1 Introduction

There are two broad streams of modern scholarship on the linguistic effect of human migrations and contacts, each identified with its own type of contact language. One stream of scholarship is focused on the vernacular varieties that English has morphed into in the wake of its global spread. These varieties are now commonly known as non-native, indigenized, world, or New, Englishes. The other stream is concerned with the emergence of pidgins and creoles, which started off as lingua francas in multilingual communities of various sorts. Although English is the lexical source of most extant pidgins and creoles, other languages, among them French, Malay, and Portuguese, have their share. As distinct types of contact languages, New Englishes, pidgins, and creoles are not associated with the same contact conditions. Nevertheless, they all contain lexical and grammatical innovations that can only be attributed to language contact.

Through a cursory reading of a few recently published compilations in the relevant literature, one can sense a different theoretical tenor between the two streams of scholarship. Linguists who study New Englishes tend to focus on the ethnolinguistic and sociolinguistic peculiarities of English in the colonial and postcolonial settings, and works on novel grammatical features are largely descriptive (Kortmann & Schneider 2004; Kachru et al. 2006; Hickey 2010; Kirkpatrick 2010). How the features emerged does not figure prominently on the research agenda. Pidgin and creole specialists are also interested in the ethnolinguistic and sociolinguistic aspects of the new language varieties, but they have been preoccupied with the question of origin. Within the general conceptual and analytical framework of structural linguistics, various theories have been put forth to account for the emergence of pidgins and creoles, arguing primarily from the perspectives of Universal Grammar (Bickerton 1981, 1984; and contributions in DeGraff 1999), the lexifier language (Mufwene 1994, 1996, 2001, 2010; Corne 1999; Chaudenson 2001), and the linguistic substratum (Muysken 1981; Keesing 1988; Singler 1996; Lefebvre 1998; Siegel 1999, 2008b; Aboh 2006), without losing sight of the complex interaction among all three factors. Furthermore, scholars have tried to link pidginization and creolization to imperfect second language learning, with

varying degrees of success or skepticism (Andersen 1983b; Lefebvre, White & Jourdan 2006; Siegel 2008a, 2008b; Sharma & Rickford 2009), and to account for contact-induced linguistic change in terms of surface structural congruence (Weinreich 1964; Siegel 1999), parallel grammaticalization (Bruyn 1996; Heine & Kuteva 2003, 2005), or pattern replication (Matras 2009; Matras & Sakal 2007). Contact linguistic theorizing, however, has been dominated by the debate among superstratists, substratists, and universalists, especially since the pioneering work of Weinreich (1964). It is not surprising that the genesis debate has been informed mainly by linguistic evidence from pidgins and creoles, and develops in tandem with the development of theoretical linguistics. Scholarly interest in pidgins and creoles and in New Englishes has followed somewhat different trajectories.

This book is an attempt to merge these two streams of scholarship by taking a contact-theoretic approach to the emergence and stabilization of Singapore English, among the most studied New Englishes in recent literature. New Englishes offer unique contact phenomena, linguistic or otherwise, that help shed light on the formal mechanisms that enable foreign grammatical constructions to transfer and take root. In this chapter, I first establish a typology of contact languages that includes New English, and then outline the components of the theoretical model to be developed in the subsequent chapters.

1.1 Singapore English and New Englishes

The typological status of Singapore English, and indeed that of New Englishes generally, is controversial, and the view of it as a contact language changes over time. Following DeCamp (1971), the pioneer group of scholars described Singapore English as a pidgin, with the typical speech continuum ranging from the acrolect, through the mesolect, to the basilect (Platt 1975, 1977a; Tay 1979; Platt & Weber 1980). Platt (1975) labels the basilectal variety of Singapore English a creoloid, which has creole-like features but lacks the usual pidgin predecessor. Arends et al. (1994) classify the vernacular as an extended pidgin. These views are clearly influenced by the work on pidgins and creoles at the time. Others adopt Ferguson's (1959) notion of diglossia and classify Singapore English as the low, colloquial variety in diglossic opposition to Standard English, the high, formal variety that one learns in school (Gupta 1989). This is the theoretical basis for the dichotomy between Singapore Standard English and Singapore Colloquial English, which is still used in recent works on the vernacular.

In Kachru's (1982, 1985) geographically based model of World Englishes, the world is arranged in three concentric circles. In the inner circle are countries such as Britain, America, Australia, and New Zealand, where English is

the native language. These countries saw large-scale settlement of English-speaking people, with constant dialect contact, but contact with other languages was limited, especially when compared with the extensive contact in the outer and expanding circles (Mufwene & Pargman 2003; Trudgill 2004; Wolfram & Schilling-Estes 2006; Mufwene 2009). Singapore is placed in the outer circle, along with India, where historically English functioned as the language of colonial administration with insignificant settlement by English-speaking people. The rest of the world belongs to the expanding circle, where English is by far the most popular foreign language being taught and learned. The three-circle model is viewed as an alternative to Quirk's (1985) model in terms of acquisition order. For Quirk (1985), English is the native and first language in the inner circle, the second language in the outer circle, and the foreign language in the expanding circle. In the outer circle, but not the expanding circle, English also functions as an official language, in addition to the local languages. Indeed, Singapore has extensive bilingualism, with English as the common second language among the three main ethnic communities (Tay 1979; Pakir 1991; Thumboo 2001).

More recently there have been attempts to approach New Englishes from different perspectives. Schneider (2003, 2007) argues for a dynamic model of what he calls "post-colonial Englishes," tracing the development of English over a five-stage trajectory. English in Australia and New Zealand has reached the final stage, where it has a differentiated repertoire of regional and social dialects. Singapore English is placed at Stage 4, the stage at which the endo-normative, or locally derived, linguistic norms would have emerged and stabilized. Alsagoff (2010) proposes the cultural orientation model, arguing for a culturally relevant perspective on Singapore English, linking its internal variation to the multilingual and multicultural mosaic that is Singaporean society. Inspired by the indexical approach to variationist studies (Eckert 2008), Leimgruber (2012, 2013) proposes the indexical model, according to which the linguistic variables of Singapore English are indices of social meanings or stances. For both Alsagoff (2010) and Leimgruber (2013), whether Singapore English is a pidgin, creole, or New English is not an interesting question. It is the linguistic tool with which the speakers navigate the social and cultural mosaic of Singapore.

The diverse views of Singapore English share one common assumption, namely, Singapore English is the result of English having accumulated lexical and grammatical features from the local languages over the past 200 years. This is explicitly stated as one of the defining criteria of creoloid in Platt (1975:373). Schneider's (2003) dynamic model likens the emergence of New Englishes to dialect birth, assisted through the midwifery of the languages in the new contact ecologies. Dialect birth, of course, presupposes the existence of a common language before dialect differentiation sets in. From the perspective

of contact linguistics, the models of New Englishes and of Singapore English surveyed above can all be considered superstratist, despite the models' differences in orientation and emphasis.

It is the genesis of pidgins and creoles that is more controversial in the contact linguistics literature. For Thomason (1997, 2001) and Thomason and Kaufman (1988), pidgins result from broken transmission, with limited lexical and structural resources. They are not the native or primary language of the speakers. This view is shared to a varying extent by substratists (Keesing 1988; Lefebvre 1998) and in different ways by universalists (Bickerton 1981, 1984), but uniformly rejected by superstratists (Mufwene 1991, 2001; Chaudenson 2001). The superstratist orientation compels us to see New Englishes and English-based pidgins and creoles as dialects of English that emerge naturally under different contact conditions, without any break in transmission. The continuity of transmission in contact languages, New Englishes included, is assumed in other theoretical models as well, from the model based on pattern replication (Matras & Sakal 2007; Matras 2009) to models based on contact-induced grammaticalization (Bruyn 1996; Heine & Kuteva 2003, 2005).

It is clear from our current state of knowledge that New Englishes are distinct from English-lexified pidgins and creoles. How to integrate them into an inclusive typology of contact languages is not a straightforward matter. The typology proposed in Thomason and Kaufman (1988) and Thomason (1997) is based on the historical information concerning lexical and grammatical transmission, adult or child acquisition, and communicative roles. It contains three distinct types: pidgins, creoles, and bilingual mixed languages. Pidgins involve broken transmission, have no native speakers, and are limited in communicative role. They become creoles when they acquire native speakers and serve the primary communicative functions in the communities. Bilingual mixed languages are the result of contact between two languages where there is extensive bilingualism. Typological work that relies on morphosyntactic properties of contact languages, such as that proposed in McWhorter (1998), is ill-equipped for the task. There is no single list of grammatical features that could uniquely define pidgins, creoles, and mixed languages as distinct types. New Englishes are not addressed at all by creolists. From the perspective of transmission, the models proposed by Kachru (1982, 1985) and Schneider (2003, 2007) provide us with interesting, albeit different, perspectives on New Englishes, but their models ignore pidgins and creoles that are well known in the contact linguistics literature.

Chaudenson (1977) examines the nature of the economic life of a community and its effect on the emerging contact language. Two types of communities are distinguished: exogenous and endogenous. Exogenous communities are agricultural plantations which are reconstituted with linguistically diverse groups of slaves or indentured laborers and are far removed from the

geographical origin of any of the constituent groups. Obviously, such communities need lingua francas. By contrast, endogenous communities are not reconstituted, and commerce is the main economic activity. Linguistically, however, the agricultural or commercial activities of the multilingual population are not as important as the absence or presence of the main substrate language(s) in the community.

Exogenous contact languages arise in communities that do not have a common, numerically dominant, substrate language, whereas endogenous contact languages arise in communities in which the indigenous populations in contact may continue to speak their mother tongues. In both communities, it is assumed that the lexifier is not the scholastic variety, but the vernacular, typically nonstandard, variety. Chaudenson's (1977) superstratist approach is further developed by Mufwene (1996, 2001), who, influenced by population genetics, analogizes languages to species, and the emergence of contact languages to speciation. According to the population genetics model, the lexifier language evolves into different varieties, or "sub-species," in adaptive response to the various factors in its new environments. The lexifier language draws grammatical features from the feature pool, the analogue of gene pool, available in its new contact ecology. The selection and recombination process gives rise to a contact variety that bears the imprint of the languages that contribute to the feature pool. This typology of contact languages is defined in terms of the socioeconomic characteristics of the communities. Both pidgins and creoles have evolved from the nonstandard vernacular varieties of the lexifier language. Pidgins emerged in trade colonies, in which contacts between speakers of the lexifier and those of indigenous languages are sporadic; whereas creoles emerged in settlement communities with multilingual yet relatively stable populations engaged in various political, social, and commercial activities that sustain the communities. This typology has been extended to include New Englishes as well. Like English-lexified creoles, New Englishes emerged in former British colonies that resemble endogenous settlement communities, but they have evolved from scholastic varieties of English (Mufwene 1994, 2001, 2009).

I summarize this typology of contact languages in Table 1.1.

The definition of a community in terms of trade and settlement patterns is not as clear-cut as it might seem. For example, Tok Pisin is considered a pidgin or expanded pidgin, but its ecology is not characterizable as a trade ecology (Keesing 1988; Arends et al. 1994; Mühlhäusler 1997; Romaine 1999). Nevertheless, community and lexifier serve as good sociolinguistic criteria for typing contact languages. It is worth noting that the typology displayed in Table 1.1 is superstratist, which recognizes only a limited role for the substrate languages in influencing the grammatical development of the contact languages. It compels us to view English-lexified pidgins, creoles,

Table 1.1 *The typology of contact languages according to Chaudenson (1977, 2001) and Mufwene (1996, 2001)*

	Pidgin	Creole	New Englishes
Community	trade	settlement	endogenous
Lexifier	vernacular	vernacular	scholastic

Table 1.2 *A revised typology of contact languages*

	Pidgin	Creole	New Englishes
Community	trade	settlement	endogenous
Lexifier	vernacular	vernacular	vernacular, scholastic
Substratum	varied	varied	constant

and New Englishes as dialects of English, the result of English adapting to new linguistic ecologies.

There is, however, one important sociolinguistic factor that distinguishes New English ecologies from pidgin and creole ecologies. In pidgin or creole communities, the heritage languages of the migrants are often diverse and their communicative roles are limited. This is especially true in exogenous communities. The nature of the linguistic substratum shows considerable variation across extant pidgin/creole ecologies (Mufwene 2001, 2008). The newly created pidgin or creole quickly assumes the role as the primary means of communication in the community. This is not the case for the endogenous ecologies that produce New Englishes. In India and Singapore, two former British colonies, English has enjoyed the status of a prestige language since the early days of colonial rule. While it fails to displace the languages of the local population, it remains the language of the elite, playing an indispensable function in government, education, and commerce. The same languages, English and the languages that form the linguistic substratum, have been in constant contact since the Big Bang, as it were, when English started on the journey of dialect birth or speciation in new linguistic ecologies. It is this contact dynamic, not the putative scholastic English, that sets New Englishes apart from English-lexified pidgins or creoles. We recast the Chaudenson–Mufwene typology in Table 1.2.

New Englishes are distinguished from pidgins and creoles not in terms of the lexifier language, whether vernacular or scholastic, but in terms of the contact dynamic – they emerge and develop in ecologies with the same mix of languages and the same linguistic and sociolinguistic relationships among the

languages. In this respect, typical pidgin or creole ecologies have linguistically and sociolinguistically varied linguistic substrata (Mufwene 2001, 2008). In population genetics terms, a New English ecology provides a gene pool which is continuously replenished from the same linguistic stock. As a result, New Englishes undergo grammatical restructuring under the constant and often antagonistic tension from the contributing and competing languages, and it is possible that the impact of the substratum is more pronounced in New Englishes than in English-lexified pidgins and creoles which emerge in contact ecologies with diverse linguistic substrata. In Hawai'i, for example, the initial group of Chinese laborers were brought to the sugarcane plantations in the mid-1800s, and were joined at the turn of the twentieth century by laborers from Japan and the Philippines (Roberts 1998, 2000; Siegel 2000), diluting the effect of the Chinese substratum on Hawai'i Pidgin/Creole English. Pointedly, this is not the case for ecologies that give rise to New Englishes. Indeed, it is the continued presence of the linguistic substratum that determines the nature and extent of substrate influence, regardless of the type of contact language that emerges.

The national identifiers – "Singapore" in Singapore English and "Indian" in Indian English, for example – may refer to countries where they are spoken, but more importantly, they identify the languages which have midwived and nourished the birth and development of the New English variety. Since each New English ecology has its unique history and unique language mix, New Englishes tend to differ from each other in the kinds of substrate features that have been appropriated into their grammars.

1.2 The thematic description of the book

The present book is an in-depth analysis of the grammar of Singapore English, focusing on the grammatical constructions that have been traced to Chinese or Malay, the main languages of the contact ecology. I develop a usage-based theory of contact-induced grammatical restructuring. The basic ideas of the theoretical model have been discussed in a series of papers I have published (Bao 2001, 2005, 2009, 2010a, 2010b). These ideas will now be integrated into a coherent narrative. Here I briefly describe the terminology, the data collection methods, and the theoretical assumptions on which the analytic model is built.

1.2.1 Terminology

To the extent possible, I follow common terminological conventions. The unmodified English is used as a broad and inclusive term to refer to the English language without the usual geographic affiliations. When it is used in opposition to New English, the term refers to the native English spoken in

Britain and its former settlement colonies of North America, Australia, and New Zealand. So English does not have Chinese-sourced grammatical constructions that are commonly found in the English vernacular heard in Singapore. The term SINGAPORE ENGLISH is also used in two senses. In the broad sense, it refers to English used in Singapore across all domains of life, from government to education to informal banter in coffee shops. In the narrow sense, it refers to the vernacular English spoken in informal contexts, which is also known as Singlish or Singapore Colloquial English. Given the focus of the book, the term is used mostly in the narrow sense to refer to the vernacular that is acquired at home (Gupta 1994). At the present time, the vernacular is the dominant home language spoken by a third of households in Singapore (*Census of Population* 2010).

In the contact linguistics literature, FEATURE is a cover term for lexical items, phonological contrasts, or morphosyntactic constructions, a practice which I will follow. It is nearly synonymous with the term CONSTRUCTION in Construction Grammar, where constructions are any linguistic units that are paired with some sort of meaning (Goldberg 1995). Since my primary focus is the grammar of Singapore English, I will use the term CONSTRUCTION when the structural aspect of a substrate feature needs to be foregrounded; otherwise, I will use the two terms interchangeably.

Also interchangeable are the terms TRANSFER and APPROPRIATE, without the usual theoretical commitment associated with them. The two terms are metaphors, linking the source and target languages of relevant grammatical features. There is general agreement in the literature that substrate features exist in the contact language, be it pidgin, creole, or New English. There is, however, a clear difference in perspective between the substratist and superstratist accounts of the phenomenon. Substratists tend to see features transferring from the substrate language to the contact language, whereas the superstratists see the lexifier appropriating features from the substrate language as it evolves into the contact language. Following Mufwene (1994, 2001) and Schneider (2003), I see New Englishes as dialects of English. This position is not incompatible with the substratist theory that I will develop in the following chapters. I argue that the substrate-influenced grammatical restructuring that Singapore English has undergone supports a theory that places the substratum at the core of analysis and explanation. I will use the two terms, where appropriate, in describing the features in Singapore English that are derived from the linguistic substratum. A substrate feature may transfer to, or be appropriated by, the contact language.

When a substrate feature transfers to the contact language, it needs to be "fleshed out," or expressed, with suitable materials from the lexifier. I call the process EXPONENCE, and the morphosyntactic means of expressing the transferred feature, EXPONENT. The exponents of substrate features vary, ranging

from lexical items to abstract morphosyntactic structures or categories. Take past tense and reduplication, for example. The past tense is exponenced as *-ed* in English, as zero in Chinese, and as *-ed* or zero in Singapore English. Similarly, the so-called tentative aspect is exponenced through verbal reduplication in Chinese (*zǒu-zǒu* 'walk a bit') and to a rather limited extent, in Singapore English (*walk-walk*). It does not exist in English.

1.2.2 Data

The data used in this book are collected from published works, field notes, computerized corpora, and intuitions of native or fluent speakers. Corpus data and intuition data are complementary, despite the obviously different philosophical underpinnings between the two types of data and the two modes of collection (Wasow & Arnold 2005; Bisang 2011). Computer corpora reveal the usage pattern of a given grammatical feature, whereas intuitions reveal the full range of formal properties of the feature, some of which may not be represented in a corpus. The corpus that we will rely on for usage data is ICE-SIN, and to a lesser extent ICE-GB, respectively the Singaporean and British components of the International Corpus of English.[1] We will also make occasional use of data from SCoRE, the 4-million-word corpus of classroom discourse collected from the country's primary and secondary schools by researchers at the National Institute of Education (Hong 2009). The data from these corpora corroborate the data from published sources, and from native-speaker intuition.

Intuition has been a source of data in linguistic analysis since the generative turn in structural linguistics (Chomsky 1965, 1986, 2000). For generative linguists, the object of linguistic inquiry is the I-language, which is the grammar internalized in the speaker's mind that can only be accessed through introspection. Linguistic data are elicited through introspective grammaticality judgment. Grammatical intuition reflects the stable properties of the speaker's I-language and offers a unique window into the structural possibilities of the language regardless of use. In this book, we rely on the intuitive judgments of native or fluent speakers to map out the complete range of structures associated with a given grammatical subsystem or construction, including those structures which are seldom used and may only be attested in a sufficiently large corpus.

[1] The International Corpus of English, or ICE, was first proposed by Sidney Greenbaum (1988). It is a collection of corpora from countries where English is used as a vernacular or as an official language, usually also as an important lingua franca in the latter case. To facilitate comparisons, the components of the ICE follow the same design structure. Each country corpus is composed of five hundred 2,000-word texts, for a total of 1 million words. The texts are grouped into three major registers: DIALOGUE (private: 100 texts; public: 80 texts), MONOLOGUE (scripted: 50 texts; unscripted: 70 texts), and WRITING (printed: 150 texts, nonprinted: 50 texts) (Greenbaum & Nelson 1996). We will use the 200,000-word PRIVATE DIALOGUE, which samples spontaneous conversations.

However, intuition-based data need to be handled with caution. A typical Singaporean speaks English and one of the three so-called mother tongues: Chinese, Malay, and Tamil, as well as other languages such as Hindi and Bengali. Given the widespread bilingualism, it is often impossible to determine if the informant refers to the "right" grammar when making grammaticality judgments on Singapore English data. A grammaticality or acceptability judgment on a Singapore English utterance may reflect the informant's competence in Singapore English, in Chinese, or in English; or in any combination of the three languages. Consider the data in (1):[2]

(1) a. Big one

b. ?John one
'John's'

c. ??John give Mary scold.
'John was scolded by Mary.'

d. *John one book
'John's book'

In English, (1b,c) are both ungrammatical. Among Singaporean informants, the grammaticality status of (1) is not clear-cut. Example (1a) is unequivocally grammatical and (1d) ungrammatical. Example (1b) is grammatical for many, who are quick to label it basilectal. Example (1c) is acceptable for only a small number of informants, who will conjure up a scenario in a Chinese-dominant coffee shop or playground where one might be able to encounter such utterances. The Singapore English data have exactly the same structure as the Chinese counterparts shown in (2), all of which are grammatical:[3]

(2) a. dà de
big PRT

b. Zhāngsān de
Zhangsan PRT
'Zhangsan's'

[2] In the book, Singapore English data are presented in English orthography and will be glossed when the meanings are obscure. The question mark and the asterisk in (1) represent different degrees of acceptability by native-speaker informants.

[3] A brief note on transcribing Chinese data is in order. Mandarin data are transcribed in pinyin with lexical tone marks, without indicating the effect of tone sandhi, as in (2). In Mandarin, grammatical morphemes are often pronounced with the so-called light tone, which is typically analyzed as toneless in Chinese linguistics. Such morphemes are transcribed without tone marks. For the sake of consistency, the Mandarin data cited from published sources will be re-transcribed in pinyin if the sources use idiosyncratic transcription schemes. For lack of a standard romanization scheme, data from Chinese dialects are cited from sources verbatim or transcribed in Roman alphabet without tone marks.