of linguistic patterns (Matras 2009). At its most common, in lexical borrowing – the borrowing of words – lexical items are borrowed and assimilated to the borrowing language. English is notorious for having a vast number of loanwords, estimated to comprise some three quarters of its vocabulary, with the majority from French and Latin, mostly after the Norman Conquest in 1066 and the subsequent contact between Norman French and English then. Lexical borrowing can involve a change in segmental phonology and the addition of lexical tone, as in English *cheese* becoming Cantonese[1] 芝士 *zi1si2*, and *taxi* becoming 的士 *dik1si2*. While lexical borrowing is used by monolingual as well as bilingual speakers, structural borrowing (Thomason and Kaufman 1988) entails intensive contact and widespread bilingualism. Nevertheless, it is recognised that any feature in any aspect of language structure can be borrowed, in the right circumstances. For example, with the majority of Uyghurs living in what is now the Autonomous Region of Xinjiang in western China, whose language belongs to the Turkic group of the Altaic language family of central Eurasia, Altaic case suffixes are found in Xinjiang Chinese. Structural borrowing can be so intense as to lead to deep grammatical change in a language (see Section 1.2.3).

1.2.2 Code mixing

Code mixing is the practice of mixing together features from different linguistic systems, in particular in describing more stable situations

[1] The romanisation system for Cantonese used is Jyutping (粵拼 *jyut6 ping3* [jyːt̚ pʰɪŋ˧]), which was developed by the Linguistic Society of Hong Kong in 1993; its formal name is *The Linguistic Society of Hong Kong Cantonese Romanization Scheme*. The name Jyutping is a contraction consisting of the first Chinese characters of the terms *Jyut6 jyu5* (粵語 'Yue language') and *ping3 jam1* (拼音 'phonetic alphabet'). In Jyutping, tone numbers are assigned to the nine tones in Cantonese. Tone 1 has a pitch contour of 55 or 53 where, in the Asianist system of tone level numbers, 5 represents a high pitch and 1 a low pitch; a pitch level of 55 would thus be a high level tone. Tone 2 has a pitch contour of 35.

where multiple languages are used, and emphasising a more hybrid form. This contrasts with code switching, which is usually used to indicate a speaker's movement from one grammatical system to another – that is, an alternation of languages within a conversation, usually at semantically or sociolinguistically meaningful junctures, which is associated with particular pragmatic effects, discourse functions or associations with group identity. Note, however, that the terms code switching and code mixing are sometimes used more or less interchangeably, and/or with code mixing used to encompass both practices. Research has focused on situational and contextual motivations for switching, as well as on the structural characteristics of code switching, aiming to identify general patterns. Code mixing, as a hybrid form that draws from distinct grammar, is very commonly observed in bi- or multilingual speakers and societies, which suggests that a rigid distinction of grammatical systems in the brain is not the natural state of bi/multilingual communities (Sridhar and Sridhar 1980; Muysken 2000). Such stabilised code mixed systems resemble lects (i.e. any distinguishable variety of a language) in their own right – they are also termed mixed languages (but see Section 1.3.4 below) or fused lects – and thus code mixing can be seen as one route to contact language formation (Auer 1999). Mixed codes will be discussed in greater detail in Chapter 2. And as we will see in the case of Sri Lanka Malay (SLM) in Chapter 4, code mixing is indeed the driving force behind the change that leads to the emergence of a new vernacular.

1.2.3 Contact languages

One of the most interesting research areas in the field of language contact concerns those contexts in which intense language contact between mutually unintelligible languages leads to substantial grammatical restructuring. Particular attention has been given to the phenomenon of language genesis – i.e. the 'birth' of a new language from a multilingual contact situation. The richest inventory of such cases known to us today comes from the centuries of Western colonial expansion. In this period Western nations settled and exploited various regions of the world, in particular the Caribbean, the Pacific and the Southeast Asian regions. In these contexts, trade languages referred to as pidgins (or jargons), as well as indigenous vernaculars known as creoles, emerged; these present a challenge from various points of view. Contact languages are usually studied along the following lines (e.g. in Holm 1988, 1989; Arends, Muysken and Smith 1994; Sebba 1997; Bakker and Matras 2013; Michaelis et al. 2013; Velupillai 2015).

1.2.3.1 *Pidgins*

Pidgins are contact vernaculars which emerge in situations where communities with mutually unintelligible languages are in close and repeated contact in specific situations and need a medium for mutual communication. At the time of their formation, pidgins are not the mother tongue of any speakers but a secondary language, typically used in certain limited contexts or as a lingua franca when communicating with speakers of other languages than their own. Pidgins are characterised by very simple and variable grammatical rules in which traits of the various input languages combine, by a limited vocabulary and by elimination of many grammatical devices such as number and gender (DeCamp 1971). Pidgins may exist for a long time, typically for as long as the contact situation exists; in the case of trade, they often fade with the dissolution of the trading community that created them.

Several typical situations are recognised in which pidgins arose, such as the category of trade and nautical pidgins. The earliest known recorded pidgin, Lingua Franca, was in fact a maritime or nautical pidgin, which was spoken from the time of the Crusades at the beginning of the second millennium and used in ports along the Mediterranean coast until the nineteenth century. Other well-known examples are Russenorsk (Lunden 1978), which developed in Russian–Norwegian fishing communities that met seasonally in coastal areas during the whaling season, and China Coast Pidgin – also known as Chinese Pidgin English. This emerged in Canton (present-day Guangzhou in southern China) in ports where trading with Europeans was permitted, to facilitate communication between Western and Chinese merchants, and survived for more than two centuries (Ansaldo, Matthews and Smith 2010, 2012). A trade pidgin which did not involve seafaring was Chinese Pidgin Russian (Shapiro 2010). The well-known China Coast Pidgin serial verb construction *look see* calqued on Cantonese *tái gin* 'look-see' = 'look' is shown in example 1.1 (Ansaldo, Matthews and Smith 2012: 82).

(1.1)

My wantchee look see counta
'I want to check the accounts'

Interactions between colonial people and locals led to another category of domestic workforce pidgins, such as Butler English in India, which emerged some 200 years ago in interactions between local domestic staff and their masters in colonial households. There are also multilingual workforce pidgins, the most well-known of these being plantation

pidgins, which emerged as a consequence of mass importation of labour of workers from a variety of linguistic backgrounds to plantation sites. Examples of these are Tok Pisin (Smith 2002), initially primarily used on the plantations in the German territories of New Guinea, and Hawai'i Pidgin English (Sakoda and Siegel 2003), which arose in urban centres and in interaction with Anglophones and replaced Pidgin Hawai'ian as the main language used between plantation workers of different origins. A further category of mine and industry pidgins includes Fanakalo in South Africa, which originated about 200 years ago but is still used in workforce situations such as on farms, in mines and in domestic employment (Mesthrie and Surek-Clark 2013: 34).

A pidgin is typically distinguished from a jargon. A jargon can be regarded as an unstable/early/primitive/incipient/rudimentary pidgin or pre-pidgin. As a contact variety that is highly variable and lacks a stable set of norms, it represents essentially an individual solution and is subject to individual strategies. A pidgin can also be seen to develop into an extended or expanded pidgin when it becomes the main means of interethnic communication and is used in a wider variety of domains than at its origin. With such a transformation it develops a larger set of structural norms, becomes more stable and may even become a mother tongue for some of its speakers. A recent term that has been suggested is pidgincreole (Philip Baker in Bakker 2008: 131), indicating its affinity with both pidgins and creoles.

1.2.3.2 Creoles

Creoles are nativised contact languages that emerge from the restructuring of different – usually mutually unintelligible – varieties, a consequence of situations of intense language contact, when people of diverse ethnocultural and linguistic backgrounds were brought together and formed a distinct community. The term is primarily used for the languages that developed as products of European expansion and colonisation of the world between the sixteenth and eighteenth centuries, and which share the sociohistorical background of slavery, deportation and eventual emancipation that is found in the Caribbean and the Pacific and Indian Oceans. Creoles combine traits of the various input languages; they also display novel features not easily traceable to a specific source. Typically this involves a combination of features from the lexifier or superstrate, with features from one or several substrate languages. The superstrate or lexifier is the language usually associated with higher prestige as well as late arrival. The substrate usually represents the dominant language of the groups in subordinate position.

A fundamental divide between types of creoles has to do with whether the speakers were indigenous to the region (Chaudenson 1979). Exogenous creoles were formed in situations where none of the population groups in contact with each other were indigenous to the area. Instead, it was usually the mass importation of labour, typically of people from diverse backgrounds, that led to various languages in contact. The most typical example of this is a plantation creole; these arose in the contact of the colonial plantation economy, which required large-scale importation of labour – either slave labour, characteristically in the Atlantic, or indentured labour, in the Pacific – with European settlers numerically few but dominating the societies. Examples of plantation creoles are found in various regions. Those in the Caribbean include English-lexified Trinidad English Creole in Trinidad, French-lexified Haitian Creole in Haiti and Berbice Dutch in Guyana; in southern North America there is English-lexified Gullah; off the West African coast there are the Portuguese-lexified Cape Verdean Creoles in the Republic of Cape Verde; and in the Indian Ocean there is French-lexified Mauritian Creole in the Republic of Mauritius. Maroon creoles are languages of slaves who had escaped from plantations and formed settlements based on having a common ethnic origin or on having come from the same plantation. Well-known Maroon creoles include the English/Portuguese-lexified Saramaccan of Suriname and the Spanish-lexified Palenquero in northern Colombia.

Endogenous creoles (also called fort creoles), on the other hand, typically developed through contact between immigrant settlers, usually engaged in trade, and the indigenous population of the areas. This context was typically found along the West African and Indian coasts, when European merchants set up trading centres for their activities in the area. Examples include Guinea-Bissau Kriyol in Guinea Bissau and southern Senegal, Diu Indo-Portuguese on the island of Diu to the west of India, and Papia Kristang in Malacca in western Malaysia, all Portuguese-lexified. A most up-to-date and detailed discussion can be found in Velupillai (2015).

The social dynamics, as well as the structural typologies, change somewhat when we look at languages such as Melanesian Pidgin, Solomons Pidgin, Tok Pisin and Baba Malay. On the one hand, economic migration and/or political exile enter the picture as an alternative to slavery; on the other hand, we find contexts in which European colonial languages play a lesser role. We will explore such contexts in this book, for example in the case of Sri Lanka Malay in Chapters 4 and 6.