Bilingualism, language contact and change: The case of Bengali and English in India

by

Tridha Chatterjee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in the University of Michigan
2015

Doctoral Committee:

Professor Marlyse Baptista, Co-Chair
Professor Acrisio Pires, Co-Chair
Professor Madhav Deshpande
Professor Sarah G. Thomason

Acknowledgements

This dissertation would never have been possible without the immense support and guidance of my co-advisor, Marlyse Baptista. Her role in my professional development has been tremendous. Her high standards for methods, analysis, and writing have been vital in the completion of this dissertation. My other guide and mentor, my co-advisor, Acrisio Pires has influenced my graduate career immensely with his continuous encouragement, support and insights on my work. I owe a deep gratitude to Sally Thomason for her indispensible guidance and insightful comments on my project, which have made me a better writer and clearer thinker. Thanks also to Madhav Deshpande for his expertise in Indo-Aryan languages and interest in my work.

In addition to my committee, many other members of the Linguistics Department have supported me over the years. My first mentor in graduate school, Carmel O'Shannessy helped me immensely in understanding the basics of linguistics research and academic writing. I also thank Andries Coetzee for being an approachable graduate chair; the members of the SoConDi discussion group for their valuable feedback on my research; the past and present support staff of the Linguistics department, Sylvia Suttor, Tim Moore, Jen Nguyen and Sandee Petee for their help. My research was also supported by financial assistance from Rackham Graduate School and the Linguistics Department; the staff of CSCAR for help with the quantitative analysis; Jeffrey Martin, librarian for South Asian Studies in Hatcher Library for helping me locate Bengali plays; the undergraduate students who helped code the English data.

I am indebted to many graduate students in the Linguistics Department for their companionship and guidance: Harim Kwon and Mike Opper, for always being there, Sujeewa Hettiarachchi, my connection to South Asia, Jae Young Shim, for making me laugh so often, Erica Beck for her help and support during my move to Ann Arbor, Cameron Rule, for agreeing to read my dissertation, Jon Yip, for sharing my frustrations regarding teaching, David Medeiros, for always making sure that I did not feel left out in a group, and Miyeon Ahn, Jiseung Kim and Yiwen Zhou for being such warm people.

I thank my friends in Ann Arbor outside of Linguistics: Bratati Kahali, Kohinoor Dasgupta, Samopriyo Maitra, Mukesh Patir, Liduina Kimbu, Surma Talapatra, and Madhumita Dutta for their kindness and warmth, for helping me overcome homesickness, and treating me to delicious home-cooked meals. My dearest friends in India, Tansushree Bhattacharya, Ayesha Gautam, and Saswati Chatterjee have always expressed a keen interest in my work and without their help and that of several other relatives and acquaintances in Kolkata and Durgapur, I would have never been able to complete my fieldwork.

Finally and most importantly, I am immensely grateful and appreciative of my family's support throughout my graduate studies. Without my husband, Soumya De's support and motivation, I would never have started graduate studies, applied to the University of Michigan or survived its rigors. His continuous encouragement has been indispensible for my progress in graduate school. My father and mother, Tushar and Lakshmi Chatterjee and my sister, Deepsikha Chatterjee contributed immensely to my well-being and education. Their faith in me has always given me the strength to carry on in life. Ann Arbor, my home for the last six years, also deserves mention for being such a kind, welcoming and safe city and for helping me grow up in many ways.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

ACC	Accusative
CL	Classifier
DAT	Dative
DEF	Definite
FUT	Future
GEN	Genitive
HAB	Habitual
HON	Honorific
IMPFV	Imperfective
LOC	Locative
NEG	Negator
P	Person
PFV	Perfective
PL	Plural
PROG	Progressive
PRS	Present
PST	Past
PTCP	Participle
SG	Singular

Abstract

This dissertation examines the nature of bidirectional influences in Bengali and English in the understudied bi/multilingual setting of West Bengal, India. Close contact between the two languages has resulted in extensive bilingualism in a section of the population. In order to get a better understanding of language change and development in this community, the following research questions are investigated: i) What types of contact effects are observable in Bengali today? ii) Which grammatical features in bilingual Bengali-English have changed under contact and which ones result from language-internal developments? iii) How can we ascertain that new properties found in bilingual Bengali-English are contact-induced? iv) What are the grammatical differences between monolingual and bilingual Bengali? These questions are addressed from both diachronic and synchronic standpoints. Two corpora of spoken monolingual and bilingual Bengali, collected and compiled through fieldwork in India, were compared with nineteenth-century Bengali plays for a set of linguistic features, following the methodology of Thomason (2001). Further, a corpus of English speech was collected from Bengali-English bilinguals and analyzed to evaluate what types of changes occurred in English. This investigation has also been complemented with a quantitative analysis of the distribution of linguistic features in the corpora, following a variationist methodology (Poplack & Levey 2010). The findings indicate that English influence has led to extensive code-switching, heavy lexical borrowing in Bengali, and morphosyntactic changes in the domain of bilingual Bengali-English complex verbs, formed with English nouns or verbs and the 'do' verb from Bengali. Change was also identified in Bengali equational sentences, where a new meaning and position of the *hocche* 'be' verb, which was unattested in the nineteenth-century Bengali plays, were argued to emerge from multiple causation. The English data reveals slight divergences in the use of articles and progressive forms, which have resulted partially from interference from Bengali, but are mostly conditioned by bilingual speakers' different proficiency levels in English. These findings provide evidence for overall stability in the grammar of both languages (specifically for the subject pool that was tested in this dissertation), despite the presence of extensive bilingualism in the community.

Chapter 1: Introduction

Several types of outcomes may result when languages come in contact with each other. These outcomes may involve subtle phonological or morphosyntactic changes in the grammar of one or both of the languages, or more obvious changes such as lexical borrowing and code-switching between the languages (the use of two or more languages in a single utterance). Much research has been produced in the past decades, developing frameworks to account for such kinds of contact phenomena, attempting to distinguish contact-induced changes (involving processes such as interference and transfer) from internally-motivated ones; identifying changes that emerged from multiple causation and proposing constraints on code-switching (e.g. Haugen 1950, Weinreich 1953, Poplack 1980, Disciullo, Muysken and Singh 1986, Van Coetsem 2000, Thomason & Kaufman 1988, Myers-Scotton 1993a, 1993b, Muysken 2000, Backus 1996, Thomason 2001, Poplack & Levey 2010). Although these works have furthered our understanding of how languages behave when they come in contact, many challenges and puzzles remain.

India is one of the areas in the world where there is more widespread language contact, between different Indian languages and between Indian languages and English. There is a sizable body of research investigating language contact between English and languages such as Hindi, Punjabi, Marathi and Tamil, often focusing on contact-induced changes that have occurred in these languages due to the long-term influence from English (e.g. Pillai 1968, Bhatia 1982, Kachru 1975, 1994, Annamalai 1989, Romaine 1986, Pandharipande 1990, Muysken 2000,

Puri 2011). Indian languages have also been shown to affect the variety of English that is spoken in India, also known as Indian English (IndE), with effects on its article system, copula use, agreement marking including number, tense etc. (e.g. Selinker 1972, Agnihotri et. al. 1984, 1988, Sand 2004, Sharma 2005a, 2005b, Sedlatschek 2009).

Bengali, a prominent (eastern) Indo-Aryan language spoken in India and Bangladesh[1] has been in contact with English for more than two hundred years in the Bengal region corresponding to eastern India and Bangladesh, leading to the emergence of extensive Bengali-English code-switching among part of the population of English-educated Bengalis (especially in informal contexts), which I will identify as bilingual Bengali speech at different points in this dissertation. Yet, there has been very little research on Bengali-English contact. Given this fairly intense contact, it is reasonable to expect that Bengali and English may have mutually affected each other, possibly leading to changes at the lexical, phonological and morphosyntactic domain of both languages. We could also expect to find that bilingual Bengali speech has undergone different degrees of contact-induced change, which may have extended to monolingual Bengali (spoken by monolingual speakers of Bengali). The investigation of this contact situation poses different challenges, such as ascertaining and distinguishing the subtle changes (in addition to the arguably obvious ones, such as lexical borrowing) that have taken place due to contact, internally motivated changes and changes occurring from multiple causation. In addition, examining the direction of change in the languages of contact is also a concern.

In this dissertation, I attempt to shed light on the nature of contact between Bengali and English in West Bengal, India by specifically investigating the different types of changes

[1] The Bengal province was partitioned to create West and East Bengal in 1905 by the British administration in India. The East Bengal province became a part of Pakistan in 1947 (when India gained independence and Pakistan was born out of India), and came to be identified as East Pakistan. In 1971 East Pakistan seceded from Pakistan and became a new country, Bangladesh.

observable in bilingual Bengali, as used by proficient Bengali-English bilinguals, in comparison to the corresponding monolingual varieties of English and Bengali. First, I carry out a synchronic comparison of bilingual speech by bilinguals with monolingual Bengali speech by Bengali speakers who have little to no exposure to English. Second, in conjunction with the synchronic comparison, I undertake a diachronic comparison of monolingual and bilingual Bengali to an older variety of monolingual Bengali, as represented in Bengali plays from the nineteenth century, which were written prior to the development of intense contact with English. The objective is that diachronic comparison will help us establish instances of contact-induced change that have taken place in Bengali (following a framework proposed in Thomason 2001). We not only expect bilingual Bengali to have undergone many changes in relation to the older variety of Bengali, but also expect some amount of change in monolingual Bengali given the regular interaction between monolingual and bilingual speakers and the possible diffusion of contact features from bilingual to monolingual Bengali speech. Third, I investigate changes in the English spoken by bilingual Bengali-English speakers to ascertain possible contact effects from Bengali. Through this investigation, I first try to establish the changes that have taken place in Bengali and English, including the direction of change (English to Bengali and Bengali to English) and then attempt to identify whether they have emerged from contact, internal developments within the languages or from multiple causation.

This dissertation not only fills a gap in the literature on language contact in India, but it also addresses some of the challenges in investigating change by using synchronic and diachronic methods of comparison to examine the bidirectional nature of change in this contact setting. The specific research questions I investigate in this dissertation are listed below (1.1), and the organization of this dissertation is laid out in section (1.2).

1.1. Research questions

1) *Given the contact between Bengali and English, what are the types of contact-induced changes that are observable in Bengali today?*

   i) Are there changes that are easily observable in speech?
   ii) Are there changes that are more subtle in nature?

Question (1) examines firstly the contact-induced changes in Bengali that are easily noticed. These would include changes in the lexicon such as the borrowing of English vocabulary into Bengali, the use of loan translations or calques in Bengali, changes at the morphosyntactic level such as the use of bilingual verbs (consisting of English lexical nouns or verbs forming complex verbs with Bengali helping verbs such as *do* and *be*) and the emergence of extensive code-switching between Bengali and English. Secondly, it aims to investigate other subtle changes that may have occurred in Bengali, primarily in the grammar of the language. In order to determine these changes, I first investigate properties of Bengali grammar (as represented by nineteenth-century Bengali plays) prior to the increase in contact with English (which would have resulted from the introduction of English as a language of instruction in the education system of India, as discussed in chapter 2). Then I systematically compare those features to modern Bengali speech and to English (following Thomason's (2001) framework discussed in chapter 3). These research questions are addressed in chapters 5 and 6 of this dissertation.

2) *What are the grammatical features in bilingual speakers' speech that have changed under contact and what are those changes that do not have any evidence of being contact-induced?*

   a. *How can we ascertain that the new properties found in bilingual Bengali-English speech are contact-induced?*

With question (2) I address the issue of how to ascertain that the changes found in bilingual speech are indeed contact-induced, and the theoretical framework of contact-induced change that

I use (Thomason 2001) will help in this investigation (I describe this framework in detail in chapter 3). It also takes into account those changes that cannot be argued to be only contact-induced and are thus likely to be internally-motivated or to result from multiple causes. I discuss this issue in chapter 6 in relation to the copular predicates of Bengali equational clauses.

3) What are the grammatical differences between monolingual and bilingual Bengali?

Question (3) examines how modern monolingual and bilingual Bengali differ from each other. As mentioned before, many Bengali speakers in West Bengal primarily communicate in bilingual speech (in informal contexts), which is partially due to their formal education in English. There are also people with very little exposure to English education primarily speaking monolingual Bengali, which shows little borrowing from English and almost no code-switching. Given this linguistic scenario, we expect monolingual and bilingual speech to differ from each other, not just in terms of the degree of borrowing and code-switching but also in terms of the range of morphosyntactic changes that have taken place in each one of them. Chapters 5 and 6 address this question by providing evidence for the changes that have taken place in monolingual and bilingual Bengali and developing a precise analysis of these changes.

4) What is the direction of change in this contact situation?

   a. Are there changes that are detectable as occurring from English onto Bengali?
   b. Are there changes that are detectable as occurring from Bengali onto English?

Question (4) addresses the likelihood that contact between Bengali and English has led to bidirectional change in the contact situation. The term bidirectional simply refers to mutual contact-induced changes of two languages on each other (e.g. Pavlenko & Jarvis 2002). By using this term I do not imply that both English and Bengali have led to mutual changes in terms of the

same set of features (I address this issue in chapter 3). To determine whether the English spoken by bilingual speakers has been influenced by Bengali, I examine in chapter 7 a corpus of monolingual English speech data (introduced in chapter 4 on methodology) for the properties of articles and progressive aspect. I investigate these particular features because they have been discussed in several studies as showing divergences in Indian English.

1.2. Organization of the dissertation

This dissertation is organized into 8 chapters, and each chapter highlights distinct components of the project. In chapter 2, I delve into the socio-historical background of contact between English and Indian languages in India and specifically focus on the contact between English and Bengali in the Bengal region. I discuss the social, cultural and literary changes that took place in Bengali as a result of this contact. In chapter 3, I review several relevant studies of language contact and change and their theoretical underpinnings in order to contextualize the research questions that I investigate. Chapter 4 introduces the general methodology I have used in the investigation of the research questions. In this chapter, I provide details on the different corpora I examine in this dissertation including how they were collected and compiled. I also briefly discuss the method I used to code the features for the quantitative analysis. In chapter 5, I first discuss the contact-induced changes in Bengali that have taken place at the lexical level, the occurrence of code-switching and then investigate in detail changes at the morposyntactic level in the use of bilingual verbs. The focus of this chapter is therefore on changes that are clearly identifiable in speech. In chapter 6, I focus on changes that are more subtle in nature, examining changes in copular predicates of Bengali, which I analyze as resulting from multiple causation. In chapter 7, I focus on the English spoken by bilingual Bengali-English speakers in West Bengal and investigate the use of articles and progressive aspect in this variety of English and

whether they show evidence of changes. I argue that there are minimal divergences in English in these domains, which are related to different proficiency levels in English, that is, speakers with higher proficiency in English are less likely to have divergences than those who are not as proficient. In chapter 8, I provide the conclusion of this dissertation.

Chapter 2: The socio-historical background of contact between English and Indian languages

2.1. Introduction

In this chapter, I provide a socio-historical background to the contact between Indian languages and English, specifically focusing on the Bengali-English contact situation. First, I outline the relevant historical events that led to the inception and development of contact between English and Indian languages. Then I narrow my focus on the situation of Bengali-English contact, including the social and cultural ramifications of that contact in the Bengal region.

This chapter is organized as follows: In section (2.2), I discuss the factors that brought English speakers to India at the beginning of the seventeenth century including the linguistic situation in India prior to the arrival of the British. Then in section (2.2.2), I discuss the historical events that led to the dissemination of the language in India, through the administrative policies of the East India Company and exchanges between the British and Indians in areas of trade, military and governance. This section also concentrates on the important milestones in English education in India, with education in English being the most important tool for expansion of the language in the Indian subcontinent. The status of English in India after Indian independence (in 1947) and its overall linguistic impact on Indian languages is described in section (2.2.3). In section (2.3), I turn to the Bengal region and discuss the historical events (2.3.2) resulting in contact between English and Bengali and how the contact has continued thereafter. In section

(2.3.3), I describe some of the influences of English education and language on the overall social, cultural and literary situation in Bengal. I focus on the social, cultural and literary developments in Bengal in addition to the linguistic changes because these three spheres underwent concurrent influences from the English language and culture, and they in turn impacted each other. Also, given that I use literary texts for the diachronic investigation of linguistic changes in Bengali (following the framework of Thomason 2001), this discussion will justify my choice of specific literary texts and genres for the investigation. In section (2.4) I provide the conclusion of this chapter.

2.2. History of English in India

In this section, I first account for the historical events that led to the introduction of English in India and then discuss the linguistic situation of India when the first English speakers arrived. Then I outline the factors that led to spread of the language in the Indian subcontinent and its political and social status after Indian independence to the present.

2.2.1. English and its predecessors in India

A few merchants from London, Britain who were granted a charter on 31st December, 1600 by Queen Elizabeth I for trade with regions in the East, were the first people to introduce the English language to India. This charter led to the establishment of the East India Company (henceforth, EIC), which traded and governed in the Indian subcontinent till 1858, after which the British government took sole responsibility of governance of India. English was the third European language that came to India after Portuguese and Dutch. By the time the English traders came to India, the Portuguese had already gained military control over parts of the western and eastern coast of India (see map in appendix 1). Portuguese influence in India had

paved the way for the formation of Portuguese-based pidgins, which served as the primary means of communication between Indians and Europeans and also between Europeans in India (Sailaja 2009: 96-98). Unlike the traders of the EIC whose sole aim at first was to set up trade relations with India, the Portuguese were involved in proselytization and missionary activity in addition to trade and military control. By the time the British arrived in India, the Portuguese had already introduced Christianity in the subcontinent. To propagate Christian creeds, the Portuguese had also started imparting religious education through the diffusion of religious texts printed in the Roman script. The influence of the Portuguese language at the time was to the extent that all Europeans who came to India had to learn Portuguese in order to operate in the region. The EIC alone had a repository of over two hundred Portuguese dictionaries and employed many interpreters to translate between English and Portuguese (Sailaja 2009: 96-98; Krishnaswamy & Burde 1998: 8).

2.2.2. The dissemination of English in India

After gaining permission from the Moghul emperor in Delhi (the most powerful dynasty in India at the time), the EIC set up its first trading post in Surat in Gujarat (western India) in 1612. Within a few years, other trading centers were set up in the regions comprising Madras in the South (1639), Bombay[2] in the West (1668) and Hoogly (Hugli) in Bengal in the East (1681). Up to 1700 the EIC was still trying to gain a foothold on the Indian subcontinent, although it was primarily interested in trade and faced stiff competition from the Portuguese, Dutch and French companies and resistance from many Indian rulers against the expansion of their trade activity. At first, use of the English language was limited within the British trading posts where it was

[2] Bombay was given to Britain in 1662 by Portugal as dowry for King Charles II's wedding to Catherine of Braganza (Sailaja 2009: 97).

used by the British merchants, military personnel, and the few Indian traders, translators, interpreters and peons working in the EIC (Sailaja 2009: 97-98; Krishnaswamy & Burde 1998: 10).

Up until the turn of the seventeenth century, British missionaries were not encouraged by the EIC to come to India. However, when the charter for trade was renewed and a 'missionary' clause was added in the charter in 1698, India received an influx of British missionaries. By this time Catholicism had gained some standing in India thanks to Portuguese missionaries and the EIC subsequently authorized British missionaries to come to India in order to spread Protestantism. The spread of Protestantism also included evangelical education, which took place either through Portuguese or Portuguese-based pidgins or the local Indian languages. But these initial attempts were not very successful as religious books were available either in English or in Portuguese and were unintelligible to the Indian pupils. The first European school that was set up in India was at Fort St. George (the British settlement in Madras) in 1670 by a Scottish preacher who imparted religious education to British and Portuguese children and a few children of the Indians who worked in the Madras Presidency. Between 1670 and 1714, this school imparted religious education in Portuguese, which got replaced by English in 1715 by Rev. William Stevenson. This school was the first of its kind to impart education (mainly consisting of the teaching of the Gospel) in English and was followed by the setting up of similar schools in Bombay (1719) and Calcutta (1739). Till the 1750s English education barely reached the Indian population since it was meant for Europeans, Anglo-Indians and only a few Indians working for the EIC; the Company was also not very attentive towards education of the 'natives', and its sole interest was in trade and economic and military expansion. Therefore, up to the mid-eighteenth

century (roughly 1757) English was generally considered a foreign language in the Indian subcontinent (Sedlatschek 2009: 8-9; Sailaja 2009: 97-98).

Till the mid-eighteenth century, other European powers such as the Dutch, Portuguese and French were competing with the British to gain economic and military supremacy in India. However, events turned in favor of the British when in 1757, the EIC won the Battle of Plassey over the ruler of Bengal and became the leading power in India by bringing the entire eastern region of India under its control. This left the other European forces with small units in Goa, Bengal, Pondicherry and Yanam. Following the Battle of Plassey, the military power of Bengal (present day Bangladesh and West Bengal), Bihar and Orissa came into the hands of the Company by 1765, and British merchants were allowed to trade freely in the country. After a few years of financial irregularities and some reorganization within the EIC, the India Act of 1784 passed in Britain gave the EIC and the British Crown joint responsibility for governance of India. The same act again prohibited the entry of British missionaries in India and the EIC turned most of its efforts into acquiring territories across India, continuing military and economic expansion. At this time, the EIC did not take any further interest in the education system than what had already been in place (in small private schools scattered over different parts of India). However, the expansion of military control led to the employment of more Indians in the EIC. Still, formal education was hardly available in English at the time and administrative work was conducted mainly in Persian, because the Indians who were employed in the EIC administration at the time either already knew Persian (given that it was the language used in the Mughal courts), or learned the English language from their British peers (Sedlatschek 2009: 9-10; Sailaja 2009: 97-98; Dutta 2002: 48-50).

The early phase of English use in India led to the formation of a few pidgin varieties of English. A pidgin, often characterized as a simplified language, develops as a means of communication in a contact setting between two or more groups of people who do not share a common language. Three English-based pidgins, *Babu (Baboo) English*, *Butler English* and *Boxwallah English* have been documented as having sprung up during the early phase of British colonization in India. Babu English was used by the Bengali clerks (referred to as *Babus* or *Baboos* in Bengali) who worked in the EIC. This style is characterized by "excessive stylistic ornamentation, politeness and indirectness" [3] (Kachru 1994: 509). Butler English, another English-based pidgin, was spoken by the servants working in the Madras Presidency. Boxwallah English was the variety that both British and Indian traders used in commercial interactions with each other inside the British trading posts. However, use of these varieties became increasingly restricted when Indians started learning formal English in schools, and subsequently, English became the primary medium of communication between the two groups (Sailaja 2009: 97-98).

Irrespective of the lack of interest in education by the EIC, some educationists and missionaries were in favor of providing education to Indians. The second half of the eighteenth century and later years were marked by an expansion of English (Western) as well as Oriental scholarship. In 1778, the first English printing press was established in Hoogly, in Bengal and English newspapers were circulated in the three Presidencies, Bengal, Madras and Bombay. By 1813, some more private English schools had been set up in different parts of the country, which were usually run and owned by British widows, pensioners or missionaries. These schools however lasted only as long as the funds did and did not provide long-term education to students. They catered mostly to the children of the Europeans or Anglo-Indians and the few children of

[3] Even though some scholars have characterized *Babu English* as a pidgin, this variety is not like a prototypical pidgin.

wealthy Indians who were able to afford the fees for these schools. Alongside these efforts, some British educationists saw the importance of oriental scholarship. For example, in 1784, Sir William Jones established the Asiatic Society, an institution dedicated to Asian studies. Warren Hastings, the first Governor General of India who supported the cause of providing education in the Indian languages set up the Calcutta Madrasa in 1781 for education of Muslims and the Sanskrit college in Benares in 1791 for education of Hindus. Some missionaries who felt responsible for the children of Indian converts to Christianity also started providing them with religious education in the Indian languages (Sailaja 2009: 97-98; Krishnaswamy & Burde 1998: 10-12).

In 1813, when the charter was again renewed, Charles Grant, a former director of the EIC and other Anglicists (people who favored western education) argued in favor of English education for the Indians in order to uplift them to the 'light of Christianity' and provide them with the benefits of a liberal Western education. They also recommended making English the official language of the EIC and the British government in India and argued that educating Indians in English would make the EIC's administrative enterprise easier. Among the Anglicists were a few Indians who believed that a liberal Western education was the cure for the evils of Indian society. However, some others favored the idea of educating Indians in their vernaculars first and then gradually introducing them to English if at all. This group called themselves the Vernacularists or Orientalists. There were many advocates within both groups, and therefore when institutions such as the Fort William College (1800), Hindoo (Presidency) College (1817), and private schools such as The Hindoo School (1817) and the Hare School (1818) were established in Calcutta, classical languages like Sanskrit, Arabic and Persian were taught alongside English (Dutta 2002: 48-50; Sailaja 2009: 97-98).

Gradually, the EIC started taking responsibility for the education of Indians and allocated funds for that purpose. Committees such as the General Committee for Public Instruction, which was formed in 1823, managed issues related to education. Upon recommendation from this Committee, the British courts instructed the three Presidencies to make Western sciences and literatures more accessible to Indians and stated the need to train more Indians to prepare them for administrative jobs in the government. The committee however realized the difficulty in teaching science through translations in Indian languages and wanted the medium of instruction to be English. In 1835, Lord Thomas Macaulay became the Head of the Committee for Public Instruction and vociferously argued in his 'Minute on Indian Education' for the introduction of English as the medium of instruction in higher education. The 'Minute' was approved by a resolution on March 7 the same year and officially introduced English in higher education in India, a policy that continues to this day. This policy encouraged "subtractive bilingualism" (Bhatia & Richie 2006: 791), the goal of which was to provide Indians with an education of Western literature and science in English and turn them into monolingual speakers of English. Ultimately, it was designed to benefit the British government financially by making more Indians eligible for employment in the government, a more cost-effective strategy than employing British people (Sailaja 2009: 97-98; Krishnaswamy & Burde 1998: 10-12). Starting from this period, education in English became accessible to a larger section of the population.

From this point in time to the turn of the nineteenth century, English education saw a major expansion. The 'Magna Carta of India Education' in 1854 made recommendations for the creation of universities, teacher training institutes, more English medium schools, as well as elementary and middle schools with vernacular instruction. These recommendations were put into effect with the creation of different types of educational institutions across the country. Also,

the establishment of railroads, industries, post and telegraph, and the inclusion of Indians in the civil services, created a need for more and more Indians who could work in these sectors and this maintained a steady chain of demand and supply of education in English (Sailaja 2009: 97-98; Krishnaswamy & Burde 1998: 10-12).

In 1885, the Indian National Congress was formed. This organization gave educated Indians belonging to different parts of the country (who spoke several different first languages) a common platform to discuss and implement strategies to free India from British rule (it was ultimately instrumental in bringing about Indian independence). This organization played a big role in fostering nationalistic impulses among its members as well as in the general population and in all this, the English language served as a link language. The Indian National Congress was also successful in getting administrative power from the British government in a number of areas and this included bringing under its fold the existing department of education. Around the 1920s when there was much discussion about what the official language of the country should be after independence, opinions were divided over both English and Hindi/Hindustani (because it was the first language of almost 40% of the Indian population). However, because M.K. Gandhi was in favor of Hindustani (a mix of Hindi and Urdu), which he thought would unite Hindus and Muslims, Hindi was a more popular choice than English (Sailaja 2009: 97-98; Krishnaswamy & Burde 1998: 10-12; Dasgupta 1976; Gandhi 1936).

2.2.3. The role of English after Indian independence

When India gained independence in 1947, nationalistic sentiments were at their peak and so Hindi was declared as the official language of the country. However it was agreed that Hindi was not ready for use as a language in several administrative settings and so English was adopted alongside Hindi as the other official language of the country for a period of eighteen years,

during which technical vocabulary in Hindi would be developed. However, when the time for the transition from English to Hindi arrived in 1965, there was vehement opposition to this changeover, primarily from the southern Indian state of Tamil Nadu, and therefore English, considered a more neutral language in this context, was adopted as an associate official language of the country in 1967 (Dasgupta 1976; Bhatia & Richie 2006; Sedlatschek 2009, Thomason 2001).

Committees that were set up after Indian independence to oversee educational matters have shown ambivalent stances towards the use of English. Although everyone acknowledges English as not being native to India, it is viewed as a necessity for the country because of its international prominence and prestige and also because it is deeply embedded within Indian society. Therefore, language planning policies in India have tried to counter the dominance of English over Indian languages, while still maintaining English. These efforts have led to policies being formulated that would maintain balanced multilingualism in the country. An outcome of this effort is the *Three-Language Formula* whereby students have to be taught three languages in school. These include firstly, the first language or official Regional Language of the state in which the school is located; Hindi or any other modern Indian language if the first language is Hindi; and English. Although on paper the Three-Language Formula still exists, this policy has not been uniformly implemented in the country (Dasgupta 1976; Bhatia & Richie 2006).

Since India's independence from British rule, English has overcome its primary role as a tool of colonization and is now used in the domains of education, law, government, science, and technology, media and in inter-personal communication in urban areas. As a result of English being widely used in the education system (both before and after independence), it has become a second language for most educated people across urban India. This has led to widespread

bi/multilingualism between English and Indian languages. However, people who are not literate or who have not been schooled in English, especially those living in rural areas, have little to almost no knowledge of it. Some sources say that only about 5 to 10 percent of the Indian population actually speaks English (Joshi 1994; Bhatia & Richie 2006: 791).

The effect of English becoming a second language to a portion of the population of Indians has resulted in a process of 'Indianization' of English. 'Indianization' of English is characterized by the occurrence of divergent properties that are specific to the English spoken in India. One of the factors that may have influenced the occurrence of these divergent properties is contact with Indian languages (e.g. Kachru 1994; see also Kachru 1975, which discusses several grammatical features in Hindi that have emerged from contact with English). Additionally, code-switching between English and Indian languages is a regular occurrence in the day-to-day speech of English educated bilinguals in India (Bhatia & Richie 2006). The impact of English on Indian languages has been observed in various domains of grammar. I will review some of these studies in Chapter 3 on language contact and change. The historical events that led to the spread of English in India, which I have discussed in this section, are summarized in table 2.1 below. The table also shows the events that took place in Bengal, which I will discuss in more detail in the next section.

Table 2.1: Timeline of important events leading to the spread of English in India

Year	Important events in India	Events relevant to Bengal
Before 1600	Economic and military control by Portuguese	
1600	Charter by Queen Elizabeth I to trade in India; East India Company (EIC) set up	
1612	First English trading post in Surat	
1640		First Englishmen to arrive in Bengal
1670	First European school set up in Madras by Scottish preacher	
1681		First English military base in Hoogly
1698		Fort William built in Calcutta
1739		First English school in Calcutta
1757		Battle of Plassey between Nawab of Bengal and British; Britain becomes the leading European power in India
1765		EIC takes military control of Bengal, Bihar and Orissa
1778		First English printing press in Hoogly
1784	India Act: EIC and British crown jointly take over governance of India	Calcutta becomes capital of British administration; Asiatic Society (for Asian studies) set up by Sir William Jones in Calcutta
1781-1818		Several educational institutions are founded in Bengal
1835	English officially introduced in Indian higher education	
1854	Magna Carta of Indian Education	Publication of first play in Bengali
1858	British crown takes sole responsibility for Indian administration	
1905		Partition of Bengal into West and East Bengal provinces
1947	Indian independence (Hindi chosen as official language of India)	
1967	English becomes associate official language of India alongside Hindi	

2.3. Background of Bengali and English in Bengal

In the previous section, I sketched the historical events (as summarized in table 2.1 above) that led to and maintained contact between English and Indian languages before and after Indian independence. Contact between the British and Indians took place first through trade and subsequently through British military expansion, administration and education. In this section, I first provide background about Bengali and its development within the Indo-Aryan language family (2.3.1) and then explore the history of English in Bengal, since the main focus of this dissertation is the linguistic impact of English on Bengali and vice versa. My specific focus is on the historical events that brought English to this region and led to its spread across the population through educational institutions. I also explore some of the social and cultural impacts of English on the Bengali sociocultural, literary and linguistic domain.

2.3.1 Background on Bengali

Hindi, which is the most widely spoken indigenous language in India, is followed by Bengali (Bangla) as the second most widely spoken language (Bhatia 1982: 235-236). Bengali is the official language of the state of West Bengal and Tripura and is used for day-to-day communication between people across these states. Additionally, there is a sizable population of Bengali speakers in the states of Delhi, Assam, Bihar, and Karnataka. As is true for other parts of the country, English is spoken by the educated population of West Bengal, especially in the state's urban centers such as the cities of Kolkata (Calcutta) and Durgapur. There has also been an impact of Hindi in West Bengal in the past few decades, from Bollywood movies (Hindi language movies), national and regional media and educational policies which have been responsible for teaching Hindi across many schools in West Bengal. Due to migration of

working class people from the neighboring states of Bihar, Jharkhand and Uttar Pradesh where Hindi is spoken, there is also a sizable Hindi speaking population in the state.

Bengali belongs to the New Indo-Aryan (NIA) division of Indo-Aryan languages, which developed in three stages comprising Old Indo-Aryan (OIA, approximately from 1500 BC to 500 BC), Middle Indo-Aryan (MIA, ca. 500 BC to roughly about 1000 AD) and New Indo-Aryan (ca. 1000 AD to present). Indo-Aryan languages are a subgroup of the Indo-Iranian branch of the Indo-European language family. The earliest recorded forms of Old Indo-Aryan include the Vedic dialect, which is the language of the *Rig Veda*[4] with Classical Sanskrit being a later form of the Vedic dialect and closely related to it. Middle Indo-Aryan comprises the medieval dialects of Indo-Aryan, which descended from Old Indo-Aryan dialects such as Sanskrit. These Middle Indo-Aryan dialects were used in Buddhist and Jain inscriptions and scriptures and the term 'Prakrit', which literally means 'natural' is often applied to these languages because they had relatively simplified grammar and morphology as compared to Classical Sanskrit, which literally means 'refined'. The Prakrit dialects have several sub-divisions, which include varieties such as 'Ardhamagadhi' and 'Pali' that were used in religious texts. The later stages of MIA consist of the 'Apabhraṃśa' dialects, which literally means 'corrupt'[5] in Sanskrit; these dialects are genetically the closest predecessors of New Indo-Aryan languages such as Hindi and Bengali (Masica 1991: 51-52). By the 10th century AD, Bengali had developed into an independent language. Bengali has also come in contact with several European languages since the 16th century, including Portuguese, French, Dutch and English. Of all the languages that have come in contact with Bengali, it is with English that contact has lasted the longest, leading to stable

[4] The earliest and one of the most sacred texts of Hinduism, consisting of hymns.
[5] These dialects have been described as being corrupt because they characterized a gradual erosion of morphological complexity, as compared to the older varieties of Vedic dialects. However, this is not a scientifically appropriate way of describing language change.

bilingualism for a sub-section of the population in Bengal.

2.3.2. The arrival and spread of English in Bengal

When the first Englishmen arrived in Bengal in the beginning of the seventeenth century, the Portuguese had already set up a trading center in Hoogly in southern Bengal, along the Gangetic plain. Around 1640, a few English ships arrived in the Bengal region and the ruler of Bengal at that time, Shuuja Khan gave permission to the EIC to set up factories in Ballasore (present day Orissa) and Hoogly. This was in gratitude to Mr. Boughtan, an English surgeon who had cured his ailing daughter. For the next few years while Shuuja Khan ruled in Bengal, British merchants had permission to trade freely in Bengal. After the death of Shuuja Khan, The EIC had to obtain an imperial permission from the Mughal ruler of Delhi in 1681 to set up its first military base in Hoogly. This military base turned Bengal into a separate trading center, prior to which it was a part of the Madras Presidency. The English received further permission to fortify their factories, therefore Fort William was constructed in 1698 around the trading post of Hoogly (Stewart 1903: 393). After the construction of Fort William, the British built the city of Calcutta by integrating three villages (Kolikata, Shutanuti and Gobindpur) along the Hoogly delta, and it continued to be the Capital of British administration till 1911. During this time, the British had regular interactions with local Bengali merchants, artisans, rulers and administrators. However, these interactions were limited to the Mughal durbars (courts) or British trading centers, which also housed Indian soldiers, servants and clerks. At first these interactions usually took place in Portuguese or Portuguese-based pidgins or through English-based pidgins such as Boxwallah English that had sprung up in the trading centers. Often interpreters and translators were needed in these exchanges (Sedlatschek 2009: 9; Krishnaswamy & Burde 1998: 10-13; Marshall 1988: 54-58)

While the EIC concentrated on military and economic expansion, the 1698 charter allowed missionaries to come to India, which led to the gradual spread of Protestantism and the setting up of religious schools in the three presidencies of Calcutta, Madras and Bombay. The first English religious school was set up in Calcutta in 1739 by the Society for the Promotion of Christian knowledge, which aimed at teaching the Gospel to Protestant children of Europeans, Anglo-Indians and to children of those Indians who were employed in the EIC. This society also set up a circulating library in Calcutta in 1709, the first of its kind in the subcontinent (Sailaja 2009: 99-100).

This situation persisted till the mid-eighteenth century, when the EIC clashed with the Nawabs of Bengal in the Battle of Plassey in 1757 and the subsequent victory of the EIC over the rulers of Bengal resulted in the British becoming the most powerful military and economic force in the subcontinent. As a result, the other European powers such as the French, Dutch and Portuguese were rendered relatively powerless and got confined to small areas in the west and southern coast of India. After the Company and the British government got joint political control over Bengal in 1784, Calcutta became the center of trade and commerce and the capital of the British administration in India (Dutta 2002: 48-50).

2.3.3. English education and its sociocultural, literary and linguistic impact in Bengal

In this section, I first briefly discuss the educational changes that were introduced in India after the official establishment of British administration in India. Then I go on to describe some of the impacts of British administration, education and culture on the sociocultural, literary and linguistic situation of Bengal. I include the sociocultural and literary domains in my discussion because I investigate linguistic changes in Bengali through the examination of Bengali literary texts. Also, the social, cultural and literary domains in Bengal experienced simultaneous

influences from English education and culture and in turn influenced each other. Therefore a discussion of linguistic change in Bengali is incomplete without an exploration of the sociocultural and literary changes in Bengal. This broader discussion is also crucial for understanding my choice of particular literary texts for the study of change.

British administrative and educational policies from the late eighteenth century on began to impact the linguistic situation of Bengal in addition to affecting the socio-political and cultural climate. Firstly, British educationists and missionaries took an active interest in education in India and helped establish institutions for studying oriental languages as well as English and other European languages. By this time, Bengal, especially Calcutta, had become the seat of British administration and this turned it into a hub of educational institutions. Some of the most prominent institutions that came up include the Asiatic Society (1784), the Calcutta Madrasa (1781, promoting education for Muslims), Fort William College (1800), the Hindoo College (1817, later renamed as the Presidency College) and private schools such as the Hindoo School (1817) and the Hare School (1818) (Dutta 2002: 48-52). The first English printing press was also established around this time in 1778 in Hoogly, which started the circulation of English newspapers in the British presidencies (Sailaja 2009: 98). These institutions especially the Hindoo College became quite popular among both rich and poor Indian students particularly because both English and Indian languages were taught and poor students got scholarships to attend these colleges. Also students who graduated from these institutions often got recruited in the EIC (Stewart 1903: 430).

The introduction of Western literary and scientific education and the influence of Christian missionaries are considered catalysts for many social and religious reforms in India. Alongside these social reforms, some cultural, linguistic and literary changes also took place in

Bengali starting in the 1780s and continuing onto the late-nineteenth century (De 1919: 58). This period is usually referred to as the *Bengal Renaissance* due to an 'awakening' in Bengali society, culture and thought that was characterized by an appreciation for Western liberal and scientific education, politics and law and a simultaneous celebration and glorification of ancient Indian literature and culture. One of the highlights of the Bengal Renaissance was the abolition of many heinous socio-religious practices such as Suttee (also spelled Sati),[6] infanticide, polygamy and child-marriage. These socio-religious reforms were made possible by Indian intellectuals and reformers like Raja Ram Mohan Roy, Ishwarchandra Vidyasagar and Keshab Chandra Sen, with help from British administrators and missionaries (De 1919: 345-357).

The nineteenth century saw several developments in the city of Calcutta and in the social life of the Bengali elite who came in contact with British culture in Calcutta. Apart from the educational institutions set up by the British, they also set up several play-houses or theatres for their own entertainment. These theatres were frequented by members of the Bengali elite whose imagination was captured by the novelty (for the Bengali audience) of dramas and plays being performed there.[7] Within a few years, the rich, elite Bengali patrons of the English theaters, with help from a few Englishmen and Europeans, began conceptualizing and producing plays in Bengali specifically for a Bengali audience. These performances took place in theatres that were built solely (at first) for the entertainment of the Bengali population of Calcutta. The first plays to be produced in these theatres were translations of English and Sanskrit plays, but the resulting success and popularity of these productions prompted the Bengali elite to write and produce original plays in Bengali (Guha-Thakurta 1930: 51-53).

[6] The funereal ritual that forced Hindu women to burn themselves in the funeral pyre of their deceased husbands.
[7] Bengali literature at the time did not have an established theatrical tradition. The dramatic tradition of Bengali till that point in time consisted mainly of performing *Jatras* (folk theatre) of mythological and religious topics. The *Jatras* however were not published and so did not thrive as a literary genre (Guha-Thakurta 1930: 51-53).

The cultural influence of English theatres in conjunction with English education resulted in a burst of literary production in Bengali, a language that had hitherto been unpopular in the literary domain because of the dominance of classical languages like Sanskrit and Persian (De 1919: 46-53; Dutta 2002: 49-56). Also, under the Muslim rule, Bengali literature had not received much patronage; barring the sparse production of an ornate style of poetry and songs[8] dealing with religious themes and translation of Sanskrit literature into Bengali, the language did not have a distinguished literary tradition (De 1919: 40-42). Furthermore, Bengali literature underwent a dull period during the political upheaval in Bengal from 1757 (Battle of Plassey) to the establishment of British administration through the EIC in 1784 (De 1919: 46-53). Therefore, the first half of the nineteenth century was a period during which existing Bengali literature was consolidated, textbooks and grammars were prepared, older texts of Bengali and texts in other classical languages (such as Sanskrit) were reprinted. British missionaries such as William Carey, Nathaniel Halded and Henry Forster were instrumental in these accomplishments. The first Bengali press was also set up during this period in 1800 in Srirampur by William Carey (De 1919: 345-357; Dutta 2002: 56-57).

The focus on consolidating the Bengali language and the simultaneous exposure to English literature subsequently brought about changes in Bengali literature with the introduction of new genres such as plays (as discussed above) and novels. It was only after the 1840s that original Bengali plays were written and performed. Arguably, the first play to have been written in Bengali is *Kulin Kula Sarbaswa* by Ram Narayan Tarkaratna, published in 1854 (Guha-Thakurta 1930: 51-53), which I am using for my diachronic investigation of linguistic change in

[8] Some of these literary works include Bharat Chandra Ray's *Annadamangal*; religious (Bhakti) poems and songs (*Ramprasadi*) by Ramprasad Sen; *Manasa*, *Dharma-mangal* and *Mahabharat* (translation) by Kasidas and translation of the *Ramayana* by Krittibash Ojha.

Bengali. Also, till the mid-nineteenth century, a diglossic relation existed between two varieties of Bengali, a colloquial variety used mostly for spoken purposes and a formal variety used mainly for literary purposes. From the 1850s on this hierarchy was opposed by several playwrights and novelists who started using the colloquial variety in their writing. The play *Nil Darpan* (1860), which I am also using as part of my diachronic investigation, is a product of this period in Bengali literature (Chatterjee 2007: 55).

With the introduction of English in schools, more people in Bengal started learning the language and becoming bilingual in Bengali and English. As the contact situation intensified, Bengali not only acquired several English loanwords (as discussed in Chatterji & Grierson 1926) but was also possibly influenced structurally by English. For example, Yates & Wenger discuss in their grammar of Bengali (1885) how Bengali relative clauses became similar to their English counterparts[9] (see also Pillai 1968). However, the possible linguistic influences of Bengali and English on each other have not been explored in much detail so far. With widespread multilingualism in the twentieth century, code-switching between Bengali and English has also become very common among the English educated elite of West Bengal, India. Furthermore, with the ever-growing importance of English as a global language, use of English is expanding with great speed throughout India, including in West Bengal, especially in the media and in inter-personal communication.

Investigation of language change through the examination of literary texts

As briefly mentioned in chapter 1, the linguistic analysis of change in Bengali in my dissertation is based on the methodology of Thomason (2001), which provides five criteria to be followed in order to establish contact-induced changes in a language. This framework is also

[9] I discuss specific linguistic changes in Bengali as a result of English contact in greater detail in chapters 5 and 6.

grounded in an assumption that the socio-historical forces surrounding a contact situation are crucial in shaping the outcomes of that situation. One of the criteria proposed in this framework (which I discuss in detail in the next chapter) involves showing that the features being investigated in a receiving language were not present in that language before contact with the source language occurred. In addition, one must also provide evidence that the shared features (between the source language and receiving language) were present in the source prior to contact with the receiving language (Thomason 2001: 91-94). Therefore, to identify changes in Bengali, I have used older Bengali plays and compared them with modern Bengali speech, to find out whether the grammatical features that seem to be innovations in modern Bengali speech were indeed absent in a prior stage of Bengali. The analysis of change in Bengali is therefore heavily dependent on the choice of appropriate texts for this purpose. For the purpose of this dissertation, I examined in detail only two Bengali plays as part of the diachronic investigation. But, I plan to expand this corpus in future research.

Revisiting the socio-historical events of nineteenth-century Bengal, therefore, gives us several insights into the trajectory of the development of Bengali literature, which is particularly relevant for the selection of the 'right' texts when analyzing language change. This point has also been emphasized by Poplack & Levey (2010), who highlight the importance of choosing appropriate historical varieties as a baseline for the comparison between the older and the contemporary variety of a language. They warn of the pitfalls associated with the study of language change when a linguistic variety that is chosen for the diachronic investigation is an 'idealized invariant standard'. When that standard variety is compared to a contemporary colloquial variety, the differences between the standard and the colloquial variety could be construed as changes, when in fact they are just cases of dialectal variation (Poplack & Levey

2010: 394-395). The literary texts that I have chosen for the diachronic comparison between older Bengali and contemporary Bengali are aimed at avoiding these very pitfalls. For example, in contemporary times, people primarily speak colloquial Bengali, and to control for the variation, I chose a historical text (*Nil Darpan*) that was written in a colloquial variety.

The choice of literary texts (plays) from the mid-nineteenth century for the diachronic comparison is arguably not an ideal choice, given that contact with English had already been established by that time. However, because the contemporary corpus consists of oral production data, a genre that is more likely to reflect naturalistic Bengali speech, as it was produced at the time is desirable and plays are appropriate in that regard. Because the genre of playwriting developed as a result of English literary influence on Bengali literature, the earliest plays I can use for my investigation are from the mid-nineteenth century. Despite this caveat, the nineteenth-century plays I have consulted have consistently shown the absence of certain features that frequently occur in modern Bengali speech; examples of these include code-switching, bilingual verbs and copular predicates (as discussed in chapters 5 and 6). These texts therefore provide useful windows into the nature of older Bengali before significant effects of contact with English could have taken place.

2.4. Conclusion

In this chapter, I first traced the historical events that brought English to India including Bengal. Then I focused on the events that led to the diffusion of English in India as well as the linguistic and educational policies of first the British and then the Indian government that helped in sustaining the language in India. Further, I explored some sociocultural impacts that English had on Bengali society, literature and culture, while also addressing socio-historical aspects that

may have affected the Bengali-English contact situation and the analysis of linguistic change and development in this context. Finally, I made brief reference to the framework (as proposed by Thomason 2001) I have used to ascertain change in Bengali (which I return to in chapter 3) to emphasize how using older texts is crucial in bringing insights into the linguistic analysis of contact-induced change. In that context, I provided a justification for using particular texts in the diachronic analysis of Bengali.

In the next chapter, I review some key studies in the literature on language contact and change and present the foundational theoretical background to the research questions being investigated in this dissertation.

Chapter 3: Literature review on language contact and change

3.1. Introduction

In this chapter, I present an overview of some of the literature on language contact, specifically those studies that are most relevant to the research questions of my dissertation. I start by further addressing the main issue that I am tackling in my dissertation, the possibility of bidirectional contact-induced change in Bengali and English in West Bengal, India. In section (3.2), I consider language contact and change and briefly discuss the types of changes that may affect languages, which include language-internally motivated changes, contact-induced changes and those that arise from multiple causation (3.2.1). After introducing these three types of changes, I focus on particular processes of language contact (3.3), such as the different types of contact-induced changes, namely, *interference, transfer, borrowing,* and mechanisms that result in those changes such as code-switching. All the types of changes that I examine in this section directly relate to the research questions of this dissertation. In section (3.4), I discuss the framework proposed by Thomason 2001 to ascertain if particular changes that have been identified in a language are indeed contact-induced. In this section, I also discuss the framework proposed by Poplack & Levey 2010 that I take into account to analyze the changes that I have established as being contact-induced. In the last section (3.5), I revisit the process of bidirectional influence in language contact and discuss some studies on contact between Indian languages and English in India.

My primary aim in this dissertation is to examine how two or more languages in a contact situation influence each other; I am specifically interested in bidirectional influences that Bengali and English may have had on each other as a result of over two hundred years of contact in Eastern India. Pavlenko & Jarvis (2002:192) defined the term bidirectional influence as the "two-way interaction between the two linguistic systems of an L2 user (i.e L1 influence on the L2 and L2 influence on the L1)". The use of this term does not mean that both languages in the contact situation have influenced each other in terms of the same set of features, or that these influences are restricted to bilingual speakers. I take the term more broadly to mean influence of both languages on each other, even when they involve different sets of features. The influence of one language on another may occur in all areas of grammar, including the lexicon, phonology, morphology, syntax and semantics. However, this dissertation has focused on specific changes, including borrowing of lexical items and grammatical changes that have taken place in Bengali and English.

3.2. Language contact and change

Language change can be defined in view of at least two different perspectives. From the point of view of formal syntax, a change is the acquisition of different grammars across two or more generations of speakers, while from the perspective of historical linguistics, a change is an innovation that takes place in an individual's grammar but subsequently spreads to the speech of other adults (Pires & Thomason 2008: 33-34). Although these two points of view are not completely incompatible with each other, I will not be delving into that discussion here (see e.g. Pires & Thomason 2008 for detailed discussion on this). I will use the term 'change'[10] the way it

[10] My examination of change involves both contact-induced change as well as language-internal developments.

is used in historical linguistics to mean an innovation that developed in the lexicon or grammar of an individual, but has subsequently spread to other speakers. For example, in the case of Bengali, I will consider a feature to show change only if it is frequently used by different speakers and is in some way different from the way it occurs in an older stage of the language (also considering for instance the use of monolinguals and bilinguals, which I return to later). It is also important to note here that according to scholars like Thomason (2001), a change in frequency of use of a particular feature is also considered to be a change in the language and I too adopt this definition of change and consider frequency change as a change in the language.

It is also important to consider how language change can be investigated. This question can be approached from two different angles, which are: examining language at a given point in time, that is studying the synchronic aspects of language, or studying the development of a language over time, that is, a diachronic study. However, neither of these two approaches alone can provide a complete picture of language change. Therefore, by investigating both diachronic and synchronic aspects of language change in my dissertation, I try to provide a fuller account of change in Bengali and English. The synchronic investigation of Bengali and English will be much more detailed in this dissertation, but the additional diachronic investigation of Bengali will also be very relevant and will help us get a larger picture of the development of Bengali over time.

Languages can change primarily due to two forces. These forces may operate in isolation or in combination with each other to result in changes. One of these is language internal pressures such as analogic change that take place when one element of a language becomes similar to another element in that language. For example, the plural for the word *cow* in Old English was *kine* but it later became *cows* by modeling the more common plural marking process

of adding *–s* at the end of nouns (Hock 2003: 442). The other way by which a language may change is through the influence of languages with which it is in contact. These kinds of changes are referred to as contact-induced changes. Sometimes however, both these forces may work in combination to result in a change in the language. Such changes are believed to result from multiple causation (Thomason 2001: 62). In this dissertation, I report on all three types of changes in Bengali and English. In the next section, I discuss contact-induced changes in more detail.

Contact-induced change

Many scholars of language contact today would agree that change in the grammar of one or both languages is the most common outcome of language contact. However, the field of language contact has been witness to opposing views in terms of the degree to which languages can influence each other grammatically. In the earliest studies on contact, scholars like Müller (1871) and Whitney (1881) denied the possibility of (any) grammatical interference between languages in contact. Others had the opposite view regarding grammatical interference, such as Hugo Schuchardt (regarded as the father of creole studies), who stated that "there is no limit in principle to the influence which one morphological system may have upon another" (Schuchardt 1928, as quoted in Weinreich 1953: 29).

Uriel Weinreich, who provided one of the first comprehensive accounts of language contact in 1953, defined language contact as the alternate use of two or more languages by the same individual, and identified the need for an examination of both psychological and socio-cultural factors to clearly understand the behavior of languages in a contact situation (1953: 1-4). Among later studies on language contact, Thomason (2001) defines contact-induced change as "any linguistic change that would have been less likely to occur outside a particular contact

situation is due at least in part to language contact" (62). This broad definition by Thomason includes not only the changes that emerge as a result of properties being directly imported from one language onto another (such as morphemes, syntactic patterns or both) but also changes that take place specially in dying languages as a result of speakers shifting to other languages, though such changes may not be direct imports from the dominant language. The third type of change, as discussed in Thomason (2001), encompasses subsequent changes that take place in a receiving language (RL) as a result of an earlier direct transfer of properties from a source language (SL) into the RL. This definition also does not rule out changes that emerge from "multiple causation", those that result from the combination of both external and internal causes in the language (62-63).

Even though many scholars today agree that change is the most frequent outcome of contact (a notable exception being Poplack & Levey (2010)), within language contact studies there is no clear consensus on how to ascertain and classify the outcomes that occur as a result of contact. In the next section, therefore, I discuss the different types of contact-induced change and the terminologies associated with them, as proposed by some scholars of language contact. I also discuss some of the mechanisms that occur in contact situations (such as code-switching), which are relevant to the research questions in this dissertation.

3.3. Processes and mechanisms of contact-induced change

Language contact may result in the occurrence of different types of changes. However, there is a lot of contention about what is a change and what constitutes processes or mechanisms of contact. It is not clear whether commonly discussed processes in the field of language contact such as *borrowing* and *interference*, which I discuss below, are types of changes or rather mechanisms that lead to changes. While some scholars treat *borrowing* and *interference* as

contact-induced changes and *code-switching* as a mechanism by which change occurs (e.g. Thomason 2001), others would call *borrowing* a process (Haugen 1950) that is not necessarily a change. Another issue is that scholars often refer to similar phenomena with different names and to different phenomena with the same name and this has resulted in some confusion in the field. My aim, therefore, will be to examine these phenomena based first on the definitions provided by scholars and then provide my own working definitions for these terms in the way I will be using them in my dissertation. The list here is limited only to those concepts that are most pertinent to the research questions I investigate in this dissertation.

3.3.1. Interference and Transfer

In this section, I look closely at the terms *interference* and *transfer*, which are used very frequently in the language contact literature to refer to contact-induced changes in general. I focus on the terms *interference* and *transfer* in the same section because there is a tendency within the field of language contact to use these terms interchangeably to refer to the same phenomenon.

The term *interference* was first introduced by Weinreich (1953), for whom it is a general linguistic change that occurs very frequently in language contact in the native language of a bilingual speaker due to pressures or influences from the foreign language. Weinreich defined *interference* as "deviation from the norms of either language" resulting in a "rearrangement of patterns" in the structures of linguistic systems. It is not akin to a case of *borrowing* a word in a language, because lexical *borrowing* is simply an addition to "the loosely patterned domains of language" (1953: 1). For Weinreich, *interference* is a cover term for linguistic change in general and *borrowing* or *transfer* are particular types of *interference*. *Borrowing* or *transfer*s occur when there is an actual use of a non-native (source language) element in the native (recipient)

language. So, only if a word or morpheme that didn't belong to a language originally is used in that language with the same phonetic form as was in the non-native language, it is a case of *borrowing* or *transfer*. However, if speakers are using a pattern in their native language that they modeled upon another language, such as a word order pattern which was not native to their language, then that would not be a case of *borrowing* or *transfer* but of *interference* in general (Weinreich 1953: 7).

Weinreich mentioned three types of grammatical *interference* that are possible between the two languages in contact. On the one hand, there can be a direct *transfer* of morphemes or words from the source language to the recipient language. For example, the use of American Yiddish /nit er bʌt ix/ 'not he but I' (Weinreich 1953: 30). On the other hand, there are cases where a language adopts the patterns of another language such as word order. The sentence *he comes tomorrow home* is an example of this type of interference where the word order of German is applied to the morphemes of English (Weinreich 1953: 30). The third type includes instances where there is a change in the function of "indigenous" morphemes by extension of the function of equivalent morphemes in the source language or a loss of distinction in the function of morphemes in the recipient language by replication of the function of morphemes in the source language. For example, the word *ministro* 'cabinet official' in Colorado Spanish acquired the new meaning of 'Protestant ecclesiastic' on the model of the English word *minister* (Weinreich 1953: 48, 64).

Weinreich also distinguished between *interference* that occurs in speech as opposed to that which occurs in language; the former occurs when a bilingual introduces a non-native element into their language by virtue of their knowledge of another language, while the latter happens when the non-native element has become a part of the grammar of monolingual

speakers. This distinction is theoretically relevant because studying one instead of the other provides different results to the historical linguist (11-12). It is also important to note that for Weinreich, *interference* is a linguistic change and not a mechanism. *Interference* usually takes place with the help of mechanisms like direct *transfer* or *borrowing* of non-native material, *replication* or *modeling* non-native material in the native language or *extension* of functions in the native language based on functions in the non-native language.

Some subsequent studies on language contact such as Thomason & Kaufman (1988) and Thomason (2001) adopted the term *interference* and used it to mean contact-induced change in general much like the way Weinreich used it. For example, according to Thomason & Kauffman (1988) and Thomason (2001) linguistic *interference* falls into two categories, which include *borrowing* and *interference through shift* or *shift-induced interference*. *Borrowings* constitute cases where speakers of a language incorporate non-native linguistic material into their native language. Importations of vocabulary or structure from a source language to a recipient language would count as a borrowing, for example English derivational affixes *–able/ible* were borrowed into English from French loanwords and subsequently spread to native English vocabulary. Similarly, languages may also borrow structure, such as structure that leads to distinctions in word order (Thomason 2001: 70). Usually, the first non-native materials to get integrated in a language are words and often these borrowed words are treated as stems and take on affixes just like native words. Depending on "long-term cultural pressure from the source language" onto the receiving language, borrowing of structure may follow lexical borrowing. *Interference through shift* is the other type of *linguistic interference*, which typically occurs when speakers are shifting to a second language and in the process fail to learn that language (target language or TL) perfectly. This is called *imperfect learning*, and unlike borrowing, in cases of *shift-induced*

interference, structural changes are usually the first to occur, and may be followed by lexical incorporations from the first language of the shifting group (Thomason & Kaufman 1998: 37-39; Thomason 2001: 67-68, 74-75). These studies discuss the case of Indian English as an example of *interference through shift*, which shows very little occurrence of Indic and Dravidian vocabulary (Thomason & Kaufman 1998: 21) but several instances of syntactic *interference* from these languages (such as in the use of articles, tense, aspect, agreement marking, copula deletion etc.; see e.g. Sharma 2005b). Indian English, however, is not a typical case of *interference through shift* because in this case, speakers of Indian English did not shift to English but continue to speak their native languages (Thomason 2001: 74).

For Thomason (1997, 2001), *borrowing* and *shift-induced interference* are contact-induced changes that take place with the help of linguistic *mechanisms* like code-switching (alternate use of two or more languages in a single utterance). Other mechanisms, including *code-alternation*, *passive familiarity*, *second-language acquisition strategies, bilingual-first language acquisition* and *conscious resistance (deliberate decision)*, may also lead to contact-induced changes (I address some of these in chapters 5 and 7).

Van Coetsem (2000) considered the term *interference* or *transfer* to be a source of confusion in the language contact literature and therefore proposed two other terms, *borrowing* and *imposition* to refer to distinct phenomena in language contact. He reserved the term *transfer* not for a specialized meaning but to generally mean "transmission of materials or elements from one language to another" (51), and the way Van Coetsem used the word *transfer* corresponds to the way Weinreich (1953), Thomason & Kaufman (1988) and Thomason (1997, 2001) used the term *interference*, which is to refer to contact-induced change in general. For Van Coetsem, there are two types of *transfer* and language dominance in a bilingual is an important factor when it

comes to identifying the two types, and so, depending on which language is dominant for the speaker, *borrowing* or *imposition* may occur. For example, if the Recipient language (RL) is dominant for the speaker then the RL speaker will use their agency to introduce elements from the Source language (SL) into the RL. This is different from *code-switching* in that code-switching is a synchronic phenomenon where the switch is happening in terms of language production, whereas *borrowing* or *imposition* corresponds primarily to a diachronic process in Van Coetsem's terms. This phenomenon is referred to as *borrowing*. On the other hand, if the source language (SL) is dominant for a speaker, then through the agency of the SL speaker, SL material will be introduced in the RL and this is called *imposition* (49). *Borrowing* thus refers to *RL agentivity* and *imposition* to *SL agentivity*. *SL agentivity* or *imposition* in fact correspond to the term *shift-induced interference* of Thomason & Kaufman. However, Van Coetsem's terms have not enjoyed much popularity in the field of language contact and many scholars continue to use the term *interference* to refer to contact-induced change in general.

Although the term *interference* has been used to refer to contact-induced change in general, there has been a tendency to use it for changes that show up particularly in the L2 or second language of the bilingual speaker (Winford 2005: 374). However, the term *interference* has gradually been replaced by the term *transfer* in the Second Language Acquisition (henceforth, SLA) literature because of a negative connotation attached to the former (Sankoff 2001:639, Jarvis & Pavlenko 2008:3), but many scholars continue to use these terms to refer to the same phenomena (such as Thomason 2001). According to Van Coetsem (2000: 34), the association of the term *transfer* primarily with SLA studies is responsible for a separation of the field of SLA from the larger field of language contact, even though SLA is very much a part of language contact study.

As mentioned above, in SLA studies, *transfer* is used to refer to "carry-over" features from the L1 or first language to the L2 or second language. This means that L2 learners use their first language features (phonology, morphology or grammar) to learn the L2, and the L1 may either act as a basis for constructing the L2 grammar or that the L2 learner has not yet been able to recognize differences between the L2 and L1 grammars, keeping features of the L1 in their L2. Transfer may also be *positive* or *negative*. For example, if L1 and L2 features match and the L2 learner is successful in quickly acquiring the L2 feature then that is a case of *positive transfer*. However, if there is a mismatch between the L1 and L2 features and the L2 learner applies the rules of the L1 feature in learning the L2 feature then that would be a case of *negative* transfer. For example, if an L2 learner is unsuccessful in acquiring the article system of English due to a mismatch in the article systems between English and their L1, that would be a case of *negative* transfer (Odlin 1989, as mentioned in Siegel 2008: 108).

It seems therefore that the terms *interference* and *transfer* are often used to refer to the same phenomena, although *transfer* is used more widely in the SLA literature whereas *interference* is more commonly used in the general language contact literature. In my discussion on contact-induced change in Bengali and English, I will be using the term *interference* to refer to structural changes that appear in English of the bilingual speakers in West Bengal and for which I have sufficient evidence to argue that they have been caused by contact. If I am unable to provide evidence of contact for a given change, I will not use the term *interference* and treat it simply as a general case of language change.

3.3.2. Borrowing

The term *borrowing* figures very prominently in the work of Haugen (1950; 1969) who used it to refer to linguistic diffusion in general and defined it as "the attempt by a speaker to

reproduce in one language patterns which he has learned in another" (1969: 363). Haugen clearly identifies *borrowing* as a historical or diachronic process (1950: 213) such that at one point in time, a borrowed word must have appeared as an innovation which later spread in the language. Therefore in Haugen's view a thorough analysis of *borrowing* requires a diachronic investigation because if one has to prove that an item in a language is borrowed, it must be shown to have existed in one of the languages prior to contact and then after contact with the other language took place, it passed onto the other language (217). However, Haugen's analysis of linguistic borrowing is mainly limited to the lexical domain and he provides classifications of different types of lexical borrowings, which include *loanwords*, *loanblends* and *loanshifts*. *Loanwords*, according to Haugen "show morphemic importation without substitution", such as the use of American English word *whip* in American Norwegian. Loanblends show "morphemic substitution as well as importation" and are also referred to as *Hybrids* by Haugen. For example, the use of the word *plum pie* as [bluməpai] by Pennsylvania German speakers is a combination of the English borrowing *pie* with the native word [blumə] which means 'plum'. *Loanshifts* refer to what is commonly known as loan translations or calques, where speakers import "a particular structural pattern, viz. the combination of the two constituents into a compound expression with a new meaning of its own not derivable by a simple addition of the two parts". For example, the use of German *Wolkenkratzer,* which is modeled upon English *Skysraper* (213-214). I will focus on lexical borrowing in Bengali in chapter 5 and address some of these terms there.

In some subsequent studies, scholars have used the term *borrowing* to refer to instances where material from the non-dominant language of the bilingual (usually the non-native or second language) is introduced into the dominant language (usually their native or first language). Thomason (2001) uses the term *borrowing* to refer to instances where speakers of a

language introduce linguistic materials from their second language into their first language (some examples of these were provided earlier on p. 38). In these instances, the notion of *imperfect learning* does not play a role. This is because fluent speakers of a source language are taken to be the ones who introduce borrowed materials into their language (Thomason 2001: 67-68). As mentioned in the previous section, Van Coetsem's (2000) use of the term *borrowing* (or *RL agentivity*) also corresponds to Thomason's use of the term, because, for Van Coetsem *borrowing* occurs when SL material is introduced into the RL by speakers who are dominant in the RL. In contrast, as I mentioned in the previous section, for Weinreich, *borrowing* was a particular mechanism of contact-induced change that involved a direct transfer of material from one language onto the other.

Thomason & Kaufman (1988) make some predictions about *borrowing* as well as *shift-induced interference* and state that *borrowing* typically starts with the lexicon and then goes on to affect the phonology and syntax. *Shift-induced interference*, on the other hand, typically starts with phonology and syntax and may have some morphological features and lexical importation at a later stage. According to the *borrowing scale* of Thomason & Kaufman (1988) and Thomason (2001), non-basic vocabulary is the easiest to borrow when contact between languages is casual, but with the increase in intensity of contact, various kinds of features can be borrowed. Finally, in very intense contact situations, all sorts of features become vulnerable to borrowing (Thomason & Kaufman 1988; Thomason 2001:69-70). However, Thomason (2001) warns us that these borrowing scales are a matter of "probabilities, not possibilities" and are likely to work better for typologically distant languages (74). In the case of Bengali and English, which are typologically distant and have been in fairly intense contact with each other, it can be

observed that both basic and non-basic vocabulary, some phonology, and possibly morphosyntactic features have been borrowed.

In sum, based on the discussion presented here, it seems that the term *borrowing* is usually reserved for a particular type of contact-induced change (Haugen 1950, Thomason & Kaufman 1988, Thomason 2001, Weinreich 1953, Van Coetsem 2000) while the terms *interference* and *transfer* have been used as cover terms for contact-induced change in general (Thomason & Kaufman 1988, Thomason 2001, Weinreich 1953, Van Coetsem 2000, Siegel 2008). In my discussion of Bengali and English, I will use the term *borrowing* only to refer to lexical importation of material from the other language. For any other structural change for which I have sufficient evidence to argue that they are contact-induced, I will use the term *interference* as a cover term.

In the next section I turn to some discussion of code-switching, which is of relevance for this dissertation given that the bilingual data I investigate consists of extensive code-switching. Also, *borrowing* and *code-switching* have often been linked together in the code-switching and language contact literature. In chapter 5, I address the issue of Bengali-English code-switching as being a possible mechanism of contact-induced change in Bengali.

3.3.3. Code-switching

The phenomenon of *code-switching* (henceforth CS), the use of two or more languages in a single utterance, has received much attention in the language contact literature in the past thirty years perhaps because of its frequent occurrence in many different bilingual areas across the world and also because the ability to switch back and forth between two languages is quite remarkable in terms of cognitive processing. In this section, I discuss some prominent models

that have tried to explain the mechanism of CS and how it relates to specific contact-induced changes such as *borrowing*.

Instances of lexical borrowing may seem identical to single-word code-switches in some cases. This apparently identical nature of CS (single-word switches) and borrowing (as shown in example 1 below) has prompted scholars like Poplack and colleagues to propose tools to distinguish between the two processes as well as enumerate constraints that regulate CS (e.g. Poplack 1980, Poplack, Wheeler & Westwood 1989; Sankoff, Poplack & Vanniarajan 1990; Poplack & Meechan 1995; Poplack & Dion 2012). Consider the example from Bengali-English CS data below, where we cannot be certain of whether the word *responsible* is a borrowing or a code-switch from English in Bengali.

(1) O ækebarei *responsible* na
3SG not at all responsible NEG
He/she is not at all responsible.

However, this distinction becomes easier to make when borrowing and CS are studied both synchronically and diachronically. Studies about *borrowing* in relation to CS have also shed light on how the two processes are connected in terms of the diachrony and synchrony of language. For example, CS is a synchronic process and *borrowing* a diachronic one, and acts of borrowing often start with code-switches (e.g. Haugen 1969).

There has been extensive research on CS in the past few decades, which has produced some prominent models attempting to explain CS phenomena. For our present purposes, CS is an important point of inquiry because it frequently occurs alongside contact-induced changes in many bi/multilingual contact situations, and is one of the major mechanisms behind such changes (as discussed in Thomason 1997, 2001). In this dissertation, I consider contact-induced change by analyzing Bengali and English in CS data, and as I will discuss in chapter 5, the

bilingual code-switched corpus shows a higher frequency of occurrence of contact-induced features. Also, the feature of bilingual complex verbs (discussed in chapter 5) occurs most frequently in code-switched conversations. Below, I discuss some of the prominent models of CS that were proposed by Poplack (1980) (and later by Poplack, Wheeler & Westwood (1989), Sankoff, Poplack & Vanniaajan (1990) and Poplack & Meechan (1995)); Myers-Scotton (1993a) and Muysken (2000). Of these three, I will use Myers-Scotton's and Muysken's models of CS as descriptive frameworks for the CS patterns in Bengali-English.

Very broadly, CS is defined as the use of two languages in a single utterance. It is generally agreed that there are two types of CS, inter-sentential switching (the switch from one language to another at a clause or sentence boundary) and intra-sentential switching (the switch from one language to another within a sentence) (Romaine 1995: 122). Most CS research has focused on intra-sentential switching. Within this area, several models and constraints have been proposed to account for various types of CS data. Researchers such as Muysken (2000) and Myers-Scotton (1993a, 2002) have provided models that distinguish types of CS. Some models such as Di Sciullo, Muysken & Singh (1986) have proposed formal constraints on CS, attempting to account for limitations on switching at various points in a sentence. Other researchers have proposed broader models that also include constraints on code-switching (e.g. Poplack 1980, Poplack, Wheeler & Westwood 1988, Poplack & Meechan 1995).

Much research on intra-sentential switching has focused on accounting for the places where switching can occur. One of the earliest and most influential of models of CS was proposed by Poplack and associates (e.g. Poplack 1980, Poplack et al. 1989, Poplack & Meechan 1995). This model dominated the field of CS for a long time and influenced later models such as the one proposed by Muysken (2000). According to Poplack's model, the process of single-word

switches is different from the process that produces multiple word switches. The model states that the former falls into the broad category of borrowing and insertion, while the latter is characterized by CS or alternation. Poplack's model distinguishes borrowing from CS (especially regarding single-word switches) primarily on the basis of morphosyntactic and phonological integration into the host language. If an element from the source language is integrated into the recipient language morphosyntactically and phonologically, then it is a loanword. This definition implies a difference between established loanwords and borrowings that are not integrated morphosyntactically and phonologically into the host language. According to Poplack's model, the latter are characterized as 'nonce loans'. This model considers borrowing as a process that includes both established loans and nonce loans, and only alternation is considered as a real switch in a language. The alternation from one language to another may not always be smooth. These transitions might sometimes be flagged, where mixing is specially marked, for instance with hesitation on the part of the speaker or insertion of dummy elements or fillers.

One model that I will briefly consider in relation to Bengali-English CS data and more extensively in relation to the bilingual verbs of Bengali-English was proposed by Muysken (2000). This model makes three distinctions in terms of the types of CS that can occur in different language pairs. These three types of CS are insertional switching, alternational switching and congruent lexicalization. In connection in particular with the notion of insertional switching, I will also consider the role of two concepts often used in CS research, which are embedded language and matrix language. These terms were used by Myers-Scotton (1993b) in trying to distinguish the role of the two languages involved in code-switching (Myers-Scotton did not coin these terms). The embedded language (EL) is usually taken by Myers-Scotton to be the non-dominant language of the bilingual speaker, and she argues that the EL usually plays a

lesser role in CS. This language supplies lexical items to be inserted into the grammatical frame of the other language or the matrix language. The matrix language (ML) is usually the dominant language of the bilingual and is taken by Myers-Scotton to play a more prominent role in CS. The ML provides the grammatical frame for the insertion of EL lexical items (Myers-Scotton 1993b: 6).

Insertional switching occurs when lexical items or even whole constituents from one language are inserted into the grammatical frame of the other language. The sentence below from Bengali-English CS is an example of insertional switching, where English is the EL and Bengali the ML. In the example below, the EL elements are *intention* and *clear,* which are inserted into the Bengali grammatical frame.

2(a) amar *intention* kintu khub *clear*
1SG.GEN intention but very clear
But, my intentions are very clear.

Alternational switching, a concept made popular by Poplack and associates, occurs when stretches of conversation from both languages alternate. These stretches can be as short as a phrase to as long as a few sentences. The example below, again from Bengali-English illustrates this type of switching.

2(b) ami oi chele-ta-ke cini *who dropped out of school*
1SG that boy-DEF-ACC know.1P who dropped out of school
I know that boy who dropped out of school.

In the above example, the first stretch of the sentence is from Bengali and the next stretch from English. There is alternation between the two languages in the same sentence. Congruent lexicalization occurs when both languages to a large extent share a grammatical structure and that structure gets filled by material from either language (Muysken 2000: 3). The example

below is from Dutch-English CS where both Dutch and English (closely related languages) share a grammatical frame for part of the sentence, and this frame is filled lexically with elements that belong to both Dutch and English.

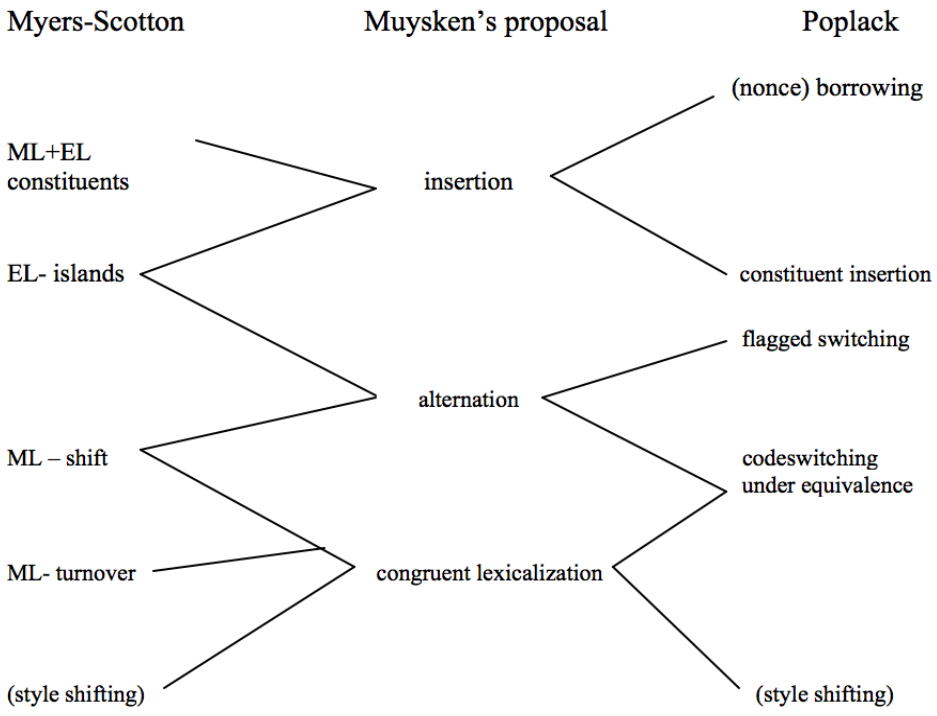
2(c) Weet jij [whaar] Jenny is?
Do you know where Jenny is?
(Crama and van Gelderen 1984, cited in Muysken 2000: 5)

In the above example, the sequence *whaar Jenny is* is shared by both English and Dutch (Dutch *whaar* is very close to English *where*) and filled with elements from either language. The three types of switching proposed by Muysken are not mutually exclusive because they may co-occur in the same conversation. A particular CS community can show all these three types of switching. Also, each of these patterns of switching may turn into other patterns, depending on a change in the bilingual situation. That is, insertional switching may turn into alternational switching or congruent lexicalization, in case there are changes in the bilingual setting such as changes in language dominance, proficiency, attitudes etc.

Muysken's (2000: 3) model does not include only one type of switching (unlike Myers-Scotton's model (1993a) or Poplack's model (1980)) because for him there are various kinds of processes taking place in bilingual speech which cannot be accounted for by only one type of switching. Instead his model includes three types of switching to account for various types of CS data. It is not necessary for mixing between two languages to show all these three processes, though they may co-occur. This model is useful because it does not distinguish between CS and borrowing. For example, in bilingual complex verbs, it may not be clear at first glance whether the EL verbal or nominal element is a nonce loan or not. The classification of this element as a nonce loan or established loan is often difficult and may not be reliable, because that would

involve an investigation of factors such as frequency, phonological and morphological integration, as well as sociolinguistic factors such as speaker intention, setting, topic and proficiency (Muysken 2000). By using insertional switching, the distinction between single word code-switches, borrowings and nonce loans becomes less relevant. Myers-Scotton's and Poplack's models have some overlaps but they make use of different terminology to refer to similar or identical processes. Some of the terminological differences and similarities become more apparent through figure 3.1 below, proposed in Muysken (2000: 32).

Figure 3.1 Muysken's CS model in relation to Myers-Scotton and Poplack's models



In this section, I described some of the common processes and mechanisms of language contact which will be relevant for the discussion on contact-induced change in Bengali and

English in later chapters. Among the processes of contact, I considered different kinds of contact-induced change (*interference, transfer* and *borrowing*), as discussed in previous studies. I also considered briefly the issue of CS, which has been described as a mechanism of contact-induced change. I presented a description of some prominent frameworks that have accounted for CS patterns and the ways in which CS is related to other contact-induced changes such as *borrowing.* In the next section, I turn to the issue of how to ascertain that contact-induced change has taken place in a language and how to subsequently analyze the changes that have been established as being contact-induced.

3.4. Ascertaining contact-induced change

In this section, I first discuss the framework proposed by Thomason (2001), which I will adopt to establish the existence of contact-induced change in Bengali. Then I focus on the quantitative aspect of the framework proposed by Poplack & Levey 2010) that I will be using to analyze the changes that I investigate in Bengali from a synchronic perspective. I do not propose to integrate the two frameworks in my analysis, because I will be using them for different purposes, the first one (Thomason) for a diachronic analysis and the second (Poplack & Levey) for a quantitative analysis of contact-induced change.

To make it possible to ascertain and make a strong case for contact-induced change Thomason (2001)'s theoretical framework provides five criteria that should be satisfied in order to demonstrate that contact-induced change has indeed taken place. This framework is also grounded in an assumption that the socio-historical forces surrounding a contact situation are crucial in shaping the outcomes of that situation; given the 'right' circumstances, any change can occur.

In order to establish change in the first place, Thomason argues that the first step is to examine the (receiving) language as a whole. One might be interested in investigating only a few structural features but a case for contact induced change becomes more robust if it can be shown that much structural interference has occurred in the RL. Secondly, an SL has to be identified. In contact scenarios where people have shifted completely to the dominant language, and there are no sufficient records of an SL, it might be hard to identify an SL. Not only is there a need for the identification of an SL, but enough contact should also have taken place for structural interference to occur in the RL. Third, shared structural features between the SL and the RL have to be identified. However, when changes take place from the SL to the RL, structural features belonging to the SL may not occur in the same way in the RL as they do in the SL. Fourth, it is important to show that the shared features in the RL were not present in that language before contact with the SL occurred. Fifth, one must also provide evidence that the shared features were present in the SL prior to contact with the RL (Thomason 2001: 91-94).

Thomason's framework also takes into consideration the issue of *multiple causation*, those changes that may emerge from a combination of factors that include language internal developments as well as contact-induced changes. Multiple causation is always a real possibility in a contact situation and an examination of a given change in a contact situation must take into account potential internal and external motivations. An example of multiple causation provided by Thomason is that of the development of palatal consonants in three neighboring languages, Yimas, Alamblak and Enga of the Sepik River basin in northern Papua New Guinea. These three languages share a palatal series of consonants that are otherwise rare in non-Austronesian languages of New Guinea. These consonants are old in Enga but are innovations that have taken place in Yimas and Alamblak. Given that the three languages form part of a Sprachbund, at first

glace, this feature seems like a likely candidate for a contact-induced change in Yimas and Alamblak. However, this feature arose independently in Yimas and Alamblak through an internally-motivated phonetic change from an apical consonant such as /t/ + /i/ or /y/. Therefore, it seems plausible that the palatal consonants arose in Yimas and Alamblak from both internal motivations as well as through interference from Enga (62-63). In the analysis of copular predicates in Bengali (chapter 6), I consider the possibility of change from multiple causation.

In the next section, I turn to the discussion of how to analyze the changes that have been established as having emerged from contact.

Analyzing contact-induced change

Thomason's 2001 framework provides a solid methodology for the diachronic investigation of contact-induced change. However, after establishing that a change in a language is contact-induced, it also needs to be analyzed synchronically. To investigate the synchronic aspect of contact-induced change, I use the variationist quantitative methodology proposed in Poplack & Levey (2010). Although Poplack & Levey also provide a methodology for ascertaining contact-induced change in terms of diachrony, I have not integrated the diachronic frameworks of Poplack & Levey and Thomason. I use the two frameworks in a complementary manner, in that Thomason's framework provided the methodology for the diachronic investigation of contact-induced change and Poplack & Levey's framework provided a quantitative methodology for the synchronic analysis of change.

Poplack & Levey's (2010) methodology is geared towards ascertaining the linguistic and social variables that govern changes in languages. A main principle of the variationist approach is that when speakers use a particular variable frequently in their speech, the choice of that

variable is not random but can be predicted from linguistic environments and social characteristics of the speaker. For Poplack & Levey, "the key theoretical construct of this framework is the linguistic variable […] which comprises a set of variants among which speakers alternate in expressing a given meaning or function". Analysis of this nature shows that variation is structured and "can be inferred from the distribution and conditioning of competing variants in discourse" (398).

In addition, before proceeding to the analysis, Poplack & Levey discuss the need to use or collect linguistic data from the appropriate speech community that shows "intensity of contact", greater "length of contact", relative difference between the "status of the languages" and a bigger "size" of the community. They provide arguments in favor of using the *variable rule analysis* framework and the multiple regression analysis procedure to investigate which linguistic (e.g. phonological, morphological, syntactic) and extra-linguistic (e.g. age, minority status, proficiency) factors contribute to the choice of one variant over the other and whether such choices are statistically significant as compared to other variants (400).

Poplack & Levey analyze three features that have been claimed to be candidates of contact-induced change (attrition of subjunctive in Ottawa-Hull French, preposition stranding in the same variety of French and relative marker in Quebec English) according to the variationist methodology and argue that in each of the cases, contact was not the cause behind the occurrence of that variant. In case of the subjunctive, the use of the variant was class-based and it did not constitute a change in the language; in case of preposition stranding the variant is dependent on the lexical identity of the preposition while the use of relative markers in Quebec English is related to the age of the speakers. They argue that for each of these features, contact had been invoked to explain the variation but the variation was actually conditioned by other social and

linguistic factors. They make this argument only on the basis of three studies, which does not constitute evidence to discard contact-induced change as a rare phenomenon. However, they urge scholars to be careful with regard to variation in language, which often ends up being interpreted as change in the language.

In this section, I discussed the issue of ascertaining and analyzing contact-induced change and introduced the frameworks I will use to determine and investigate change in Bengali and English. In the next section, I address the issue of bidirectional influence in language contact and highlight the nature of mutual interaction between languages in contact.

3.5. Bidirectional influences

The issue of bidirectional influences is a primary focus of this dissertation, where I examine possible morphosyntactic changes that the languages Bengali and English have had on each other due to longstanding contact.

Traditionally studies in bilingualism and second language acquisition have discussed linguistic influence as being unidirectional, that is when languages are in contact, it is the native language or the first language of the bilingual speaker that has been argued to have an effect on the second language of the bilingual speaker (e.g. Dechert & Raupach 1989, Gass & Selinker 1992 and Kellerman & Sharwood Smith 1986, as mentioned in Pavlenko & Jarvis 2002). In more recent times, several studies (as pointed out in Jarvis & Pavlenko 2008) have argued against the notion of unidirectionality in language transfer by showing that both the L1 and L2 can influence each other. In other words in bilingual speech, there is evidence of *crosslinguistic transfer*. *Crosslinguistic transfer* is defined as the influence of a person's knowledge of one language on that person's knowledge or use of another language" (Jarvis & Pavlenko 2008:1).

An instance of *crosslinguistic transfer* comes from a study by Pavlenko & Jarvis (2002) on oral narratives from 22 Russian-English bilinguals who learnt English as adults. This study provides evidence to suggest that the effect of *transfer* is observable in both the L1 and L2 of the speakers. As part of this study, subjects were shown two films without any dialogue and were asked to narrate the films both in Russian and English (195-196). The narratives were later analyzed in terms of morphosyntax, lexis and semantics and nine categories were considered for the analysis: linguistic framing (of the narrative), semantic extension, lexical borrowing, tense/aspect, case marking, loan translation, subcategorization, article use, and word order. The results of their analysis show that some categories such as framing, semantic extension, loan translation, subcategorization, article use and word order show an effect of the L1 on the L2, whereas L2 influence on L1 is also evidenced in framing, semantic extension, lexical borrowing, case marking, loan translation, and subcategorization. While some linguistic categories are more susceptible to either L1 influence (article use and word order) or L2 influence (lexical borrowing and case marking), bidirectional transfer is seen in framing, semantic extension, loan translation, and subcategorization. Overall this study establishes the fact that bidirectionality in language transfer should be researched further to get a more complete picture of how the bilingual brain functions and to what extend the two languages can influence each other.

Pavlenko & Jarvis's (2002) study also suggests that bidirectionality of (synchronic) transfer shows the dynamic and flexible nature of the bilingual brain and of the ability to control two languages. And this supports Cook's (2006) and Grosjean's (1998) proposals which have argued in favor of a multi-competence framework which views a bilingual not as being two monolinguals in one body but someone who possesses a "distinct compound state of mind" and has a unique and "complete linguistic system" (Jarvis & Pavlenko 2008: 17). Research on

transfer in terms of bidirectionality has experienced a decent amount of progress in recent times as discussed in Jarvis & Pavlenko (2008).

As we will see in the case of Bengali and English, there is some evidence of bidirectional influence but this does not weaken the nature of the competence that the bilingual speakers have in the two languages. The speakers are in control of a linguistic system that lets them communicate with ease within and outside this community, even though from a monolingual perspective their competence in the two languages may seem partially inadequate. Below, I discuss some literature on the influence of English on Indian languages as background for my investigation of Bengali-English contact.

Influence of English on Indian languages

Some contact-induced changes in Indian languages have been documented as a result of long standing contact between those languages and English. In the case of Hindi, Kachru's (1977) study and Bhatia's (1982) study on the *Englishization*[11] of Hindi discuss how Hindi has been affected in terms of phonological and lexical borrowings as well as morphosyntactic changes resulting from contact with English. They also report that CS between Hindi and English is very frequent in urban areas of the country, which is supposedly developing into a mixed code called Hinglish (a combination of Hindi and English). These studies also discuss how bilingual verbal compounds such as observe *karna* (do) 'to observe' have become common in Hindi because of extensive CS. Other studies have reported on Hindi-English CS in the media, films and the urban context (e.g. Bhatia 1982, 1999, Bauer 2008; Si 2011). The frequent use of bilingual compound verbs in CS between Indian languages and English is also reported in

[11] Kachru uses this term to refer to English influences on Hindi as a result of contact.

Annamalai 1978, 1989, Romaine 1986, 1995, Moravscik 1975, Wohlgemuth 2009, Pillai 1968, and Muysken 2000.

Apart from lexical borrowing, use of bilingual verbs and CS, other morphosyntactic changes in Hindi have also been reported by Bhatia and Kachru in areas such as impersonal constructions, where impersonal (passive) constructions similar to English, e.g. *dekha gaya hai* "It has been seen", have become more common, arguably due to English influence, instead of the active forms such as *dekha hai* '(we) saw', which were traditionally more frequent in Hindi. Also, Hindi has SOV word order, but SVO constructions, supposed to have been influenced by English, are used for stylistic purposes. Indirect speech and passivization with *dvara* 'by' are also argued to be due to English influence[12]. Kachru also mentions that a construction with the post head modifier *jo* 'who' which is a type of relative clause construction in Hindi is considered to be an influence from English. Consider the two examples below:

3 (a) vah larka jo ja raha hai mera bhai hai
DEM boy which go IMPFV is 1SG.GEN brother is
The boy who is going is my brother.

(b) jo larka ja raha hai vah mera bhai hai
which boy go IMPFV is DEM 1SG.GEN brother is
Lit: Which boy is going, that is my brother.
The boy who is going is my brother.

According to Kachru, sentence (3a) is a new construction in Hindi (arguably from English influence) because traditional Hindi relative clause constructions are like the one shown in (3b). However, these studies are not well grounded or extensively supported because it is not clear how these changes were established and to which extent they have spread (refer to Thomason's 2001 criteria in section 3). More recently, though, Puri's (2011) detailed work involving a

[12] Passivization with *dvara* is probably not a contact-induced change in Bengali, because one of the nineteenth-century plays shows occurrence of this structure.

diachronic investigation of Hindi argues how contact with English has led to a change in the relative-correlative (RC-CC) clause structure of Hindi. Originally, Hindi had structures where the RC-CC would adjoin to a main clause. Puri's investigation of old and modern Hindi texts show that the RC-CC construction now can both adjoin and embed to the main clause because of the influence of English.

Studies on Tamil-English and Punjabi-English have also shown how these languages have been affected by English in terms of heavy lexical borrowings from English, increase in use of bilingual verbal compounds, and more use of CS with English (Annamalai 1978, 1989; Romaine 1986).

However, there has been very little research regarding Bengali-English contact (in India and Bangladesh). A few studies have reported lexical borrowings into Bengali from English (Chatterji & Grierson 1926, Pillai 1968, Banu & Sussex 2001, Thompson 2010). In a grammar dating back to 1885, Yates and Wenger also briefly mentioned how English has influenced the relative clause constructions of Bengali. For example, in English a sentence with a relative clause would be "Here is the boy whom I saw", while in Bengali, the same sentence would be:

4 Je chele-ke ami dekh-ech-i she eikhane
REL PR boy-ACC 1SG see-PFV-1P DEM here
Literally: which boy I saw (he) is here
Here is the boy whom I saw.

Translated literally, the sentence would be "Which boy I saw he is here" and saying it in any other way would sound odd. However, due to the influence of English, Yates and Wenger argue that Bengali developed relative-demonstrative sentences like the ones used in English as shown below (Yates and Wenger 1885:100-101):

5 Ei shei chele ja-ke ami dekh-ech-i
This DEM boy REL-ACC 1SG see-PFV-1P
This is the boy whom I saw.

Similar to the argument for some English effects on Hindi, this claim about Bengali is also not well grounded. Therefore, I investigate the nature of Bengali-English contact by investigating changes both in monolingual and bilingual speech, considering the frequency of variation in both, as well as in the older variety to determine whether they are indeed contact-induced. I address this issue in chapters 5 and 6.

In addition, some studies on Indian English (IndE) have argued for the presence of interference effects from Indian languages on IndE. Research on IndE shows the presence of some properties that seem to diverge in non-native varieties of English in general such as (articles, progressive forms, omission of copulas, lack of agreement among other features). The explanations for the occurrence of these properties have ranged from interference (or transfer) effects of L1 to pedagogical lapses on the part of instructors and to discourse universals. Early work on non-native varieties of English has mainly explained the occurrence of non-standard features in terms of interference effects from the L1 or of unsuccessful acquisition of the target language. In chapter 7, I investigate potential changes in the English spoken by bilingual Bengali-English speakers of West Bengal and explore relevant studies on IndE in more detail.

3.6. Conclusion

In this chapter, I introduced some key concepts in the field of language contact that are relevant to the research questions of my dissertation. At first I discussed the issue of contact-induced change in general including changes that arise from multiple causation. Then I addressed the different processes of contact (interference, transfer, borrowing) as well as mechanisms (code-switching) that lead to those changes. I provided details about the ways in which these terms have been used to label contact-induced changes and how I will use some of those terms in

my analysis of change in Bengali and English. I also introduced the language contact framework of Thomason (2001) which I will use to ascertain change, and the framework of Poplack & Levey (2010), which I use to analyze synchronic variation in connection with change. I addressed the topic of bidirectional influences between languages in contact and discussed some studies on contact between Indian languages and English. This discussion aimed at supporting the relevance of the research questions in this dissertation and providing analytical background for my investigation of Bengali-English language contact in India.

Chapter 4: Methodology: Data collection and coding

4.1. Introduction

In this chapter, I discuss the methodology I have used to investigate the research questions of my dissertation. The primary aim of this dissertation is to ascertain and analyze changes in Bengali and English as a result of bidirectional influences from each other. As described in chapters 1 and 2, as a result of widespread contact with English, English-educated Bengali speakers primarily speak a "mixed code" involving at least some code-switching between Bengali and English. However, Bengali speakers who have had little exposure to English education primarily speak monolingual Bengali, which includes few lexical borrowings from English with almost no CS. Due to the length and intensity of contact with English, including frequent CS between the languages, we expect the speech of bilingual Bengali-English speakers to show contact-induced effects from English, which may not show up as extensively in the speech of monolingual Bengali speakers. Even so, we expect the speech of monolingual Bengali speakers to show at least some amount of change (in comparison to Bengali as it was spoken when contact with English had not yet become as intense) because monolingual speakers interact with bilingual speakers regularly and some features from bilingual speech may have diffused into the speech of monolinguals. Therefore, methodologically, I analyze changes in Bengali by doing a synchronic comparison between a bilingual Bengali-English corpus and a monolingual Bengali corpus, which I use as a control variety. Additionally, I carry out a diachronic comparison between an older variety of Bengali (as represented by

Bengali plays published in the nineteenth century, used as a historical control variety) and the bilingual and monolingual Bengali corpora, following the framework proposed by Thomason (2001) on ascertaining contact-induced change, which I discussed in chapter 3. I also address the possibility of changes in English by carrying out a quantitative and qualitative investigation of a corpus of production data in English collected from bilingual Bengali-English speakers. In the sections below, I describe all the four corpora that I have compiled to analyze in this dissertation (4.2), the procedure of the data collection (4.3) and the coding and analysis of the data (4.4).

4.2. Description of corpora

4.2.1. Nineteenth-century Bengali corpus

For the diachronic investigation of contact-induced changes in Bengali, I used a corpus compiled from two Bengali plays published in 1854 and 1860. For the purpose of this dissertation I examined only two plays but in future I plan to expand this corpus by including more texts. I have used Bengali plays from this period firstly because it was around this time that drama as a genre developed in Bengali literature (Guha-Thakurta 1930; Chatterjee 2007; Rakshit 2013). As discussed in chapter 2, prior to the nineteenth century, the Bengali theatrical tradition consisted of performances of folk-theatrical forms called *Jatras*, musicals that were based on mythological and historical subjects (Guha-Thakurta 1930). These forms were usually performative and hardly ever published. Nevertheless, in the nineteenth century, Bengali writers started writing Bengali plays that emulated the European plays in English that they saw staged and performed at theatres in Calcutta. At first, the plays written in Bengali were translations of Sanskrit and English plays. However, within a few years original Bengali plays were also produced. Another notable change in Bengali literature took place around the mid-nineteenth

century when colloquial Bengali started being used for literary purposes. A diglossic relation has existed between two varieties of Bengali, a formal and a colloquial one, which were used for literary and spoken purposes respectively. From the late 1850s on, a new literary tradition developed when some plays (and novels) started getting written in colloquial Bengali (Chatterjee 2007: 222). Therefore, plays from this period would be one of the earliest sources of written data that would be more comparable to present day spoken (colloquial) Bengali.

I used plays for the diachronic investigation because the synchronic investigation of language change is based on speech data and plays too would presumably provide a good reflection of oral speech. Therefore, the diachronic comparison can be relatively more uniform and accurate if a genre that provides a good reflection of speech is used. However, because the genre of playwriting in Bengali developed as a result of English influence, plays that were written prior to the development of contact with English do not exist. So, the only texts we can use were written after contact with English was already under way. One of the two plays I have examined for the diachronic comparison is *Kulin Kula Sarbaswa,* written by Ramnarayan Tarkaratna and published in 1854. This play is arguably the first original play to have been written in Bengali and it deals with the practice of polygamy among the high-caste Brahmins of Hindu society (Guha-Thakurta 1930). The other play I have used is *Nil Darpan*, written by Dinabandhu Mitra and published in 1860. *Nil Darpan* 'The Indigo Planting Mirror' represents a political protest by poor peasants in Bengal against an exploitative regime of British indigo planters.

4.2.2. Bilingual Bengali-English corpus

As part of the synchronic comparison, I use two corpora. The first is a bilingual Bengali-English corpus of production data that I collected from bilingual speakers of Bengali and English

in the city of Kolkata in West Bengal, India in 2010. The data was collected from speakers for whom Bengali is the first language and English the second language, which they learnt from age 4 in English instruction schools. They also have Hindi as part of their linguistic repertoire because most high schools in India provide about three years of Hindi language instruction.[13] Knowledge of Hindi also gets reinforced through its status as official language of the country and its extensive use in the media.

This data was collected from 30 speakers (the ages of the speakers ranged from 20 to 45 years) and consist of recordings of 12 informal bilingual conversations that last between 30 minutes and an hour. However, I was able to use data from only 27 of the 30 speakers because one of the recordings had a high level of noise in the background. This bilingual corpus corresponds to an estimate of 93,538 words. In all the conversations, the participants code-switched between Bengali and English, though there were differences in the degree of code-switching, with a few conversations consisting of more use of English than Bengali and others consisting of an equal amount or lesser use of English than Bengali. Apart from CS, there were also stretches in the conversations that were entirely in English, but there were very few stretches in the conversations that were entirely in Bengali, that is there were usually a lot of English lexical insertions (or borrowings) in Bengali sentences.

4.2.3. Monolingual Bengali corpus

The monolingual Bengali corpus was compiled through fieldwork I did in the cities of Kolkata and Durgapur in West Bengal, India in 2012. The corpus consists of recordings of

[13] I discussed the three-language formula in relation to language planning in India in chapter 2. The teaching of Bengali, English and Hindi in most high schools of West Bengal has been enforced as part of the language planning efforts in India.

informal conversations (about an hour long) from 14 subjects who have Bengali as their first language. The corpus corresponds to an estimate of 51,516 words. The ages of the speakers ranged from 22 to 60 years. Of the 14 speakers, ten speakers have about 0 to 4 years of exposure to English in high schools, that is, they were not educated primarily in English but received only second language instruction in English. The other 4 monolingual Bengali speakers have only 2 years of elementary school experience where they received no instruction in English. However, all of these monolingual Bengali speakers interact regularly with bilingual speakers. As a result of this regular linguistic interaction we expect their speech to show some diffusion effects from the speech of bilingual speakers. Hindi is also part of the linguistic repertoire of some of these monolingual Bengali speakers, who are exposed to the language at least through the media[14].

4.2.4. Monolingual English corpus

The monolingual English corpus was compiled from English production data that I collected in Kolkata, West Bengal, India in 2012. This corpus consists of recordings of two sets of data, narrations of two stories based on picture books and individual sociolinguistic interviews between the subjects and the interviewer. Both the narrations of the stories and the interviews were conducted entirely in English. The data were collected from 16 bilingual speakers of Bengali and English. These speakers speak Bengali as their first language and learnt English as a result of being educated in English instruction schools since age 4. The ages of the speakers ranged from 23 to 65 years. These speakers also learned some Hindi as part of their linguistic repertoire.

Story narrations

The story narrations are based on the picture books *A boy, a dog and a frog* by Mercer Mayer

[14] Hindi language instruction is usually not enforced in many public schools run by the government in West Bengal.

and *A boy, a dog, a frog and a friend* by Mercer and Marianna Mayer. The subjects were instructed to narrate the stories in the most natural way they could. There were no specific instructions on how the first story was to be narrated but they were asked to narrate the second story in the past tense. Henceforth I will refer to the first story, *A boy, a dog and a frog* as Story A and the second story, *A boy, a dog, a frog and a friend* as Story B. This corpus of story narrations therefore consists of 32 narrations, 16 for Story A and 16 for Story B, corresponding to an estimate of about 25,133 words.

Interviews

The interview data consists of 16 sociolinguistic interviews conducted entirely in English[15] where I asked the subjects questions about their life experiences, such as their experiences from childhood, memories and impressions about people from different places, impressions about the city they reside in and about their jobs. The interviews lasted from thirty to forty minutes. The interview data corresponds to a total estimate of 91,783 words.

For this dissertation the entire corpus of story narrations was transcribed, coded and analyzed, whereas a random selection of a block of about 100 lines (one line corresponds to about 8 words in CLAN) was coded and analyzed from each one the interviews. This was done to make the analyzed data from each story and interview comparable in size.[16]

4.3. Procedure of data collection

In this section, I describe the procedure I followed to collect the data. I first obtained

[15] The list of questions that were asked during the sociolinguistic interviews is provided in appendix 5.

[16] The monolingual Bengali corpus and the monolingual English corpus were collected and compiled as part of a project with my co-advisor, Acrisio Pires, aimed at investigating the morphosyntactic properties of English and Bengali as spoken by Bengali-English bilinguals (some of the features and phenomena that were the focus of that project are also investigated in this dissertation).

signed permission from the subjects to do the recordings, through informed consent forms. Once subjects provided their consent to participate in the study, they were asked to fill out questionnaires that asked them to specify details about their ethnolinguistic background. These questions included details on how many languages they spoke, when and how they learnt the languages and also the domains in which they use Bengali and/or English and/or Hindi in their daily lives. The ethnographic questionnaires are provided in appendices 3 and 4. The bilingual speakers answered a written questionnaire in English, whereas for the monolingual Bengali speakers, I asked them the questions orally in Bengali and wrote down their answers.

The bilingual Bengali-English corpus and the monolingual Bengali corpus consist of recordings of informal conversations and each conversation had between two and three participants. The participants in each conversation knew each other beforehand because they were part of the same peer group and therefore it was easy to initiate conversations between them. The conversations were recorded using a Marantz Professional recorder in places such as homes, university buildings, classrooms or offices. Participants were also asked to speak in their most casual and usual way, the way they usually converse with their peers.

For the monolingual English corpus data collection, the subjects also took a written Proficiency test in English after the questionnaire, with the goal to assess their proficiency in English. The proficiency test consisted of two written cloze test passages from the Cambridge Proficiency Exam. After these steps, the subjects were asked to narrate the stories based on the two picture books (details above) and then to participate in the sociolinguistic interview with me.

4.4. Coding of the data

The bilingual Bengali-English data, monolingual Bengali and monolingual English data were all transcribed in CLAN and then transferred and coded in Microsoft Excel. I used the

variationist methodology as discussed in e.g. Poplack & Levey 2010 and Tagliamonte 2012, to ascertain the quantitative distribution of the variables in the corpora. The first step of variation analysis aims to determine how often the variants of a variable occur in the body of data. This step consists of providing counts and frequencies of each of the variants of the dependent variable (Tagliamonte 2012: 11-12). Following this methodology, I first coded all the different variants of the features. Therefore, for the complex verb structures, I coded all the Bengali and bilingual complex verbs; for the copular predicates I coded all the null and overt equational copulas (*hocche*) of Bengali and for the articles and progressive forms in the monolingual English corpus I coded all the divergent and non-divergent occurrences of these features. After the coding of these variables, I determined the frequency of occurrence of each of these variables. Then based on the frequencies, I ran chi-squared tests of statistical significance on the data firstly to ascertain which of the variants of a particular variable had a significantly higher rate of occurrence than the other variants, and secondly, to determine if these differences were related to their language background. I provide details on the coding of the individual features in the following chapters, which analyze these features.

4.5. Conclusion

In this chapter, I discussed the methodology that I have used in this dissertation to examine the research questions. I described the four bilingual and monolingual Bengali and English corpora I collected and compiled to investigate the research questions. I also described the procedure of the data collection and the coding of the data for the analysis. In chapter 5, I analyze the complex verb structures, in chapter 6, I analyze the copular predicates in Bengali[17]

[17] I have worked jointly with my co-advisor, Marlyse Baptista on some aspects of this project.

and in chapter 7, I investigate the articles and progressive forms in the English of the bilingual speakers.

Chapter 5: Contact-induced changes: From the lexicon to morphosyntax

5.1. Introduction

In this chapter, I investigate evidence of changes that have taken place in Bengali considering the speech of monolingual and bilingual speakers and the degree of change in both varieties. The changes I discuss in this chapter include those that are more clearly noticeable, including changes at the lexical level comprising borrowing of English words into Bengali and the use of loan translations in Bengali (section 5.2). Then I discuss the mechanism of code-switching between Bengali and English as observed in the speech of bilingual speakers (section 5.3). Following that, I provide evidence of change in Bengali at the morphosyntactic level, focusing on the identification and analysis of bilingual verbs. I analyze the verbal structures in modern monolingual and bilingual Bengali in comparison to nineteenth-century Bengali and discuss the ways in which the complex verbs of modern Bengali differ from those of the older Bengali texts. I argue, following some previous analyses of bilingual verbs in other language pairs, that the way in which bilingual verbs are reanalyzed in modern Bengali constitutes a case of morphosyntactic change.

5.2. Changes at the lexical level

In this section, I discuss English influences on Bengali at the lexical level. The most easily observable influences of a source language on a recipient language are importations of vocabulary. As a result of the length of contact between Bengali and English,

there is no dearth of English loans in Bengali, both in monolingual and in bilingual Bengali speech.

5.2.1. Borrowing of English lexical items in Bengali

It may be difficult to ascertain in different cases, by examining synchronic data alone, if the uses of single English words are borrowings into Bengali or code-switches between English and Bengali. Single-word code-switches may and often do turn into loanwords, but if we are examining the speech of bilingual speakers in our corpus it may be difficult to identity whether some of the English words used in their speech have already become a part of the Bengali lexicon or not. The examination of monolingual Bengali speakers' speech (who have very little knowledge of English) might reveal more in terms of which words of English constitute true borrowings and have become a part of the Bengali lexicon. In assessing loanword status, I refer to Haugen's definition of borrowing, as a working definition to examine the monolingual and bilingual Bengali data. Haugen (1950: 212) defined borrowing along the following terms:

> **(**1) EVERY SPEAKER ATTEMPTS TO RE-PRODUCE PREVIOUSLY LEARNED LINGUISTIC PATTERNS in an effort to cope with new linguistic situations. (2) AMONG THE NEW PATTERNS WHICH HE MAY LEARN ARE THOSE OF A LANGUAGE DIFFERENT FROM HIS OWN, and these too he may attempt to reproduce. (3) If he reproduces the new linguistic patterns, NOT IN THE CONTEXT OF THE LANGUAGE IN WHICH HE LEARNED THEM, but in the context of another, he may be said to have 'borrowed' them from one language into another (212).

Haugen's definition of borrowing also takes into account the issue of *substitution*, wherein speakers substitute or replace some aspects of the borrowed element with aspects of their own grammar. Haugen provides the example of American Norwegian speakers using the English word *whip* [hwip], but pronouncing it as [hypp-] as a result of substituting aspects of their own phonology in the English word (213). Such a process is also observable in the speech of

monolingual Bengali speakers while using English loanwords. For example, in sentence (1a), the word *office* is adapted in Bengali as *apish* because Bengali lacks the alveolar fricative /*s/* and the labiodental fricative /f/ phonemes. Also, *post* is adapted as *posht* since Bengali doesn't have the word final -*st* cluster. In bilingual speech however, as shown in (1b), the word *office* is used without any phonological adaptation in the language.

1(a) *posht apish*-e jiggesh kor-te hɔ-b-e (from monolingual speech)
post office –LOC question do-IMPFV be-FUT-3P
… have to inquire at the post office.

(b) æk bɔchor-i hoe-ch-e *office*-e (from bilingual speech)
one year-FOC be-PFV-3P office-LOC
It has just been a year in the office. (Bilingual Speech)

But at least in bilingual speech, phonological adaptation might not always be an accurate indicator of loanword status.[18] Another criterion that has been mentioned with regard to loanword identification is morphosyntactic adaptation. If a word from a source language is used in a receiving language, it may get morphosyntactically integrated in the receiving language by getting attached to case markings or other grammatical inflections. However, this criterion might also not be completely reliable because bilingual speakers may use source language words in their speech by morphosyntactically integrating them into the receiving language, but that does not mean that those words have spread into the speech of other bilingual and monolingual speakers, in which case we cannot tell with certainty if they are cases of true lexical borrowings. This is not to say however that phonological and morphosyntactic integration in a receiving language are completely unusable criteria for identifying loanwords. Indeed, in most cases they

[18] For example, when Indian English speakers pronounce English words with alveolars /t/ and /d/, they use the retroflex /*ʈ*/ and /*ɖ*/ instead, which would be cases of phonological adaptation. So, based on this criterion alone we cannot be certain if English words such as *tire* or *damage* in bilingual speech are in fact borrowings or code-switches.

may accurately identify loanwords, but in other cases however, they may yield wrong results. I will return to these cases in section (5.3), which deals with CS between Bengali and English. Therefore, one must be cautious while employing these to identify loanwords. However, measuring the relative frequencies of occurrence of source language words in the speech of monolingual speakers would perhaps be able to tell us with more accuracy about the loanword status of different source language words. This measure would tell us how common the words are in the speech of monolingual speakers.

Apart from English loanwords such as *post office* (which have been borrowed because such services were introduced in India by the British government and Bengali didn't have equivalent terms for them prior to British rule), other common English words have also been imported into Bengali and occur regularly in the speech of monolingual Bengali speakers as shown in (2a) and (b) below.

2 (a) tarpor unun dhɔra-te æk ghɔnta *time* lag-e (from monolingual speech)
then stove hold-IMPFV one hour time take-HAB.3P
It then takes an hour's time to light the stove.

(b) *daily* kaaj kor-te ja-i
daily work do-IMPFV go-HAB.1P
(I) go for work daily.

Examples (2a) and (2b) are from two monolingual speakers of Bengali who had no formal education in English and had only 2 years of elementary school education in Bengali. Despite this background and the fact that the words *shomɔe* 'time' and *roj* 'daily' exist in Bengali, these particular speakers used the English counterparts for these words. This is presumably because through interactions with bilingual speakers these English lexical items have made their way into their speech. The fact that these words occurred in the speech of (completely) monolingual

speakers gives us sufficient evidence to argue for these words to be cases of borrowings from English, rather than code-switches.

English lexical items occurring in the speech of monolingual Bengali speakers is not limited only to culturally non-native objects or technological innovations such as *computer, laptop, telephone, mobile phone, televisions, television channel, airplane* etc., but also extend to words for which there are Bengali counterparts such as *time* and *daily*.

When it comes to the speech of bilingual speakers, the use of English words is widespread and occurs for both basic and non-basic vocabulary items as shown in the sentences below. In (3a) we find the word *skin*, which may be taken as basic a vocabulary item (given its reference to a body part) and in (3b) we find the word *account*, which is from non-basic vocabulary.

3 (a) amar *skin*-e problem hɔ-e
1SG.GEN skin-LOC problem be-HAB.1P
My skin has problems.

(b) prothom *account*-ta jɔkhon khul-echi-l-am
first account-DEF when open-PFV-PST-1P
When I had opened the first account….

The examples above are only a couple of representative examples of English words that are used in Bengali. We find numerous other cases of borrowings widespread in the speech of bilingual speakers, such as *laptop, internet, website, office, university, building, table, chair, phone, ball, paper* etc. There is also use of compound words such as *special effects, TV channel, main paper, photo frame, service provider,* adjectives such as *cool, plain, beautiful, proud, stupid, simple,* and adverbs such as *specifically, actually, basically, possibly, hopefully.*

As discussed in the borrowing scale of Thomason & Kaufman (1988) and Thomason (2001) when contact between languages is at the initial stage, non-basic vocabulary items are

usually the first elements to get borrowed. With increase in the intensity and duration of contact, more and more vocabulary items are prone to get borrowed. A similar trend is evident in Bengali, where we find that in the speech of monolingual Bengali speakers, there are almost no instances of the use of English words in the domain of basic vocabulary. For non-basic vocabulary however, there is ample evidence of English influence. In case of bilingual speakers, English words are found in the domain of both basic and non-basic vocabulary.

Unlike modern Bengali, we find only eight English words in the 1860 play and none in the 1854 play. The English words occurring in the 1860 play are *school, college, magistrate, inspector, government, mark, commission* and *doctor*. The use of all of these words is not surprising given that this play was published when the British had already established their presence in India, which made way for the introduction of these terms. Also, some of these words appear in the play with some phonological changes. For example 'school' is written *iskul*, 'college' as *kalej*, 'government' as *gɔbment* and 'doctor' as *daktar*. As I mentioned above, according to the borrowing scale, when contact between two languages is at the initial stages, borrowing is usually limited to the non-basic vocabulary. Therefore, the fact that only eight English words appear in the 1860 Bengali play, for concepts that were introduced in India as a result of that contact, suggests that English influence on Bengali at that stage was only limited to non-basic vocabulary. This is of course assuming that the plays provide a relatively accurate reflection of oral speech of the mid-nineteenth century. Bengali now has a native counterpart of all these eight English words, although even monolingual Bengali speakers rarely use them.

5.2.2. Loan translations from English

Another type of English lexical influence on Bengali includes the use of loan translations or calques. Loan translations consist of those English lexical items that are not directly imported

in Bengali (definition of the phenomena of loan translation was provided on p. 42, drawing from Haugen 1950). Instead the English phrases are translated word-for-word to be used in Bengali. Examples of these cases include *kalo taka* 'black money', *kalo bajar* 'black market', *chiruni tollashi* 'combing operation', *prodhan montri* 'prime minister', *mukkho montri* 'chief minister', *duro dɔrshon* 'television', *mɔgoj dholai* 'brainwash' etc. Some of these concepts (such as prime minister) are not necessarily new to Bengali or the speakers of the language, but the translations and the resulting combinations from those translated words are used in place of English words. For example, few speakers of monolingual Bengali today would use *duro dɔrshon* for the word 'television', but Bengali news anchors (from the official government channels) currently use *duro dɔrshon* for the word 'television' while delivering the news. Since governmental television channels aim at projecting Indian identity by using Indian languages, they refrain from directly using English words. For the same reason loan translations played an important role after India gained independence from British rule. At that time these languages were being promoted for use in official spheres (in place of English) and the vocabulary of Indian languages was being enhanced to make them adequate for use in these settings (Sridhar 1992).

5.3. Code-switching between Bengali and English

Now, I turn to the issue of code-switching (CS) between Bengali and English. As mentioned earlier, CS occurs very frequently in bilingual speech; when bilingual speakers of Bengali and English communicate in informal contexts, they do so by constantly going back and forth between the two languages. In fact, it would not be wrong to say that bilingual speech (the mode of communication used in informal contexts by bilingual speakers fluent in both Bengali and English) constitutes CS between Bengali and English. In this section, I describe some of the patterns that are observable in CS between Bengali and English. I do not intend to take up a

formal analysis of CS in this discussion, because that is not a goal of this dissertation. I address on CS for two reasons. First, if one compares the Bengali plays of the nineteenth century to modern bilingual speech, the first thing one observes is the very sparse use of English in the plays. Therefore, CS is a mode of communication that most likely[19] developed among Bengali speakers after these plays were published, and therefore it constitutes a type of 'change' in Bengali (at least regarding language use by bilinguals) insofar as it was completely absent in the older texts. By discussing CS, I intend to provide the reader with a sense of what this mode of communication looks like and in what contexts it is mostly used. Second, CS is one of the mechanisms by which contact-induced language change may take place. The constant back and forth between the languages not only introduces English words in Bengali (over time) but may also prime the use of English grammatical properties in Bengali (I will argue that these properties are observable in the use of bilingual verbs and copular predicates). In the discussion here, I use the frameworks proposed by Myers Scotton (1993a, 1993b) and Muysken (2000) as descriptive tools to understand Bengali-English CS patterns.

CS in Bengali and English is observable at both the intra-sentential and inter-sentential level. Intra-sentential switching is defined as the switch from one language to another within a sentence and inter-sentential switching is defined as the switch from one language to another at a clause or sentence boundary (Romaine 1995: 122). The two examples below illustrate this. Sentence (4a) shows intra-sentential switching and (4b) shows inter-sentential switching.

4 (a) *second account* ta khola-r pɔr ɔnek kichu ho-e ge-l-o
second account-DEF open-GEN after many few be-PFV.3P go-PST-3P
Many things happened after opening the second account.

[19] Although it may have been the case that the author of the 1860 play was just using monolingual speech and had no interest in portraying the bilingual speech of the time, if it existed.

(b) *I didn't go.* Ami ghum-ie por-echil-am
1SG sleep-PFV.PTCP fall-PFV.PST-1P
I didn't go. I had slept off.

According to Myers Scotton's CS framework, speakers who code-switch usually do not have both languages on the same footing. One of the languages is usually more dominant than the other and the dominant language is called the *matrix language* (ML), while the non-dominant one is the *embedded language* (EL). The ML usually provides the grammatical frame for the insertion of elements from the EL. In Bengali-English CS, Bengali usually serves as the ML and provides a structure for the insertion of English elements. These elements can either be single-word switches or multi-word switches. Multi-word switches have been referred to by Myers Scotton as *EL islands* (example (5a) illustrates this). Excessive use of EL islands may change the ML and EL distinction. That is, the EL may become the ML and serve as a grammatical frame for the insertion of ML items. We find evidence of this even in Bengali-English CS (as shown in 5b). In the majority of CS cases, however, Bengali is the ML.

5(a) dekh-te ho-b-e kono *possibility of renewal* ach-e kina
see-IMPFV be-FUT-3P any exist.PRS-3P if
(I) have to check if there exists a possibility of renewal.

(b) *The kittens were born on the* rasta
street
The kittens were born on the streets.

Usually when *EL islands* from English appear in Bengali-English CS, Bengali postpositions and clitics attach to these elements, which serve to integrate them morphosyntactically in the language (as shown in 6a below). In that regard, the EL islands are assimilated as native elements. However, when the EL islands are larger, the degree of morphosyntactic integration may decrease, as shown in (6b), where we expect the word *time* to have a genitive marking

(*time-er por* 'time-GEN after') attached to it but that is missing. This is most likely because *certain point of time* is larger chunk and it is not being treated as an insertion in the language but more like an alternation to English.

6 (a) *government of India*-r *subsidy* dewa bɔndho kɔra uchit na
GEN give stop do should NEG
The government of India should not stop providing subsidies.

(b) Ekta *certain point of time* por ektu *discouraging* ho-e ja-e
One-CL certain point of time after bit discouraging be-PFV.PTCP go-HAB.3P
After a certain point of time, it becomes a bit discouraging.

As I mentioned before, it is difficult to identify (only on the basis of a synchronic study) if single words from English in bilingual speech have become a part of Bengali or are used only in an ad-hoc way (what Poplack calls a *nonce* borrowing). Certainly, as I pointed out above, very few English words occurred in the nineteenth-century plays. However, only on the basis of the three corpora I considered one cannot be certain whether these words have become a part of the Bengali lexicon. Muysken (2000) deals with similar cases by identifying them as instances of insertional switching, putting aside a decision as to whether they could constitute instances of borrowing. Insertional switching occurs when lexical items or even whole constituents from one language are inserted into the grammatical frame of the other language.

The emergence of CS in this community is related not only to bilingual speakers' proficiency in both languages but also to other social factors. Firstly, Bengali and English have different social prestige associated with them. English as a language of higher education in India has more worth in terms of giving people employment opportunities. So, the need to be fluent in English is more important than the need to maintain a 'pure' use of Bengali (there are no contexts in which one is required to speak entirely in Bengali). However, when it comes to

formal contexts such as jobs or education one cannot switch freely between English and Bengali and has to communicate entirely in English. Therefore it is imperative for people to become and remain fluent in English. The use of English also indexes an upper class identity. Therefore, speakers may have initially started using CS to index this identity, though at present this is the default mode of communication for most bilingual speakers in my corpus; similar observations have been made for other parts of India (Kachru 1978). Many bilingual speakers in fact would find it very difficult or even impossible to speak without switching to English since several Bengali words have fallen out of use in their repertoire.

The degree of CS is also related to the topic of discussion and informal exposure to English. For example, topics such as international politics, Hollywood movies, academic discussions, discussions related to jobs, use of technology contain many more instances of CS than when daily chores are being discussed. Therefore speakers are less likely to switch in some conversations than in others. The amount of exposure to English language movies, media, literature, and books is also an important predictor of the degree of CS. For example, participants in my corpus who worked for English newspaper publishers, or were pursuing graduate studies in English literature were more likely to switch to English and to use long stretches of English in their conversations. However, speakers otherwise not familiar with these environments are less likely to use as much English in their speech. Use of CS is also related to other linguistic and social factors such as, for emphasis, humor, dampening directness, bonding and sounding authoritative (e.g. Gardner-Chloros 2009).

In this section, I described some patterns of CS in Bengali and English and also discussed some social factors that may have given rise to CS and contributed to its maintenance in this community. In the next section, I discuss the use of bilingual verbs in both monolingual and

bilingual speech, a change that has occurred in modern Bengali at the morphosyntactic level, which may have been caused at least in part by the frequent use of CS in this community.

5.4. Complex verb structures

In this section, I first discuss the complex verb structures of modern Bengali and then compare their structures with that of older Bengali to show the changes that have taken place in these structures in modern Bengali (following the methodology of Thomason 2001). I address the differences between all three corpora in terms of the quantitative distribution of these structures. I also argue that the way these structures are reanalyzed in modern Bengali constitutes a morphosyntactic change in the language, which may have been caused partly due to the extensive use of CS in this community.

Complex verb structures have been shown to occur frequently in Indo-Aryan languages (Hook 1974, Abbi & Gopalakrishnan 1991, Butt 2003, 2010), such as Urdu, Hindi, Bengali, Marathi, Gujarati, among others languages in the world. These structures contain two or more than two elements such as nouns, verbs or adjectives, which together predicate as a single unit, that is, "their arguments map onto a monoclausal syntactic structure" (Butt 2010:2). Their structure can be 'Noun + Verb' (N+V) or 'Verb + Verb' (V+V), making them complex structurally but not necessarily so in terms of semantics, as I discuss below. I use the term 'complex verb' to refer to the N+V and V+V structures together because of their complex structures but will use separate terms for the individual N+V and V+V structures. Independently, the N+V structures have been called *conjunct verbs* (Bhattacharyya et.al 2006) and V+V structures are called *compound verbs* (Thompson 2010).

In an N+V structure (conjunct verb), the noun is followed by a light verb, such as 'do' or 'be' which bears inflections and turns the N+V structure into a single predicate. For example,

bikri kɔra ‘sale do’ is a complex verb in Bengali, which together means ‘sell’. The V+V structure (compound verb) consists of two or more verbal elements, where the first verb or the main verb in the construction provides the main semantic content to the construction. This verb is followed by a light verb that bears inflections, affects the aktionsart of the joint predication and adds, instead of its inherent lexical meaning, more subtle semantic information such as benefaction, suddenness, volitionality, forcefulness etc. (e.g. Abbi & Gopalakrishnan 1991, Butt 2003, Basu 2010). The choice of light verb lends a slightly different meaning to the construction; for example, *dewa* ‘give’ has a benefactive reading while *newa* ‘take’ indicates that the action is directed towards the subject. An example of the V+V structure is *bhule gechi* ‘forget go-PFV’ meaning ‘(I/we) have forgotten’ and together these two verbs predicate as a single element. The light verbs that occur in a V+V construction include *dewa* ‘give’, *newa* ‘take’*, jawa* ‘go’*, bɔʂa* ‘sit’*,* etc. (see table 5.2 below). These light verbs do not contribute their standard verbal meaning, although they are identical in form to a main verb in the language. Some scholars, such as Butt (2003), do not make a distinction between light verbs such as *kɔra* ‘do’ and *hɔwa* ‘be’ (that occur in N+‘do’ structures) and light verbs such as the ones mentioned above that occur in V+V structures. However, in this chapter I argue that there are some crucial distinctions between these two types of light verbs (see also Basu 2010, which deals with the semantics and event structure of compound V+V verbs in Bengali).

5.4.1. The verbal system of modern monolingual and bilingual speech

The verbal system of modern Bengali (both monolingual and bilingual speech) consists of different types of verbs, including simple and complex ones. The complex verbs can be categorized into two main types, *monolingual verbs* and *bilingual verbs*. The monolingual verbs consist of N+V+(V) and V+V structures where all the components are from Bengali. The

bilingual verbs also consist of N+V+(V) and V+V structures, but crucially in these verbs the main element that provides the semantic content to the construction is from English. Table 5.1 below summarizes and provides examples of these different structures.

Table 5.1: Verbal system of modern Bengali

Verbs types	Sub types	Components of verbs	Examples of each type Example Gloss Translation
Simple Verbs		1 verb	bɔla 'speak' to speak
Monolingual Complex Verbs	Conjunct verbs	N + *kɔra/hɔwa* N + *kɔra/hɔwa* + V	bikri kɔra 'sale do' to sell jiggesh kore newa 'question do take' to ask for oneself
	Compound verbs	V + V	khe newa 'eat take' to eat for oneself
Bilingual Complex Verbs	Conjunct verbs	N [Eng] + *kɔra/**h**ɔwa* N [Eng] + *kɔra/**h**ɔwa* + V	drawing kɔra 'drawing do' to draw manage kore jawa 'manage do' to manage continuously
	Compound verbs	V [Eng] + *kɔra/**h**ɔwa* V [Eng] + *kɔra/**h**ɔwa* + V	deny kɔra 'deny do' to deny decide kore phela 'decide do throw' to decide (completely)

Modern Bengali not only has single-word verbs, which I call simple verbs (following Thompson 2010) such as *khæla* 'play', or *ʂona* 'hear' just like in English, an example of which is provided below in (7). These verbs consist of a single word denoting an action or a state. Apart from simple verbs, there are the complex verb structures, which will be the focus of my discussion in the next section.

(7) ʂe gach theke **poɽ-ech-e** **(simple verb)**
3SG tree from fall-PFV-3P
He/She has fallen from the tree. (Thompson 2010)

5.4.1.1. Conjunct verb structures of modern Bengali (N+ kɔra/hɔwa 'do'/'be' & N+ kɔra/hɔwa 'do'/'be' +V)

Conjunct verb structures are two-part and three-part verbs that occur in addition to simple verbs in modern Bengali. In these structures, nouns (or adjectives) occur alongside helping verbs such as *kɔra* 'do' or *hɔwa* 'be'. The nouns provide the semantic content to the conjunct verb and the helping verbs turn the constituent N+V into a verbal predicate. The helping verb also bears inflection. An example is *cheshta kɔra* 'to try', formed by the noun *cheshta* 'trial' or 'trying' and the verb *kɔra* 'do'. When a conjunct verb has *hɔwa* 'be' after the noun, it is a passive construction an example of which is *nihɔto hɔwa* 'killed be' or 'was killed'. However, in this chapter, I focus only on the structures with the actional helping verb 'do'. Modern Bengali shows many occurrences of these N+V structures in addition to simple verbs. The frequency of occurrences of these verbs in modern Bengali is provided in table 5.3 at the end of this section on modern Bengali complex verbs. It must also be noted that Bengali (modern and older Bengali) does not have any V + *kɔra* 'do' structures. For example, *kine kɔra* 'buy do' (intended meaning 'to buy') would be an impossible structure because *kine* is a verb in Bengali and does not need a helping verb to express the action of buying.

Some conjunct verbs also have a third element, occurring after the noun and the helping verb, forming an N+'do'+V structure. This third element is a verb that alters the aktionsart of the complex verb and in some instances provides additional subtlety of meaning. This verb has been referred to as a *vector verb* (Dasgupta 1977, Ramchand 1990), or it has been treated as part of the class of *light verbs* (Butt 2003), including verbs such as *kɔra* 'do' and *hɔwa* 'be'. In this chapter, I make a distinction between helping verbs and vector verbs because, as I argue in my discussion on bilingual Bengali-English verbs, these two types of verbs often perform different functions and have different properties. I use the term *helping verb* to refer to verbs such as *kɔra* 'do' and

hɔwa ‘be’ that turn N+V constituents into verbal predicates and reserve the term *vector verb* (following Dasgupta and Ramchand) for verbs such as: *dewa* ‘give’*, newa* ‘take’*, b*ɔʂa *‘sit’, otha* ‘rise’*, tola* ‘lift’*, a*ʂ*a* ‘come’ etc. that affect the aktionsart of the complex verb and provide further semantic nuances to it. Examples of two-part and three-part conjunct verbs from modern Bengali are shown below.

8(a) robi lok-er ɔnek **upokar** **kor-ech-e** **(N+‘do’)**
Robi people-GEN much help do-PFV-3P
Robi has helped people a lot.

(b) bepar-ta na jene **jiggesh** **kor-e** **boʂ-ech-i** **(N+‘do’+V)**
thing-DEF NEG knowing question do-PFV.PTCP sit-PFV-1P
Without knowing, (I) suddenly (and unintentionally) asked him/her about the thing.

The conjunct verbs in each of the examples (8a) and (b) function as a single predicational element including the noun *upokar* ‘help’ or *jiggesh* ‘question’ providing the main semantic content of the construction. The *kɔra* ‘do’ verb, functioning as a verbalizer, carries tense, aspect, and person inflections as shown in (8a). The third verb or vector verb in example (8b) alters the aktionsart of the construction and adds subtle semantic content. For example in this sentence it indicates that the action of *asking* was done suddenly and unintentionally. Unlike in example (8a), where *kɔra* ‘do’ carries inflections, in (8b) it is in perfective participle form and the vector verb carries inflections of tense, person and aspect. It must be noted that not every noun adjacent to a verb forms a conjunct verb. In regular object-verb constructions, the nouns are only arguments of the verbs (8c), and together do not form a complex verb construction (for relevant distinctions see Bhattacharyya et. al. 2006: 347-350).

(c) uni amar theke **ʂahajjo** **ni-ech-e-n** **(O V)**
3SG.HON 1SG.GEN from help take-PFV-3P-HON
He/She has taken help from me.

5.4.1.2. Compound verb structures of modern Bengali (V+V)

Modern Bengali speech also has compound verbs as shown in (9a) and (b). In these structures, two verbs adjacent to each other in a (V+V) structure express a single event. The first verb (or the main verb) is in the non-finite perfective participle form and the second verb is a light verb that carries tense, aspect, and person inflection (Dasgupta 1977, Ramchand 1990, Basu 2010, Paul 2003, Abbi & Gopalakrishnan 1991, Butt 2003). The first verb is often referred to as the *pole* and the second verb as the *vector verb* (Dasgupta 1977, Ramchand 1990). The example below (9a) illustrates a compound verb. These verbs differ from simple verbs in terms of subtle semantic information and aksionsart or aspectual information. In (9a), the vector verb indicates that the action of falling has been completed. However, in comparison, the simple verb *poɽe* 'fall-PFV.PTCP' in (9b) despite carrying an inflection expressing perfectivity, is not able to clearly convey a sense of telicity.

9(a) ʂe gach theke **poɽ-e** **gæ-ch-e** **(V+V)**
3SG tree from fall-PFV.PTCP go-PFV-3P
He has fallen from the tree (completely). (Thompson 2010)

(b) ʂe gach theke **poɽ-ech-e** **(Simple verb)**
3SG tree from fall-PFV-3P
He has fallen from the tree.

Studies on the compound verbs of Bengali have considered 12 to 16 verbs that can occur as vector verbs (e.g. Paul 2003, Thompson 2010, Basu 2010). Table 5.2 below provides a list of all of the verbs that have been classified as vector verbs.

Table 5.2: Bengali lists of vector verbs

Verbs from different analyses	Paul's (2003) list	Thompson's (2010) list	Basu's (2010) list
jawa 'go'	x	x	x
cɔla 'move'	x	x	x
aʂa 'come'	x	x	x
otha 'rise'	x	x	x
tola 'lift'	x	x	x
bɔʂa 'sit'	x	x	x
phæla 'throw'	x	x	x
dewa 'give'	x	x	x
newa 'take'	x	x	x
pɔɽa 'fall'	x	x	x
rakha 'keep'	x	x	
daɽano 'stand'	x	x	
bæɽano 'roam'	x		x
mɔra 'die'	x		x
ana 'bring'	x		
chara 'leave'	x		

There have been several proposals about the status of the vectors in compound verbs. Hook (1974:94-97) and others have argued that vector verbs of Indo-Aryan languages are semantically empty and have undergone a process of *grammaticalization*. I use the term *grammaticalization* to refer to the process by which lexical items such as nouns or verbs change into items that serve only a grammatical function in a language, often through semantic bleaching. Others such as Abbi & Gopalakrishnan (1991) have argued that vector verbs are multi-functional and have functions such as *aspectual*, *adverbial,* and *attitudinal*. For example, a vector such as *gæche* 'go-PFV' performs an aspectual function where it only affects aktionsart, whereas some vectors such as *boʂeche* 'sit' have an adverbial function that mark suddenness. Basu (2010: 44) suggests that vector verbs have undergone different stages of grammaticalization and that some vector verbs such as *gæche* 'go-PFV' are more semantically

bleached possibly indicating a higher degree of grammaticalization than other vector verbs such as *bæɽacche* 'roam' that are less vacuous semantically.

Vector verbs of Bengali can be best analyzed in terms of the grammaticalization parameters proposed by Heine and Kuteva (2003). Firstly, these verbs of Bengali only show signs of desemanticization (loss of meaning). For instance, when *gæche* 'go-PFV' occurs as a vector verb (9a), it does not add any meaning of *going* to the compound verb. It only acts as a grammatical item that carries inflections and affects the aktionsart of the construction. Therefore, in that context, it has undergone semantic loss. However, when it occurs as a simple verb in the language, it retains its full meaning. In sum, this is in line with factors that are typical of the process of grammaticalization (Hopper and Traugott 1993).

However, in terms of the other mechanisms of grammaticalization such as *decategorization* (loss of structural properties such as the ability to bear inflections, case etc.) and *erosion* (loss of phonetic material), the vector verbs of Bengali do not show either type of loss, but only semantic loss (see also Heine and Kuteva 2007 for additional parameters of grammaticalization). Some vectors such as *gæche* 'go-PFV' and *pheleche* 'throw-PFV' display a complete loss of semantic content, while most vectors such as *dewa* 'give'*, newa* 'take'*,* bɔʂa *'sit', otha* 'rise'*, tola* 'lift'*, aʂa* 'come' do not display such a loss. Instead, there is a reduction of semantic content. Based on these properties, it can be said that vector verbs as a whole are on a grammaticalization path but do not display every property of grammaticalization (see also Basu 2010: 44). It is not the case that every grammaticalized item displays all the properties of grammaticalization (e.g. desemanticization, decategorization, erosion, as proposed in Heine and Kuteva 2007). An item can display just one of these semantic parameters and still be viewed as grammaticalized.

Some studies have argued that in the compound verbs of modern Bengali the poles and vectors display a semantically compatible requirement (Paul 2003: 8). Vectors such as the ones in table (5.2), when they occur in a compound, add a distinct yet subtle meaning to the compound verb. That is why they pair only with those poles with which they have some semantic matching. For instance, the verb *nece* 'dance-PFV.PTCP' is a motion verb. So in (10a) it pairs with a vector *bæɽacche* 'roaming', which is also a motion verb. If the motion verb *bæɽacche* 'roaming' pairs with a stative verb like *ʂue* 'lie' the pairing results in an ungrammatical combination (9b).

10(a) me-ta ʂara din **nec-e** **bæɽa-cch-e** **(V+V)**
girl-DEF whole day dance-PFV.PTCP roam-PROG-3P
The girl is dancing around the whole day. (Paul 2003: 8)

(b) *meta ʂara din **ʂu-e** **bæɽa-cch-e** **(V+V)**
Girl-DEF whole day lie-PFV.PTCP roam-PROG-3P
Intended: The girl is lying around the whole day. (Paul 2003: 8)

It has also been argued that vectors such as *gæche* 'go-PFV' do not seem to have a strict semantic compatibility requirement, in that they occur productively with all kinds of verbs (unlike other vector verbs such as 'rise' or 'lift'). These verbs do not add any distinct lexical meaning to the compound, possibly because of semantic bleaching; their only function as shown in (9a) is to modify aspectual information (Paul 2003: 8-9, Basu 2010: 44). An example of this is *bujhe gæche* 'has understood', *bujhe* is 'understand-PFV.PTCP' and *gæche* is 'go-PFV'. Although *gæche* 'go-PFV' is a motion verb, *bujhe* 'understand-PFV.PTCP' is not, and in spite of that, this pairing is grammatical. However, the feature of semantic compatibility needs to be examined in more detail to determine other properties of these verbs that lead to the semantic match. I leave this matter to future research.

5.4.1.3. Bilingual complex verb structures of modern Bengali {N[Eng] + 'do'/'be' + (V); V[Eng] + 'do'/'be' + (V)}

Another type of complex verb in modern Bengali (both bilingual and monolingual speech) is the bilingual complex verbs (which are completely absent in the nineteenth-century Bengali texts, as I discuss below). These bilingual verbs of modern Bengali combine English nominal or verbal lexical elements in the bare form with helping verbs from Bengali. These verbs in modern Bengali consist of either two-part or three-part structures and the first lexical element is always from English, combining with helping verbs such as *kɔra* 'do' or *hɔwa* 'be' from Bengali. Sometimes vector verbs may also occur after the helping verbs. The examples below show the range of lexical elements from English that can occur in these structures. Example (11a) illustrates English nouns, (11b) English verbal nouns, (11c) English phrasal verbs and (11d) an English simple verb. All of these nominal and lexical elements occur beside the *kɔra* 'do' verb from Bengali. I use the term verbal noun to refer to elements such as *shopping* and *skipping*, and phrasal verbs to refer to elements such as *build up, work out*, *hang up*, as used by Romaine (1986).

11(a) ʂei moment-e **operation** **kor-l-o** **(N[Eng]+'do')**
that moment-LOC operation do-PST-3P
In that moment, (he) did the operation (surgery).

(b) ʈui o-r ʂoŋge **shopping** **kor-b-i**? **(Verbal noun[Eng]+'do')**
2SG 3SG-GEN with shopping do-FUT-2P
You will be shopping with her?

(c) cinema-ʈa je bhabe **build up kor-echil-o** **(Phrasal V[Eng]+'do')**
cinema-DEF COMP way build up do-PST.PFV-3P
The way they had built up the movie…….

(d) ora bol-ech-e application-ʈa **renew** **kor-b-e** **(V[Eng]+'do')**
3PL say-PFV-3P application-DEF renew do-FUT-3P
They said (they) would renew the application.

The Bengali elements that occur in the bilingual verbs are the helping verb *kɔra* 'do', in active structures, and *hɔwa* 'be', in stative structures, in line with Romaine's (1986) observation of Punjabi-English bilingual speech. Each complex verb together expresses one single event, the English nominal or verbal element, which is in the bare form, provides lexical meaning and the helping verbs bear tense, person, and aspect inflections.

These bilingual complex verbs occur in both the modern monolingual and the bilingual corpus data that I examined. As we would expect, they occur much more frequently in the bilingual corpus than the monolingual one (see table 5.3 below). One might find it surprising that these verbs occur even in the speech of monolingual Bengali speakers, given their lack of broad knowledge of English. However, if we consider the way in which the speech communities of these two groups are organized, it is not surprising. The monolingual speakers from whom I collected the speech data interact with bilingual speakers on a daily basis (such as in the workplace or marketplace) and their speech has most likely been affected by that of bilingual speakers. So, although the monolingual speakers do not have extensive knowledge of English, these constructions are making inroads into their speech via interaction with the bilingual speakers. Also, the majority of the bilingual verbs in the monolingual speakers' speech consists of combinations such as *phone kɔra* 'to phone/call' or *post kɔra* 'to post' in which the noun has been borrowed from English into modern Bengali. Table (5.3) provides the number of occurrences of complex verbs (both bilingual and monolingual ones) in the modern Bengali corpora. The last column in the table corresponds to N/V[Eng]+'do' constituents, in which the N/V are lexical items in English that cannot be clearly identified as either a noun or a verb. For example, *shift kɔra* 'to shift' where *shift* can be analyzed either as a noun or a verb. I have included both verbal nouns and phrasal verbs in this category.

Table 5.3: Number of occurrences of complex verbs in modern Bengali

Speech Type	N[Beng]+'do'	N[Eng]+'do'	V[Eng]+'do'	N/V[Eng]+ 'do'
Monolingual Bengali speech	358	23	12	39
Bilingual Bengali speech	284	70	342	206

I will analyze these structures and comment on the frequency of occurrence of some of them in section 5.4.3 below. But before that, I turn to the verbal system of the nineteenth-century Bengali texts in order to show ways in which they differ from the complex verbs of modern Bengali.

5.4.2. The verbal system of nineteenth-century Bengali

The examination of verbs in the older Bengali texts reveals the occurrence of both simple verbs and complex verbs, with the complex verbs consisting of *conjuncts* (Noun + *kɔra/ hɔwa* 'do'/'be'; Noun+ *kɔra/ hɔwa* 'do'/'be'+ Verb) and *compounds* (Verb + Verb). These structures are similar to the monolingual complex verbs of modern Bengali discussed in section (5.4.1.1). The properties of these verbs are outlined in table (5.4) below.

Table 5.4: Verbal system of nineteenth-century Bengali

Bengali Verbs	Sub types	Components of verbs	Examples of each type Examples Gloss Translation
Simple Verbs		1 verb	pɔɽa 'read' to read dækha 'see' to see
Complex Verbs	Conjunct verbs	N+ *kɔra*/*hɔwa* N+ *kɔra*/*hɔwa* +V	snan kɔra 'bath do' to shower kɔed kore 'prison do take' to imprison dewa for someone else
	Compound verbs	V + V	ʂue pɔɽa 'sleep fall' to fall asleep dekhe newa 'see take' to see for oneself

5.4.2.1. Conjunct verb structures of nineteenth-century Bengali (N+ 'do'/'be'; N+'do'/'be' +V)

The conjunct verbs of older Bengali are identical in structure and semantics to the monolingual conjunct verbs of the modern Bengali corpora. These structures have nouns occurring alongside the verbs *kɔra* 'do' and *hɔwa* 'be' from Bengali and in some cases have a vector verb occurring at the end, that is they may have two parts or three parts just like the monolingual verbs of the modern Bengali corpora. Sentences (12a) and (12b) illustrate the two-part and three-part conjunct verbs of nineteenth-century Bengali**.**

12(a) apni baɽi ʤa-n ami prottoho pɔtro **prerɔn kor-i-bo** **(N+'do')**
2SG.HON house go-3P.IMP 1SG soon letter sent do-FUT-1P
You go home, I will send the letter soon. (1860 Bengali play)

(b) æksɔto taka agami maʂ-e **porishod kor-ia di-b-o** **(N+'do'+V)**
one hundred rupee next month-LOC repay do-PFV.PTCP give-FUT-1P
I will repay one hundred rupees next month (to someone). (1860 Bengali play)

In (12a), which has no vector verb, the helping verb *koribo* 'do-FUT' bears inflection of tense, person and aspect. In the absence of a vector verb, it is not clear who the intended beneficiary of the action is. The vector verb in (12b) *dibo* 'give-FUT' makes it clear that the subject/speaker will repay the money to someone else and also indicates that the action of repaying will be complete by adding telicity to the predicate. In a two-part conjunct verb, the helping verb bears tense, person, and aspect inflection. In a three-part conjunct this verb is in the perfective participle form, whereas the vector verb bears tense, person, and aspect inflection.

5.4.2.2. Compound verb structures of nineteenth-century Bengali (V + V)

There is also evidence of monolingual V+V or compound verb structures in nineteenth-century Bengali texts. In these structures, two verbs are also adjacent to each other and express a single event. The main verb occurs in the non-finite perfective participle form and the second

verb (light verb) carries tense, aspect, and person inflection. Example (13b) below shows a complex verb, in contrast to a simple verb in (13a).

13(a) ami nijer chokh-e **dækh-l-am** **(simple verb)**
1SG my own eyes-INST see-PST-1P
I saw (it) with my own eyes. (1854 Bengali play)

(b) etoda poribar na khat-i pej-e **mor-e** **ʤa-b-e** **(V+V)**
this big family NEG eat-IMPFV get-PFV.PTCP die-PFV.PTCP go-FUT-3P
This big family (of mine) will die (completely) without food. (1860 Bengali play)

In (13a) the simple verb *dækhlam* 'saw-PST' carries inflections, while (13b) has a compound verb, in which the pole verb *more* 'die' is in the perfective participle form, and the vector verb *ʤabe* 'go-FUT' carries inflections. Without the vector verb in (13a), the action of seeing does not have a telic interpretation despite the past marker on the verb. In (13b), the vector verb undergoes semantic bleaching and loses the meaning of 'go' in order to convey telicity of the predicate. Also, in (13b) perfective aspect is marked on the pole; although the pole is always in perfective aspect, the vector can have any aspect marked on it such as progressive, completive or perfective.

Example (14) below shows a vector that not only affects the aktionsart of the construction, but also adds subtle semantic content to the construction, like example (6b) from modern Bengali. Here, the vector verb *neben* inherently means 'take-FUT' but when it occurs as a vector, it indicates that the action is self-directed and for the sole benefit of the subject. Therefore the vector verb *ʤabe* 'go-FUT' is slightly different from *neben* 'take-FUT' because the latter adds meaning to the construction that the former doesn't (see also Basu 2010).

(14) choto ʂaheb æmon maal pel-e to **lup-e** **ne-b-e-n** **(V+V)**
younger Sahib like goods get-IMPFV EMPH grab-PFV.PTCP take-FUT-3P-HON
If the younger Sahib gets goods like these, he will grab it (completely for himself).
(1860 Bengali play)

The verbs of the nineteenth-century texts show similar properties to the modern Bengali ones in that they can occur independently as simple verbs, pole verbs or vector verbs, and when they occur as vectors, they often lose their inherent meaning and add a slightly different meaning to the structure. Examples from the historical texts include *diache* 'give-PFV', which shows benefaction, *niache* 'take-PFV' which indicates doing something for oneself, *boṣiache* 'sit-PFV' adding suddenness, *pheliache* 'throw-PFV' indicating completion.

However, the crucial way in which the nineteenth-century verbs differ from modern Bengali is in the bilingual verbs. The older texts show no evidence of bilingual (N**[Eng]**+'do') or (V**[Eng]**+'do') structures, suggesting lack of contact effect from English in these structures in the nineteenth century[20]. As a result, the modern Bengali bilingual verbs constitute changes, which I address in the next section. Table (5.5) below provides the number of occurrences of complex verbs in the older texts and the modern Bengali corpora.

Table 5.5: Occurrences of complex verbs in older Bengali texts and modern Bengali

Language type	N[Beng]+'do'	N[Eng]+'do'	V[Eng]+ 'do'	N/V[Eng]+ 'do'
19th c. Bengali texts	322	0	0	0
Modern Monolingual Bengali speech	358	23	12	39
Modern Bilingual Bengali speech	284	70	342	206

In this section, I showed that two types of verbs occur in the older Bengali texts, simple verbs and complex verbs. I also discussed the function of each of the verbs in these structures and how conjunct verbs 'N+V' differ from compound verbs 'V+V'. However, unlike the modern

[20] I must reiterate that no sign of bilingual speech in the older texts does not mean that bilingual speech was not emerging among the highly educated Bengali speakers of that time. My observation of English words in the texts is an indication that English was already making inroads.

Bengali corpora, the complex verbs of the nineteenth-century texts show no occurrences of bilingual verbs and therefore indicate no evidence of English contact effects in that domain. In the next section (5.4.3), I will focus specifically on the bilingual verbs of modern Bengali and compare them to the bilingual verbs of other language pairs, addressing their analysis as changes emerging from contact effects of English.

5.4.3. Analysis of bilingual complex verbs in modern Bengali

In this section, I first review literature on bilingual verb structures in general and then propose an analysis considering the syntactic and semantic properties of these verbs in Bengali-English speech, in line with some previous analyses of similar structures. I restrict the discussion of bilingual verbs to complex verbs with 'do' helping verb only. Some of the changes I discuss in this section are applicable to complex verbs with 'be' helping verbs also.

Apart from Bengali, bilingual complex verbs also occur frequently in CS between other Indian languages and English. Romaine (1986) reports the occurrence of these verbs in bilingual Punjabi-English speech. In her data, the complex verbs consist of English nouns, verbs, verbal nouns (such as *lobbying*) or phrasal verbs (such as *pick up*) alongside Punjabi helping verbs such as *karna* 'do' and *hona* 'be'. Romaine (using the term *operators* for these verbs) shows that 'do' and 'be' distinguish the stativity of the structures, in that 'do' occurs in actional constructions and 'be' in stative ones. The operators modify the English nominal or verbal elements and bear inflections. Romaine uses the term *compound verb* to refer to these structures and says that there is a tendency for English verbs to occur more frequently than English nouns in such structures.

Annamalai (1978), in his analysis on Tamil-English bilingual speech, distinguishes between structures such as *try pannu* 'try' (V+'do') and *business pannu* 'do business' (N+'do'). He argues that bilingual Tamil-English V+'do' structures are different from bilingual N+'do'

structures because accusative case can be optionally added to *business*, but not to *try*, and so he argues that *business* is a noun and not a verb. In addition, dummy verb (Annamalai uses this term for helping verbs) 'do' is added after a main verb as a carrier of inflection. In a later study Annamalai (1989) furthers a hypothesis about balanced bilinguals and imbalanced bilinguals (people who are stronger in Tamil than in English) in their use of N+'do' and V+'do' structures. He says that groups of bilingual speakers can be distinguished on the basis of the types of mixed compounds they use. This is because imbalanced bilinguals conform to the N+'do' structures of native Tamil by using structures such as 'reservation do' whereas balanced bilinguals would tend to use structures such as 'reserve do', which are changes in bilingual speech, because in monolingual Tamil a V+'do' construction cannot occur.

Muysken (2000) also distinguishes bilingual N+'do' structures from bilingual V+'do' structures in many language pairs across the world. He argues that the insertion of foreign or native nouns alongside helping verbs and the use of foreign verbs alongside helping verbs in the recipient language constitute different processes. In the noun insertion strategy, there is a process of insertion of the foreign noun in the position of a native noun. However, the occurrence of a foreign lexical verb beside the helping verb from a matrix language (usually considered to be the dominant language of the bilingual speaker) is a process of adjunction of the foreign verb to a matrix language verb 'do'. He argues that this process of verb adjunction is an example of alternational code-switching between the lexical verb from the embedded language (usually the non-dominant language of a bilingual) and the helping verb from the matrix language. The foreign verbs in the complex verb structures are not borrowings, because they are not phonologically adapted to the recipient language. They also happen to be quite productive in the recipient language (Muysken 2000:193-206).

It has also been postulated that some bilingual complex verbs emerge as a result of the grammaticalization of the helping verbs. That is, these constructions initially start as N+'do' constructions but later go on to become V+'do' constructions (Muysken 2000: 213, Annamalai 1989: 51). In his investigation of three generations of Turkish immigrants in the Netherlands, Backus (1996) found that more proficient bilinguals of Turkish and Dutch had more V+'do' constructions than less proficient bilinguals or mostly monolingual speakers, who had more N+'do' constructions (similar to Annamalai's observation of Tamil-English bilinguals). The first generation of Turkish immigrants that Backus studied consisted of people who were dominant in Turkish and knew minimal Dutch. These people immigrated to the Netherlands as adults. The second generation of Turkish immigrants were born in Turkey and immigrated to the Netherlands at a very young age. The third generation of immigrants consisted of people who were born in the Netherlands to Turkish immigrant parents. The second and third generation speakers were bilingual in Turkish and Dutch. When these people spoke Turkish, Backus (1996) found that the third generation speakers had mostly V+'do' constructions in their speech. The root verbs were from Dutch and the 'do' verbs were from Turkish. Backus (1996) found that the first generation of Turkish speakers used mostly N+'do' and the second generation of speakers had a mix of both N+'do' and V+'do' constructions in their Turkish. This finding lends support to the hypothesis that the V+'do' constructions derive from their nominal counterparts.

There have also been opposing claims about the status of the foreign verbal element in the complex verb structures. Scholars such as Moravcsik (1978) have argued that if a language borrows a verbal element from another language, then it will be borrowed in the form of a nominal element. According to Moravcsik, verbs cannot be borrowed in the form of verbs, and a borrowed verbal element will always behave like a nominal element in the recipient language,

since they require a helping verb to appear beside them, which then verbalizes the construction.[21] More recently, Wichmann and Wohlgemuth (2005) and Wohlgemuth (2009) have argued that verbs are often borrowed as non-verbs in recipient languages, and not necessarily as nouns as Moravscik has claimed. They argue that the foreign elements that occur in these bilingual structures can either be nouns or verbs (that is, they are identical in form to the nouns and verbs of the foreign language) but the recipient language does not specify their part of speech (Wichmann and Wohlgemuth 2005: 2-5, Wohlgemuth 2009: 104-111).

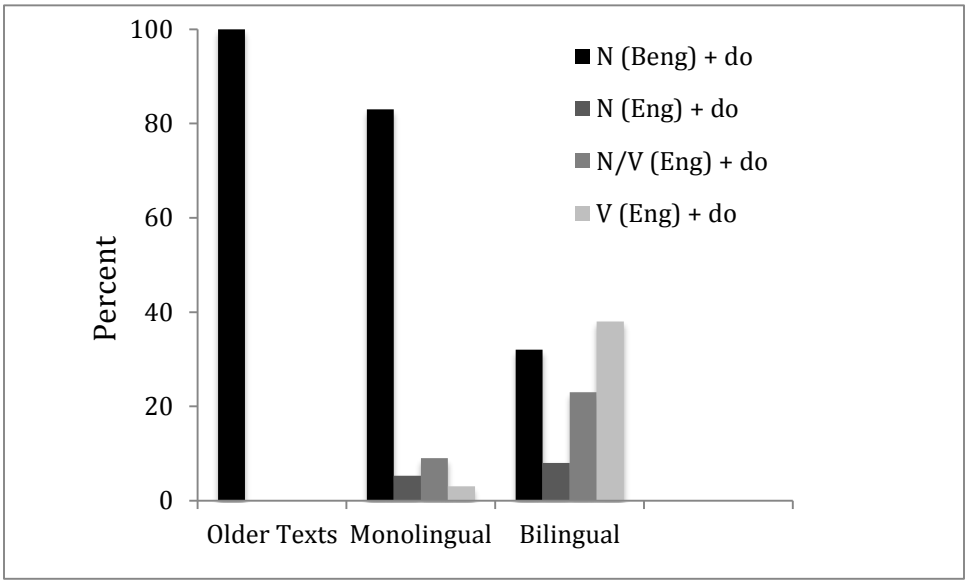
The literature discussed above shows two different strands in the analysis of bilingual verbs. While Muysken (2000), Romaine (1986), Backus (1996) and Annamalai (1989) distinguish between N+'do' and V+'do' constructions structurally, for others, these two are similar structures. Below I try to analyze the bilingual verbs in modern Bengali as a domain showing influence from English. I partially draw from the analysis of Muysken (2000), Romaine (1986) and Annamalai (1989) in my explanations of the bilingual verbs of Bengali.

What some of this literature has highlighted is a structural distinction between an N+'do' construction and a V+'do' construction in bilingual speech. This is relevant for Bengali-English bilingual verbs also because they are parallel, as I will discuss below, to the bilingual verbs in Punjabi-English and Tamil-English speech and to other instances discussed by Muysken (2000) and Backus (1996). In line with these observations for other language pairs, Bengali-English bilingual speech shows a higher frequency of occurrence of English verbs than English nouns (as also shown by Annamalai 1989 in case of balanced and imbalanced bilinguals of Tamil and English). Figure 5.1 below provides the frequency of occurrence of these structures in the

[21] This generalization is surprising, given that in some languages verbs can be borrowed as verbs, such as French Quebecois verbs such as *tier* 'to tie (one's shoe laces)' + French infinitival suffix *–er*. It's usually presented as a borrowing from English due to phonological adaptation, but it is a case of a verb being borrowed as a verb (Marlyse Baptista, p.c. 2015).

monolingual and bilingual corpora in relation to the older texts. As seen from the figure, the bilingual corpus has many more occurrences of bilingual V+'do' structures than the monolingual corpus (bilingual 38 % vs. monolingual 2.8%). Based on the results of a chi-squared test, it was found that this difference is very significant statistically, X^2 (1, N = 447) = 46.481, $p < .001$. The difference between the use of V+'do' and N+'do' in the speech of bilingual speakers alone is also quite large (V+'do' 38% vs. N+'do' 7.7%).

Figure 5.1: Frequency of complex verbs (with 'do' helping verb) in 19th c. texts, modern monolingual and bilingual corpora



The use of complex verbs (X+do structures) in the bilingual corpus has the following distribution: N[Beng]+'do' 32%; N[Eng]+'do' 7.7%; V[Eng]+'do' 38%; N/V [Eng]+'do' 22.8%, showing that almost 70% of the complex verbs involving X+do structures in bilingual speech consist of lexical elements from English. This indicates a clear case of a contact effect in this domain, where English lexical elements (nominal and verbal) are allowed to freely occur alongside the helping verbs of Bengali, especially in the speech of the bilinguals. In comparison to the nineteenth-century texts where there was no case of bilingual X+do structure, in modern

bilingual speech we find that 70% of the X+do structures include English elements. In addition, in modern monolingual Bengali, about 17% of the X+do structures are bilingual.

Based on the studies by Muysken (2000), Backus (1996), Annamalai and Romaine (1996), one could also argue for another dimension to the morphosyntactic change in the X+do verbal structures, comparing modern Bengali to the older texts. This change lies in the occurrence of the 'V[Eng]+'do' structure; in native Bengali (as represented in the nineteenth-century texts), only Bengali nouns (and adjectives) can grammatically occur in complex verbs before the 'do' verb, while in modern Bengali different English elements including nouns, verbs, verbal nouns and phrasal verbs are allowed as part of complex verbs. I argue that the English verbs can now combine with Bengali helping verbs as a result of a structural reanalysis in modern Bengali speech. This is because Bengali furnishes the X+do structure, but the native X+do construction only licenses a noun in the place of X, but in contact with English, the co-occurrence of English verbs with Bengali 'do' in V+'do' is not only licensed but also far exceeds the occurrence of English nouns in N+'do' complex verbs (as evidenced in bilingual speech).

Muysken (2000) has argued that in languages that have a noun + do construction monolingually, the 'do' verb usually selects[22] the elements that occur before them. This is true for languages such as Bengali or Hindi where only nouns and adjectives can occur before a helping verb corresponding to 'do'. For Muysken, this selectional requirement is a property of the government relation where the governing element (in this case the helping verb) provides a slot for the insertion of only certain elements. Whereas selectional/subcategorization requirements of heads can be encoded without reference to government, in the case of monolingual Bengali (as represented in the nineteenth-century texts), the helping verb 'do' can

[22] In syntactic terms, the selectional requirement stated by Muysken has been standardly treated as (syntactic) subcategorization.

be argued to select/subcategorize for a noun or adjective, but not for a lexical verb, blocking a V+'do' possibility in nineteenth-century Bengali. The restructuring in modern Bengali would have expanded those possibilities, by adding verbs to the subcategorization options of the helping verb 'do'.

However, scholars such as Moravcsik (1975, 1978) have argued that in the process of verb borrowings, the recipient language only selects nouns and therefore there must be a government relationship blocking the foreign verb and the recipient language *do* verb. However, in bilingual Bengali-English, other dependent lexical elements can occur before the 'do' verb, as shown above in (11a), (11b), (11c) and (11d). In these examples, both English nominal and verbal elements can be selected/subcategorized by the helping verb *kɔra* 'do'. Therefore, if *kɔra* 'do' only selected a noun in the case of modern Bengali, we would not have found these different nominal and verbal elements in modern bilingual and also monolingual speech.

Muysken argues that in bilingual speech, the operator or helping verb such as 'do' does not select its complements (the foreign elements) that can occur before it, suggesting that there must be only a relation of adjunction and modification (instead of complementation) between the foreign element and the 'do' verb. Muysken provides examples of Sarnami-Dutch (a bilingual variety used by Surinamese immigrants in the Netherlands) bilingual verbs where all different kinds of Dutch elements (such as Dutch verbs, prepositional particles, nouns, verb phrases and other phrases) can occur before Sarnami operators, suggesting that a government relation between the Dutch elements and the Sarnami operators does not hold. For the government relation to exist, the operator must select its complement (which is why the operator always selects nouns in native cases). But in bilingual V+'do' structures, Muysken argues for a modification relation between the non-native verb and the helping verb 'do'. This is because, if a

noun occurs before a 'do' verb, it receives a thematic role from it, whereas a non-native verb modifies the 'do' verb. Muysken argues that the interpretation resulting from this modification relation is that for V+'do' structures, one is 'doing an action', whereas for N+'do' structures, the interpretation is of 'performing an action, that of X' (Muysken 2000: 201-202)

Without resorting to this specific proposal by Muysken, which has several unclear points, I would like to propose a hypothesis about why more English verbs than nouns are used in the bilingual X+do structures. Intuitively (based on my judgments as a native speaker of bilingual Bengali-English) a distinctive, yet subtle interpretation seems plausible for Bengali-English V+'do' complex verbs, in examples such as *advertise kɔra* 'to advertise', as opposed to N+'do' complex verbs such as *advertisement kɔra* 'doing advertising'. In the former case, the action of *advertising* is being done, whereas the latter seems like a more indirect, periphrastic way of expressing the action of *advertising*. Semantically, the latter structure indicates that one is doing an action, which happens to be advertising. Extending this interpretation to examples with other verbs or nouns, it seems that in Bengali-English bilingual verb structures, some English nouns are incapable of contributing the same semantic information to the complex verb as English verbs are able to. Three examples are provided below which attempt to illustrate this subtle difference. In (15a), the monolingual conjunct verb is *ɔsshikar kɔra*, 'denial do' which means *denying* in English. If, however, a bilingual speaker wants to use an English form in the construction, then *denial kɔra* might not be the preferred choice (15b), because intuitively, it seems like a more complex, periphrastic way of expressing the act of denying something. In (15b), the meaning of the complex verb corresponds to 'doing an act of denial'. The structure, *deny kɔra* 'to deny' (15c) seems to be a more straightforward way of expressing the act of denying.

15 (a) chatro-ta aarop-ta-ke **ɔsshikar kor-ech-e** (**N[Beng]+'do')**
Student-DEF accusation-DEF-ACC denial do-PFV-3P
The student has denied the accusation.

(b) chatro-ta accusation-ta-ke **denial kor-ech-e** **(N[Eng]+'do')**
Student-DEF accusation-DEF-ACC deny do-PFV-3P
The student has denied the accusation.

(c) chatro-ta accusation-ta-ke **deny kor-ech-e** **(V[Eng]+'do')**
Student-DEF accusation-DEF-ACC deny do-PFV-3P
The student has denied the accusation.

Other examples of English nouns and verbs that have these subtle meaning differences include *permit kɔra / permission kɔra*; *react kɔra /reaction kɔra*; *decide kɔra /decision kɔra*; *object kɔra /objection kɔra* among several others.

This brings us to one of the possible reasons why we find more English verbs in the X+do structures of bilingual speech. Firstly, given that the bilingual speakers have more English verbs than English nouns in the bilingual X+do structures (see figure 5.1), there may be something special about verbs that nouns do not appear to contribute in this specific case. Secondly, because bilingual speakers have the option of choosing between two grammatical structures that have subtle meaning differences, it may be the case that they prefer structures that appear to be more semantically direct than others. As I discussed regarding (15), some bilingual combinations involving English nouns do not have the same referential meaning as combinations involving English verbs. This connects with what Annamalai (1989) had proposed about balanced bilinguals using more English verbs than nouns in Tamil-English bilingual speech. Because balanced bilinguals have more access to different English constructions, they might be using those that are more semantically applicable, while the monolingual Bengali speakers having little knowledge of English use the Bengali nominal counterparts of English verbs.

However, if bilingual N+‘do’ is structurally and semantically different from bilingual V+‘do’, one could ask why these bilingual speakers don't use English verbs directly with Bengali inflections (without resorting to the helping verb construction). At present the X+do structure is the only way in which Bengali adds new English verbs to the language, because when English verbs get used, they appear only in that position. So, if the bilingual situation had further intensified and became more pervasive, we would perhaps have verbal structures with English verb roots combining directly with Bengali inflections (see Wichmann and Wohlgemuth 2005 and Wohlgemuth 2009 for more discussion on verb formation strategies and their relation to intensity of contact).

Could one argue instead that, because the English verbs only occur in the X slot in the X+do construction, Bengali treats these elements as nouns? If this were the case, despite the fact that there may be subtle semantic differences between bilingual N+‘do’ and V+‘do’, Bengali might still be treating the English verb forms as nominal elements, which is why there is a need to integrate them as nouns into the X+do structure. The differential use of English nouns and verbs could then be restricted to the hypothesis I discussed above about the subtle semantic differences between bilingual N+‘do’ and V+‘do’ structures. However, this hypothesis would predict that there should be no structural differences between the N+‘do’ and V+‘do’. I leave for future research a more extensive investigation of the structural/grammatical differences between the V+‘do’ and N+‘do’ complex verbs in modern Bengali.

In this section, I discussed the bilingual complex verb structures in modern Bengali and argued that they occur in modern Bengali due to contact effects from English, and Bengali supplies the X+do frame for the occurrence of these elements. I also attempted to provide an analysis of these verbs taking into account syntactic and semantic considerations drawing

partially from previous accounts of bilingual verbs. Below I turn to the three-part bilingual verbs, which had not been reported in the literature on bilingual verbs till now.

5.4.4. Three-part bilingual complex verbs
(N/V+ 'do' +V)

The literature on bilingual complex verbs has reported the occurrence of only two verbs. However, some bilingual verbs in modern Bengali have a three-part structure that consists of a bare English nominal or verbal element, a Bengali helping verb and a Bengali vector verb. This is similar in structure to the monolingual conjuncts with the exception of the English noun or verb before Bengali 'do'. This structure differs from two-part complex verbs with respect to meaning and aspectual nuances. The examples below show a two-part (16a) and a three-part complex verb (16b).

16(a) o Ritayon-ra **shift** **kor-ech-e** **(V[Eng]+'do')**
oh Ritayon-PL shift do-PFV-3P
Oh Ritayon and folks have shifted.

(b) o Ritayon-ra **shift** **kor-e** **gæ-ch-e** **(V[Eng]+'do'+V)**
oh Ritayon-PL shift do-PFV.PTCP go-PFV-3P
Oh Ritayon and folks have shifted (completely).

The three elements in the three-part complex verb together express a single event (16b). In these three-part bilingual verbs, the helping verbs are in perfective participle form and the vector verbs carry tense, person and aspect inflection, affect the construction's aktionsart and often provide additional semantic information. Vectors such as *gæche* 'go-PFV' and *pheleche* 'throw-PFV', which have lost their lexical meanings completely, can only affect the construction's aktionsart. It has been hypothesized that they are on a more advanced path towards grammaticalization and therefore perform only a grammatical function in the construction. For instance, in (16b) the

vector *gæche* ‘go-PFV’ does not add any extra meaning of going to the construction but affects the aktionsart (aspectual properties) by adding telicity to the predicate.[23] Other vectors such as *dewa* ‘give’ and *newa* ‘take’ also seem to have partially lost their lexical meanings but they add an additional meaning that is usually different from their own inherent meanings, in addition to contributing to the aktionsart of the complex predicate. For instance in (17) the vector *dieche* ‘give-PFV’ introduces a benefactive meaning, whereas by itself, *dieche* ‘give-PFV’ indicates handing an object to someone, but as a vector it adds a benefactive meaning. In this respect, the vector *dieche* ‘give-PFV’ affects the argument structure of the sentence since it implicitly introduces another argument to the predicate. In addition, it also affects the aktionsart of the construction by adding telicity to the predicate.

(17) ora park-ta **renovate** **kor-e** **die-ch-e** **(V+‘do’+V)**
3PL park-DEF renovate do-PFV.PTCP give-PFV-3P
They have renovated the park (completely for someone else).

Helping verbs such as ‘do’ or ‘be’ are similar to vector verbs in that they can both carry inflection. In two-part complex verbs, (N/V+‘do’) the helping verb ‘do’ carries inflection, in three-part constructions (N/V+‘do’+V), the ‘do’ is in the perfective aspect and other inflections are instead carried by the vector verb. However, the role of the helping verb ‘do’ is just to turn the N+‘do’ constituent into a predicate while vector verbs perform additional functions such as affecting the aktionsart of the construction and sometimes adding subtle semantic information. Bengali helping verbs append to bare nouns and adjectives in monolingual Bengali and to bare nouns and verbs in bilingual Bengali-English. Also, Bengali helping verbs can follow both native and non-native elements while vectors follow only the native verbs that are in perfective

[23] Other kinds of subtle aspectual distinctions also arise for progressive and habitual forms, which is in line with the aspectual subtleties provided by the perfective forms.

participle form. Only the verbs *kɔra* 'do' and *hɔwa* 'be' belong to the class of helping verbs while there are 12 to 16 verbs that occur as vector verbs. Table (5.6) summarizes the relevant similarities and differences between these verbs in three-part complex verbs:

Table 5.6: Similarities and differences between helping verbs and vector verbs

	Helping verbs	Vector verbs
Similarities	Carry inflection for tense, person and aspect	
Differences	Perform a verbalizing function	Alter the aspect and/or semantics of the construction
	Append to bare forms	Append to perfective forms
	Append to nouns, adjectives and verbs	Append only to verbs
	Class of 2, only 'do' and 'be' [24]	Class of 12 to 16 verbs
	Append to both native and non-native elements; native elements are always nouns or adjectives	Append only to native verbs

Property of semantic compatibility

The principle of semantic compatibility that affects poles and vectors in monolingual complex verbs may also be a relevant property to consider in some of these three-part bilingual verb structures. It appears that there should be semantic matching between the Bengali vector verbs and the English nominal or verbal elements. So, on the one hand, with a Bengali vector verb such as *gæche* 'go-PFV' and *pheleche* 'throw-PFV', any English nominal or verbal element can occur. I hypothesize that this is because verbs such *gæche* 'go-PFV' and *pheleche* 'throw-PFV' have lost their own inherent meanings and have become grammaticalized.

[24] In modern Bengali, other helping verbs are possible, such as the ones corresponding to 'take', 'cut', 'give' etc. (Thompson 2010). Pillai (1968) says that 'give' can also occur as a helping verb in bilingual complex verbs. However, if 'give' occurs as a helping verb in a bilingual complex verb, it is not clear if the construction is a complex verb or a regular object verb construction.

(18) O tui already **apply** **kor-e** **phel-echi-s?**
Oh 2SG already apply do-PFV.PTCP throw-PFV-2P
Oh, you have already applied (completely)?

On the other hand, with a Bengali vector such as *boṣeche* 'sit-PFV', the English nominal or verbal elements is more semantically restricted. This is because *boṣeche* 'sit-PFV' is a telic verb, which means that the verb has a specific endpoint. As a vector, the verb *boṣeche* 'sit-PFV' means doing something suddenly. Therefore it cannot pair well with verbs such as *maintain* (19a) and *develop* that do not have definite endpoints, contrary to *delete* (19b).

19(a) *ami amar GPA **maintain** **kor-e** **bos-ech-i**
1SG 1SG.GEN GPA maintain do-PFV.PTCP sit-PFV-1P
I have suddenly maintained my GPA.

(b) are Pritam gaan-ta **delete** **kor-e** **bos-ech-e**
So Pritam song-DEF delete do-PFV.PTCP sit-PFV-3P
So, Pritam has (unintentionally) and suddenly deleted the song.

However, as is the case for the semantic compatibility matching in monolingual verbs, this property in bilingual verbs also needs to be examined in further detail.

Syntactic tests on bilingual complex verbs

Ramchand (1990) used some syntactic tests on Bengali compound verbs (V+V) to determine if the two verbs display syntactic unity. I have applied some of these tests to the three-part complex verbs below, to show that the three parts form a constituent.

Adverb intrusion test

Insertion of adverbs in between the three elements of the bilingual complex verb results in ungrammaticality. In (20a), the adverb is placed before the complex verb. This is the position where adverbs should occur with respect to complex verbs. From this position, adverbs take scope over the entire complex verb and modify all the three verbs of the construction. However,

if adverbs are placed between any of the three elements of the compound verb, the sentences become ungrammatical (20b).

20(a) Professor solution-ta **clearly explain kor-e di-l-o**
Professor solution-DEF clearly explain do-PFV.PTCP give-PST-3P
The professor explained the solution clearly.

(b) Professor solution **explain (*clearly) kor-e (*clearly) di-l-o**
Professor solution explain do-PFV.PTCP give-PST-3P
The professor explained the solution clearly.

Negation test

Negation occurs after the complex verb (21a), taking scope over the whole complex verb. If negation is placed in any other position the sentence becomes ungrammatical (21b). Therefore, it is not possible to negate only part of the complex verb.

21(a) Professor solution-ta **explain kor-e di-l-o na**
Professor solution-DEF explain do-PFV.PTCP give-PST-3P NEG
The professor did not explain the solution.

(b) Professor solution-ta **(*na) explain (*na) kor-e (*na) di-l-o**
Professor solution-DEF NEG explain NEG do-PFV.PTCP NEG give-PST-3P
The professor did not explain the solution.

Coordination test

If there are two English nominal or verbal elements in a sentence then each nominal or verbal element should have its own Bengali helping verb and vector verb. Two English nominal or verbal elements cannot be coordinated with only one Bengali helping verb and vector verb. Therefore, sentence (22b), which has two English elements combined with only one helping verb and vector verb is not a well-formed sentence. Sentence (22a) is well formed because both English verbs *simplify* and *explain* have their own Bengali helping verbs and vector verbs.

22(a) Professor solution-ta **simplify** **kor-e** **di-l-o**
Professor solution-DEF simplify do-PFV.PTCP give-PST-3P

ar **explain** **kor-e** **di-l-o**
CONJ explain do-PFV.PTCP give-PST-3P
The professor simplified and explained the solution.

(b) *Professor solution-ta **simplify** ar **explain** **kor-e** **di-l-o**
Professor solution-DEF CONJ do-PFV.PTCP give-PST-3P
The professor simplified and explained the solution.

Question and answer test

The question and answer test needs to be modified slightly in order to test bilingual complex verbs. This is because if a sentence containing a complex verb is turned into a question, then the entire complex verb or only the Bengali verbs can be used as the answer to that question (23b). However, using only the English element as the answer is not acceptable (23c). Therefore the test does not work in exactly the same way for a bilingual complex verb as it does for a monolingual compound verb, because in a monolingual compound, both verbs have to be used as the answer to the question. The test is still constrained in case of the bilingual verbs, because using only the English element is not allowed (23c).

23(a) Professor ki solution-ta **explain** **kor-e** **di-l-o**?
Professor Q solution-DEF explain do-PFV.PTCP give-PST-3P
Did the professor explain the solution?

(b) hæ **(explain)** **kor-e** **di-l-o**
Yes explain do-PFV.PTCP give-PST-3P
Yes (he) explained it.

(c) *hæ **explain**
Yes explain
Yes explained.

These tests also apply successfully on two-part complex verbs, which shows that two-part complex verbs also act as one syntactic unit. The table below 5.7 summarizes the similarities and differences between two-part and three-part complex verbs in bilingual Bengali-English.

Table 5.7: Similarities and differences between two-part and three-part bilingual verbs

	Two-part complex verbs	Three-part complex verbs
Similarities	Express a single event and form a single syntactic unit	
Differences	No vector verb	Vector verb present
	Without the vector there is a difference in meaning from the three-part structure	Vector adds meaning and affects aktionsart
	Consist of two elements	Consist of three elements
	Helping verbs bear inflections of tense, person and aspect	Helping verbs in perfective participle form. Vector bears additional inflections.
	Semantic compatibility between English element and helping verb is unnecessary	Requires semantic compatibility between English element and vector verb

5.5. Conclusion

In this chapter, I first discussed the changes that have taken place in Bengali at the lexical level in the form of importation of English lexical items and the use of loan translations in Bengali. I described lexical borrowings in Bengali as constituting the most clearly observable signs of contact. Then I discussed the emergence of CS in Bengali and English and the different CS patterns associated with it. I focused on CS because it is not only a result of contact but is also a mechanism by which contact-induced changes generally take place in languages.

Second, I focused on changes in the morphosyntactic domain in the use of bilingual verbs. I described the verbal system of modern Bengali, which consisted of two types of verbs, simple verbs and complex verbs. These complex verbs can have an (N+V) and (N+V+V)

structure, called conjunct verbs, or a (V+V) structure, called compound verbs. The complex verbs also comprise monolingual and bilingual types. In the monolingual type, all the elements in the N+V+(V) structure are from Bengali. In the bilingual type, the first element is always in English, leading to the structures as N[Eng]+'do'(+V) or V[Eng]+'do'(+V). Next, I discussed and compared the verbal system of nineteenth-century Bengali texts, which also has the simple verbs and complex verbs distinction akin to those in the modern Bengali corpora. However, the nineteenth-century Bengali texts I examined lacked any occurrence of bilingual verbs. I argued that the occurrence of bilingual verbs in modern Bengali constitutes evidence of contact effect in Bengali. In particular, the X+do complex verb pattern of Bengali provides a slot for the freer insertion of elements from English. These changes in modern Bengali speech are possibly causing a restructuring of the verbal system of the language, given the high frequency of occurrence of the bilingual verbs. I suggested an analysis of this structure in modern Bengali partially in line with previous analyses of bilingual verbs, by considering both the structure and semantics of these structures. The N+'do' structures appear to show subtle semantic differences from the V+'do' structures. The structural and semantic differences between the V+'do' and the N+'do' structures might be linked to the bilingual speakers' use of more V+'do' than N+'do' structures. Because V+'do' structures seem to be semantically less straightforward that N+'do' structures, bilinguals are possibly using more of these structures (which they have knowledge of given their fluency in English) as opposed to the N+'do' structures, which appear to convey a slightly different referential meaning. In comparison, monolingual speakers have very few bilingual structures and among these they have more N+'do' than V+'do' bilingual structures. I also discussed the existence of three-part bilingual verbs in modern Bengali, a structure that had not been reported in the literature on bilingual verbs before. The three elements together express

one single event in three-part bilingual verbs; I also showed with the help of syntactic tests that all three elements are part of the same constituent. Finally, I discussed in what ways the two-part complex verbs differ from their three-part counterparts, and the differences between helping verbs and vector verbs in the three-part complex verbs.

In the next chapter, I turn to the discussion of copular predicates of Bengali. I argue that even though the changes in copular predicates can be construed as resulting from contact, language-internal tendencies in fact offer a stronger motivation for their occurrence. Therefore I will take these changes as examples of changes resulting from multiple causation.

Chapter 6: Changes in the copular predicates of Bengali

6.1. Introduction

In the previous chapter I examined changes in Bengali at the lexical and morphosyntactic level. The changes that I discussed included those that are clearly observable in bilingual and monolingual speech. In this chapter I examine a type of syntactic change that is more subtle in nature and therefore more difficult to identify than lexical changes. I examine the copular predicates of Bengali, in equational sentences as well as the existential verbs of Bengali, and focus specifically on the equational copula *hocche* 'be-PRS.PROG-3P' due to its non-canonical sentence-medial position (1b), in contrast to other verbs (1a) which canonically occur in sentence-final position, given that Bengali is a verb-final language.

1 (a) ami bhaat kha-cch-i
1SG rice eat-PRS.PROG-1P
I am eating rice.

(b) Jolpole ho-cch-e ek-ta international terminus
Jolpole be-PRS.PROG-3P one-CL international terminus
Jolpole is an international terminus.

As I have done for other changes in Bengali, my investigation of the copular predicates also involves both diachronic and synchronic data. Diachronically, I examined nineteenth-century Bengali texts to trace the historical development of the *hocche* copula. For the synchronic comparison, I used the two sets of modern Bengali corpora consisting

of monolingual Bengali and bilingual Bengali-English speech. As I discuss below, Bengali also displays constructions in which the equational copula may be present or absent.

The aim of this chapter is to answer the following research questions regarding the behavior of the copula in Bengali: 1) Does the sentence-medial position of the Bengali copula result from a contact effect from English (given the extensive history of contact between the two languages) or from a language-internal development? 2) Do other copulas or semi-copulas in Bengali display the same sentence-medial position? 3) Do equational constructions displaying overt or null copulas occur in free variation? If not, what are the factors that regulate whether the copula is overt or null? 4) Does the *hocche* copula occur as regularly with NPs, APs and other predicates or does it occur with such phrases with different frequency rates? If so, how can the distinctive behavior of the *hocche* copula be accounted for?

The chapter is organized as follows: In section 6.2, I briefly recapitulate the diachronic and synchronic methods I used to study the copular predicates of Bengali. Section 6.3 provides a detailed discussion of copulas in general and of Bengali copulas in particular, with a focus on equational copulas. With regard to the *hocche* copula specifically, the syntactic environments in which it occurs are examined and compared with its zero counterpart in the synchronic and diachronic corpora. Section 6.4 examines whether the current medial position of the copula *hocche* is the result of contact-induced change from English or from language-internal development. In section 6.5, the development and behavior of the *hocche* copula is accounted for using a theoretical framework based on Pustet's (2003) universal hierarchy of copularization. Section 6.6 provides a summary and main conclusions of this chapter.

6.2. Methodology

In order to investigate the differences between monolingual Bengali and bilingual

Bengali in terms of copular constructions, two corpora consisting of monolingual and bilingual speech have been examined. The older variety of Bengali to which the modern Bengali data is compared, consists of plays written in Bengali that were published around the mid-nineteenth century. I should reiterate that I am somewhat constrained in my choice of plays because this genre developed in Bengali as a result of English influence. Consequently, the earliest plays I have at my disposal were written after initial contact with English had been established, as I motivated in detail in chapter 4. In the next section, a thorough introduction to the typology of copulas in Bengali is provided with a specific focus on the equational copula *hocche*, due to its distinctive behavior.

6.3. Copular predicates in Bengali and other languages

This section is organized as follows: First, a basic working definition of the copula is provided to differentiate it from other verbs, including semi-copulas. In section 6.3.2, some key studies on copulas in general and on the Bengali copula in particular are reviewed. Existential verbs and equational copular constructions are contrasted to highlight the distinctive behavior of *hocche* in terms of position and semantics. In 6.3.3., the environments in which the copula *hocche* appears are studied and it is considered whether the structures in which *hocche* is overt and those in which it is absent truly occur in free variation or whether identifiable variables regulate its presence vs. absence. In section 6.3.4, the behavior of the copula in the modern Bengali dataset (including both monolingual and bilingual corpora) is compared to the older plays. The final section 6.5 proposes an analysis of the behavior of the Bengali copula, using Pustet's (2003) theoretical framework of universal copularization hierarchy.

6.3.1. What is a copula?

The cross-linguistic behavior and categorical identification of copulas is a notoriously difficult issue to muster in linguistics. Copulas display distinct behaviors along a number of dimensions in many languages. They behave differently from other verbs at the semantic, morphological and syntactic levels.

At the semantic level, given that copulas typically do not add a semantic value (at least in terms of independent predicational properties) to the event denoted in a given discourse, some linguists consider them to be semantically vacuous (Hengeveld 1992: 32). Some researchers assume that their only function is to carry tense, mood or/and features.

Morphologically, in a number or languages, the copula behaves differently compared to other lexical verbs in instantiating distinct paradigmatic forms; for instance the French verb *être* is morphologically richer in displaying five distinct forms in the present tense (*je suis* 'I am', *tu es* 'you are', *il est* 'he is', *nous sommes* 'we are', *vous êtes* 'you are', *ils sont* 'they are') in the present tense compared to the first conjugation *–er* verbs for which there are only three phonologically distinct forms through the conjugation (i.e., *parler* 'to talk': one single phonological form in *je parl**e*** 'I talk'/*tu parl**es*** 'you speak'/ *il parl**e*** 'he speaks'/*ils parl**ent*** 'they speak', in addition to *nous parl**ons*** 'we speak' and *vous parl**ez*** 'you speak').

Syntactically, some languages have several copulas depending on the type of predicates involved, others have one only and others none. As is discussed in section 6.4, languages like Bengali not only have overt and null copulas but their copula also appears in sentence medial position, whereas all other verbs in the language abide by the canonical SOV word order by appearing sentence finally.

Copulas display much variation cross-linguistically; they instantiate distinct morphosyntactic and syntagmatic properties, may co-occur with the full set or just a subset of parts of speech in a given language as their dependent predicates, may follow different development paths and have distinct etymological roots (Pustet 2003).

For the purpose of this chapter, Pustet's (2003) definition of a copula is adopted: "A copula is a linguistic element which co-occurs with certain lexemes in certain languages when they function as predicate nucleus. A copula does not add any semantic contents to the predicate phrase it is contained in." (Pustet 2003: 5) This working definition of the copula has been kept in mind when examining (below) the full range of copulas in Bengali.

6.3.2. A taxonomy of copulas in Bengali

Bengali is endowed with a range of copulas with distinct semantic imports, reflected by the different morphological forms of copulas in the language. Cross-linguistically, this is not surprising as different copulas are known to perform different functions in the language. For instance, in languages like Portuguese and Spanish, a distinction is made between copulas expressing a permanent state (*ser*) versus a temporary state (*estar*). In contrast, as it is shown in this section, in languages like Bengali, distinct morphological forms do not express permanent versus temporary state (though Bengali makes such a distinction using a different strategy, as discussed in section 6.3) but instead a distinction between existential verbs and predicative copulas.

Regarding predicative copulas, it is important to make a distinction between three types of copular clauses, following Mikkelsen (2005). Consider the following three types of copular clauses, as described in Mikkelsen (2005: 1):

(2) The lead actress in that movie is Ingrid Bergman (specificational)

(3) Ingrid Bergman is the lead actress in that movie (predicational)
(4) She is Ingrid Bergman (equative)

Her analysis argues for a taxonomy of copular clauses based on some clear definitional criteria allowing us to identify (2) as being a specificational clause, (3) as a predicational clause and (4) as an equative clause.

In Mikkelsen's analysis, predicational clauses can be identified on the basis that they express a property believed to hold of the entity denoted by the subject (for instance, in (3), the property is that of lead actress). In other words, predicational clauses tell us something *about* the referent of the subject. In contrast, specificational clauses do not predicate a property of the subject referent; all the subject does is introduce a variable (the x such that x is the lead actress and the post-copular expression provides a value for that variable). A specificational clause does not tell us anything additional about the referent, just who and what the referent is. Equatives, in contrast, involve two expressions denoting the same individual (Mikkelsen, 2005: 12)

In Bengali, the copula involved in specificational, predicational and equative clauses is *hocche* and has the peculiarity of occurring sentence medially when endowed with such functions. Consider the Bengali copular constructions in (5), (6) and (7), equivalents to the English examples in (2), (3) and (4).

Specificational clause in Bengali

(5) oi chayachobi-r prodhan naika ho-cch-e Ingrid Bergman
That movie-GEN main actress be-PRS.PROG-3P Ingrid Bergman
The lead actress in that movie is Ingrid Bergman.

Predicational clause in Bengali

(6) Ingrid Bergman ho-cch-e oi chayachobir prodhan naika
Ingrid Bergman be-PRS.PROG-3P that movie-GEN main actress
Ingrid Bergman is the lead actress in that movie.

Equative clause in Bengali

(7) O ho-cch-e Ingrid Bergman
3SG be-PRS.PROG-3P Ingrid Bergman
She is Ingrid Bergman.

Interestingly for this purpose, Bengali copulas seem to behave in such a way that sentences (5), (6) and (7) can be collapsed into what could be labeled "equational copulas" in the sense that whenever two noun phrases are equated as denoting the same referent, *hocche* appears sentence-medially. This section will show that the Bengali language marks a distinction between equational copulas and semi-copulas.

The copula *hocche* in Bengali has been the topic of numerous studies, including Sableski (1956), Ferguson (1972), Bai (1986), Dasgupta & Ghosh (2006) and Thompson (2010). Bengali being a verb-final language, it has been observed repeatedly (most notably in Ferguson, 1972) that while other verbs and semi-copulas (existential, predicational, Mikkelsen, 2005) occur sentence-finally like lexical verbs, the equational copula *hocche* is the only verb in the language (and by extension the only copula) that occurs sentence medially in the pattern NP COP NP. This fact has been consistently reported but no explanation has been proposed as to why the copula *hocche*, when overt (it can also be absent), is the only verb occurring in this position. Given that the equational copula occupies the same position as the English copula, a reasonable question to ask is whether the length and intensity of contact between the two languages led to the emergence of sentence-medial *hocche*. Considering these facts, this chapter investigates if the NP COP NP structure is a contact-induced change that occurred in Bengali under the influence of English? Does the NP no-copula NP (henceforth, NP NP) structure convey the same semantics as NP COP NP? Do NP NP and NP COP NP carry the same range of temporal, aspectual and modal interpretations (for instance can NP NP mean 'he was my father')? Are NP

NP and NP COP NP restricted to noun phrases or can one find in Bengali NP AP for instance; in other words, can constructions with no copulas take on other types of predicates?

In order to answer these questions, this chapter investigates the semantics, distribution and diachronic development of the copula *hocche*. The synchronic examination of modern monolingual and bilingual Bengali shows that the bilingual corpus reveals significantly more occurrences of equational *hocche* than the monolingual corpus, leading to a potential interpretation of the sentence medial position of the copula as being contact-induced. Also as is discussed in section (6.3.2.4) diachronically, no sentence-medial *hocche* is attested in the old texts that were examined. These texts attested only the NP NP structure. This leads to a close examination of several paths of evolution for copulas as proposed by Ferguson (1972).

> a) If languages alternate between overt and null copulas diachronically, the null form of the copula will prevail in a language; in contrast, languages that display 'NP ø NP' diachronically, can develop a copula over time.
>
> b) For languages that display a null copula at first and develop a copula over time, the use of that copula may be to mark identity or equivalence with a noun complement.
>
> c) When a language with a zero copula inserts a pro-copula it does so in between the subject and complement regardless of the normal position of verbs in the language.

This chapter supports in part Ferguson (1972)'s generalizations by showing that the sentence-medial copula *hocche* is a recent development in Bengali and as such, marks identity or equivalence. However, a distinct labeling from Ferguson's 'pro-copula' is used for this copula because as will be argued, it behaves like a genuine verb and not a pronoun. The ultimate objective is to offer a cogent explanation for why the Bengali copula *hocche* behaves like it does, using the framework of Pustet (2003). In this sense, this chapter fills a gap in the research on Bengali copulas and presents a more fine-grained picture of their behavior.

In the next subsection, a detailed account of copulas in Bengali is provided, comparing specifically equational copulas, which may or may not involve the overt expression of a copula with existential verbs, in order to highlight the distinct behavior of the equational copula in terms of its position and semantics. Also, as we will see in section 6.4, the equational copula *hocche* in modern Bengali has acquired expanded contexts of use, including sentences where previously existential verbs were used. Therefore, the section below on existential verbs will highlight the contexts where such verbs are generally used in modern Bengali.

6.3.2.1. Existential sentences

In present tense existential sentences, the existential verb *ach* 'exist-PRS' is used as shown in (8a). This verb occurs only in the present tense and it is unable to take on future or past tense inflections, as shown by the ungrammaticality of examples (8b) and (8c). For this reason, it is at times referred to as a defective verb (Thompson 2010). Since Bengali is a head-last language, these verbs occur in the sentence-final position.

8 (a) bhogoban ach-e 'existential'
God exist.PRS-3P
There is a God.

(b) * bhogoban achi-lo 'existential-PST'
God exist-PST-3P
There was a God.

(c) * bhogoban ach-be 'existential-FUT'
God exist-FUT-3P
There will be a God.

Existential constructions also involve locatives, as shown in (9). These sentences indicate that there is a location (a home in this case) where the subject is.

9) O ghor-e ach-e 'locative'
3SG home-LOC exist.PRS-3P
He is at home.

In past tense existential sentences (including existential, locative and possessive, as summarized in Table 6.1 below), a variant of the verb *ach*, which is *chilo* 'exist-PST-3P' is used, as shown in (10):

(10) O ghor-e chil-o 'locative-PST'
3SG home-LOC exist.PST-3P
He was at home.

The other type of existential sentence includes possessive sentences indicating possession of some entity by the subject, as in (11):

(11) Amar ek-ta gari ach-e 'possessive'
1SG.GEN one-CL car exist.PRS-3P
I have a car.

Table 6.1 shows the forms and distributional patterns of the different existential verbs in Bengali. As the table shows, for the future tenses, neither *ach* 'exist-PRS' nor *chilo* 'exist-PST-3P' is used. Instead, the verbs *hɔbe* 'be-FUT-3P' or *thakbe* 'stay-FUT-3P' is used.

Table 6.1. Forms and distributional patterns of different Bengali existential verbs

Tense/Type of Existential Sentence	Locative	Possessive	Existential
Present	ache 'exist.PRS-3P'	ache 'exist.PRS-3P'	ache 'exist.PRS-3P'
Past	chilo 'exist.PST-3P'	chilo 'exist.PST-3P'	chilo 'exist.PST-3P'
Future	thakbe 'stay-FUT-3P'	hɔbe/thakbe 'be-FUT-3P' 'stay-FUT-3P'	hɔbe/thakbe 'be-FUT-3P' 'stay-FUT-3P'

In summary, existential constructions involve the use of overt verbs or semi-copulas translatable as 'exist', 'be' or 'stay'. They all behave like regular verbs in that they abide by the Bengali canonical SOV word order. I focus next on equational copulas. As mentioned before, these constructions are labeled as such because they establish an identity or equivalence between

two noun phrases. Equational *hocche* will be examined in detail and analyzed in the environments in which it is present or absent.

6.3.2.2. Equational sentences

Present tense equational sentences with nominal predicates may or may not involve an overt copula. No overt copula linking the two nouns is necessary, as shown in (12a); when the copula is overt, it occurs in the non-canonical sentence-medial position, as shown in (12b) and (12c).

12 (a) chele-ta amar chatro ‘no overt copula’
Boy-CL 1SG.GEN student
The boy (is) my student.

(b) chele-ta **ho-cch-e** amar chatro ‘overt copula’
Boy-CL be-PRS.PROG-3P 1SG.GEN student
The boy is my student.

(c) Tar baba **ho-cch-e** ei bisshobiddaloy-er oddhapok
3sg.GEN father be-PRS.PROG-3P this university-GEN professor
His father is a professor at this university. (Thompson 2010)

However, in the past tense, a copula is obligatory and occurs in sentence-final position, restoring the canonical SOV word order of Bengali. This verb is *chilo* ‘exist-PST’ and is the same verb that appears in equational and existential sentences (c.f. 10).

(13) chele-ti amar chatro **chil-o** ‘overt copula in past tense’
Boy-CL 1SG.GEN student exist.PST-3P
The boy was my student.

Given these facts, a couple of questions naturally come to mind: why does the copula appear when it may be optionally absent? When it appears, why does it occur in the non-canonical sentence-medial position and only in the present tense?

The literature on equational constructions in Bengali does not provide satisfactory

answers to these queries. Some scholars state that the copula occurs to provide "a bit more structure and weight" to equational sentences with just a subject and a complement (Thompson 2010: 507). Others have called it a "pro-copula" because its primary function is non-copulative (Ferguson 1972: 95; Thompson 2010), and there are studies which state that there is no difference in meaning between sentences that have the copula and those that do not (Ferguson 1972, Sableski 1956). It has also been called a "positive polarity copula" because if these sentences are negated or turned to questions, they not only do not require a copula but are in fact grammatically ill-formed with the copula. Other studies on Bengali copulas have omitted this construction altogether (Paul 2009, Bhattacharya 2005).

This chapter suggests a possible scenario accounting for the behavior and the occurrence of the sentence-medial position of *hocche*: it is possible that the presence vs. absence of the copula may have to do with information structure. When the topic of an utterance is introduced for the first time, the copula *hocche* is overt but when the same entity appears in the discourse subsequently, then *hocche* is optional and is no longer necessary. A second analysis is that the equational copula appears in the sentence-medial position to signal a complementary distribution with a homophonous *hocche* verb that carries different semantics. In other words, there may be two lexical entries in the Bengali language: a *hocche* copula corresponding to the equational 'be' under study and another *hocche* that appears in sentence-final position but that has distinct semantics and behaves like a full lexical verb.

Indeed, when *hocche* occurs sentence-finally, it loses its copulative function and takes on a range of different meanings. In (14), sentence-final *hocche* takes on the meaning of 'become':

14) Tar baba ei bisshobiddaloy-er oddhapok ho-cch-e-n
3SG.GEN father this university-GEN professor be-PRS.PROG-3P-HON
His father is becoming a professor at this university. (Thompson 2005)

In (15), it assumes the meaning of auxiliary ‘be’, in (16), it is interpretable as ‘occur’, and in (17), as ‘happen’. Such examples give us ground to speculate that the different positions are motivated by distinct semantics, setting apart present tense equational *hocche* from all the others.

(15) amar bari toiri ho-cch-e ‘be’
1SG.GEN house built be-PRS.PROG-3P
My house is being built.

(16) Prottek din rail durghɔtona ho-cch-e ‘occur’
every day rail accident be-PRS.PROG-3P
Rail accidents are occurring everyday.

(17) onushthan-ta kaal ho-cch-e ‘happen’
event-CL tomorrow be-PRS.PROG-3P
The event is happening tomorrow.

Such observations would be in keeping with Pustet (2003) who clearly differentiates genuine equational copulas (sentence-medial *hocche*) from semi-copulas (sentence-final *hocche*) that typically behave like full lexical verbs.

For the sake of comparison, it should be added that the occurrence of *ho-* ‘be’ as a copula in sentence-medial position is attested in only two other Indian languages, Assamese and Oriya, which are both Eastern Indo-Aryan languages as is Bengali. As in Bengali, in Assamese a zero copula is common in equational present tense clauses. However, a ‘be’ verb in sentence-medial position also acts as a copula as shown in (18) below.

(18) ra:m-ar sola:to-r da:m hal/haise exa taka
Ram-GEN shirt-CL-GEN price COP hundred rupees
The price of Ram’s shirt’s is a hundred rupees. (Nath 2009)

Table 6.2 below summarizes how different scholars have characterized *hocche* and table 6.3 summarizes the distributional and semantic range of *hocche* in Bengali. Further discussion about the proposals summarized in table 6.2 is included in section 6.4.1 below.

Table 6.2. Characterization of *hocche* by different scholars

Ferguson (1972)	Dasgupta & Ghosh (2006)	Thompson (2010)
Pro-copula	Positive Polarity Copula	Pro-copula
Indicate emphasis, hesitation or carry stylistic values	Mood marker	To add emphasis and weight to the equational sentence

Table 6.3 Distributional and semantic range of *hocche* in Bengali

Form of *hocche*	Position in the sentence	Function	Meaning
hocche	Sentence-medial; between two NPs	Copulative	*Is, am, are*
hocche	Sentence-final	Semi-copula	occur
hocche	Sentence-final	Semi-copula; inchoative	become
hocche	Sentence-final	Semi-copula	happen

In the next section, I analyze the remaining environments in which *hocche* may occur.

6.4. Linguistic environments governing the occurrence of *hocche*

The previous sections discussed occurrence of the copula *hocche* in modern Bengali primarily with nominal predicates in equational sentences. This section provides a discussion of the other types of phrases with which this copula occurs in today's Bengali, namely prepositional and adjectival phrases. The distribution of *hocche* in both the monolingual and bilingual Bengali-English corpora is considered and the environments in which the copula is absent are also examined.

As discussed in section 6.3.2.2, scholars such as Sableski (1956), Ferguson (1972), Bai (1986), Dasgupta & Ghosh (2006) and Thompson (2010) have all reported on the behavior of the *hocche* copula with nominal phrases, as seen in example (19).

(19) Tar bhai ho-cch-e ɔnchol prodhan 'NP hocche NP'
3P.GEN brother be-PRS.PROG-3P regional head
His brother is the regional head. (Monolingual Bengali)

In the monolingual and bilingual Bengali corpora, the occurrence of this copula is found with prepositional and adjectival phrases as well, as shown in (20) and (21).

(20) Shæoɽaphuli ho-cch-e Hoogly-te 'NP hocche PP'
NAME be-PRS.PROG-3P NAME-LOC
Sheoraphuli is in Hoogly.

(21) East Bengal-er khæla ho-cch-e shaŋghatik 'NP hocche AP'
East Bengal-GEN match be-PRS.PROG-3P charged up
The East Bengal match is charged up.

However, it should be noted that in comparison to noun phrases, the occurrence of the *hocche* copula with prepositional and adjectival predicates is infrequent. Out of all the sentence-medial occurrences of *hocche* in bilingual speech, only about 10% of those occur with prepositional predicates and only 2% occur with adjectival predicates. In monolingual speech, only 18% of sentence-medial *hocche* occur with prepositional predicates and only 6% with adjectival predicates. Table 6.4 below provides a summary of the frequency of occurrences of the *hocche* copula in its different syntactic environments.

Table 6.4: Frequency of *hocche* in different environments

Predicate Type	Monolingual Bengali	Bilingual Bengali
'NP hocche NP'	76%	88.59%
'NP hocche PP'	18%	9.39%
'NP hocche AP'	6%	2%

In the next section, the environments in which the copula *hocche* is present and those in which it is absent are examined.

6.4.1. 'NP hocche NP' vs. 'NP NP'

In this subsection, it is examined to what extent the construction displaying an overt

copula ‘NP hocche NP’ and the one with no copula ‘NP NP’ are semantically equivalent and whether or not they occur in free variation.

Before focusing on these two structures, let’s briefly recall that besides ‘NP NP’ structures, as shown in (22) below, Bengali also displays copula-less structures with PP, as in (23) and AP, as in (24).

(22) Chele-ti chatro ‘NP NP’
The boy student
The boy is a student. (Ferguson 1972)

(23) ʂe ghɔr-e ‘NP PP’
He home-LOC
He is at home.

(24) lok-ti dhoni ‘NP AP’
The man wealthy
The man is wealthy. (Ferguson 1972)

Previous studies have not provided a satisfactory account of why ‘NP *hocche* NP’ may occur in some environments while ‘NP NP’ may occur in others. On the one hand, one could assume that they occur in free variation. On the other hand, some scholars have argued that they carry different semantic imports. Ferguson (1972) argues that sentences with the *hocche* copula seem to carry more emphasis than those without. Sableski (1956) makes a similar proposal while Thompson (2010) suggests that the occurrence of the *hocche* copula adds more weight to the sentence, compared to its copularless counterpart. Dasgupta & Ghosh (2006) propose that *hocche* is possibly a mood marker and it occurs in sentences to fulfill a modality function.

The examination of sentences with the *hocche* copula and without a copula does not show any difference in emphasis. In other words, the sense is that sentences with *hocche* do not convey more emphasis than those without. Instead, one could hypothesize that the occurrence of the *hocche* copula may be related to information structure. It seems that *hocche* occurs when some

new information about an NP that has already been introduced in the discourse is provided. In other words, it is proposed that the role of *hocche* is to introduce new information in the discourse. Consider the sentence in (25): In the first part of the sentence, a reference to the *city* is established, which informs us that the city lies beneath the mountain. In the second part of the sentence, the name of the city (which is new information for the hearer) is introduced and *hocche* appears sentence-medially. The copula *hocche* seems to signal new information about the NP, in this case, *the name of that city*.

(25) pahar-er niche je **shohor-ta** ach-e,
mountain-GEN below which city-DEF exist.PRS-3P
The city, which lies below the mountain,

shei shohor-ta-r naam ho-cch-e Tirumala
that city-DEF-GEN name be-PRS.PROG -3P Tirumala
The name of that city is Tirumala.

Similarly, in the sentence in (26), the topic of the sentence being the association FETO, it's only after further information is introduced (in this case, the fact that it is an association of engineers) that *hocche* appears in (26c).

26 (a) age shadharonoto montri as-t-o,
before generally minister come-PST.HAB-3P

kintu ei bar **FETO**-r meeting-e CM ash-ch-e
but this time FETO-GEN meeting-LOC CM come- PRS.PROG-3P

Generally before, a minister used to come, but this time the CM is going to attend the FETO meeting.

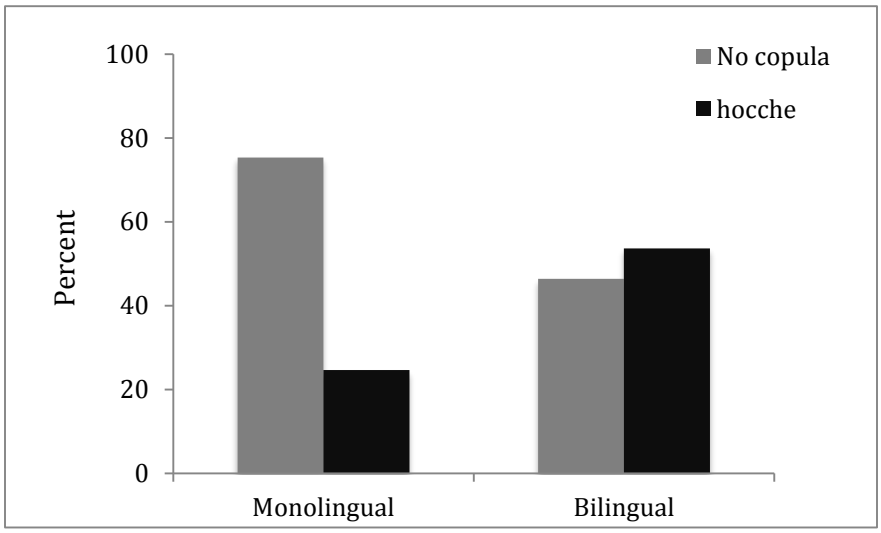
(b) **FETO** ta ki?
FETO DEF what
What is FETO?

(c) **FETO** ho-cch-e ek-ta association engineers-der
FETO be- PRS.PROG-3P one-CL association engineer-PL.GEN
FETO is an association of engineers.

In summary, intuitively it seems that the role of *hocche* is to introduce new information; once that is done, the copula-less structure is used. This is however just a hypothesis, as this matter needs to be investigated further before any solid claims can be made on this issue.

A basic quantitative analysis of the data is provided below. Crucially, for the synchronic data, the monolingual and the bilingual corpora differ significantly from each other with respect to copula usage with NPs. The 'NP NP' structure occurs much more frequently in the monolingual corpus than in the bilingual one. Figure 6.1 below summarizes the frequency of the *hocche* copula in comparison to absence of copula in modern monolingual and bilingual Bengali.

Figure 6.1: Frequency of *hocche* in comparison to null copula within nominal predicates



The frequency of *hocche* in monolingual Bengali within nominal predicates is about 25%. This frequency is out of the total set of eqautional sentences with nominal predicates. This frequency is roughly half the rate in the bilingual corpus, which is about 53%. A chi-squared test confirmed that this frequency distinction was statistically significant: *hocche* has a significantly higher rate of occurrence in the bilingual corpus than in its monolingual counterpart, X^2 (1, N = 400) = 32.556, $p < .01$.

Such results could lead us to assume a contact-induced explanation for the sentence-medial *hocche* copula. If the behavior of sentence-medial *hocche* is contact-induced, this would explain why it occurs at a higher rate in the bilingual corpus than in the monolingual corpus. The constant contact between bilinguals and monolinguals would then explain why sentence-medial *hocche* can be found in monolingual speech, though to a much lesser degree. However, such assumptions alone do not hold when taking into consideration other data, which will be discussed in the next section.

6.4.2. A diachronic investigation of nominal predicates in Bengali

In the previous section, the synchronic corpora of Bengali revealed that the sentence-medial copula *hocche* occurs not only with noun phrases but also (though less frequently) with prepositional and adjectival phrases. One would not expect to find *hocche* being used with PPs and APs, as such phrases typically occur with the existential verb *ache*. This points to the fact that *hocche* is currently competing with *ache* in domains where *ache* is still dominant. This competition may underlie language change. It was also found that the bilingual corpus has a significantly higher rate of occurrence of sentence-medial *hocche* compared to the monolingual corpus.

Now we focus on to the older variety of Bengali consisting of data from two Bengali plays. The examination of equational, specificational and predicational sentences in older texts yielded no occurrence of the verb *hocche* in the sentence-medial position. Sentences (27) and (28) below do not have any verb or element connecting the two noun phrases. All equational, specificational and predicational sentences sentences in the old texts display the 'NP NP' structure. There are no overt copulas in medial position. The number of occurrences of copula-less structures and those with *hocche* in sentence-medial position are shown in table 6.5.

(27) Isshor din-er rokkhɔk
God poor-GEN savior
God (is) the savior of the poor. (1860 play)

(28) choto bodhumata amar khuki
younger daughter-in-law 1SG.GEN young girl
(Our) younger daughter-in-law (is) my child. (1860 play)

Table 6.5: Occurrence of *hocche* and null copula with nominal predicates in 19th century texts

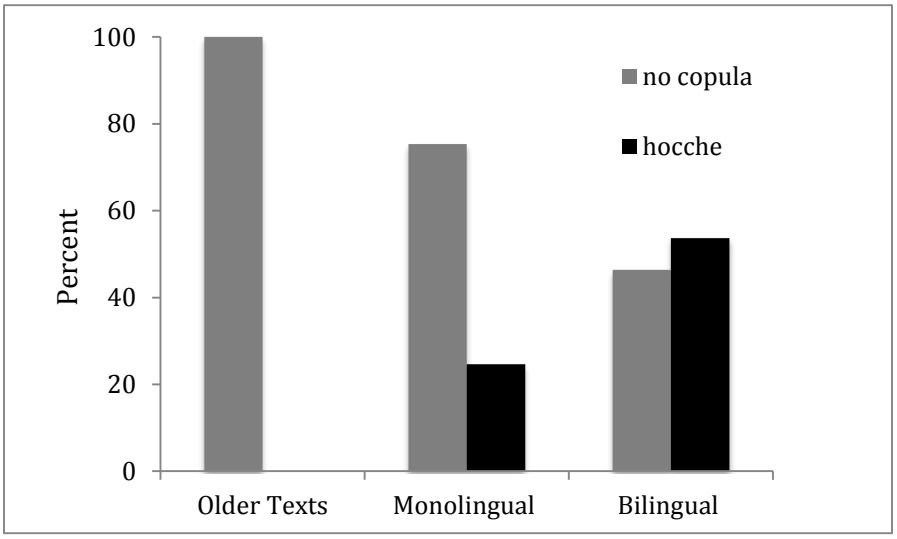
Copula Type	1854 Play	1860 Play
Hocche copula	0	0
No copula	76	84

In these older texts, the word *hocche* appears to have only one function, which is that of a regular auxiliary *be* verb. Also, whenever *hocche* occurs as a *be* verb, it does so only in the sentence-final position, as shown in (29).

(29) gram-e bohu din dhore durghotona ho-cch-e
village-LOC many days for mishap be-PRS.PROG-3P
For several days, mishaps have been occurring in the village (1860 play)

These findings indicate that the auxiliary 'be' verb has acquired additional functions in modern Bengali, which is that of a copula verb, a property not attested at a prior stage of the language. Figure 6.2 below provides the frequency of the 'NP NP' structure and the copula *hocche* with noun phrases in the diachronic and synchronic corpora. The distribution in figure 6.2 below simply adds the diachronic data to the frequency distribution that was shown before for the synchronic data in figure 6.1.

Figure 6.2: Frequency of *hocche* and no copula with nominal predicates in older and modern Bengali



6.5. A theory of copularization (Pustet 2003)

To summarize, the previous section compiled synchronic and diachronic data to account for copula usage in Bengali. The two modern Bengali corpora attempted to identify discrepancies in usage between the two groups of speakers. The results showed that while bilingual speakers used the NP *hocche* NP construction more than monolingual speakers, leading one to consider the possibility that such pattern may be contact-induced, the fact that the medial position of the copula was not maintained across tenses, moods and aspects undermined such a hypothesis. One can postulate instead that such a position may be the result of an internal development. The following questions motivate the next steps in the chapter: 1) How can we account for the distributional patterns observed in Bengali copula usage? 2) What motivates the observed copularization patterns in terms of the categories of lexemes the copula is compatible with[25]? Are there factors that may have triggered the emergence of the Bengali copula in a sentence-medial

[25] As already discussed, *hocche* occurs more often with noun phrases than with adjectival phrases or prepositional phrases.

syntactic position? In order to answer these questions, several tenets of Pustet's (2003) hierarchy of copularization framework will be considered, as discussed in the next subsection.

Pustet's (2003) main assumptions of the hierarchy of copularization

This section lays out the core assumptions of Pustet's (2003) hierarchy of copularization, as her framework will guide the proposed analysis of the Bengali copula behavior. Pustet's hierarchy seeks to answer the following questions:

1) What are the distributional patterns of copularizing versus non-copularizing lexemes that can be found in the languages of the world?

2) What lexical classes are compatible with copulas?

3) What are the semantic motivations for copularization?

In order to answer these questions, she conducted a survey of 131 languages. Although it does not include Bengali, her framework accounts for a lot of the Bengali data that this chapter has discussed, as is noted below. Focusing on nominal, adjectival and verbal phrases, the taxonomy in (30) emerges from the 131 languages Pustet surveyed.

(30) a. Non-copularizing languages (Tagalog)
b. Fully copularizing languages (Bambara)
c. AN languages (German)
d. AV languages (Burmese)
e. Split-N languages
f. Split-A languages
g. Split-V languages

(30b) designates languages in which nouns, adjectives and verbs all take copulas whereas in contrast, languages in (30a) do not. In Pustet's model, AN (30c) symbolizes parallelisms in the treatment of nominals and adjectivals and AV (30d) shows parallelisms in the treatment of adjectives and verbs. To be more precise, in AN languages like German, the adjective and the

noun display the same behavior in allowing a copula. In AV languages like Burmese, the adjective and the verb have a uniform behavior in prohibiting a copula. Split-N in (30e) refers to languages in which some nouns may appear with the copula while others do not. Split-A in (30f) refers to languages in which some adjectives may occur with the copula whereas others do not and finally Split-V in (30g) refers to the same phenomenon for V.

The observations of copularization patterns of nouns, adjectives and verbs in 131 languages led Pustet to formulate a universal, implicational hierarchy of copularization, following the schema in (31):

(31) Nominals > adjectivals > verbals

The universal hierarchy of copularization that Pustet proposes is unidirectional in that:

> "any lexeme that is located to the left of the cut-off point between copularizing and non-copularizing lexemes in the lexicon of a given language receives a copula; any lexeme that is located to the right of this cut-off point does not combine with a copula. In other words, copularization is unidirectional: if any lexical class of lexemes other than nominal copularizes, the adjectival class does; only if both nominals and adjectivals copularize, verbals may also copularize". (Pustet 2003: 73).

An important generalization of Pustet's unidirectional copularization scale that she makes explicit is that:

> "if a language has a copula which combines with only one of the three semantic macro-classes [nouns, adjectives, verbs], this copula will combine with the class of nominals; if a language has a copula which combines with two of the macro-classes, it will combine with nominals and adjectivals. Only if a given copula combines with all three of the basic semantic macro-classes, will it combine with verbals. Languages in which verbals copularize but nominals and adjectivals do not, are not documented in the sample. Languages in which verbals and adjectivals copularize but not nominals do not exist either. Put differently, copularization 'starts' at the left-hand pole of the scale and 'proceeds' to the right, but never vice-versa." (Pustet 2003: 78).

Based on Pustet's copularization scale, Bengali clearly fits among the AN languages in that its noun and adjective predicates display a uniform behavior in their ability to copularize but not its verbal predicates. In other words, the *hocche* copula is never found coupled with another verb, only with nouns and to a lesser extent with adjectives, as clearly shown in table 6.4. This leads us to our next questions: why does *hocche* occur significantly more often with nouns than with adjectives? In addition, as shown in Figure 6.2, why is the absence of copulas so dominant in monolingual speech, in keeping with an early stage of the language in which only the NP NP structure (with no copula) was attested?

We might find a tentative account of some of these empirical facts in Pustet's time-stability parameter which she borrows from Givón (1979).

The semantic parameters of copularization

This subsection lays out Pustet's semantic parameters of copularization as well as her definitions of nouns, verbs, and adjectives; all definitions are pivotal to her theory. Pustet (2003: 11) provides three subdefinitions of the lexical macro-classes nouns, verbs and adjectives.

(32) Subdefinition 1 (Pustet 2003: 10)

"Prototypical nouns designate things, entities, or concrete objects. Prototypical verbs designate actions, events, and processes. Prototypical adjectives designate properties."

Subdefinition 2 (Pustet 2003: 11)

"Nouns tend to function as arguments. Verbs tend to function as predicates. Adjectives tend to function as attributes."

Subdefinition 3 (Pustet 2003: 11)

"In discourse, nouns tend to combine with grammatical categories that denote case, number, definiteness/indefiniteness, and gender/noun class; verbs tend to combine with grammatical categories that denote person, tense, modality,

and voice; adjectives do not show such categorical preferences."

We should emphasize subdefinition 2 in particular, as it equates nouns with arguments, verbs with predicates and adjectives with attributes. Pustet's view of these macro-classes interlocks with Croft's (1991) view of markedness in which frequency interacts with structural complexity[26].

Bengali models fairly well this view of markedness, in that the language (monolingual Bengali specifically) reflects a strong preference for copula-less equational constructions, as shown in Figure 6.2. This preference is also attested in the older texts that were examined. When the copula is used, it is used primarily with noun phrases, much less with adjectives and not at all with verbs, as shown in table 6.4.

In summary, there are no copularizing verb predicates in Bengali, just copularizing constructions with nouns and adjectives. As adjectives copularize, it was rightly predicted following Pustet's hierarchical scale that nouns copularize as well. However, what still needs an explanation is why nouns copularize more frequently than adjectives. An explanation of this state of affairs may be found in four interlocking lexical semantics parameters that Pustet designs: dynamicity, transitivity, transiency and dependency. In this chapter, the parameters that have been capitalized on are transiency, transitivity and dependency but nevertheless all four have been introduced.

Dynamicity according to Pustet is:

> "a lexeme receives the specification [+dynamic] if it involves at least two disparate component states of affairs, the specification [-dynamic] if it involves homogeneous component states of affairs only, and the specification [+/-dynamic] if it can be interpreted as either involving disparate or homogeneous states of affairs. Thus, concepts such as 'to grow', 'to jump' are classed as

[26] I am aware that Croft's view of markedness is controversial. It does, however, serve the purpose of this particular case.

> [+dynamic], concepts such as 'round' and 'tall', however, as [-dynamic]. Lexemes such as Spanish *entristecerse* 'to be/to get sad' have both dynamic and non-dynamic readings." (Pustet 2003: 97).

This parameter becomes particular useful when dealing with minimal pairs like 'happy' and 'rejoice' in which 'happy' is considered [-dynamic] whereas 'rejoice' is considered [+dynamic] in addition to connotating agency in the latter and no real agency in the former (Pustet 2003: 98).

The concept of transience attempts to capture the distinction in many minimal pairs between permanence and non-permanence. As Pustet puts it, "this semantic parameter may manifest itself in terms of a state versus personality feature contrast", as in the Indonesian minimal pair *berani* 'courageous' (in a given situation or inherently/permanently), vs. *pemberani* '(inherently/permanently) courageous' (2003: 105). The contrast between permanence and non-permanence is expressed differently, depending on languages. For instance, in Turkish, adjectives generally do not copularize. However, adjectives like *utangaç* 'shy' that may denote the meaning of being inherently, permanently shy, can take the copula *dir*. In a number of other languages (including Estonian, Indonesian, Swahili or Japanese), the same adjective may convey a temporary or permanent property when occurring in the absence of a copula or may express a permanent property only when occurring with an overt copula. For instance, Japanese *noroi* 'slow' may refer to a temporary or permanent adjective with no copula but may refer to a permanent property when used in the presence of a copula (Pustet 2003: 106).

The concept of transience seems to interact with a third semantic/structural parameter, that of transitivity. Lexical minimal pairs show a contrast between lexemes that are classified as transitive because they obligatorily require the presence of two arguments, whether or not these arguments are expressed. As Pustet (2003: 114-115) puts it "lexical minimal pairs can be established in which the same semantic type of noun phrase is coded both as a valence-bound

argument, and as an argument which is not valence-bound and therefore dispensable. In English verbal-adjectival minimal pair *to fear* vs. *afraid* illustrates this:

(33) He fears poisonous reptiles
*He fears

(34) He is afraid of poisonous reptiles
He is afraid

As illustrated by the examples in (33) and (34), "poisonous reptiles" is optional in (33) but is obligatory in (34). Such examples clearly show the connection between syntactic and semantic transitivity. Other examples of minimal pairs in English are 'to know/to be aware of', 'to regret/to be sorry, to resent/to be resentful) and the same transitivity minimal pairs can be found in a variety of languages, including Estonian, German, Spanish, Swahili or French (*convoiter* 'to covet', *avide* 'covetous/greedy') (Pustet 2003: 116).

Pustet identifies a fourth semantic parameter that correlates with minimal lexical pairs involving presence or absence of copula. This last parameter is dependency, which is based on the distinction between object concept and property concept. In some languages, the presence of the copula signals that the lexeme is a referential object while the absence of the copula refers to the property concept. For instance, in Lakota, Pustet notes that the reduplicated form of c^{h}asmú 'sand' which is 'c^{h}asmúsmu 'sandy' does not copularize when it is a property concept. The object c^{h}asmú 'sand', can appear with a copula and is a mass noun (Pustet 2003: 121):

(35) Lé c^{h}asmú héc^{h}a
this sand COP
'This is sand'

(36) Pahá ki líla c^{h}asmúsmu
hill DEF very sandy
'The hill is very sandy' (Pustet 2003: 121).

Pustet's (2003: 81) account of why nominals lend themselves more readily to argument function, verbs to predicate functions and adjectives to attribute function is based on Givón's time-stability parameter. The semantic parameter behind the notion of time-stability is that of permanence. As Pustet summarizes it:

> "arguments generally function as the given points of reference in a clause, whereas the predicate typically specifies distinctions pertaining to variation in time, such as tense and aspect. By virtue of their intrinsic semantic content, which usually includes the feature [-time, -stable], verbals are predestined for predicate position. Thus, the predicate worthiness of individual lexemes may be interpreted as a function of their time-stability value." (Pustet 2003: 81).

As copulas are also subject to the markedness principle and can lead to changes in the markedness of some predicates (meaning that they add to the morphosyntactic complexity of predicates), this would account for why copularizing predicates are less frequent in discourse than non-copularizing predicates.

I propose that the set of semantic parameters laid out in this section may account to a large extent for the behavior of copulas in Bengali. Lexemes that normally refer to permanent states (like nouns) copularize more frequently than lexemes that normally refer to temporary states like adjectives[27]. Givón's notion of time-stability and Pustet's concept of transience in which permanence plays a crucial role arguably contribute to accounting for copula behavior in Bengali. As shown in table 6.4, the fact that *hocche* occurs much more frequently with nouns than with adjectives may be due to the fact that nouns typically refer to permanent properties. In contrast, many adjectives refer to temporary states (though not necessarily of course). Such contrast between noun phrases and adjectival phrases is captured in examples such as (37):

[27] The proposal that nouns refer to permanent states and adjectives refer to temporary states does not hold for all nouns and adjectives, which Pustet acknowledges (2003: 86). She uses these distinctions only for prototypical nouns and adjectives.

(37) contrast between noun ‘sand’ and adjective ‘sandy’ in Bengali:

(a) eta ho-cch-e bali
DEM be-PRS.PROG-3P sand
This is sand.

(b) eta balukamoy
DEM sandy
This (is) sandy.

It goes without saying that much more empirical evidence needs to be provided in order to establish that Bengali builds upon distinctions such as transience, transitivity and dependency and the distinction between a permanent versus a temporary property to trigger the realization of the overt sentence-medial copula. One way of determining whether this is the case would be to examine whether there is a contrast in Bengali between adjectives that express a permanent property (hence, favor the presence of a copula) versus a temporary property (hence, favor the absence of the copula) but this matter has to be left for further research.

6.6. Conclusion

This chapter has tried to account for the distinctive behavior of the *hocche* copula in Bengali, which occurs in the sentence-medial position (obligatorily) unlike other verbs in the language. Based on the diachronic and synchronic investigation, the key findings from this chapter are the following: Bengali *hocche* occurs much more frequently with nouns than with adjectives (and not at all with verbs) and does so much more frequently in bilingual speech than in monolingual speech. A combination of information structure requirements with Pustet’s semantic parameters of transience, transitivity and dependency may account for the occurrence versus absence of *hocche.* Diachronically, sentence-medial *hocche* is not attested in the older texts that were examined, pointing to a recent development in the language. The fact that it occurs predominantly in the sentence-medial position in bilingual speech could potentially

reflect a contact-induced change. However, the fact that it does not occur in other tenses points rather to a language-internal development. Such change may have been motivated by the fact that sentence-final *hocche* takes on a different meaning in the language, leading to the emergence of two distinct lexical categories: the genuine sentence-medial copula *hocche* and the sentence-final semi-copula *hocche*. Of course, the fact that Bengali is in close contact with English may have reinforced the change in word order in that direction, pointing to multiple causation.

Chapter 7: Changes in the English of bilingual speakers

7.1. Introduction

Both chapters 5 and 6 provided evidence for changes that have occurred in Bengali, and together they address one of the two directions in the investigation of bidirectional influences in the languages of a contact situation. This chapter considers the other direction, that is changes that may have taken place in the English spoken by bilingual Bengali-English speakers, possibly as a result of contact effects from Bengali.

In this chapter, I specifically examine potential divergences that may have occurred in the use of articles and progressives aspectual forms in the English of bilingual speakers and explore the factors that may have led to these divergences, focusing primarily on potential interference from Bengali and on second language factors such as proficiency level in English. This investigation will also consider to what extent the observed differences in this variety of English are similar to or different from what have been claimed to be general patterns in Indian English. The specific investigation of the use of articles and progressive forms has been taken up because these features have been identified in several studies as frequently diverging in varieties of Indian English (Sailaja 2012: 363). This chapter is organized as follows: The next section (7.2) presents an overview of research on Indian English, with a focus on the use of articles and progressive forms. It also addresses some of the outstanding issues in the existing literature with regard to the use of these features. Section (7.3) lays out the methodology

that was used to code and analyze the data. Section 7.4 presents the descriptive and quantitative analyses of results, some discussion on the qualitative aspects of the observed features, and an explanatory analysis of the results. Section 7.5 presents the conclusion.

7.2. Review of studies on Indian English

This section presents a review of research on Indian English, with a specific focus on the features of articles, tense and aspectual forms.

Studies on Indian English (IndE) have argued that there is high variability concerning its multiple regional and social varieties (e.g. Schneider 2007: 168; Sedlatschek 2009: 26). Variability within IndE may partly be due to it being a second language variety, since it is usually learned in addition to a first language. The first language for most Indians is usually some regional Indian language, whereas English is learned primarily in educational institutions [28] (Sailaja 2012). English is the language of higher education in India but it is also used as the primary language of instruction in numerous so-called English-medium private and public schools (from grades 1 to 12) across India. In addition, several government run public schools provide primary instruction in regional Indian languages (from grades 1 to 12) and provide English instruction as an additional language only for 3 to 5 years in high school (usually between grades 5 and 8, after which only regional languages are used). The number of years of exposure to instruction in English in different types of academic institutions has been argued to be an important factor in determining proficiency in the language (Agnihotri et.al. 1988).

[28] This is not the case for Anglo-Indians who have both British and Indian ancestry, which are not among the subjects in the current study. Anglo-Indians usually report English to be their first language. Additionally, some families in urban India where parents speak two different Indian languages might use English as the language of communication at home. In such cases, children belonging to these families would grow up with English as one of the first languages, possibly in a situation of simultaneous bilingualism (2L1). The subjects included in the current study all reported being sequential bilinguals, learning English as their second language.

Similarly, the influence of a substrate language (sometimes referred to as the *substrate hypothesis*), which could be one or more Indian languages, has also been invoked to explain variation in IndE (e.g. Kachru 1994, Sharma 2005a, Sedlatschek 2009, Sailaja 2012). Other second language and social factors such as the age at which English was acquired and the relative prestige of the languages spoken by the bilinguals/multilinguals may also affect the level of English proficiency of different speakers.

The variation in the levels of proficiency in English spoken in India has been described as a "cline of proficiency" (Kachru 1994: 509; Sailaja 2009: 18). Kachru takes 'educated South Asian English' (spoken by Indians who have had many years of exposure to English in academic settings) to be at one end and "Broken English' (spoken by people who have only had rudimentary access to English) to be at the other end of the proficiency cline. Educated South Asian English has also been referred to as 'standard Indian English' (a variety or varieties that have been accorded the highest social prestige, as exemplified by the speech of news anchors).

Articles and progressive forms are among the grammatical features that have been reported as diverging in IndE. For example, some studies have argued that there is rampant omission of definite and indefinite articles as an IndE feature (Spitzbardt 1976, as cited in Sedlatschek 2009: 197). Progressive forms have been reported as being widely used with stative verbs (e.g. Sailaja 2012: 363). These features have also been argued to be properties that diverge in non-native varieties of English in general (e.g. Sharma 2005a). The explanations for this divergence have ranged from interference (or transfer) effects from the L1, to pedagogical lapses on the part of instructors and to the role of other second language acquisition strategies. Some previous studies on the use of articles and progressive forms in IndE are reviewed below and the

range of explanations that have been suggested for the occurrence of divergences in these features is also discussed.

7.2.1. Articles

With regard to the use of articles, Agnihotri et al. (1984) examined 336 Hindi and Punjabi L1 speakers' control over the English definite and indefinite article in relation to their forms and functions. They tried to establish a hierarchy of difficulties that these L1 speakers experience with article use and discussed the possible causes of these difficulties (115). The subjects were tested on a passage where all articles, except sentence-initial ones were omitted and their task was to add the definite and indefinite articles in the passage wherever the subjects thought they were necessary (116). The results indicated that the subjects' control over the indefinite article *an* is better than the indefinite article *a*. The authors suggest that this is because students are aware of the fact that *a* precedes words that begin with consonants and *an* precedes words that begin with vowels, though they admit that is somewhat of a doubtful suggestion (119). However, there is also a strong tendency among the subjects to omit articles, which explains why people gave mostly correct responses when a null article was expected. The authors concluded that the subjects did not necessarily know where null articles are supposed to be used, but their general tendency to omit articles perhaps resulted in their correct responses in this category. They also found a tendency for these L2 speakers to use definite articles in place of indefinite articles (118-120).

In case of the definite article *the*, Agnihotri et al. pointed out that it could indicate four functions in the passage that was tested. It could function to specify a phrase/clause, an immediate surrounding, retrospective reference or superlative degree. Results in this category showed that the subjects gave more correct responses when *the* was being used to specify

superlative degree and the least correct responses when *the* was being used to specify a phrase/clause (121-123). Also, as the complexity of the linguistic environment in which the article occurred increased (i.e. how complex the NP is), there was a tendency to omit the definite article, though the authors do not provide any explanation for why this may be the case (123). Their results also suggested that students are not aware that only nouns require articles, because they often inserted articles before lexical items of other grammatical categories (125). The authors concluded that these errors occurred because of pedagogical lapses on the part of high school instructors who failed to instruct students in terms of the correct use of articles (127). However, they did not consider whether these errors occurred because of the subjects' L1 interference on their L2s.

A study by Sand (2004) was based on international corpora of English including different text categories (such as conversations, public dialogues, monologues, student essays, informational writing, newspaper texts and fiction) consisting of several contact and native varieties of English. It found no quantifiable 'underuse' of the definite article for varieties whose substrate does not contain definite articles such as IndE (290). Several Indian languages such as Hindi, Punjabi and Gujarati do not have a definite article and instead mark definiteness with word order. Nouns placed at the beginning of sentences receive a definite interpretation while those in the sentence-medial position receive an indefinite interpretation (Sharma 2005a: 537-538). Therefore if there is an effect of L1 transfer, then substantial omission of definite articles would be expected in all of the different types of texts in the IndE corpora. However, this is not what Sand found. The study reported that there was a text-specific distribution of definite articles in both native and contact varieties of English, where informal spoken genres such as

conversations had the lowest rate of definite article use whereas student essays, press texts and informational writings had the highest rate of article use (289-91).

In terms of indefinite articles, Sand found that the frequencies of article omission had a different pattern in the contact varieties. For both native and contact varieties of English, there was a drop in frequency of usage of indefinite articles in conversations as opposed to written genres. The rate of use of indefinite articles in the conversations for the contact varieties was lower than in the native varieties (an average of 15.3 articles per 1000 words in the IndE Corpus as opposed to 26.6 articles per 1000 words in the Great Britain Corpus used by Sand). In the written text types such as press texts and informational writings, however, there was little variation in the IndE Corpus as compared to the Great Britain Corpus. However, student essays in the IndE Corpus showed a very low frequency of indefinite article use (13.5%), which Sand relates to "individual speakers' or writers' level of competence or stylistic preferences" (294). Overall these findings do not support the substrate hypothesis, and instead show that if there is a substrate effect at all (which would predict omission of articles), it is only in the case of conversations.

Another study on IndE article use by Sharma (2005a) proposed the effect of a combination of factors on the use of articles, such as substrate influence when the substrate (L1) has an overt form and the intervention of universally available discourse knowledge when the substrate lacks an overt article. Sharma's study reported the use of English by 12 adult first generation Indian immigrants to the US who were all L2 speakers of IndE and L1 speakers of an Indian language. These speakers varied in terms of their proficiency in their L2, with some speakers who considered IndE on an equal footing with their L1 and others who were late learners of L2 IndE.

Sharma formulated four hypotheses about article use in IndE. The first hypothesis dealt with the transfer of first language positional marking. For example, in Hindi (which lacks a definite article) definiteness is determined based on where nouns occur in a sentence. Sentence initial nouns have a definite interpretation while sentence-medial ones get indefinite interpretation. This first hypothesis was only partially supported because results showed that the null definite is used only in the case of clause-initial theme NPs (550). In clause-final nouns, the indefinite article *a, an* was used which was not what the hypothesis had predicted.

Hypothesis two dealt with a direct transfer of the L1 article system. Several Indian languages do not mark definiteness with an overt article but do so by word-order or case marking. In case of the indefinite article, the L1s use the numeral *one* with a specific indefinite meaning. This hypothesis was also partially upheld when the specific indefinite NP bore an overt article matched by L1 influence (*I'm staying in a house with three other students*). Also, non-specific indefinite NPs (*I want to spend some time in my village, definitely if I get a chance*), which are null marked in the L1 of these speakers, were null marked in English at a much higher rate. However, a second prediction of this hypothesis was not clearly supported, that the definite article will be absent in IndE because the L1 lacks it. If the L1 did not have an overt form, that did not necessarily translate into a lack of overt form in the L2 (551-52).

Hypothesis three dealt with the intervention of a universal tripartite system of article marking when there is a mismatch between the L1 and L2 article systems. This predicts that in case of [+ specific reference, + hearer-known] entities, the NPs will have a definite article. In case of [+ specific reference, - hearer-known] entities, NPs will have an indefinite article, while in case of [- specific reference, - hearer-known] and [- specific reference, + hearer known] entities, NPs will have no article. This hypothesis was partially supported only for more

proficient speakers who marked the distinction between [+hearer-known] and [-hearer-known] because the Standard English system has this distinction. However, for less proficient speakers, this hypothesis did not hold. Therefore, the first, second and the third hypotheses were only partially confirmed (553).

Hypothesis four predicted that "Speakers' use of articles will vary according to the relative givenness or newness of the NP in question" (541), that is, it tested the overall discourse status of the noun (i.e. whether the noun has been introduced earlier in the discourse). In terms of hypothesis four, IndE seemed to "extend the scale of assumed familiarity to include the option of null use" of articles. That is, the definite article *the* was increasingly omitted when the relative familiarity of the NP increased. "Brand new NPs" in a discourse – a classification proposed by Prince (1981) – had the lowest rate of article omission, whereas "evoked NPs" had the highest rate of article omission. In addition, IndE speakers systematically used overt articles to disambiguate NPs (554-57). Finally, divergence in the article system was also sensitive to the degree of NP modification and quantification. That is, if the noun had been modified by adjectives, numeral modifiers or quantifiers, that favored deletion of articles (similar to what Agnihotri et. al. 1984 reported).

In sum, according to Sharma, the article system of these IndE speakers emerged from both L1 transfer and discourse knowledge (the status of an element in the overall discourse, that is, whether an element has already been introduced in the discourse or not). For instance, because there is an overt specific indefinite article in the L1 of the speakers (*ek* 'one' in Hindi for example), for specific indefinite nouns in English an overt form is used. However, if the L1 lacks a definite article, then there is an "imposition of other discursively available pragmatic principles" for the use of English articles (558).

Sedlatschek's (2009) study on IndE (from a corpus that mainly included press texts (40 texts), broadcast materials (40) and student essays (10)) focused on several features including the use of articles. This study provided a comparative analysis of article use in IndE and in British (BrE) and American English (AmE). The study argued that there was no general tendency to overuse or omit articles in IndE as had been widely claimed by several previous studies. Instead, the cases of article use that did not conform to Standard English usage were often related to "text type, mode and style" of the data. For example, in the case of press materials, IndE showed a higher use of definite articles than BrE or AmE, whereas in the student essays there was a tendency for article omission. Sedlatschek's findings corroborated some of Sharma's findings in that definite articles were often dropped when the discourse disambiguated the noun, either by establishing a reference for it earlier or by uniquely identifying it with the use of modifiers. This goes in line with Sharma's (2005a) claim that discourse familiarity of the NP is often responsible for the deletion of articles. Also, Sharma's and Agnihotri et al.'s argument about the correlation between the presence of NP modification and the dropping of articles held true for Sedlatschek's corpus as well (205-206). Overall this study showed that IndE speakers use articles innovatively often to suit specific discourse functions but these innovations did not diverge in significant ways from the norms of Standard English article use; both definite and indefinite articles in IndE, "broadly conform(s) to the rules of Standard English article use" (227). Sedlatschek also argued that no satisfactory one-dimensional explanation, such as drawing on the L1 interference hypothesis, could be provided for the slight divergences in article use.

In this section some previous studies on article use in IndE were discussed including the range of explanations that they have provided in terms of why article use in IndE diverges from standard varieties of English. Agnihotri et. al (1984) discussed the issue of pedagogical lapses in

the unsuccessful acquisition of articles, while Sharma (2005a) argued that articles in IndE seem to emerge from both L1 interference effects and available discourse knowledge. However, both Sand (2004) and Sedlatschek (2009) pointed out that L1 interference is not sufficient to explain divergences in article use. While Sedlatschek (2009) pointed out that divergences in IndE articles are not significant, Sand (2004) showed that indefinite articles are likely to have more divergences than definite articles in the corpus data considered in that study.

Three of the four studies discussed also showed a connection between the omission of articles and NP modification and discourse disambiguation. That is, if the noun has been modified by adjectives or numerals, articles are likely to be omitted, which is also true if the noun has already been introduced and a reference for it has been established. Some of these considerations will be addressed in the qualitative discussion of articles divergences in the variety of English spoken by Bengali-English bilinguals. In the next section, I turn to the discussion of tense use in IndE.

7.2.2. Tense and aspect

This section discusses research on the use of tense and aspect in IndE, with a special focus on the progressive forms.

In a study on tense in Indian English, Agnihotri et al. (1988) examined the use of tense by 356 English language learners in relation to 19 socio-psychological variables associated with language learning. The aim was to test the performance of the subjects in the use of English tenses and determine a scale of difficulty in their acquisition. The subjects were presented with questionnaires that required them to fill in the blanks in a series of sentences with missing tenses. They were tested on the present, past and future tenses including the indefinite, perfect, continuous, and perfect continuous aspects within each tense category. Based on their

performance in the test, the subjects were divided into three groups, good (group A with a score above 60%), average (group B between 33% and 43%) and poor (group C with a score below 16%); the average performance of correct use of tenses among these three groups was 31.69%. The subjects in each group did not have much intra-group variation in the use of the past, present and future tenses, that is, group A subjects performed well in the use of both past, present and future tense while group C subjects scored poorly in all of those categories. However, for all the groups, subjects had a slightly higher average performance in the present tense than in the past and future tenses. In terms of the continuous forms of the present tense, results showed that group A scored about 63%, group B about 46%, and group C about 17%. But within the continuous forms, the present continuous form such as *These days I am learning French at the University of Delhi* elicited more correct responses, with almost 95% for group A, 68% for group B and about 16% for group C, as compared to the present perfect continuous form such as *What a huge house it is! I have been painting it since morning and it is not even half finished!* In this category, group A had about 56% correct responses, group B only about 6%, and group C subjects got no correct answers. This trend also showed up in in the past perfect and future perfect continuous forms.

Based on these results, the authors proposed an implicational hierarchy of difficulty in tense and aspect use, which showed that the continuous forms of the present, past, and future tenses were easier to acquire for the subjects than the present perfect, past perfect and future perfect continuous forms. The study also found that out of the 19 socio-psychological variables that were tested as possible predictors for achievement in tense use, 8 variables correlated significantly. These variables included type of college that the subjects attended (elite colleges as opposed to non-elite ones), type of school (elite high schools with English instruction as

compared to schools which provide instruction in vernacular languages), scores in English exams in high school, overall score in high school exams, major in college, claimed control over English, exposure to English and patterns of language use in daily life. The variable that correlated most significantly was schooling, that is the type of public school the subjects attended. Because this variable relates to factors such as the number of years the subjects received English language instruction in schools, whether or not these schools had English language instruction and the quality of English instruction in those schools, it is reasonable that having attended a good public school would correlate to good English skills. It is also interesting that 'exposure to English' was a significant predictor of language achievement. What the authors mean by 'exposure to English' is whether the subjects had non-academic exposure to the language in terms of watching English films, television, reading English novels etc. The fact that this correlated significantly with language achievement indicated that besides formal academic exposure to English, informal exposure may also have aided successful acquisition of English. The study also showed that variables such as gender, socio-economic status, family size, attitudes to mother tongue did not have a significant correlation with performance in tense.

Over-extension of the progressive marker has often been reported as a marked feature of IndE (e.g. Sailaja 2012). A study by Sharma (2009) on the typological diversity in New English varieties such as Indian English and Singapore English found that there is a connection between over-extension of the progressive marker and substrate interference. Sharma found that the subjects in her study used the progressive marker for non-delimited habituals (*Every week I'm calling [my parents],* 181) and statives (*Some people are thinking it's a bad job*, 181), which we would not expect in standard varieties of English. Sharma argued that these divergences occurred because the English progressive marker *-ing* has been extended in IndE as a global imperfective

marker. Hindi, which is the substrate language of the speakers examined in Sharma's study, obligatorily marks all finite clauses as perfective or imperfective. Perfective in Hindi is marked with *–ya*. Imperfective is marked with either *rahna*, to indicate an interpretation corresponding to progressive ('He is writing a letter') or *ta* for interpretation as statives ('You love music') and habituals ('I drive'). Because there is a pressure from the substrate to obligatorily mark imperfectivity, IndE speakers arguably extend the English *–ing* to both progressive forms (where *rahna* would be used in Hindi) and to statives (where *-ta* would be used in Hindi) (185). Attested IndE sentences such as *I'm driving my car everyday* and *You are liking music a lot* constitute evidence for this explanation by Sharma.

As with the use of articles, there seems to be a genre or register specific distribution even for divergent use of progressive forms. In a study on register variation in IndE, Balasubramanian (2009) reported that the progressive forms with stative verbs occur more frequently in Spoken Indian English than in Written English (16% vs. 2.8%). Different genres of spoken English such as conversations, spoken academic English and spoken news also have higher rates of stative uses of progressives than Written English such as written news, written sports news, entertainment news, travel, fiction and correspondences. Only in case of written academic English (such as student writings) are stative uses of progressives relatively high (149-51). This goes in line with what Sedlatschek also reported for student essays, which tended to omit more articles than other genres such as press texts.

In this section, studies on the use of tense and aspect in IndE and the specific types of divergences associated with it were reviewed. The studies mentioned here argued that IndE speakers tend to use the progressive form with stative verbs, also linking its variation to different registers of IndE. In general it seems to be the case that non-standard uses of the progressives

were more common in spoken registers and in only certain written registers such as student essays. Some of these points will be addressed in the analysis of English progressive forms by Bengali-English bilinguals.

The next section on methodology describes how the data was coded and analyzed.

7.3. Methodology

This section presents details about the coding of the articles and progressive forms. General details about the corpora that are analyzed in this chapter have already been provided in chapter 4 on the methodology.

7.3.1. Corpora

For the investigation of changes in the use of articles and progressive forms in this variety of English, a corpus of monolingual English production data was analyzed. This corpus includes narrations of two stories based on picture books and sociolinguistic interviews with each subject. The first story is referred to as Story A and the second story as Story B, and subjects were asked to narrate them in this order. The subjects were specifically asked to narrate Story B in the past tense, while no specific instructions were provided on which tense to use in the narration of Story A.

Procedure of the data collection

The procedure of data collection was as follows: the subjects were first provided with a questionnaire in English that asked them for details about their ethnolinguistic backgrounds. The questions not only aimed to ascertain their linguistic background but also the domains in which they use English in their daily lives. Following the questionnaire, they were administered a Proficiency test in English which consisted of two written cloze test passages from the

Cambridge Proficiency Exam, the aim of this test was to assess their proficiency in English[29]. After these steps, the subjects were asked to narrate the stories based on the two picture books (details below) and then to participate in the sociolinguistic interview.

7.3.2. Coding of the data

The story and interview data were transcribed in CLAN and then transferred and coded for different variables in Microsoft Excel. Based on the scores that the subjects received in the cloze test, they were divided into two groups, High Proficiency and Low Proficiency. Participants who received a score of above 30 (out of a maximum possible score of 40) were included in the high proficiency group and those who received a score of 30 or below were placed in the low proficiency group. 10 subjects scored 30 or above in the test and were therefore placed in the high proficiency group and the low proficiency group had 6 subjects. A score of 30 was used as the cut off score for the placement in the high proficiency group because native speakers of English tested with the same proficiency tests in other studies hardly ever scored below 30 in these proficiency tests (e.g. Hettiarachchi & Pires 2015).

As mentioned in chapters 3 and 4, the variationist methodology (e.g. Poplack & Levey 2010, Tagliamonte 2012) was used to ascertain the quantitative distribution of the variables in the corpora. The first step of variation analysis entails determining how often the variants of a variable occur in the body of data. This step consists of providing counts and frequencies of each of the variants of the dependent variable (Tagliamonte 2012: 11-12). Following this methodology, all the occurrences of the features (articles and progressive forms) were first

[29] Agnihotri et. al's (1988) study on tense in IndE showed that despite having attended English instruction schools and being exposed to English for several years, there might be differences between speakers in proficiency levels. These could be related to the quality of instruction received, type of high school attended, attitudes towards English and informal non-academic exposure to the language. Therefore, even though all the subjects in our group were more or less uniform in terms of the duration of their formal academic exposure to English, the speakers could still show different proficiency levels. The results of the test showed that there were indeed different proficiency levels among the speakers despite possible uniformity in their academic exposure to English.

coded. Then, all the divergent occurrences of these variables were identified. The data was then analyzed quantitatively and then qualitatively.

7.3.2.1. Coding of indefinite and definite articles

All the indefinite (*a*, *an*) and definite (*the*) articles were coded and counts for all the tokens of these variables were obtained in the coded data. After getting counts of all the tokens of the indefinite and definite articles, all those occurrences of the articles that showed some distinctive property in comparison with the overall set of occurrences were coded, also considering whether they would be expected in standard varieties of English. The divergences consisted of three main types. The first most common pattern of divergence was a null use or omission of article where we would expect standard varieties of English to have an article *(…and it was dragging Burkey into **ø** water*. (SB8RAI51)[30]); (*He watches over the pond and sees **ø** particular frog*. (SA15SUM6)). Such cases are referred to as 'Null' or 'Omission' of article in the analysis below. The second type of divergence was coded when a definite article was used in place of an expected indefinite article or vice versa (*He goes near **the** particular pond.*[31] (SB15SUM4)); (*is now in possession of his own dog in **a** butterfly net*. (SA13SRA76)). These divergences are referred to as 'Dif' cases in this chapter. The third type of divergence that was coded were instances where an article was used but we would not expect it in Standard English (*He is not at all **a** cheerful*. SA8RAI24)); (*and then he gets ready to go for **the** fishing*. SA15UM5)). The term 'Overuse' is used to refer to these cases. The frequency of these

[30] These data codes represent which corpus (Stories S or Interviews I) the text comes from, the order of each subject on the proficiency scale, an identification of each subject with a 3-letter code (e.g. RAI), and the line number (in CLAN) to which the utterance corresponds.

[31] Note that the occurrence of *the* in this sentence is not ungrammatical. It is infelicitous given the context of the sentence in the corpus, in which the use of an indefinite article *a* was expected before *particular pond* because the noun *pond* had not been introduced in the discourse yet.

divergences was determined in line with the variationist methodology. Finally, chi-squared tests were run on the data to ascertain whether the degree to which the features differed was statistically significant. Chi-squared tests were also used to determine if proficiency in English was a factor influencing the rate of the different divergences.

Because the number of 'Dif' and 'Overuse' cases were very small in the stories, for the interview data, the 'Dif' and 'Overuse' cases were not coded separately, and were combined into the 'Dif' category only. In the interviews, several linguistic factors that might result in the use of divergent forms were also coded: animacy of the noun, presence of an adjective before the noun, whether the noun was count or mass, whether the noun was abstract or concrete and the presence of a quantifier or numeral before the noun. As with the stories, the frequency of divergences was determined for the interviews, and chi-squared tests and logistic regressions (in line with the variationist methodology) were run. These tests aimed to ascertain 1) which types of divergence were significantly higher than others and 2) which of the linguistic factors that was coded was able to predict divergence in the use of articles.

7.3.2.2. Coding of progressive forms

As with the coding of articles, for the progressive forms, all the verbal occurrences of the progressive *-ing* forms such as, *going, seeing, watching* were coded. The initial coding also included occurrences of *–ing* words that were ambiguous between a V+ing form and at least one other grammatical category. Subsequently, the cases that were used in a non-verbal context were excluded, such as, *He has a good <u>standing</u> among his peers* (noun); *The <u>talking</u> bird; I notice a <u>puzzling</u> behavior among my friends* and *The <u>standing</u> crowd applauded* (adjectives).

However, all verbal uses of V+ing were kept in the selected data set, such as *The crowd <u>standing</u> on the street was anxious; A friend saw a police officer <u>standing</u> at the corner; The*

footsteps leading to Tom's house; The cat was standing on the window; What is this master doing?; What is going on?; A boy is going with his dog; The dog was looking idle.

All the divergent forms of the V+ing occurrences were also coded, that is all those cases where we expect a particular tense form such as V+*ed* (simple past) or V+*s* (simple present) but instead found a progressive form. For example, cases such as *He **was leaving** the place and **going** back home* were coded as being a divergent V+ing form because we expected the use of a past tense in the context of the story narration (*He left the place and went back home)*. Even though this is not a divergent use per se, it was considered divergent given the specific context. As was discussed before in the case of some divergent article uses, these sentences are not ungrammatical in themselves but given the context of the utterance, they are infelicitous. In addition, when the V+ing form occurred with a verb that doesn't naturally take a progressive form such as stative or possessive predicates like *He is **having** a pet (*instead of *he has a pet)*, those were coded as divergent forms of the V+ing. Finally, if the form of the *be* verb before the progressive form was unacceptable, those cases too were coded as being divergent. For example, if instances such as *He **is watching** the entire thing* were found instead of the expected *He **was watching** the entire thing* those were coded as being divergent V+ing forms.

Additionally, for the interview data, all the *be* verbs that occurred before the V+ing forms were coded in order to ascertain if the divergence in the V+ing form had to do with the wrong use of the *be* verb. The use of *be* verbs before the V+ing forms normally corresponded to progressive aspect, and were identified as 'Be_prog' forms. [32]

[32] This coding was not done for the stories, where only some of the variables under consideration were analyzed in detail, given the limited number of divergences in V+ing forms.

7.4. Description of divergences

This section discusses all the divergences in the use of articles and progressive forms in the corpora. For each feature, examples of the different divergences are presented along with a quantitative analysis of the divergent patterns. The quantitative analysis also included tests of statistical significance, which aimed to ascertain which features diverged significantly from others. Following that, a qualitative discussion of the divergences is presented, including a discussion of the corresponding Bengali counterparts of these features to evaluate to which extent these divergences were related to interference effects from Bengali. An overall discussion of the findings and the conclusion is presented in section 7.5.

7.4.1. Articles

This section presents a discussion of the different types of divergences that occurred in the use of articles in the stories and the interviews as well as a quantitative and qualitative analysis of the patterns. The use of articles is analyzed for the story narration and then for the interviews, followed by a qualitative analysis of article divergences.

7.4.1.1. Stories

In the use of the indefinite and the definite articles, three types of divergences consistently emerged. First, there were cases where an article was omitted where it would be expected both in standard varieties of English and in other occurrences in the corpus, as shown in (1a) and (1b). In (1a), the indefinite article *a* is omitted before the NP *particular frog*. In (1b) the definite article *the* is omitted before the noun *bucket*. The second type of divergence occurred when an article was used but it is ungrammatical given that in standard varieties of English and in other occurrences in the corpus we would not expect an article in the same kind of context as

shown in (2a) and (2b). Both the articles in (2a) and (2b) are redundant. In (3) we have cases where a definite article was used in place of an indefinite article (3a) and an indefinite article was used in place of a definite (3b). The sentences in (3a) and (3b) are not ungrammatical per se but are infelicitous given the context in which they were used. This is due to the fact that in (3a) we expect the use of indefinite articles before *small boy* and *pond* given that this is the first time that these entities were introduced in the discourse. Therefore this sentence would require the use of indefinite articles instead of definite articles. Similarly in (3b) the noun *Snoopy* has already been introduced in the discourse and therefore a definite article before *poor dog* would be expected instead of an indefinite one. Therefore the use of articles in (3a) and (3b) is infelicitous in the larger context in which the utterances were produced.

1) Null use or omission of article

a) *He watches over the ponds and sees* ***ø*** *particular frog.* (SA15SUM6)
b) ***Ø*** *bucket is on his head.* (SA8RAI75)

2) Overuse of article

a) *The boy's face shows* ***a*** *disappointment.* (SA16TAN67)
b) *Then he gets ready to go for* ***the*** *fishing.* (SA15SUM5)

3) Different use of article: definite for indefinite and indefinite for definite

a) *One day* ***the*** *small boy decided to go* ***the*** *pond.* (SA4KAL1)
b) *... and Snoopy* ***a*** *poor dog also fell.* (SA5KOL30-31)

The section below delves into a quantitative analysis of these main divergence patterns in the data.

Quantitative analysis of divergences

The distribution of divergences in stories A and B for the two proficiency groups is presented below. Table 7.1 presents the frequency of divergences in indefinite articles and table 7.2 presents the same for definite articles.

Table 7.1: Frequency of divergences in indefinite articles across both stories

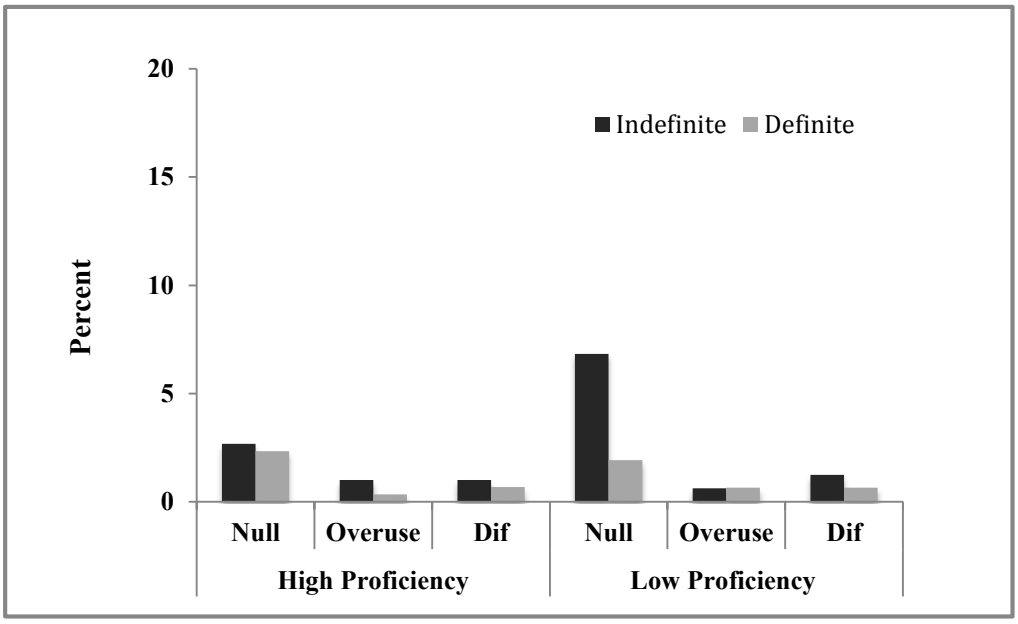
Story A & B	**High Proficiency** N = 299			**Low Proficiency** N = 161		
	Null %	Overuse %	Dif %	Null %	Overuse %	Dif %
	2.67	1	1	6.83	.62	1.24

Table 7.2: Frequency of divergences in definite articles across both stories

Story A & B	**High Proficiency** N = 1498			**Low Proficiency** N = 936		
	Null %	Overuse %	Dif %	Null %	Overuse %	Dif %
	2.33	.34	.67	1.92	.64	.64

Considering tables 7.1 and 7.2, and the summary in figure 7.1 on the next page, we see that the divergence rates in article use was extremely low. Figure 7.1 shows that the divergence rate for any of the features was less than 10%. In fact the highest rate of any type of divergence is about ~ 7% (for null indefinites). When we compare the relative divergence frequencies for indefinite and definite articles in both groups, we find that the low proficiency group had a higher frequency of omission of the indefinite article than the definite article (6.83 % vs. 1.92 %). The difference between the omission of indefinite and definite articles for the high proficiency group is relatively smaller (2.67 % vs. 2.33 %). The frequencies of other divergences such as 'overuse' and 'dif' cases are even lower (~2%) than the frequency of article omission (~ 7%).

Figure 7.1: Summary of divergence rates in article use in Stories across both groups



These results are corroborated by chi-squared tests. When the high proficiency group was examined in terms of omission of definite and indefinite articles (Figure 7.2), no statistically significant difference between the two instances of omission was found X^2 (1, N = 1840) = .117, p = .733. However, the omission rate of indefinite articles was significantly higher than the omission rate of definite articles in the low proficiency group X^2 (1, N = 1125) = 11.783, p = .001, as shown in Figure 7.3.

Figure 7.2: Omission of indefinite and definite articles within high proficiency group.

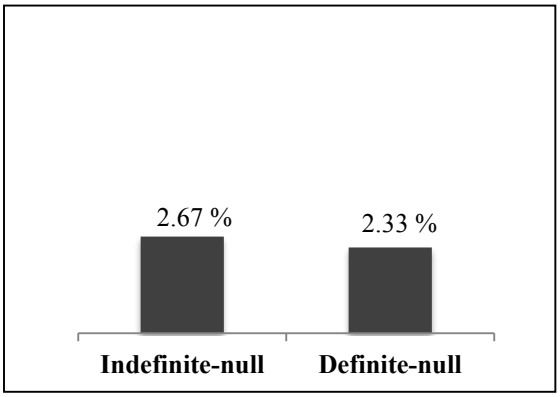
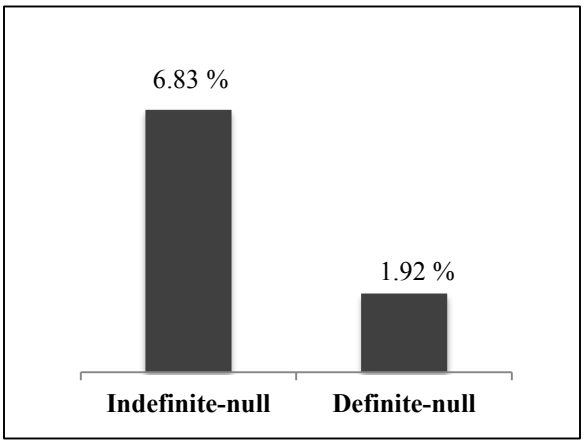
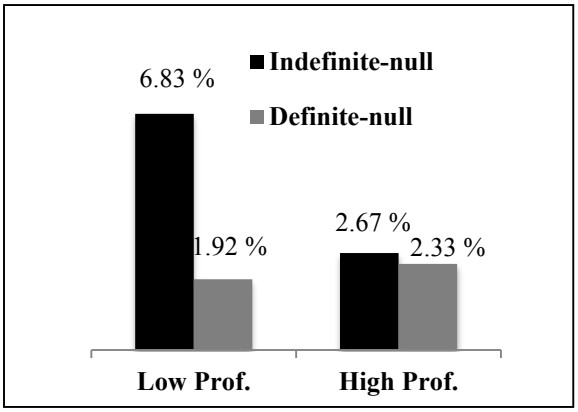


Figure 7.3: Omission of indefinite and definite articles within low proficiency group.



When the two groups were compared with each other in terms of the omission of definite and indefinite articles, the low proficiency group tended to omit indefinite articles more than the high proficiency group, X^2 (1, N = 479) = 4.156, p = .041 (Figure 7.4). There was no significant difference between the two groups in terms of the omission of definite articles, X^2 (1, N = 2486) = .438, p = .508.

Figure 7.4: Omission of indefinite and definite articles in both proficiency groups



Since the divergence rates for the 'overuse' and 'dif' cases were so low (below 2% each), differences between these divergences were not analyzed statistically. Based on the percentages provided in tables 7.1 and 7.2, we find that the high and low proficiency groups are comparable

in terms of divergences in 'overuse' and 'dif' cases. Both groups had very few occurrences of either type of divergence.

7.4.1.2. Interviews

The distribution of divergences for the articles in the interviews was similar to the stories. There was either omission of articles or different use of articles (including overuse and 'dif' cases which consisted of substituting definite for indefinite and vice versa). Table 7.3 presents the frequency of divergence for indefinite articles and table 7.4 for definite articles.

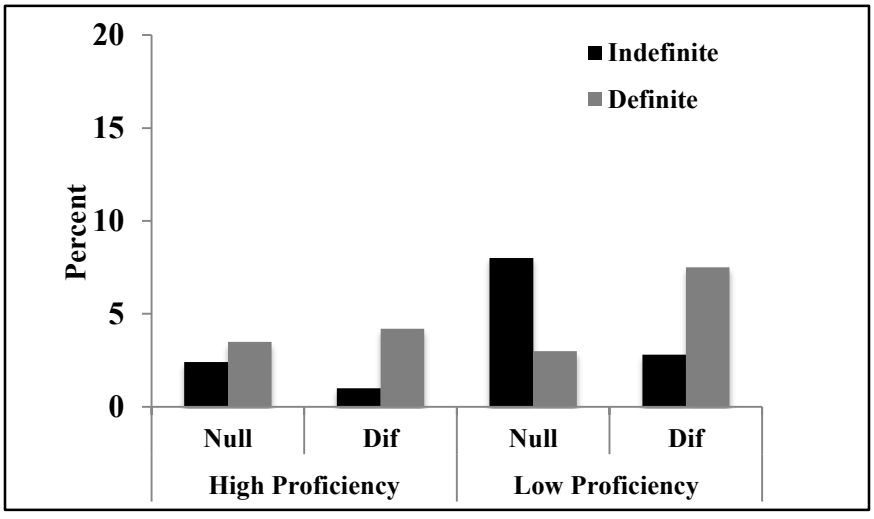
Table 7.3: Frequency of divergences in indefinite articles in interviews

High Proficiency N = 205		**Low Proficiency** N = 138	
Null %	Dif %	Null %	Dif %
2.4	1	8	2.8

Table 7.4: Frequency of divergences in definite articles in interviews

High Proficiency N = 398		**Low Proficiency** N = 259	
Null %	Dif %	Null %	Dif %
3.5	4.2	3	7.5

Figure 7.5: Summary of divergence rates in article use in Interviews across both groups



As with the frequency of divergences in the stories, we find that even the low proficiency subjects had a higher divergence rate when it came to the omission of indefinite articles in the interviews, as compared to definite articles (8% vs. 3%). For the high proficiency group, the difference was much lower between omission of indefinites and definites (2.4% vs. 3.5%). However, an interesting trend emerged in the interviews that we did not find in the stories. In the interviews, we find that the divergence rates increased for both high and low proficiency groups regarding definite dif cases. For high proficiency subjects, their divergence rate for the dif cases increased in case of definite articles (4.2 % for definite articles vs. 1% for indefinites). This increase was somewhat more substantial for the low proficiency group (7.5% vs. 2.8%).

When these frequencies were compared statistically with chi-squared tests, these results turned out to be significant. In terms of the omission of indefinite articles, the low proficiency group had a significantly higher divergence rate than the high proficiency group, X^2 (1, N = 360) = 6.140, p = .013. There was no significant difference between these groups for the omission of definite articles, X^2 (1, N = 681) = .360, p = .548. However the two groups also had a significant difference when it came to the definite dif cases. In this case, the low proficiency group had a significantly higher divergence rate than the high proficiency group, X^2 (1, N = 688) = 9.838, p < .01.

The interview results partially match the results from the story data where we found that the low proficiency group tended to omit more indefinite than definite articles. Also, as with the stories, when it came to the omission of definite articles, the two groups had no significant difference. In addition, in the interview data not only do we find that the low proficiency group tended to omit more indefinite articles, but it also had a significantly higher divergence rate for the definite dif cases. This may be due in part to the nature of the interviews, which aimed to

elicit less structured responses than the stories, which consisted of more carefully produced utterances, in addition to being a more restricted discourse domain corresponding to the picture books. Therefore, presumably, speakers were paying more attention to their speech during the narrations of the stories than the interviews. Overall, based on the results of the story and interview data, it seems that the low proficiency group had slightly less control over the use of indefinite articles than definite articles. Also, considering what had been suggested in earlier research on the effect of different genres of spoken data in the use of articles, in this examination, we do not find the subjects performing very differently in the interviews than in the story narrations. Because the story narrations were more structured in terms of production, if there was an effect of conversation type on the use of articles, we would expect the interviews to have more substantial cases of divergences. We still found a significantly higher rate of divergences for the definite dif cases in the production of the low proficiency group in the interviews, which was not seen in the stories.

The results of the chi-Squared tests were also supported by a regression analysis. When it came to the omission of indefinite articles, proficiency of the speakers seemed to be a significant predictor (p = .026). That is, whether or not the speakers were proficient could predict the omission of indefinite articles (low proficiency speakers were more likely to omit indefinite articles than high proficiency speakers). However, proficiency was not a significant predictor regarding the omission of definite articles. In terms of the dif cases, proficiency could not significantly predict whether speakers would have problems with indefinite articles, but it did so in case of definite articles (p = .001). None of the other factors that were coded for such as animacy, abstractness, quantifiers, adjectives, count vs mass were significant predictors of divergences in the use of articles.

A more detailed discussion of these statistical tests and their results is provided in the discussion section. The next section provides a qualitative analysis of the divergences and compares the Bengali determiner system with the English article system, aiming at ascertaining if the divergences in the use of English articles could have sources in Bengali.

Qualitative analysis of article divergences

This section focuses on a qualitative analysis of article divergence. The focus will particularly be on the omission of articles rather than the 'overuse' or 'dif' cases because the frequency of these cases is very low (less than 1.5%) and they consist of very few examples to allow relevant consideration of their properties. As far as omission of articles was concerned, some interesting patterns emerge. For example, articles (both definite and indefinite) were sometimes omitted when the NP contained some modifying element. These could be numerals, as shown in (4a) and (b), adjectives as in (4d), (e), (f) and quantifiers (g) and (h).

4) Omission of articles before modifiers (numerals, adjectives, quantifiers)

a) ..... *so* ***ø*** *three of us again me, Burkey, and the new friend frog.* (SB8RAI46)
b) ***ø*** *three of them, they are delighted to see one another.* (SA16TAN91).
c*) Howrah still has got foundries, which are probably* ***ø*** *best in the world.* (IN4KAL)
d) ***ø*** *main thing is that you do not have the confidence.* (IN9RAJ)
e) *and* ***ø*** *whole day we had to spend in Tashkent airport.* (IN7LAK).
f) …… *because he had* ***ø*** *severe kidney problem*. (IN9RAJ)
g) *this place only has two,* ***ø*** *couple of drawbacks.* (IN14SUD)
h) *the city has a lot of greenery,* ***ø*** *lot of places which are sports-oriented.* (IN6KUM)

Omission of articles before modifiers and quantifiers also played a role in Sharma (2005a)'s discussion of IndE data, where she found that when nouns had modifiers and quantifiers before them, articles tended to be dropped. Sharma's rationale for this omission is that the modifying elements such as *whole, lot, best, main* uniquely identify the referents of those

modifiers from a set of real-world objects. Therefore, apart from modifying a noun, they also disambiguate a particular referent and therefore the need for an article (which also functions as individualizing a referent) becomes less crucial (557-58). These observations also corroborate Agnihotri et. al's (1984) study which found a similar tendency among the subjects, that is to omit articles when the NP contained an adjective or superlative quantifier. The common point between adjectives, superlatives and quantifiers is that they make the entity they modify more specific or more identifiable, therefore making the use of the determiner less vital. The cases of omission listed here can be accounted for with this argument: omission in instances of NP modification occurred for both definite and indefinite articles as shown in the examples in (4).

Sharma's study also indicated that definite articles tended to be omitted when the nouns had already been introduced and identified in the discourse. An example she provides is *When I had completed the training, Ø ten day training at the language school*. In this sentence, the noun *training* is first introduced with a definite article, since it referred to a particular training. Also it was the first time that the noun was introduced in the discourse. However, once the introduction was done and the noun was anchored in the discourse, the definite article before *ten day training* was dropped. This was because the noun *training* was already anchored in the discourse and identified as a definite element and therefore the need for an article before *ten day training* became less crucial and could therefore be dropped.

Some of the definite article omissions in the corpus data can also be explained with the help of this discourse identification process. Consider examples (5a) and (b) where the article was omitted only after the noun was anchored in the discourse first. In (5a), the *Puppy* has already been introduced to the listener as a definite element. Once the definite interpretation has

been established, dropping of the article becomes acceptable to the speaker, because the *Puppy also sees*…. refers to the particular puppy that had been mentioned earlier.

5) Omission of article when noun has been anchored in the discourse

a) *The puppy's having a nice time taking rest near the bank. Ø Puppy also sees what is his little master doing.* (SA9RAJ33)

b) *The boy and the dog come face to face. Ø Boy tries to catch the frog with his hands.* (SA11SAY18)

Apart from article omission in the NP modification cases, omission after discourse anchoring is the other frequently occurring pattern in the data. There are other types of divergences besides these but they consist of even fewer examples, making it difficult to provide conclusive explanation for their behavior. The examples provided below consist of cases where omission of a definite article occurred before generic nouns.

6) Omission of articles before generic nouns

a) *If you go to ø mountains, people there are like very easy-going.* (IN6KUM)
b) But I feel ø sea is very dynamic. (IN8RAI)
c) *My love for ø hills started from there.* (IN16TAN)
d) *and it was dragging Burkey into ø water.* (SB8RAI51)
e) *just when I was ready to go get into ø water the dog appeared.* (SB11SAY24)

Bengali uses bare nouns in the case of generics and this might be causing omission in the use of English generics. (This seems to be a common phenomenon cross-linguistically. For example, Baptista & Gueron's (2007) examination of several creoles displayed a similar tendency of linking bare nouns with genericity). The source of these omissions may therefore be due to an interference effect from Bengali. For example, (6a) and (b) would not have had any classifiers (which are used to mark definites in Bengali) before them if they were to be translated in Bengali, as shown in (7a) and (b).

7) Use of bare nouns in Bengali with generics

a) tumi jodi **pahar-e** ja-o, okhane lok-era khub shotosphurto
2SG if mountains-GEN go-HAB.3P there people-PL very easy going
If you go to (the) mountains, people there are like very easy-going.

b) amar mon-e ho-e **shomudro** khub chonchol
1SG-DAT mind-GEN be-HAB.3P sea very dynamic
I feel (the) sea is very dynamic.

(8a) and (b) however correspond to omission of articles before proper nouns. Since *Everest* and *Black Sea* both refer to very specific referents, the need for a definite article in both cases is perhaps perceived as unnecessary, which is why the speaker omitted it. These omissions also relate to the omission of articles in case of discourse disambiguation. Because *Everest* and *Black Sea* refer to unique referents, the omission is not perceived as infelicitous by the speaker.

8) Omission before proper nouns

a) They worship Kanchanjanga. They worship ***ø*** Everest. (IN16TAN)
b) I learnt swimming in ***ø*** Black Sea (IN7LAK)

Next, the article system of Bengali is discussed in order to determine if there are interference effects from Bengali grammar onto the use of English.

The article system of Bengali

When it comes to the use of articles, Bengali operates on what Thompson calls "a need to know basis" (2010: 93), that is the distinction between the definite and indefinite is made clear only when necessary. Usually, the classifier *–ta* is used to mark definites in Bengali, as shown below.

9) Definite marking in Bengali

(a) chele-ta khub dushtu
boy-CL very naughty
The boy is very naughty.

However, bare nouns can also get a definite interpretation without a classifier such as the examples below.

(b) dekh-o aaj kemon chaad uth-ech-e
see-2P today how moon rise-PFV-3P
Look how (the) moon has risen today!

(c) ma chele-ke bok-ch-e
mother boy-ACC scold-PROG-3P
(The) mother is scolding (the) boy.

Since *moon* has a unique referent, the use of a classifier to give it a definite interpretation seems to be unnecessary. Baptista & Geuron (2007) argued that for the creoles they examined, unique entities like *moon* or *sun* are familiar and easily identifiable and could therefore appear bare. Such a generalization seems applicable here. But, in example (9c), even though there are no classifiers attached to *mother* and *boy*, they still get a definite interpretation because the context would make that clear. Therefore such a sentence would be infelicitous (though grammatical) at the beginning of a discourse, narration or conversation. However, if some background information about the *mother* and the *son* has been provided or a context has been established, then using bare nouns is perfectly acceptable, as shown below.

d) o-der bari-te ki ho-cch-e? ma chele-ke bok-ch-e
3PL-GEN home-LOC what be-PRS.PROG-3P mother boy-ACC scold-PROG-3P
What is happening at their house? (The) mother is scolding (the) boy.

In the example above, the first sentence (*what is happening at their house?*) has implicitly identified a specific family. Therefore, when *mother* and *son* are mentioned in the following sentence, it is clear as to which *mother* and *son* are being referred to. In this case, the *mother* and *son* of a specific family. Therefore, the subsequent use of bare nouns with *mother* and *son* is not infelicitous.

Indefinite marking is usually achieved by the use of the numeral one 'ek' along with the classifier *–ta,* as shown below.

10) Indefinite marking in Bengali

a) ek-ta chele-ke aaj dak-l-am
one-CL boy-ACC today call-PST-1P
I called a boy today.

Bare nouns denote not only a definite reading but may also indicate indefiniteness, as shown in (10b) and (c) below. In (10b) there is no article before *mango* and in (c) there is, but both instances have an indefinite interpretation. The example in (b) connects to what Thompson (2010) calls "need-to-know basis". The distinction between definite and indefinite is made only when necessary. So, (b) could be used as a perfectly acceptable answer to a question such as, *what are you eating?* To this one could reply with both (b) and (c). But if (b) is the answer to such a question, then it is ambiguous between a singular and plural reading. This is because *aam khacchi* 'eating mango' could mean both 'one mango' or multiple mangoes depending on whether the focus is on how many mangoes were eaten or rather on the 'mango eating' event. But *ekta aam khacchi* 'eating a mango' clearly provides a singular reading.

(b) ami aam kha-cch-i
1SG mango eat-PRS.PROG-1P
I am eating (a) mango(es).

(c) ami ek-ta aam kha-cch-i
1SG one-CL mango eat- PRS.PROG-1P
I am eating a mango.

Bengali is an atypical Indo-Aryan language in that it has a classifier system (Thompson 2010) that languages like Hindi, Punjabi or Marathi lack. Several studies on article omission in IndE have cited the lack of articles in Indian languages as a cause for the omission of articles.

The omission of articles in the English of Bengali speakers may result from factors that act in conjunction with interference from Bengali. Since Bengali has grammatical mechanisms (use of classifer *ta* for definites and *ekta* numeral 'one' for indefinites) to make the indefinite and definite distinction, unlike Hindi or Punjabi (though for Hindi and Punjabi, there are other ways to make that distinction clear), omission of articles in English may also be caused by the process of 'discourse disambiguation' as discussed in Sharma (2005a). When the discourse has established the referents, articles may be omitted. Despite this provision, speakers by and large stick to the English rules of article use and stray only when these discourse factors step in to lessen the burden of infelicity.

This section provided a quantitative and qualitative account of article divergence in the story and interview data along with a discussion of the Bengali article system in order to trace possible interference effects in this variety of English. The next section provides a discussion of all the findings in the article data and the broader implications of those results.

7.4.1.3. Discussion of findings from articles

The story and interview data revealed that the overall frequency of article divergence was quite low, and only about 7% of article uses diverged from Standard English norms and from the overall pattern of use in the corpus. For the story narrations, comparisons between definite and indefinite articles showed that the low proficiency group had a significantly higher rate of indefinite article omission than definite articles, and comparison between the two groups showed that the low proficiency group tended to omit indefinite articles more than the high proficiency group. There was no significant difference between the groups in terms of the definite article omission. Other divergent uses of articles such as 'overuse' or use of definites in place of indefinites and vice versa (dif cases) constituted less than 2% of all divergences, and the two

groups did not differ much in terms of the other two divergence types. Both groups had mostly similar divergence rates for the 'overuse' and 'dif' cases.

A very similar pattern was also found for the interview data where the low proficiency group tended to omit indefinites more than definites and between-group comparisons showed that the low proficiency group omitted more indefinites than the high proficiency group. The groups did not have any significant difference between them in terms of the definite article omission. But unlike the pattern with the story narrations, the low proficiency group had a significantly higher rate of definite dif cases than the high proficiency group in the interviews. They used the definite article where it was not necessary (overuse) and where we would expect the use of an indefinite article more often than the high proficiency group. In the stories, there was no difference between the groups in terms of definite dif cases. Increase in divergence of definite article use in the interviews may also be related to the type of task that the subjects had to do. During the narration of stories, participants may have been more careful of their speech and therefore paid more attention to what they were saying. However, the interviews were meant to elicit more casual speech. As a result, speakers might have been prone to more 'mistakes' during the interview task.

The difficulty that low proficiency subjects had with indefinite articles is similar to what Lardiere (2004) reported on her longitudinal study of English acquisition by an L1 speaker of Chinese. Lardiere reported that even though Chinese lacks articles, her subject had a reasonably good control of article use. Divergent uses of articles constituted about 20% of overall article use, and her data revealed that even though Chinese marks indefiniteness with the help of the numeral *one*, and lacks definite articles, this subject tended to omit more indefinities than definites (332-33). This is similar to what we find in this data and the reason for this tendency

may therefore emerge from other factors, such as proficiency, that operate in conjunction with interference from the L1. These findings also corroborate Sand's (2004) report on IndE article use, where she found indefinites being omitted more than definites. One hypothesis that could be further explored, regarding the tendency to omit less definites than indefinites, is that definites usually refer to entities more relevant to speakers than indefinites. Therefore, definites are somewhat more salient to speakers than indefinites and therefore get omitted less.

Because the divergences in this data are sparse, it is difficult to draw more definitive conclusions about why they might have occurred. But based on these findings and what has been described before in the work of Agnihotri et. al (1984) and Sharma (2005a), it seems that there is an occasional tendency to omit articles in contexts where referents of the articles have been uniquely identified and nouns have been anchored in the discourse. Articles are omitted only once this anchoring has been done so that there is no scope of ambiguity in identifying the referents. We also find the same principle operating in complex NPs containing adjectives and numerals. These modifiers helped in uniquely identifying the referents of articles as a result of which omission becomes less infelicitous. However, even though this option of omission is available to speakers, they rarely use it and prefer to stick to the standard norms of article use. But, unlike what Sharma (2005a) found in her study, where both proficient and less proficient speakers tended to omit definite articles when the discourse had disambiguated the noun, in the results reported here we find that high proficiency speakers omit articles very rarely and all the speakers by and large tend to obey the rules of standard English article use. However, interference effects from Bengali in the use of English also seemed to be playing a partial role, as discussed regarding the use of generic nouns.

Proficiency in English is a significant factor in predicting article divergence in this data. Low proficiency in English was a predictor for speakers to be more likely to have difficulty with articles, and more so with indefinite articles than definite articles. This is because indefinites tended to be omitted more than definites by the low proficiency group. However, apart from proficiency we do not see any other factor being able to predict divergence significantly. Even though Sharma's study reported NP modification to be a significant factor in predicting divergence, the regression analysis (where the occurrence of adjectives and quantifiers before nouns was coded) did not show any such effect. This is most likely because the divergences in this data were very few in number.

This study also supports the 'cline of proficiency' described in Kachru (1994), which suggests that there is a continuum of proficiency when it comes to English in India and the high proficiency speakers represent an acrolectal variety of English, which is closer to standard varieties. In addition, as it has been reported extensively in the second language acquisition literature, higher proficiency in a language is directly connected to the number of years of exposure to the language and the age at which it was learned. Since all the subjects in this data were exposed to English from a very early age (roughly 4 years, except for 2 subjects) and underwent several years of schooling in the language, they are all expected to be highly proficient in the language (notwithstanding their differences in proficiency level), and this shows in the frequency of their article divergences. Although the cloze test results indicated differences in proficiency among the subjects, even the low proficiency speakers in this study can be considered to be high on the proficiency scale.

From these findings it can be said the subjects have a very good control over the use of articles, most likely due to the fact that they are highly proficient in the language from being

introduced to it at an early age and having years of exposure to it. A. Also, because the speakers in general are high up in the proficiency scale they usually comply with the norms of Standard English article usage and only occasionally diverge from such rules when it is not completely infelicitous to do so. These results provide evidence to argue for stability in this variety of English, at least in the use of articles.

7.4.2. Progressive forms

This section focuses on the progressive forms in the story and interview data.

In the use of the progressive forms, three main types of divergences were noted. The first type of divergence was when we expected a different tense in place of the progressive form such as a simple past or present tense, as shown in (11a) and (b), but instead find a progressive form. In (11a) we expect the verb *attacked* in place of *attacking*. In (b) we expect the use of present tense *They go home tired* but find the progressive form *going*.

11) Use of progressive form in place of a full tense form

a) Progressive for simple past
*The tortoise again **attacking** us.* (SB4KAL44)
Expected: *The tortoise again **attacked** us.*

b) Progressive for present
*They're **going** home tired.* (SA9RAJ68)
Expected: *They **go** home tired.*

The second type of divergence that was encountered was not in the progressive form per se but in the use of the *be* verb that occurred before the verb. In these cases, the *be* verb had the wrong tense form. For example, in (12) we expect *was* instead of *is.*

12) Use of wrong tense of *be* verb
*....and the frog **is** quietly observing.* (SB8RAI30)
Expected: *....and the frog **was** quietly observing.*

As was mentioned in the case of articles, the sentences listed under (11) and (12) are not ungrammatical as is but the context of the utterances makes them infelicitous. This is because speakers would often switch tenses within the narration. So, after starting to narrate the story in the past or the present tense, they would switch to the progressive form mid-way. Then having realized the mistake they would go back to the original tense or continue narrating using the progressive form.

The third type of divergence listed in (13) had to do with the use of the progressive *-ing* form with stative verbs (such as *have*), which do not take the *–ing* form. This has been widely attested in the literature on IndE as a typical IndE feature (e.g. Sailaja 2012).

13) Over extension of progressive to stative verbs

a) *...and the new friend frog **having** some new plans to execute.* (SB8RAI47)

7.4.2.1. Stories

The frequency of divergences in the progressive forms in stories A and B across both proficiency groups is presented below. Table 7.5 presents the frequency of the divergent progressive forms in Story A and table 7.6 presents their frequencies in Story B.

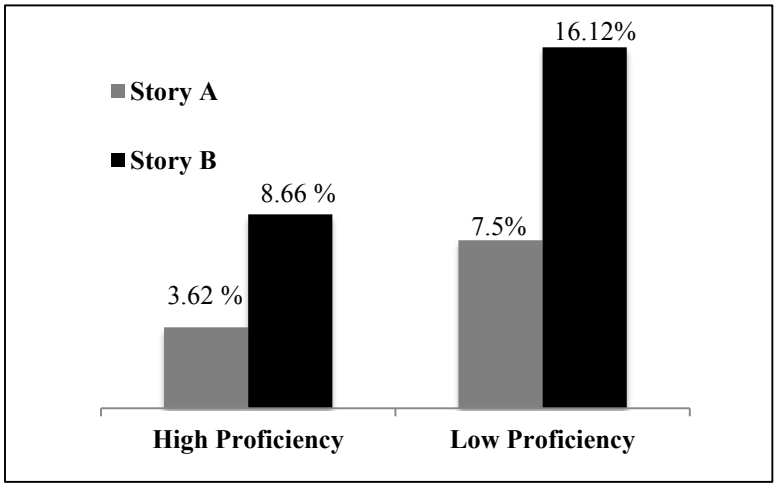
Table 7.5: Frequency of divergences for progressive forms in Story A

Story A	**High Proficiency** N = 248	**Low Proficiency** N = 160
	Dif % 3.62	Dif % 7.5

Table 7.6: Frequency of divergences for progressive forms in Story B

Story B	**High Proficiency** N = 277	**Low Proficiency** N = 155
	Dif % 8.66	Dif % 16.12

Figure 7.6: Divergence rates in progressive forms in high and low proficiency groups in Stories A and B.



From the above tables we find that the average divergence rates for the progressive forms to be about ~ 9%. As was the case in the articles, even for the progressives we find that the low proficiency group had higher rates of divergences in comparison to the high proficiency group (7.5% vs. ~4% in Story A; 16% vs. ~9% in Story B). This is true for both stories. Interestingly, we also find that the divergence rates almost doubled for both groups in Story B. Story B had to be narrated in the past tense, which led to more mistakes than Story A because the subjects had difficulty in maintaining the past tense and would often switch to the progressive forms. This

resulted in a higher frequency of divergences for both groups. These results were also supported by the chi-squared tests of statistical significance.

Based on the chi-squared tests it was found that the low proficiency group's rate of divergence was significantly higher than the high proficiency group's rate of divergence, X^2 (1, N= 910) = 6.425, p = .011, for both stories. The chi-squared tests also showed that divergences in Story B were significantly higher than in Story A for all subjects, X^2 (1, N = 910) = 8.943, p = .003. As mentioned earlier, because Story B was narrated in the past tense, the subjects experienced difficulty in maintaining the past tense form and would often switch to the progressive form. Therefore there is an effect of task-type that drove the divergences higher, and for both proficiency groups, divergence rates went up significantly in Story B than in Story A. That is, the high proficiency group had a significantly higher divergence rate in Story B than in Story A, X^2 (1, N = 558) = 4.982, p = .026, and so did the low proficiency group, X^2 (1, N = 352) = 4.467, p = .035.

The discussion of divergences in the interviews is presented below.

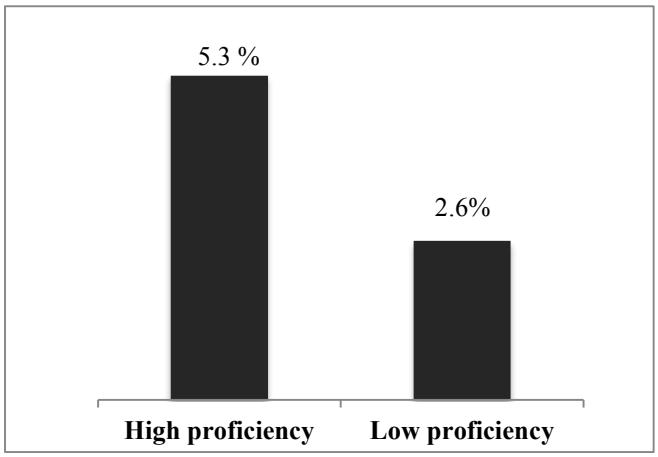
7.4.2.2. Interviews

The table and figure below present the frequency of divergences in the progressive forms in the interview data.

Table 7.7: Frequency of divergences for progressive forms in interviews

High Proficiency N = 143	**Low Proficiency** N = 74
Dif % 5.3	Dif % 2.6

Figure 7.7: Divergence rates in progressive forms for both groups in Interviews



In the interview data, we find the opposite of what the story data showed. This is because in the stories, as expected, the low proficiency group had a higher divergence rate in the progressive forms than the high proficiency group. But in the interviews we find the opposite to be the case: the high proficiency group had a higher frequency of divergence than the low proficiency group. However, there was no statistical difference in divergence rates between the two groups, as showed by the chi-square test, X^2 (1, N = 227) = .200, p = .755. Also, from the regression analysis we did not find proficiency to be a significant factor in predicting divergence (p = .720). However, the regression analysis showed that the presence of a *be* verb (what we called the Be_prog form) before a progressive form was a significant factor in predicting divergence (p = .039). That is, if there was a *be* verb before the progressive form then it is likely for an error to occur in the use of the progressive form.

Qualitative discussion of progressive forms

This section provides a qualitative analysis of divergences in the use of progressive forms in the story and interview data. First let us consider the divergences in the stories. The divergences in the stories were mainly related to an issue of maintaining the use of a particular

tense. In Story A, the speakers were not provided with any instructions on how to narrate the story, so they could choose any tense of their preference. However, even though they started the story with a particular tense, they often switched tenses mid-way. This resulted in some increase of divergence rates. In Story B where the speakers were specifically instructed to narrate it in the past tense, they started the stories in the past tense but would often switch to the progressive form during the narration, as illustrated in (14a). That is why for both high and low proficiency speakers the divergences almost doubled in story B. This suggests that the divergences had less to do with proficiency or competence in English but more to do with performance in the specific task. However, overall proficiency may also have affected their production in this task, given that completely proficient speakers of English might not have made similar mistakes in this task.

14) Switching of tense

a) *And the frog once again regained his seriousness, his meditative mood and sitting just at the center.* (SA8RAI80-82)

In (15a) the divergence lies in the wrong use of the *be* verb. In this example we expect the past tense form *was* both before *quarreling* and *looking* but instead *is* was used. This is another example of the difficulty in maintaining the use of a particular tense.

15) Wrong tense of *be* verb

a) *I was terribly afraid for my dog and I saw that the tortoise is quarreling with my dog and the frog is looking into it.* (SB15SUM15-17).

The bulk of the divergences in the stories were related to the issue of being unable to maintain a particular tense. Unlike what several studies on IndE have reported about the widespread tendency to use progressive forms with stative verbs, we in fact find only six instances in the stories and four in the interviews where the progressive *–ing* form was used with stative verbs.

An example of this divergence is provided in (16a) below. In this example, the verb *living* is used to indicate the state of being *alive,* which makes the use ungrammatical.

16 a) ... *now the uh tortoise ……. was very much* <u>*living.*</u> (SB9RAJ72)

The use of progressive forms with stative verbs amounts to less than 1% of all progressive divergences and is therefore extremely infrequent in this data. Since this data represents the speech of highly proficient speakers, the rare occurrence of this particular non-standard feature (which has been reported as a typical feature of less proficient speakers in Indian English) is not surprising.

The results from the interview data did not match what was observed in the story narration. In the interviews the overall divergence rates in the use of progressive forms were quite low at about 5%, and there was no significant different between the two proficiency groups regarding their rate of divergence. However, a closer look at the interview data showed that one particular speaker from the high proficiency group carried the bulk of all the progressive divergences, and this skewed the distribution for the high proficiency group. Therefore we cannot say that the high proficiency group overall did not perform well in the use of progressives. If we eliminate that particular speaker's output then we find that the divergence rate in the high proficiency group goes below that of the low proficiency group. It is difficult to point to the reason why this particular speaker carried the bulk of divergences. Like most other speakers he too had been exposed to English from age 4 onwards and had a score of 35 in the cloze test, which indicates that he is quite high on the proficiency scale.

As discussed earlier, most of the divergences related to the progressive forms had to do with a difficulty in maintaining tense, which could be related to a difficulty the speakers had to

adhere to the sequence of tense rule in English. A follow-up study that could be done is to investigate sequence of tense properties in Bengali to determine whether such violations arise in English due to interference from Bengali.

7.4.2.3. Discussion of findings from progressive forms

This section discusses the findings from the quantitative and qualitative analysis of the divergences in the progressive form.

The quantitative analysis showed that the overall performance of the speakers in the stories and interviews in terms of the use of progressive forms was quite good. There is enough evidence to argue that speakers had very good control of the progressive forms per se and the divergences were mainly related to the difficulty of being unable to maintain the use of a particular tense within the discourse. Therefore, if we consider the progressives forms in isolation, the use of progressives were mostly correct (except the few instances where the progressive forms were used with statives and possessives). However, within the context of the discourse, various progressive uses were infelicitous.

In the stories the low proficiency subjects had a significantly higher rate of divergence than the high proficiency subjects and these rates doubled in Story B in comparison to Story A. This is because Story B had to be narrated in the past tense and speakers would often fail to maintain the past tense and switch to the progressive form within the narration. We find that divergences in Story B were significantly higher than in Story A for both groups. Therefore the divergences were an artifact of the task that the subjects had to do (which was to narrate the story in the past tense) suggesting that they did not result from a lack of competence but in fact may result from a processing difficulty in the production of the correct tense forms for the progressive.

In the interviews the overall divergence rates in the use of progressive forms were quite low, at about 5%, and there was no significant difference between the two proficiency groups regarding their rate of divergence. There is also no particular tendency for these speakers to use the progressive with stative verbs (this type of divergence corresponded to less than 1% of all progressive form divergences). This data shows that unlike what was reported before for IndE, the use of progressives with stative verbs was quite rare in the results presented here.

7.5. Conclusion

This chapter examined the English spoken by bilingual speakers of Bengali and English in West Bengal, India particularly in terms of the use of the definite article *the* and the indefinite articles *a, an* and the progressive *–ing* form. These particular features were examined because they have been widely discussed in the literature as being divergent and being a marked feature in IndE.

The investigation revealed that divergences in the use of articles and progressive forms were quite low in both high and low proficiency groups, at about ~15%, indicating that these speakers have a very good control over the use of these features. For the divergences that did occur, we saw that the low proficiency speakers had more divergences in the use of indefinite articles than definite ones, while the high proficiency speakers did not have any more or less difficulty with either type of article. Even for the progressive forms, the low proficiency speakers had higher divergences and the divergences resulted from performance issues, since the speakers did not have any difficulty in using the progressive form in general but would often experience difficulty in maintaining the use of a particular and switch to the progressive form from the past or the present tense. Additionally, interference from the L1 may play a partial role in the use of both of these features.

There was an effect of proficiency in predicting divergences in English, in that low proficiency speakers had significantly higher divergence rates than high proficiency speakers. However, even for the low proficiency group, the divergences were less than 20%. This suggests that the low proficiency speakers in this study were on the higher end of the proficiency scale, while the high proficiency speakers represented an acrolectal variety of English, possibly very close to standard, monolingual varieties and what Kachru (1994) calls "educated South Asian English". Based on these findings regarding this particular subject pool, one can argue that this variety of English is quite stable and does not show any large-scale changes, at least for these particular features. Speakers in fact were sometimes innovative in their use of articles and progressive forms only insofar as discourse factors permitted them to do so. Even so, these innovations were few and overall complied with what we would expect in standard varieties of English.

Follow-up research could be done complementing this study to test really low-proficiency speakers in comparison to the high proficiency subjects of this study. One could also compare low-proficiency adult speakers to Bengali children in the early stages of acquiring English. It would be reasonable to expect different results from a subject pool with more distinct proficiency levels, including adults and children.

Chapter 8: Conclusion

In this dissertation, my overarching aim was to throw light on the understudied language contact situation in West Bengal, India where Bengali and English have been in contact for more than two hundred years. My interest was specifically on the bidirectional changes that may have occurred in the two languages, meaning changes in Bengali induced by English and changes in English induced by Bengali. The specific questions that have driven this research focused first on ascertaining changes in Bengali, including the most clearly identifiable changes, such as lexical borrowings, and those that are more subtle in nature at the structural level. Second, I aimed to investigate the degree to which different changes have affected bilingual speech in comparison to monolingual speech, given that bilingual speakers have had years of exposure to English and regularly code-switch between the languages and monolingual speakers either have indirect exposure to English through the speech of bilingual speakers or through a few years of instruction in the language in high schools. The third question deals with teasing apart contact-induced changes from language-internally motivated ones or those that may have emerged from multiple causation. Fourth, I investigated changes in the English spoken in this bilingual community to identify possible interference effects from Bengali.

I grounded the investigation of these research questions in a rigorous methodology consisting of synchronic and diachronic comparisons of modern monolingual and bilingual Bengali with an older variety of Bengali and a detailed investigation of English spoken by bilinguals in this community. I examined four corpora with the purpose of identifying changes in Bengali and English. First, I examined two corpora of bilingual Bengali-English and

monolingual Bengali speech data that I collected through fieldwork in West Bengal, India. I complemented this investigation with a diachronic comparison between both the monolingual and bilingual varieties of modern Bengali and an older variety of Bengali, as represented by nineteenth-century Bengali plays, following the framework proposed in Thomason (2001) to ascertain contact-induced language change. These plays were published at a time when contact between English and Bengali was not nearly as intense as it is now. Therefore, I provided a reasonable picture of what Bengali looked like prior to the development of intense contact with English. Second, I examined a corpus of English speech data collected from bilingual Bengali-English speakers (through sociolinguistic interviews and story narrations) to ascertain the changes that may have occurred in their speech from possible interference effects of Bengali or independent second language effects. I combined these investigations with quantitative analyses, which aimed to gauge the frequency rate of the different features (as discussed in Poplack & Levey 2010) in the four corpora. My goal was also to find out whether there was a significant difference in the occurrence of these features between monolingual and bilingual speech. The quantitative analyses also gave an indication of the extent of particular changes, that is whether the changes were major or minor.

Among the changes that have taken place in Bengali, I first addressed the use of English loanwords in Bengali in the speech of both monolingual and bilingual speakers. The speech of monolingual Bengali speakers, who necessarily come into contact with bilingual speakers, showed the occurrence of few English loanwords and these loanwords were sometimes phonologically adapted in Bengali. The loanwords in the speech of monolingual speakers were not used in place of Bengali basic vocabulary items, but were mostly used to complement the Bengali basic vocabulary. In comparison, bilingual speakers had many more uses of English

words in their speech, including both basic and non-basic vocabulary. From the investigation of bilingual speech alone, it is difficult to know if the English words were loanwords in Bengali or code-switches from English. This is especially difficult since both loan words and code-switches are morphosyntactically integrated in Bengali through the use of postpositional markings. One way by which they can be teased apart is by tracing them to the speech of monolingual speakers. That is, if monolingual speakers use these words in their speech, they are most likely loanwords, given that the monolinguals have minimal knowledge of English. In this context, I also addressed CS in the speech of bilinguals and described some of its patterns in Bengali-English speech. I also discussed how CS acts as a mechanism by which other contact-induced changes take place (in line with previous studies) and focused on the use of bilingual Bengali-English verbs in bilingual speech as a result of extensive CS.

I investigated the complex verb structures in modern monolingual and bilingual Bengali in comparison to the older texts and discussed how bilingual verbs (Eng N/V + Beng 'do') occur much more frequently in bilingual speech than in monolingual speech and indicated that they were non-existent in the nineteenth-century texts. I argued that these verbs constitute morphosyntactic changes in the way they are reanalyzed in bilingual speech.

The investigation of contact-induced change in a language may often reveal changes that have other sources besides contact. There may be changes that have occurred from internal developments alone or from both contact and internal developments within the language (multiple causation). I provided evidence of such a change in Bengali in the domain of copular predicates. The examination of nineteenth-century Bengali plays indicated the absence of copulas in present tense equational sentences. However, modern Bengali (both bilingual and monolingual) shows the occurrence of copulas in present tense equational clauses in the

sentence-medial position as opposed to the sentence final position where other verbs in the language occur. The frequency of this copula is significantly higher in bilingual speech than in monolingual speech. We also find that the sentence-final position can be occupied by a homophonous *be* verb, whereas this particular copula can only occur in sentence-medial position. As this distribution is considered in conjunction with the sentence-medial position of the copula (which is similar to the position of English verbs) and its absence in the older texts, it provides potential support for a case of contact-induced change. However, we also find that the copulas in all other tenses occur in the usual sentence-final position. This rules out the possibility that contact is the only cause behind this change, supporting the view that it is more likely a result of multiple causation where both language-internal factors and influence from English led to the emergence of the sentence-medial copula. We also assume that the sentence-medial position may have been reinforced by the need to disambiguate between the genuine copula and the homophonous semi-copula in Bengali.

In the investigation of changes in English, particularly in the use of articles and progressive forms, I found that the frequency of divergent uses of articles and progressive forms is not very high, in that less than 10% of the articles and 20% of the progressive forms are divergent. I also found that speakers had more difficulty in the use of indefinite articles than definite articles, possibly because definite entities are more relevant to speakers than indefinite ones. In the results regarding the use of progressive forms, the divergences were partially related to the nature of the task, where they had to narrate stories in the past tense and failed to maintain the use of the past tense. On comparing these features to the corresponding Bengali features, interference effects seemed to be playing a partial role. The divergences were also related to differing degrees of proficiency in English. That is, speakers with a higher proficiency in English

(as measured from an independent proficiency test) tended to have fewer divergences than speakers with lower English proficiency. However, even for the lower proficiency group, the divergences were quite low (less than 16%) and therefore we cannot argue for any major change in the use of articles and progressive forms in this variety of English, at least for this subject pool. These results in fact indicate that the English spoken by these bilingual speakers is overall quite stable.

The overall findings of this dissertation are quite thought-provoking. First, despite intense contact between English and Bengali for over two hundred years, including the rise of extensive bilingualism and consequently of CS, I did not find evidence of widespread contact-induced changes in the syntactic domain of either Bengali or English. I do acknowledge, however, that the results could be conditioned by my subject pool, which was made up of highly proficient speakers, even among the subject group that showed 'low proficiency' in English. A different subject pool with much lower proficiency speakers could yield different results and insights on language change. However, it is certainly not the case that these proficient bilingual speakers have deliberately resisted change. If that was the case, they would not code-switch as extensively as they do now. But the lack of major changes in the grammar of Bengali and English certainly indicates substantial stability in the grammar of both languages, at least at this stage in the language for this set of speakers. By investigating contact-effects from both directions and using a robust methodology, my findings show that the influence of English on Bengali is mostly at the lexical level with some degree of influence in the grammar (bilingual verbs and copular predicates). Similarly, evidence of changes in English is also sparse (for the specific subject pool I studied). However, these findings cannot predict what may occur in the future in Bengali and English. If contact intensifies even further, more changes may occur.

These findings also tie with some of the predictions of language contact as discussed in Thomason & Kaufman (1998: 37-39) and Thomason (2001: 69-71; 75). The predictions state that in situations of L2 influence on L1 (or borrowing situations) changes usually first affect the lexicon and then go onto affect the phonology and grammar, while in cases of shift-induced interference (L1 effects on L2) or when imperfect learning plays a role, the first changes are usually at the grammatical level and changes at the lexical or phonological level occur much later, if at all. In the case of Bengali, we find changes mostly in the lexicon and less so in the grammar, while in English, the grammar shows indication of a possible change, at least as identified in terms of variation in the use of articles and progressive forms, but hardly any changes seem to have taken place at the lexical level. Again the caveat is that these findings are applicable only in case of the specific population I studied.

One can consider the bilingual speakers of this community from the perspective of the framework of *multi-competence* as proposed by Cook (e.g. 2006) and Grosjean (1998, 2008). This framework (also referred to as a "perspective") addresses the need to study bilinguals holistically and not as two monolinguals in one body. The way society is structured in India, bilinguals speakers are required to use both their languages in different domains with equal ease. Given this scenario, bilinguals of this community display great competence in using both their languages in effective ways in their daily lives. They may not always use their language in ways that are equivalent to the use of monolingual speakers, but that does not seem to lead to communicative barriers.

This dissertation is one of the first attempts to understand the language contact situation of Bengali and English in West Bengal in a well-rounded and comprehensive way. The investigation considered in detail lexical and morphosyntactic changes that were more prominent

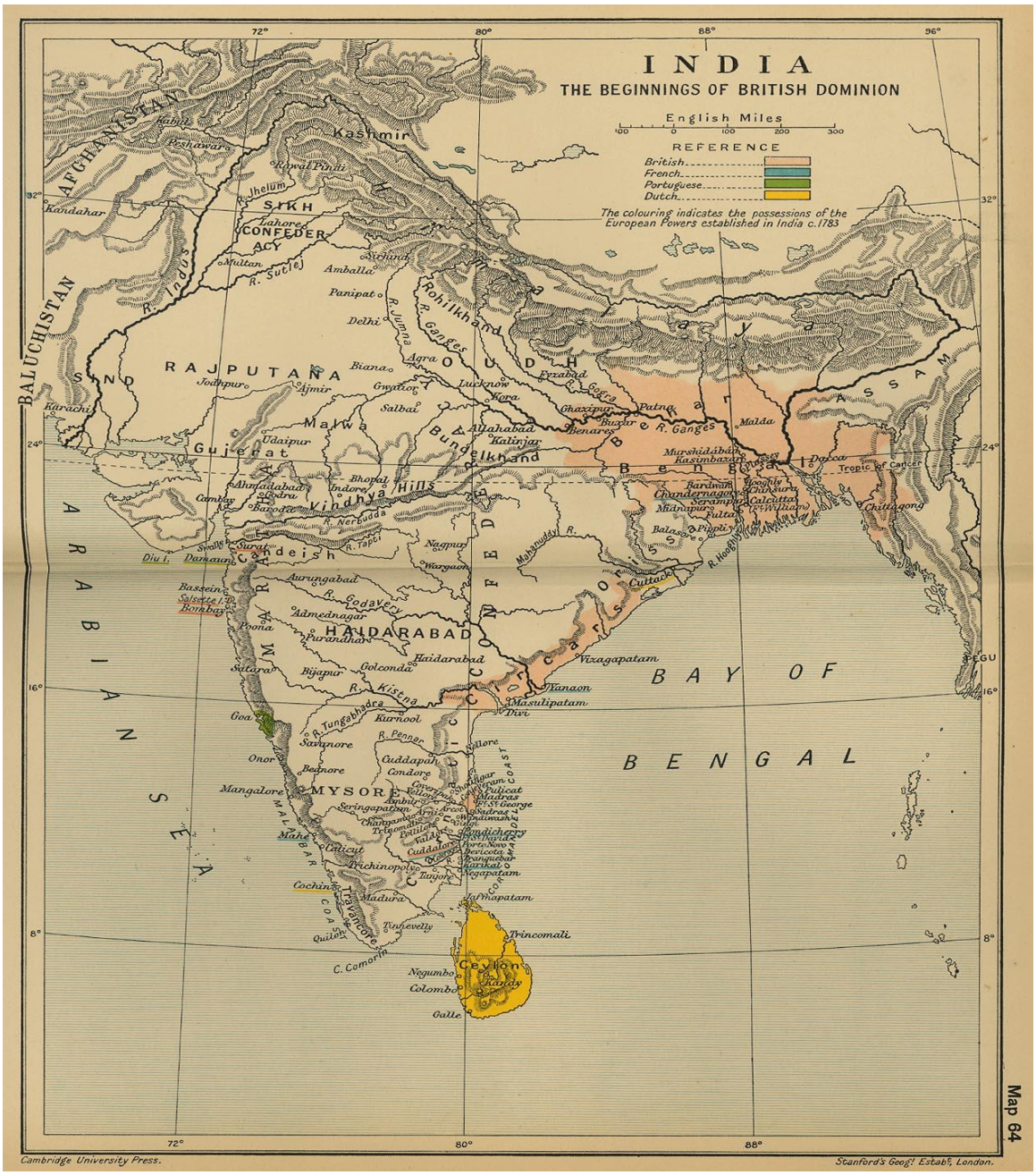
in the analysis of the corpus data. In addition, the investigation of possible changes explored the possibility of bidirectional influences between the two languages, and was supported by a rigorous methodology comprising, where necessary and possible, the analysis of both diachronic and synchronic corpora produced by either monolingual or bilingual speakers, and which were collected and analyzed as part of this research.

There are several outstanding questions that remain in the context of the research presented here. First, there are other areas of Bengali and English grammar that may have been influenced by mutual effects between the two languages. For instance, in this dissertation I have not discussed any changes at the phonological and morphological levels that may have occurred in addition to the lexicon and morphosyntax. These are areas of inquiry for future research, which is likely to yield a more complex picture of the overall impact of English influence on Bengali and vice versa. Second, further analysis can reveal much more about the features I have already investigated. For instance, a quantitative analysis of English loanwords in Bengali may reveal more about the nature of change at the lexical level. It would also be worthwhile to investigate code-switching between Bengali and English from the perspective of a theoretical framework that can contribute to a better understanding of the code-switching patterns that are found among bilinguals. There are other issues that can receive further consideration in the analysis of the copular predicates, such as the specific relation between information structure and the occurrence of the overt copula. In terms of the English spoken in this community, it would be worthwhile to carry out further investigation with speakers from a wider range of proficiency levels, which may yield very different results, and could reveal more about the patterns of use of features beyond the ones investigated in this dissertation. Notwithstanding these various areas that would benefit from additional research, this dissertation has shed light on some of the

previously unexplored aspects of this particular language contact situation in India, thereby contributing to the study of language contact and change in general.

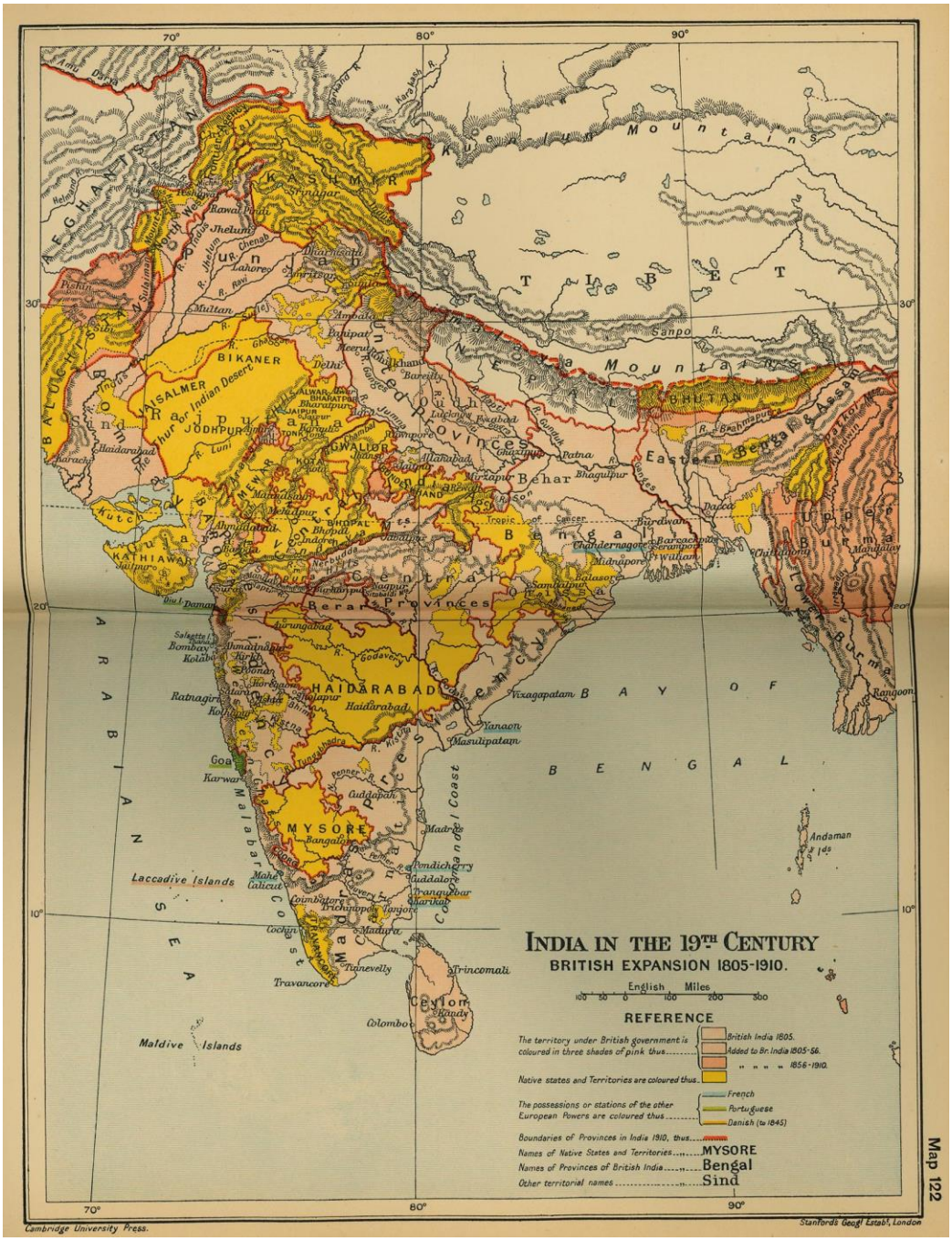
Appendix 1

Map showing the beginnings of British dominion in India
(Cambridge Modern History Atlas, 1912
University of Texas Libraries)



Appendix 2

Map showing the expansion of British dominion in India (1805-1910)
(Cambridge Modern History Atlas, 1912
University of Texas Libraries



Appendix 3

Ethnographic questionnaire used before monolingual Bengali and monolingual English production data collection:

1. What is your birth date (month and year ex: 02/1987)? _____/_____

2. What is your gender?

3. Is Bengali your native language?

   __ Yes
   __ No

4. Have you studied English? ___ Yes ___ No

   If Yes, answer the next questions. Otherwise go to question 5.

   a. When did you study English? (example: from age 16 to 20) _____

   b. What is your level of fluency in English, from 1 to 7
   (1 = beginner knowledge; 7 = perfectly native)? ____

   c. Check the box in each row that best represents your usages of English:

	Every day	3-4 times a week	Once a week	Rarely	Never
To Speak					
To Listen					
To Write					
To Read					

5. Indicate below your level of fluency in other languages, if any, from 1 to 7
   (1 = beginner knowledge; 7 = perfectly native)

   Language: _____ Level of fluency: _____
   Language: _____ Level of fluency: _____

6. What language(s) have you spoken in your home during your lifetime?
   Specify during which age periods you spoke it (e.g. ages 5-18):

   Language: _____ Age Period (from/to):_____

Language: _____ Age Period (from/to): _____

7. Have you lived in any other state outside West Bengal for more than a year? If so, in which state?

8. Please indicate your level of schooling by filling out the table below.

	City	State	Country
Elementary School			
Middle School			
High School			
College			

9. Can we contact you for additional studies of this kind? Yes/ No

If so, please include your contact information below:

Name: _____

Email (our preferred way to contact you): _____

Address: _____

Phone number(s): _____

Appendix 4

Ethnographic questionnaire used for bilingual speakers during bilingual Bengali-English production data collection:

1. Age of the consultant:
2. Gender of the consultant:
3. Education background of the participant: Tick one that is applicable

Masters degree completed/ pursuing:

Bachelors Degree completed/pursuing:

4. Is Bengali your native language:
5. Have you studied in an English medium school:
6. Have you lived in any other state apart from West Bengal for more than a year? If so, which state?

7. What languages do you speak other than English and Bengali? If so, how fluent are you in that/those:

	Native	Near Native	Intermediate
1.			
2.			
3.			

8. Do you use mainly Bengali or English to communicate with friends and family? If other languages are used for communication purposes please specify which ones they are and in what context?

9. Do you speak in any other language other than Bengali and English in any setting (family, friends, educational institutions, shopping centers, offices etc). If yes please specify:

10. Does anybody in your immediate family or close friend speak any other language other than Bengali or English natively:

Appendix 5

Questions for bilingual Bengali-English speakers during sociolinguistic interviews in monolingual English:

Instruction to informants: Participate in an individual sociolinguistic interview with the researcher, in English only, focusing on the topics:

- Describe a single trip or multiple trips that you made in childhood or as an adult and relate memorable experiences from those trips.
- Was the place that you visited different from the one you live in now? In what way?
- Would you like to go back to that place, or would they like their own place to be like that? Why?
- How are people in that place different from the people of your hometown?
- How are people from different regions or countries different in their behavior and the way they live?
-What are some things that you like and dislike about the city that you live in?
-If you had a chance to change something about your city, what would those things be?
-What was your childhood like? Were you very naughty or very serious about school?
-When you compare your childhood with the childhood of the kids of the present generation what differences do you notice?
-What do you think about the present education system of the country?
-What are your hobbies, what is your daily schedule like?
-(For older people) what other profession would you have chosen if you were not in the present profession that you are in now?
-(For younger people) what are your plans for the future in terms of choosing a profession?
- Some additional follow-up questions based on the answers the informants provided.

References

Abbi, Anvita & Gopalakrishnan, Devi (1991). Semantics of explicator compound verbs in South Asian languages. *Language Sciences,* 13(2), 161-180.

Agnihotri, Rama Kant, Khanna, Amrit Lal & Mukherjee, Aditi (1984). The use of articles in Indian English: Errors and Pedagogical implications. *International Review of Applied Linguistics in Language Teaching,* 22(2), 115-129.

Agnihotri, Rama Kant, Khanna, Amrit Lal & Mukherjee, Aditi (1988). *Tense in Indian English. A Sociolinguistic Perspective*. New Delhi: Bahri Publications.

Annamalai, E. (1978). The anglicized Indian languages: a case study of code-mixing. *International Journal of Dravidian Linguistics,* 7, 239-247.

– (1989). The language factor in codemixing. *Internationl Journal of the Sociology of Language,* 75, 47-54.

Asher, Ron E. (1968). Existential, possessive, locative and copulative sentences in Malayalam. In John W.M. Verhaar (Ed.) *The verb 'be' and its synonyms*, Part 2 (pp. 88-110). Dordrecht: D. Reidel.

Auer, Peter (1999). From code-switching via language mixing to fused lects: Toward a dynamix typology of bilingual speech. *The International Journal of Bilingualism,* 3(4), 309-332.

Backus, Ad (1996). *Two in one. Bilingual speech of Turkish immigrants in the Netherlands*. Tilburg: Tilburg University Press.

– (2005). Codeswitching and language change: One things leads to another? *International Journal of Bilingualism,* 9(3&4), 307-340.

Baksi, Shila (1983). Code-switching among Bengali-English bilinguals. Unpublished M.A. Thesis, Texas Tech Univeristy.

Bai, Lakshmi B. (1986). A note on syntactic convergence among Indian languages: The verb "to be". In Krishnamurti (Ed.). *South Asian Languages* (pp. 195-208). Delhi: Motilal Banarassi Dass.

Balasubramanian, Chandrika (2009). *Register variation in Indian English*, Studies in Corpus Linguistics, 37. Amsterdam: John Benjamins.

Banu, Rahela & Sussex, Roland (2001). Code-switching in Bangladesh. *English Today,* 2, 51-61.

Baptista, Marlyse & Guéron, Jacqueline (2007). *Noun Phrases in Creole Languages: A Multi-faceted Approach.* Amsterdam: John Benjamins.

Basu, Debarchana (2010). *Syntax-Semantics Issues for Multiple Event Clauses in Bangla*. Unpublished Ph.D. dissertation, Purdue University.

Bauer, Selket (2008). *The Sociolinguistics of English-Hindi code-switching: A case study from India*. Saarbrucken: VDM Muller.

Bhatia, Tej K. (1982). English and the vernaculars of India: Contact and change. *Applied Linguistics,* 3, 235-245.

– (1987). English in advertising: multiple mixing and media. *World Englishes,* 6(1), 33-48.

Bhatia, Tej K. & Richie, William C. (2006). Bilingualism in South Asia. In Tej K. Bhatia & William C. Ritchie (Eds.). *The Handbook of bilingualism* (pp. 780-807). Oxford: Blackwell.

Bhattacharya, Tanmoy (2005). *India not as a Linguistic Area: An insight from Minimalism*. people.du.ac.in/~tanmoy/papers/Mysore_pap_final.pdf

Bhattacharyya, Pushpak, Chakrabarti, Debasri & Sharma, Vaijayanthi M. (2006). Complex predicates in Indian languages and wordnets. *Language Resources & Evaluation,* 40, 331-355.

Butt, Miriam (2003). *The Light Verb Jungle*. http://ling.sprachwiss.unikonstanz.de/pages/home/butt/main/papers/harvard-work.pdf

Chatterjee, Lola (1992). The Landmarks in Official Education Policy: Some Facts and Figures. In Rajeswari Sunder Rajan (Ed.). *The Lie of the Land: English Literary Studies in India* (pp. 102-104). Delhi: Oxford University Press.

Chatterjee, Sudipto (2007). *The colonial staged: Theatre in colonial Calcutta*. Oxford: Seagull Publishing.

Chatterji, Suniti Kumar & Grierson, George Abraham (1926). *The origin and development of the Bengali language*. Calcutta: Calcutta University Press.

Chatterjee, Tridha (2014). Bilingual Complex Verbs: So what's new about them? *Proceedings of the 38th Annual Meeting of the Berkeley Linguistics Society* (pp. 47-62).

Cook, Vivian (2006). Interlanguage, multi-competence and the problem of the "second" language. *Rivista di Psicolinguistica Applicata, VI,* 3, 39-52.

Croft, William (1991). *Syntactic Categories and Grammatical Relations*. Chicago: University of Chicago Press.

Curnow, Timothy Jowan (1999). Towards a cross-linguistic typology of copula constructions. *Proceedings of the 1999 Conference of the Australian Linguistic Society* (pp. 1-9). Australian Linguistic Society, http://www.arts.uwa.edu.au/LingWWW/als99/proceedings.

Dasgupta, Jyotirindra (1976). Practice and Theory of Language planning: The Indian Policy Process. In William M. O'barr & Jean. F O'Barr (Eds.). *Language and Politics*. The Hague: Mouton.

Dasgupta, Probal (1977). The internal grammar of compound verbs in Bangla. *Indian Linguistics,* 38, 68-85.

Dasgupta, Probal & Ghosh, Sanjukta (2006). The Positive Polarity Copula in Bangla: An Item of unexplored origin. *Indian Linguistics*, 67, 1-4.

De, Sushil Kumar (1919). *The History of Bengali Literature in the Nineteenth Century 1800-1825*. Calcutta: Calcutta University Press.

Dutta, Sunil Kumar (2002). *The Raj and the Bengali people*. Kolkata: Firma KLM.

Edwards, Malcolm & Gardner-Chloros, Penelope (2007). Bilingual compound verbs in codeswitching: Bilinguals making do? *International Journal of Bilingualism,* 11(1), 73-91.

Emeneau, Murray Branson (1956). India as a linguistic area. *Language*, 32(1), 3-16.

Eliasson, Stig (1989). English-Maori language contact: Code-switching and the free morpheme constraint. *RUUL: reports from Uppsala University Department of Linguistics.* Uppsala, (pp.1-28).

Ferguson, Charles A. (1972). Verbs of 'Being' in Bengali with a note on Amharic. In John W.M. Verhaar (Ed.). *The verb 'be' and its synonyms*, Part 5 (pp. 74-114). Dordrecht: D. Reidel Publishing Co.

Gandhi, Mohandas Karamchand (1936). *A few obvious considerations: Our language problem,* Karachi: Anand, T. Hingorani.

Gardner-Chloros, Penelope (2009). Sociolinguistic factors in code-switching. In Barbara E. Bullock & Almeida Jacqueline Toribio (Eds.). *The Cambridge handbook of linguistic code-switching* (pp. 97-113). Cambridge: Cambridge University Press

Givon, Talmy (1979). *On Understanding Grammar.* New York: Academic Press.

Grosjean, Francois (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition,* 1(2), 131-149.

– (2012). An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16(1), 11-21.

Guha, Ramchandra (2007). *India after Gandhi*. New Delhi: Picador India.

Guha-Thakurta, Prabhucharan (1930). *The Bengali Drama: Its origin and development*. London: Kegan Paul, Trench, Trübner & Co. Ltd.

Haugen, Einar (1950). The analysis of linguistic borrowing. *Language,* 26, 210-232.

– (1969). The Norwegian language in America: A study in bilingual behavior. Bloomington: Indiana University Press.

Hengeveld, Kees (1992). *Non-verbal predication: theory, typology, diachrony*. Functional Grammar Series 15. Berlin: Mouton de Gruyter.

Heine, Bernd & Kuteva, Tania (2003). On contact-induced grammaticalization. *Studies in Language*, 27(3), 529-572.

– (2010). Contact and Grammaticalization. In Raymond Hickey (Ed.). *The Handbook of Language Contact* (pp. 86-105). Wiley Online Library.

Hettiarachchi, Sujeewa & Pires, Acrisio (2015). L2 acquisition of wh-features and syntactic constraints: Evidence for a full-access approach. Paper presented at *13th GASLA Generative Approaches to Second Language Acquisition Conference*. Indiana University, Mar 4-6, 2015.

Hock, Hans Henrich (2003). Analogical change. In Brian D. Joseph & Richard D. Janda (Eds.). *The Handbook of Historical Linguistics* (pp. 441-460). Oxford, UK: Blackwell.

Holm, John (1984). Variability of the copula in Black English and its Creole kin. *American Speech,* 59(4), 291-309.

Hook, Peter Edwin (1974). *The Compound Verb in Hindi*. Center for South and Southeast Asian Studies: The University of Michigan Press.

Hopper, Paul J. & Traugott, Elizabeth Closs (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Jarvis, Scott & Pavlenko, Aneta (2008). *Crosslinguistic Influence in Language Cognition*. New York: Routledge.

Joshi, Svati (1994). *Rethinking English: Essays in Literature, Language, History*. Delhi: Oxford University Press.

Kachru, Yamuna (1968). The copula in Hindi. In John W. M. Verhaar (Ed.). T*he verb 'be' and its synonyms*, Part 2 (pp. 35-58). Dordrecht: D. Reidel.

Kachru, Braj B. (1978a). The Englishization of Hindi: language rivalry and language change. In Gerald F. Carr Irmengard Rauch (Ed.). *Linguistic Method: Papers in Honor of Herbert Penzl* (pp. 199-211). The Hague: Mouton de Gruyter.

– (1978b). Toward Structuring code-mixing: An Indian perspective. *International Journal of the Sociology of Language* 16, (Aspects of sociolinguistics in South Asia), (pp. 24-47).

– (1994). English in South Asia. In Robert Burchfield (Ed.). *The Cambridge History of the English Language,* 5 (pp. 497-553). Cambridge: Cambridge University Press.

– (1996). World Englishes: Agony and Ecstasy. *Journal of Aesthetic Education,* 30(2), 135-155.

Korn, Herausgegeben von Agnes (2011). Pronouns as verbs, verbs as pronouns: Demonstratives and the copula in Iranian. In Herausgegeben von Agnes Korn, Geoffrey Haig, Simin Karimi und Pollet Samvelian (Eds.). *Topics in Iranian Linguistics* (pp. 53-70). Dr Ludwig Reichert Verlag.

Krishnaswamy, N. & Krishnaswamy, Lalitha (2006). *The story of English in India*. New Delhi: Foundation books.

Krishnaswamy, N. & Burde, Archana S. (1998). *The politics of Indians' English: linguistic colonialism and the expanding English empire.* Delhi: Oxford University Press.

Labov, William (1969). Contraction, deletion and inherent variability of the English copula. *Language*, 45(4), 715-762.

Lardiere, Donna (2004). Knowledge of definiteness despite variable article omission in second language acquisition. *Proceedings of the Annual Boston University Conference on Language Development,* 28, 328-339. Somerville, MA: Cascadilla Press.

Laurie, William F. B. (1888). *Sketches of some distinguished Anglo Indians,* Second edition. London: W. H Allen & co.

Lilley, Charles (2004). *Europe, 1450 to 1789: Encyclopedia of the Early Modern World*. The Gale Group Inc. http://www.encyclopedia.com/topic/Asia.aspx

MacLeod, A.G. (1967). *Colloquial Bengali grammar*. Calcutta: Baptist Mission Press.

MacSwan, Jeff (2006). Code Switching and grammatical theory. In Tej K. Bhatia & William C. Ritchie (Eds.). *The Handbook of bilingualism* (pp. 283-311). Oxford: Blackwell.

Marshall, Peter James (1987). *Bengal: The British Bridgehead*, Eastern India 1740–1828, The New Cambridge History of India. Cambridge: Cambridge University Press.

Masica, Colin P. (1991). *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Mayer, Mercer (1967). *A boy, a dog and a frog*. New York: Dial Books for Young Readers.

Mayer, Mercer & Mayer, Marianna (1971). *A boy, a dog, a frog and a friend*. New York: Dial Books for Young Readers.

Mikkelsen, Line (2005). *Copular Clauses: Specification, Predication and Equation*. Amsterdam: John Benjamins.

Mitra, Dinabandhu (1860). *Nil Darpan*. Reprinted in Bankim Chandra Chattopadhyay Rai Bahadur (Ed.). *Granthabali: Rai Bahadur Dina Bandhu Mittrer jivani o kabitta samalochana* (pp. 1-76), (1890). Calcutta: Upendra Nath Mukhopadhyaya, Basu Mati Office.

Moravcsik, Edith (1975). Verb borrowing. *Wiener Linguistische Gazette*, 8, 3-31.

– (1978). Language contact. In Joseph H. Greenberg (Ed.). *Universals of Human Langauge* I (pp. 95-122). Stanford: Stanford University Press.

Muysken, Pieter (1993). Constraints on code-switching: Indic mixed verbal compounds, *Code-switching summer school* (pp. 195-208).

– (2000). *Bilingual speech: a typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol (1993a). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

– (1993b). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.

– (1998_2000). Code-switching as indexical of social negotiations. In Li Wei (Ed.). *The bilingualism reader* (pp. 137-165). New York: Routledge

Nath, Diganta Kumar (2009). Copulas in Asamiya. *Indian Linguistics*, 82, 12-21.

Niranjana, Tejaswini (1994). Translation, Colonialism and the Rise of English. In Svati Joshi (Ed.). *Rethinking English (*pp. 47-65). New Delhi: Oxford University Press.

Pagett, Linda (2006). Mum and Dad prefer me to speak Bengali at home: code switching and parallel speech in a primary school setting. *Literacy,* 40(3), 137-145.

Pandharipande, Rajeshwari (1990). Formal and functional constraints on code-mixing. In Rodolfo Jacobson (Ed.). *Codeswitching as a Worldwide Phenomenon* (pp. 15-31). New York: Peter Lang.

Pavlenko, Aneta & Jarvis, Scott (2002). Bidirectional Transfer. *Applied Linguistics*, 23(2), 190-214.

Paul, Soma (2003). Composition of Compound Verbs in Bangla. *Proceedings of the Workshop on Multi-Verb Constructions* (pp. 1-27). Trondheim Summer School, Norwegian University of Science and Technology, Trondheim.

– (2009). Hindi hE 'to be' in Bangla – Issues in Machine Translation. Centre for Language Technologies Research Centre, International Institute of Information Technology, Hyderabad.

Pillai, M. Shanmugun (1968). English borrowings in educated Tamil. In Braj Krishnamurti (Ed.). *Studies in Indian Linguistics* (pp. 296-306). Annamalainagar: Centers of Advanced Study in Linguistics.

Pires, Acrisio & Thomason, Sarah G. (2008). How Much Syntactic Reconstruction is Possible? In G. Ferraresi & M. Goldbach (Eds.). *Principles of Syntactic Reconstruction (*pp. 27-72). Amsterdam: John Benjamins.

Poplack, Shana (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPANOL: toward a typology of code-switching. *Linguistics,* 18, 581-618.

Poplack, Shana, Wheeler, Susan & Westwood, Anneli (1989). Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. In Rodolfo Jacobson (Ed.). *Codeswitching as a Worldwide Phenomenon* (pp. 185-218). New York: Peter Lang.

Poplack, Shana & Meechan, Marjory (1995). Patterns of language mixture: nominal structure in Wolof-French and Fongbe-French bilingual discourse. In Lesley Milroy & Pieter Muysken (Eds.). *One speaker, two languages: cross-disciplinary perspectives on code-switching* (pp. 199-232). Cambridge: Cambridge University Press.

Poplack, Shana & Levey, Steven (2010). Contact-induced grammatical change. In Peter Auer & Jürgen Erich Schmidt (Eds.). *Language and Space – An international handbook of linguistic variation: Volume 1 – Theories and methods* (pp. 391-419). Berlin: Mouton de Gruyter:

Poplack, Shana & Dion, Nathalie (2012). Myths and facts about loanword development. *Language Variation and Change,* 24(3), 279-315.

Puri, Vandana (2011). The influence of English on the history of Hindi relative clauses. *Journal of Language Contact,* 4, 250-168.

Pustet, Regina (2003). *Copulas - Universals in the categorization of the lexicon*. Oxford studies in typology and linguistic theory. Oxford: Oxford University Press.

Raina, Badri (1994). A note on language and the politics of English in India. In Svati Joshi (Ed.). *Rethinking English* (pp. 279-293). Delhi: Oxford University Press.

Rakshit, Moloy (2013). Communication through public stage: A study in 19th century Bengali theatre. *Global Media Journal-Indian Edition*, 4(2), 1-18.

Ramchand, Gillian (1990). Complex predicate formations in Bangla. *Ninth West Coast Conference on Formal Linguistics* (pp. 443-452). Stanford University.

Romaine, Suzanne (1986). The syntax and semantics of the code-mixed compound verb in Panjabi/English bilingual discourse. In Deborah Tannen & James E. Alatis (Eds.), *Languages and Linguistics: The Interdependence of Theory, Data and Application* (pp. 35-49). Washington D.C.: Georgetown University Press.

– (1995). *Bilingualism*. Oxford: Blackwell.

Sableski, Julia A. (1965). Equational clauses in Bengali. *Language*, 41(3), 439-446.

Sailaja, Pingali (2009). Indian English. In Joan Beal, Patrick Honeybone, & April McMahon (Eds.). *Dialects of English*. Edinburgh: Edinburgh University Press.

– (2012). Indian English: Features and sociolinguistic aspects. *Language and Linguistics Compass*, 6(6), 359–370.

Sand, Andrea (2004). Shared morpho-syntactic features in contact varieties of English: article use. *World Englishes*, 23(2), 281-98.

Sankoff, Gillian (2001). Linguitic outcomes of language contact. In Peter Trudgill, J.K. Chambers & Natalie Schilling-Estes (Eds.). *Handbook of Sociolinguistics* (pp. 638-668). Oxford: Basil Blackwell.

Sankoff, David, Poplack, Shana & Vanniarajan, Swathi (1990). The case of the nonce loan in Tamil. *Language Variation and Change,* 2(1), 71-101.

Santorini, Beatrice & Mahootian, Shahrzad (1995). Codeswitching and the syntactic status of adnominal adjectives. *Lingua,* 96(1), 1-27.

Sedlatschek, Andreas (2009). *Contemporary Indian English: Variation and change*. Amsterdam: John Benjamins.

Selinker, Larry (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching,* 10(3), 209-231.

Sharma, Devyani (2005a). Language transfer and discourse universals in Indian English article use. *Studies in Second Language Acquisition,* 27(04), 535-566.

– (2005b). Dialect stabilization and speaker awareness in non-native varieties of English. *Journal of Sociolinguistics,* 9(2), 194-224.

– (2009). Typological diversity in New Englishes. *English World-Wide*, 30(2), 170-195.

Sharma, Devyani & Rickford, John (2009). AAVE/Creole copula absence: A critique of the Imperfect Learning Hypothesis. *Journal of Pidgin and Creole Languages*, 24(1), 53-92.

Si, Aung (2011). A diachronic investigation of Hindi-English code-switching using Bollywood film scripts. *International Journal of Bilingualism,* 15(4), 388-407.

Siegel, Jeff (2008). *The Emergence of Pidgin and Creole Languages*. Oxford: Oxford University Press.

Sridhar, Shikaripur N. (1992). Language modernization: Structural and sociolinguistic aspects. *South Asian Language Review*, 2(1), 84-101.

Stewart, Charles (1903). *The history of Bengal: from the first Mohammeden invasion until the virtual conquest of that country by the English*, A.D. 1757. Calcutta: Bangabasi.

Sunder Rajan, Rajeswari (1992). Fixing English: Nation, Language, Subject. In Rajeswari Sunder Rajan (Ed.). *The Lie of the Land: English Literary Studies in India* (pp. 8-25). Delhi: Oxford University Press.

Tagliamonte, Sali A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden: Wiley-Blackwell.

Tarkaratna, Ram Narayan (1854). *Kulin Kula Sarbaswa.* Reprinted in Satyabrata De & Sanatan Goswami (Eds.). *Kulin Kulasarbasya: A criticism with text on Ramnarayan's Drama 'Kulin Kula-Sarbasya'* (pp. 1-76), (1987). Calcutta: Pustak Bipani.

Thomason, Sarah G. (1997). On mechanisms of interference. In Stig Eliasson & Ernst H. (Eds.). *Trends in Linguistics: Studies and Monogrpahs: Language and its Ecology: Essays in memory of Einar Haugen* (pp. 181-207). Berlin: Mouton de Gruyter.

– (2001). *Language Contact*. Washington D.C.: Georgetown University Press.

Thomason, Sarah G. & Kaufman, Terrence (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

Thompson, Hanne-Ruth (2010). *Bengali: A Comprehensive Grammar*. London: Routledge.

Van Coetsem, Frans (1988). *Loan Phonology and the Two Transfer Types in Language Contact*.

Dordrecht: Foris.

– (2000). *A General and unified Theory of the transmission process in language contact*. Heidelberg: Winter.

VanGelderen, Elly (2014). The copula cycle (pp. 1-14). Unpublished paper.

Vishwanathan, Gauri (1992). English in a Literate Society. In Rajeswari Sunder Rajan (Ed.). *The Lie of the Land: English Literary Studies in India* (pp. 30-52). Delhi: Oxford University Press.

Ward, Adolphous William, Prothero, G.W., Leathes, Stanley Mordaunt & Benians, E.A. (1912) (Eds.). *The Cambridge Modern History Atlas*. London: Cambridge University Press. http://www.lib.utexas.edu/maps/historical/ward_1912.html

Weinreich, Uriel (1953). *Languages in Contact: Findings and problems*. New York: Publications of the Linguistic Circle of New York, 1. (Reprinted 1986, The Hague: Mouton.)

Wichmann, Søren & Wohlegemuth, Jan (2005). *Loan verbs in a typological perspective.* Bremen: Romanisation Worldwide.

Winford, Donald (2005). Contact-induced Changes: Classification and processes, *Diachronica,* 22(2), 373-427.

Wohlgemuth, Jan (2009). *A typology of verbal borrowings*. New York: Mouton de Gruyter.

Yates, William & Wenger, John (1885). *Bengali grammar*. Calcutta: Baptist Mission Press.