A diachronic investigation of Hindi-English code-switching, using Bollywood film scripts

International Journal of Bilingualism 15(4) 388–407 © The Author(s) 2010 Reprints and permission: sagepub.co.uk/journalsPermissions.nav DOI: 10.1177/1367006910379300 ljb.sagepub.com



Aung Si

Australian National University, Canberra, Australia

Abstract

Code-switching (CS) between an Indian language and (Indian) English is, and has long been, a normal feature of everyday speech in urban Indian society. Although much has been written about the status and role of English, and about the sociological variables associated with English usage in India, there have been, to date, no studies explicitly investigating changes in CS patterns over time. Bollywood movies are a rich source of information on the speech patterns of urban Indians throughout most of India's post-independence history. CS patterns in Bollywood movies (from the 1980s, 1990s and 2000s) were therefore investigated in this study, by means of lexical transcripts of the dialogues between characters of equivalent age and socioeconomic status. A survey of seven movie dialogues revealed that CS can be accomplished through a range of syntactic and morphological strategies. Quantitative analyses showed a massive increase in the overall use of English over this period, a trend particularly evident among young speakers. Moreover, the complexity of CS increased over the period under consideration, with 'alternations' at clause boundaries increasing in frequency at the expense of single-word 'insertions'.

Keywords

Bollywood, code-switching, diachronic, mixed code

Introduction

The phenomenon of code-switching (CS) has, in recent years, received much attention from linguists, particularly from those working with postcolonial or immigrant populations. One major feature that much of the CS research to date has in common is its focus on synchronic phenomena, i.e. the investigation of CS patterns within a speech community at a particular point in time. However, there has been a growing realization in recent years that many of the contact phenomena that were thought to have involved switching between two (or more) codes are in fact 'mixed codes' in their own right (e.g. Auer, 1998). Some authors (e.g. Gafaranga, 1997; Gafaranga & Torras, 2002) have been careful to distinguish CS, defined as instances of 'interactional otherness',

Corresponding author:

Aung Si, School of Culture, History and Language, College of Asia and the Pacific, Coombs Building, Australian National University, Canberra, ACT 2600, Australia. Email: aung.si@anu.edu.au

and characterized by deviation from a current speech medium, from 'language alternation itself as the medium', where speech participants behave as though only one language were being spoken. Such typological issues are not the focus of the current article, as a far more detailed analysis of the data would be required to correctly identify the mode of interaction. Nevertheless, it would be fair to assume that mixed codes (or indeed, instances of 'language alternation as the medium') are subject to change over time, as are 'pure' languages, but such considerations have largely been ignored by scholars of CS.

A major hurdle in carrying out a diachronic investigation of CS is the absence of longitudinal studies, even over a period of a few years. One way of overcoming this problem would be the use of printed or electronic materials that were produced at different time points, and which might be expected to depict contemporary CS patterns reasonably accurately. An ideal source of information on the northern Indian region is the vast number of Hindi films produced by the film industry centred around Mumbai, and popularly known overseas as Bollywood. Many films produced in the last four decades contain Hindi–English CS in substantial quantities, and should be extremely helpful in any study on past CS practices among native Hindi speakers.

The Bollywood phenomenon

The term 'Bollywood', used to describe the Hindi film industry based in Mumbai, is nowadays recognized by many living outside India, both emigrants of Indian origin, as well as the other inhabitants of countries with sizeable Indian populations. Approximately 1000 movies are made every year in India in a variety of languages; an astonishing output, given that only 4000 movies are produced worldwide per annum, of which 700 originate in the US (cited in Eliashberg, Elberse, & Leenders, 2006). Bollywood, the Hindi component of this industry, comprises almost half (45%) of the total Indian film market.

Indian filmmakers around the time of independence placed an emphasis on realist aesthetics in their work, often portraying 'the world of [the] underprivileged and marginalised, and which represented . . . Indian society as iniquitous and inequitable' (cited in Rao, 2006). Through the 1970s and 1980s, Hindi films continued to appeal to the underprivileged, often portraying the central character of the 'angry young man' in his (often violent) struggles against poverty and corruption (Rao, 2006). Through such devices, Rao writes, Hindi films ensured viewer identification with the working poor and lower-middle-class sensibilities.

The economic liberalization that took place in India in the 1990s brought with it rapid, unprecedented and far-reaching social and cultural changes. A major factor driving these changes was the advent of new information technologies, such as mobile phones, satellite television and the Internet (Acharya, 2004). Under these conditions, Bollywood was forced to 'reevaluate its methods of conducting business' for a variety of reasons:

The direct causes have been the rise of alternate sources of entertainment such as TV, both satellite and Doordarshan [the national television corporation], the state run media, and the Internet. On the other hand, the rise of a new middle class, made prosperous by the new jobs being created in India due to globalisation, and the Indian diaspora abroad has created an audience whose entertainment expectations have changed. (Acharya, 2004, p. 22)

Such pressures have resulted in the release of movies which no longer have the mass appeal of movies from the 1970s and 1980s. A majority of the top grossing movies post-2000 have featured 'Westernized themes', including 'foreign locations, actors, and singers; and, [a] liberal use of

English in the dialogs' (Rao, 2006). Such a trend has been matched by satellite TV channels, such as Zee TV, where CS with English is highly prevalent in a range of programmes, especially those involving audience participation and/or unscripted speech (Gardner-Chloros & Charles, 2007).

The role of English in India

Following independence, English was not only given the status of associate official language in the Constitution, but was also intended as a language to be actively taught to students in schools. These factors, along with the globalizing pressures that have seen the worldwide dominance of English in recent decades, have probably been responsible for the special status of English in Indian society, and its ongoing rivalry with the local vernaculars, particularly Hindi (Dua, 1993). English is used very liberally at various levels of government: bills or amendments to be introduced into either house of parliament or a state legislature, all acts passed by these bodies and all ordinances promulgated by the president or a state governor, and all orders, rules, regulations and bylaws issued by parliament or a state legislature are always in English (Parasher, 1991).

English has been equally visible in the education sector in India. Srivastava and Sharma (1991, p. 189) point out that:

The higher we move in education, the fewer are the languages employed as media of instruction, so much so that for higher education and technical training the only medium left for use is English.

Such forces have aided the transition of English from the status of a foreign language to that of a second language for many Indians. The evidence for this statement lies in assertions such as the following:

(i) English covers in India more than $1/4^{th}$ [sic] of the entire bilingual population (25.7%) while [the] 12 other major languages share [the] rest of the bilingual population, and (ii) in spite of the fact that English is used by less than 3% of the entire population as a necessary means of communication, it is a language that carries power and prestige, for this 3% section of the population is the most important section of Indian society. (Srivastava & Sharma, 1991, p. 190)

In addition, the leading role of English in the country's print and electronic media, as well as in the publishing sector (Parasher, 1991) bestows upon it a prestige, particularly in urban centres, not enjoyed by any other language.

Code-switching with English in the Indian situation

Interactions involving the use of lexical items originally from two of more 'codes' (dialects, variants or languages) are commonplace in India; such code-switching has been variously described in the literature. Gumperz (1961), for instance, mentions that the use of English phrases in Hindi discourse is the sign of an 'educated speaker'. While making a clear distinction between code-switching and code-mixing, Kachru (1978) explains that the former (switching from standard Hindi into the dialect Awadhi, for instance):

... may be used to express extreme anger, disapproval, in-group membership, asides, or solidarity. Code-switching in such contexts is a marker of an attitude, intensity of emotions, or various types of identities. (Kachru, 1978, p. 108)

Code-mixing with English, on the other hand, is a 'socially accepted marker of education and what may be termed "westernization" in India. It also identifies membership in a particular social class'. Kachru (1978, p. 109) cites the following example from a collection of Hindi short stories published in 1975:

Tum nahĭ jāntī, he is chairman Mr. Mehta's best friend, *yahă do cār din ko hī āye haĭ. Maine socā*, I should not miss the opportunity.

'Don't you know, he is chairman Mr. Mehta's best friend, he's only here for a couple of days. I thought, I should not miss the opportunity.'

In this example, the clause 'I should not miss the opportunity' is uttered in English, even though it could just as easily have been rendered in Hindi (especially since it is preceded by the discourse marker *maine socā* 'I thought'). This speaker is therefore not using English to fill lexical gaps in Hindi; the intention is rather to extend her register and style ranges (Kachru, 1978). Other reasons that have been put forward for code-mixing in India include:

- Register identification (e.g. administrative, political and technical registers), style identification ('sanskritized', 'persianized' or 'englishized' Hindi) and elucidation/interpretation (e.g. through the use of words like *māne* (Persian), *arthāt* (Sanskrit) for 'meaning') (Kachru, 1978).
- Neutralization: English lexical items are often perceived as being attitudinally and contextually neutral, and may be used to conceal social or regional identity (cited in Kachru, 1978).
- A range of discourse-related functions, including repeating, emphasizing, heightening contrast, creating surprise, making parenthetical remarks, teasing, challenging or reporting others' speech (Gupta, 1991).
- Reasons, suggested by speakers, such as: 'if I do not get the appropriate word in Hindi', 'easy to communicate', 'when we are short of words', 'to speed up communication', 'habit', 'unintentional', 'makes me feel comfortable', 'interesting and funny', 'scope of expression' and 'cos it gives me a feeling of Indianism [sic]' (Eilert, 2006).

Some researchers have attempted to describe Hindi–English code-switching by means of formal grammatical tools, with the aim of positing general constraints on the possible types of CS, and on the contexts in which they can occur (e.g. Bhatt, 1997; Kumar, 1986; Malik, 1994; Romaine, 1995, pp. 131–141; Singh, 1995, 1998). However, this subject lies beyond the scope of the present study, and is not discussed further.

Aims of the study

The identity issues plaguing Bollywood are likely to have a direct bearing on the present study, which aims to investigate changes in Hindi–English CS patterns among the Indian urban elite. For the purpose of this study, all instances of English use by native Hindi speakers living in India are investigated under the heading of CS.

The vast majority of the Indian urban elite, as described in the previous section, were recently (post-1990) exposed on a large scale to such technologies as the mobile phone, the Internet and satellite television. Middle-class and upper-middle-class families now have 24-hour access to western (English) programming in the form of the BBC, CNN, Cartoon Network or Discovery Channel right in their living rooms. For the many urban Indians who

already speak English as a second language, it is reasonable to assume that long-term exposure to more native (western) forms of English through the Internet or the many American, British and Australian programmes now showing on television would have lasting effects on the volume and nature of English utterances used in CS. It was hoped that through a careful choice of movies, I could compare the speech of characters of similar socioeconomic status in comparable settings, to trace changes in CS over several years. My main research question was to compare the overall level of English usage in Bollywood movies from three consecutive decades: the 1980s, 1990s and post-2000. First, however, I describe the different kinds of CS phenomena frequently occurring in Bollywood movies.

Methodology

Choice of film

A total of seven movies, spanning the last three decades of the Indian film industry, were selected, to determine changes in the prevailing linguistic situation over this period. All the movies selected were set in an urban setting, in the metropolitan centres of Delhi and Mumbai, and typically had plots where characters from an upper-middle-class background were the protagonists. Care was also taken to select movies with contemporary themes (for the period in which the movie was made). The similarity in the themes addressed in these movies is reflected in some of their titles: Dil (1990), Dil Ka Kya Kasoor (1992) and Dil Chahta Hai (2001) all deal with affairs of the heart (dil), and portray the ever popular boy-meets-girl scenario. This ensured comparability between different movies within the same decade, as well as between movies from different periods. Several popular Bollywood genres, both past and present, were eliminated through the application of these criteria. These included historical and mythological/religious movies for obvious reasons, as well as several types of action movies, which often ignored the 'home' domain, and also involved protagonists from lower socioeconomic backgrounds. In spite of these restrictions, there still remained a large number of movies, mostly falling into the 'love story' or 'family drama' genres, that were deemed suitable for inclusion into the present study.

The movies analysed in this study are grouped into three periods, indicating the date of their release. Movies were chosen to roughly cluster around the turn of each decade; details of the date of release, actors, director and scriptwriter are presented in Table 1. The movies were all predominantly in Hindi with frequent English CS, although Punjabi was also employed infrequently, depending on the ethnicity of the characters in the movies.

These seven movies can be grouped into two broad categories, namely 'art-house' and 'commercial' (marked by asterisks in Table 1). The two art-house movies are arguably the most realistic of all, both in terms of dialogue as well as storyline. These are therefore analysed separately, to obtain what is hoped will be a more accurate portrayal of Hindi–English code-switching. Both movies are set in Delhi, and recount the day-to-day occurrences in the lives of two upper-middle-class families. The separation of these movies by almost two decades should also bring to light major changes in language use patterns within the home domain. The five commercial movies, on the other hand, are slightly more diverse in their subject matter, and as they are more representative of mainstream Hindi cinema, should reveal some interesting differences between themselves and the two art-house movies. Forming a large subgroup within the commercial movies are four movies with several scenes shot in a high school or university setting. The fact that most, if not all private school and university education in India is carried out in English suggests that such centres of education might be fertile breeding grounds for novel and innovative (and often deliberate)

Table 1. Details of movies analysed in this study

Name	Year of release	Director	Scriptwriter	Principal actors	Running time
Masoom*	1982	Shekhar Kapoor	Sampooran Singh Gulzar	Naseeruddin Shah, Shabana Azmi, Saeed Jaffrey, Supriya Pathak	165 min.
Saath Saath	1982	Raman Kumar	Raman Kumar	Farooq Sheikh, Deepti Naval, Satish Shah	II8 min.
Dil	1990	Indra Kumar	Rajeev Kaul	Aamir Khan, Madhuri Dixit, Saeed Jaffrey, Anupam Kher	172 min.
Dil Ka Kya Kasoor	1992	Lawrence D'Souza	Talat Rekhi	Prithvi, Divya Bharti, Suresh Oberoi	154 min.
Monsoon Wedding*	2001	Mira Nair	Sabrina Dhawan	Naseeruddin Shah, Lillete Dubey, Vijay Raaz, Shefali Shetty, Tilotama Shome	II4 min.
Dil Chahta Hai	2001	Farhan Akhtar	Farhan Akhtar, Kassim Jagmagia	Aamir Khan, Saif Ali Khan, Akshaye Khanna, Preity Zinta	183 min.
Main Hoon Na	2004	Farah Khan	Farah Khan	Shahrukh Khan, Sushmita Sen, Zayed Khan, Amrita Rao, Naseeruddin Shah, Satish Shah	179 min.

Source: http://www.imdb.com.

code-switching phenomena. Thus, the interactions between students are also analysed separately to determine the existence of such phenomena.

Transcription and data analysis

Lexical transcriptions were produced for all movies. Code-switching phenomena were classed into two broad categories, following the scheme of Muysken (2000): insertions and alternations. A third category mentioned by Muysken, congruent lexicalization, did not occur in the movies being analysed, and was disregarded. However, it is worth noting that phenomena that could be labelled transference or convergence did occur in the data (see later), but analysis of such forms would be more appropriate in a discussion on 'Indian English'. The raw data extracted from the movie dialogues consisted of the number of turns within a particular interaction type (e.g. between two adults in the home situation) that contained at least one insertion or an alternation. The total number of insertions or alternations (tokens) within a turn did not matter – only the number of turns was counted. When both types of switch occurred in a turn, this was considered to be one count of insertion and one count of alternation. Thus, in the following exchange:

Sūri: ham to kām dhande mě aur whiskey pīne mě lag gae. varnā ham bhī beṭe ke tareh wimbledon kī tayyārī karte. Yār, DK, jo mazā apne beṭe ko javān hote dekhkar ātā hai, vo apnī javānī mě bhī nahǐ ātā thā. You must have a son, yār. DK, betā bahut zarūrī hai. 'I just got right into my work, and into drinking whiskey. If I hadn't, I'd also be preparing for Wimbledon like my son. DK, buddy, it's such a joy seeing one's son grow up – I don't think I've ever been this happy, not even in my youth. You must have a son, buddy. A son's very important, DK.'

- DK: mai caltā hŭ.
 'I'd better get going.'
- 3. Sūri: yār tū bāt karnā cāh rahā thā. Discuss kar lĕge problem. Kyā hai? 'You wanted to talk, buddy. We'll discuss the problem. What is it?'

(Masoom, 1982)

Line 1 has the English insertions 'whiskey' and 'wimbledon', as well as an alternation to English between $th\bar{a}$ and 'you', and an alternation back to Hindi between 'son' and $y\bar{a}r$. There are also the two English words 'discuss' and 'problem' inserted into line 3. In my analysis, both line 1 and line 3 would be regarded as one count each of insertion, while line 1 would also provide one count of alternation. Line 3 has two insertions, namely 'discuss' and 'problem', embedded in the Hindi frame $kar\ l ege$ 'we'll do'. This ensured that insertions, which are inherently the more numerous switch type (sentences commonly have two, three or more insertions, but only one alternation), were given the same weighting as alternations. Based on the above criteria, frequency histograms were generated for the following turn types: Hindi-only, English insertions into a Hindi sentence, Hindi/English alternation, Hindi insertions into an English sentence and English-only. Multiple pairwise comparisons for each interaction type were carried out using the chi-square test.

Results

Types of code-switching – an overview

As mentioned in the Methodology section, code-switching phenomena found in the dialogues of the movies analysed could be classified into two of the categories proposed by Muysken (2000). These were insertions (single-word or single-utterance expressions) and alternations (the simultaneous switching of several constituents), their relative abundances being strongly influenced by the interaction type, and by the decade of the film's release. A third type of switch occurred very rarely, and was confined to brief stretches of speech. This exclusively took the form of English utterances following a Hindi syntactic pattern (the syntactic transference of Clyne, 2003), as follows:²

(1) Sir, waterproofing meaning more money. (sar, waterproofing matlab aur paisā) 'Sir, the waterproofing will cost more money.'

This utterance is from the movie *Monsoon Wedding*, the speaker being the wedding planner Dube, an individual from a distinctly lower socioeconomic background than the other main characters. It is possible that this type of syntactic transference is more a hallmark of the English spoken by individuals from a certain socioeconomic stratum, but this hypothesis needs to be rigorously tested. The study of such utterances might also be more at home in a discussion on Indian English, and due to their rarity, are not discussed any further here. The other two types of code-switching, on the other hand, are described below in some detail.

Insertions

Insertions were found to most often take the form of (grammatically unintegrated) bare nouns (example 2 below), placed in the appropriate spot in accordance with the syntax of the substrate host language. Often, an English noun can also be inserted in its plural form (3). An interesting

phenomenon is the insertion of an English bare noun in a context where a plural form is called for: this is signalled by both the semantics of the sentence, as well as by inflections on the finite verb and on the copula. Thus in (4), $s\bar{a}re$ 'all', the auxiliary verb gaye and the copula $ha\check{i}$ all require the plural form 'fuses'. The speaker, on the other hand, is content to use only the bare form of the noun.

- (2) gěde lawn pe pade haĭ, sāre ke sāre. marigolds lawn on lie.PST.PL are all of.PL all 'All the marigolds are lying on the lawn.'
- (3) usko pahle das pieces ke andar kāṭ dī-jie. that.ACC first ten pieces of.PL inside cut give-IMP.POL 'First cut it into ten pieces.'
- (4) lagtā hai sāre fuse uḍ gaye haĭ. seems COP.3SG all fuses fly gone COP.3PL 'It seems that all the fuses have blown.'

It is worth reflecting a moment on the nature of these one-word English utterances within what might be considered a Hindi 'frame'. Although they have been presented here as 'insertions', and therefore as instances of code-switching, it is entirely possible that in some cases, these words might in fact be stable borrowings into the Hindi lexicon. Lipski (2005) provides some criteria for identifying English 'nonce borrowings' into Spanish, namely that they (a) are used consistently (in preference over an indigenous alternative, if any), (b) become adapted to the phonotactics and morphology of the borrowing language, (c) are no longer regarded as 'foreign' words and (d) tend to be used by speakers whose fluency in the borrowing language is much greater than that in the source language (this last section has been paraphrased to suit the Indian situation). Many 'English' words that behave in this manner can be identified in the speech of Hindi speakers from all socioeconomic strata - botal 'bottle', tren 'train', minit 'minute' and so on. These words are used exclusively to refer to the concepts they represent, have taken on Hindi phonology, are frequently used by Hindi speakers with little to no knowledge of English and are also capable of taking on Hindi inflections such as the plural: botalě 'bottles', das minițo mě 'in ten minutes'. Among the movie examples given here, 'lawn' and 'fuse' (examples 2 and 4) might be candidates for nonce borrowings (or of words gradually undergoing lexification into Hindi), mainly because of a lack of any frequently used Hindi alternatives. However, determining whether they meet the other criteria would require further research. 'Pieces' and 'change' (examples 3 and 6), on the other hand, do not meet any of the above criteria, and are undoubtedly instances of code-switching. For the purpose of this study, all instances of English usage were included in the quantitative analyses, as it was not feasible to make a ruling on each and every case.

Two or more English words are often inserted in close proximity to one another, or in a chain. However, as there is no change of code, these may not be considered instances of alternation. In (5), 'traffic jam' forms a constituent, and can be safely classed as a single insertion. In (6), however, 'Constitution change' cannot be considered a constituent, in spite of the obvious similarities to the English expression 'Constitutional change', as (the verb) change is more tightly bound to *karvāiye* 'get something done'. 'Constitution' and 'change' are therefore two separate units inserted in tandem, with 'change *karvāiye*' being used as a compound verb (for a detailed discussion on compound verbs as a feature of mixed codes, see Edwards & Gardner-Chloros, 2007).

(5) vo traffic jam mě phas gayā hǔ. that traffic jam in stuck go.PST COP.1SG 'I am stuck in a traffic jam.' (6) pahle āp Constitution change kar-vā-iye. first you.POL constitution change do-CAUS-IMP.POL. 'Get the Constitution changed first.'

English adjectives can be easily inserted into a Hindi clause, both in a predicative sense (7), as well as an attributive one (8).

- (7) $\bar{a}p$ $bh\bar{i}$ $y\check{u}$ $h\bar{i}$ tense ho $j\bar{a}$ -te ho. you. POL also thus only tense be go-1SG COP.2SG 'You get tense so easily.'
- (8) *e* Australian *chorā*, *idhar ā!*INT australian boy here come.IMP

 'You, Australian boy, come here!'

Finally, Hindi words can be inserted into an English sentence, albeit with much reduced frequency. Words that are inserted include nouns (9), adverbs (10) and adjectives (11). It is interesting to note that when Hindi insertions into English are pluralized, this is invariably done using the English plural marker -s, as in (12), indicating grammatical integration into the underlying English utterance. Similarly, some English words, when inserted into Hindi, can be pluralized using Hindi plural markers (13). However, there were no instances of Hindi nouns with Hindi plural markers being inserted into an English substrate (e.g. 14). Once again, the question of whether these words represent Hindi loans into English is relevant, and the same issues that were discussed above apply here as well. For the sake of efficiency, all Hindi words used in an English frame are analysed as instances of insertion.

```
(9) And take off that stupid topī! '...hat'
```

(10) And we got married fațāfaț.

"... straight away"

(11) Nothing is pakkā.

"... certain."

(12) He wont be singing and dancing at people's $\hat{S}ad\bar{t}$ -s.

"...wedding-PL"

- (13) mai păc ticket-ĕ lekar āyā hŭ, sājan film kī.

 I five ticket-PL bring+CONV come.PST COP.1SG sajan film of.F
 'I've brought five tickets for the movie Sajan.'
- (14) *I need five kursi-yă.

'...chair-PL'

Alternations

Alternations are another common form of CS in the present data, and are more easily characterized than the insertions. The alternations have several features in common with those described by Muysken (2000), including:

- The involvement of several constituents;
- Length and complexity;
- Structural position (occurring at clause boundaries); and
- The frequent involvement of discourse particles and adverbs, allowing bidirectional switching.

The vast majority of alternations do involve extended switches into the other languages, with multiple constituents, at clause boundaries (15, 16). In (15), there is first a switch from Hindi to English, and then from English to Punjabi (upper-case text) at the clause boundaries. However, alternations may also occur intra-sententially (16).

(15) itnī jaldī Śādī tay ho gaī, and Lalit is doing everything single-so.F fast marriage arrange be go.PST.F handedly. ŚāDī Dā KARE, SAU KāM HONDE.

marriage of do.COND hundred jobs be.PRES.PL

'The marriage got arranged so quickly, and Lalit is doing everything single-handedly. There are a hundred different things to do in a wedding.'

(16) ham kyā kar rahe haĭ is none of your business. we what do stay.PL COP.1PL 'What we are doing . . . '

Discourse markers are frequently used to cap sentences at initial (17) or final (18) positions; such markers can also be in Hindi (e.g. $nah\tilde{i}$ in (19)). In addition, kinship terms and titles also seem to have the same distribution as discourse markers, and could be considered instances of alternation (20, 21). In both examples, it would be acceptable to place $y\bar{a}r$ and $d\bar{i}d\bar{i}$ at the beginning of the utterance.

(17) *thoḍā sā is taraf muḍ-iye*, please! bit DIM this way turn-IMP.POL 'Turn a bit this way, please!'

(18) Anyway, *kuch apne bāre mě batā*. some self about in tell+IMP '...tell me about yourself.'

(19) nahĭ, I think ye bahut zyādā camkīlā hai. no I think this too much shiny.M COP.3SG 'No, I think this is too shiny.'

(20) That's not the problem, *yār*.

'... buddy'

(21) I didn't mean that, $d\bar{\imath}d\bar{\imath}$. '... sister'

Quantitative results: code-switching in Bollywood

Code-switching commonly occurred in all movies analysed, and was present in one form or another in the speech of practically all the characters. However, there were clear trends in the relative amount of code-switching from one decade to another; in general, the use of English (in the form of isolated words, phrases, clauses, sentences or entire turns) was far more prevalent post-2000 than it had been in the 1980s or 1990s. This is clearly illustrated by a comparison of the speech between middle-aged speakers (the 'adults') in a domestic situation from the art-house movies *Masoom* and *Monsoon Wedding*, which are separated by almost two decades (Figure 1). The data reveal that between 1982 and 2001, the number of turns which were composed exclusively of Hindi words fell dramatically from around 85 per cent to just under 20 per cent, while the number of turns that contained at least one English word rose from 15 per cent to over 80 per cent. An

analysis of the speech of younger speakers (typically the children of the adults mentioned above, henceforth 'young speakers') from the commercial movies reveals that while there is a similar, albeit weaker, trend of increasing English usage (Figure 2; 36% of turns in 1982 to 70% in 2004), the sharpest increase occurs post-1990, with movies from the 1980s and 1990s containing more-or-less similar proportions of Hindi and English use.

An analysis of only the turns which contain CS is highly informative. Here I focus on the interactions between young speakers, as this is the most frequent interaction type in all the movies analysed in this study, and therefore represents the most complete data set. Figure 3 shows the relative frequencies of turns containing at least one instance of CS in the art-house movies. While English insertions into Hindi are clearly the most abundant group in 1982, there is nevertheless a substantial amount of Hindi-English alternation. By 2001, however, young speakers had completely abandoned English insertions into Hindi (along with Hindi-only turns; data not shown), and were instead much more likely to insert Hindi words into English. The frequency of Hindi-English alternations was also higher than before. These results are strongly suggestive of the 'turnover' phenomenon described by Myers-Scotton (1993), where the matrix and embedded languages swap roles. Clyne, Eisikovits, and Tollfree (2002) and Clyne (2003) elaborate on this phenomenon, adding that in the Australian migrant situation, turnover may be 'associated with a shift . . . from a community language and the transfer of the function of a symbolic identity marker to an ethnolect of Australian English' (Clyne et al., 2002, p. 140). While it cannot really be argued that the young speakers in the art-house movies have abandoned their 'community' language (Hindi) altogether, it is nevertheless highly likely that current interactions between such speakers involve an ethnolect characterized by high levels of English-only, as well as Hindi insertions into English.

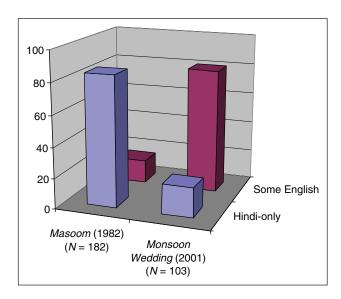


Figure 1. Overall English content in the dialogues of *adults* from the art-house movies *Masoom* and *Monsoon Wedding*

Notes: The columns represent the percentage of turns with some English and turns with Hindi-only. N denotes the number of turns analysed for each movie. The change in the distribution of turn types from 1982 to 2001 is statistically significant (p < .0001).

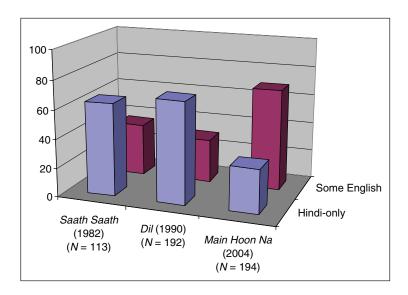


Figure 2. Overall English content in the dialogues of *young speakers* from three commercial movies from the three decades investigated in the study

Notes: The columns represent the percentage of turns with some English and turns with Hindi-only. N denotes the number of turns analysed for each movie. The change in the distribution of turn types over the three decades is statistically significant (p < .0001).

The inverse relation between English insertions and Hindi–English alternations seen in the art-house movies is apparent in the commercial movies as well (Figure 4). In 1982, the vast majority of switches (almost 85%) are English insertions into Hindi. Such switches predominate in the 1990s as well (in spite of a drop to around 65% of switches in *Dil*). However, in the 2001 and 2004 movies, there are roughly equal numbers of turns with English insertions into Hindi and Hindi–English alternation. In contrast to the 2001 art-house movie *Monsoon Wedding*, the commercial movies from the same period have a negligible amount of Hindi insertion into English sentences. Clearly, the trend in both sets of movies is towards a more English-intensive form of code-switching, exemplified by alternations in the commercial movies and Hindi insertions into English in the art-house movie.

As mentioned in the section on choice of film, four of the commercial movies revolve around high-school or university students; this provides an ideal opportunity to investigate CS patterns in the bidirectional interactions between students and their teachers. The distribution of turn types was found to change significantly across the three decades for the three student—teacher pairings (student to teacher, p = .0003; student to student, p < .0001; teacher to student, p = .0017). The trends in Figure 5 are less clear than in other interaction types, but in general, *student to teacher* interactions over the three decades exhibit decreasing English insertions into Hindi, and increasing Hindi—English alternation. More striking are the differences between student to teacher interactions and interactions between young speakers (which, in these movies, also happen to be *student to student* interactions; Figure 6). Comparing Figure 5 to Figure 6, we see that within each decade, interactions between young speakers have far more Hindi-only turns than the respective student to teacher interactions. Conversely, these latter interactions have a much greater proportion of English-only interactions than the corresponding interaction types in each decade showed

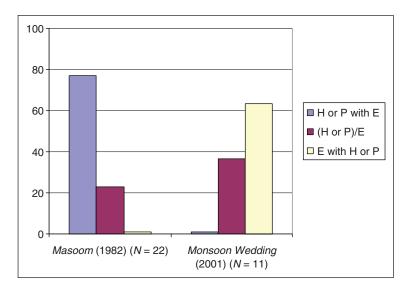


Figure 3. Proportions of code-switched turns only (excluding Hindi-only and English-only turns) in interactions between young speakers from the two art-house movies

Notes: The columns represent the proportions of the various turn types (see text for details). Legend abbreviations: H or P with E – English insertions into Hindi or Punjabi; (H or P)/E – English alternation with Hindi or Punjabi; E with H or P – Hindi or Punjabi insertions into English. Units along the y-axis indicate percentages. N denotes the number of turns analysed for each movie. The change in the distribution of turn types is statistically significant (p < .0001).

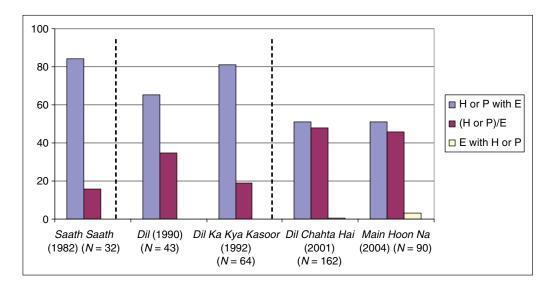


Figure 4. Proportions of code-switched turns only (excluding Hindi-only and English-only turns) in interactions between young speakers from the five commercial movies

Notes: The columns represent the proportions of the various turn types (see text for details). Legend abbreviations: H or P with E – English insertions into Hindi or Punjabi; (H or P)/E – English alternation with Hindi or Punjabi; E with H or P – Hindi or Punjabi insertions into English. Units along the y-axis indicate percentages. N denotes the number of turns analysed for each movie. The overall change in the distribution of turn types is statistically significant (p < .0001), although the movies from the 1980s and 1990s are statistically similar (p = .08).

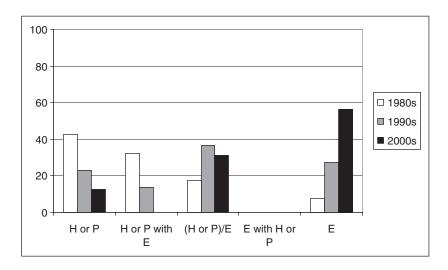


Figure 5. Code-switching patterns of the *student to teacher* interaction type from four commercial movies *Notes*: The columns represent the proportions of the various turn types (see text for details). Abbreviations along the x-axis: H or P – Hindi or Punjabi only; H or P with E – English insertions into Hindi or Punjabi; (H or P)/E – English alternation with Hindi or Punjabi; E with H or P – Hindi or Punjabi insertions into English; E – English-only. Units along the y-axis indicate percentages. Movies analysed and number of turns: 1980s, *Saath Saath* (N = 40); 1990s, *Dil* and *Dil Ka Kya Kasoor* (N = 22); 2000s, *Main Hoon Na* (N = 16).

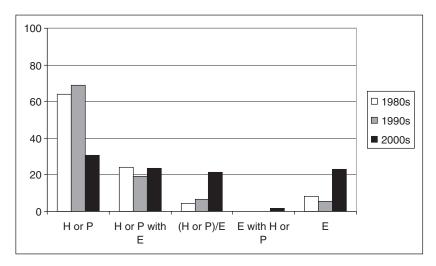


Figure 6. Code-switching patterns of the interactions between students (young speakers) from four commercial movies

Notes: Movies analysed and number of turns: 1980s, Saath Saath (N = 113); 1990s, Dil and Dil Ka Kya Kasoor (N = 422); 2000s, Main Hoon Na (N = 194). Other details as in Figure 5.

that interactions between students are indeed different from student utterances directed at teachers (1980, p = .02; 1990, p < .0001; 2000, p = .006). Teacher to student interactions were found to be more resistant to change than those discussed above, particularly for the period 1990–2000 (Figure 7). While the relative abundance of Hindi-only turns declined between 1982 and 2004, the

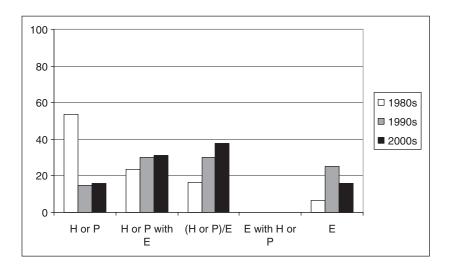


Figure 7. Code-switching patterns of the teacher to student interaction type from four commercial movies Notes: Movies analysed and number of turns: 1980s, Saath Saath (N = 30); 1990s, Dil and Dil Ka Kya Kasoor (N = 32); 2000s, Main Hoon Na (N = 32). Other details as in Figure 5.

proportion of Hindi–English alternations also seems to have risen between 1982 and 1990. However, teacher to student utterances were similar to student to teacher utterances in 1980 and 1990, only to become significantly different in 2000 (compare Figure 5 and Figure 7; 1980, p = .8; 1990, p = .49; 2000, p = .01).

These results confirm the special position that English occupies in the higher learning domain in India. As early as 1982, university students were frequently (about 30% of the time) addressing their lecturers with English-only turns. This, incidentally, was far greater than the proportion of English-only turns in speech directed towards other students (approximately 10%). This is not surprising, as it merely reflects the accepted norms of language choice when addressing teachers in a private school or university setting. The trend continues up to 2004, where over half of all student to teacher turns are in English-only, compared to just over 20 per cent of turns directed at fellow students. The speech of teachers towards students has remained more conservative – Hindi–English alternations are the most frequent switch type, closely followed by English insertions into Hindi.

In summary, the results show that there has been a massive increase in the use of English in Bollywood movie conversations for the genres investigated, with the biggest changes taking place post-1990. CS patterns have also changed over time, with dialogues from the 1980s containing a preponderance of English insertions into Hindi, and post-2000 dialogues showing a far greater proportion (compared to previous decades) of Hindi–English alternation, English-only turns and Hindi insertions into English.

On the suitability of Bollywood scripts for the study of code-switching

It could be argued that the dialogues of the movies analysed in the present study are not accurate reflections of the language use patterns among the Hindi- and English-speaking urban elite of India. After all, the output of professional scriptwriters could deviate from the 'normal' speech of the types of characters portrayed in the movies, by being overly well-crafted, ornate or stereotyped, or by over- or underrepresenting the amount of CS that such characters might engage in. In the

1960s, Einar Haugen voiced similar objections after reading the novels of Vilhelm Moberg, which purported to accurately represent the code-switched speech of Swedish immigrants to the US in the 1850s (cited in Haugen, 1977). His criticism, that the dialogues contained in the novels did not seem to be realistic examples of spoken American Swedish, was countered by the assertion that Moberg had used, as his source, the letters and diaries of an actual Swedish migrant to Minnesota. Haugen later found that these documents did not contain the types of phrases used in Moberg's novels. Haugen's criticisms stemmed from the realization that bilingual communities invariably develop a norm for speech patterns; speakers perceive these rules through 'an intuition built in through personal experience', but Moberg had, unfortunately, preferentially used the more extreme phrases that lay beyond the accepted norm.

Given such (very real) concerns regarding fictional works, it is significant that, in the present study, the dialogues of movies from the same period show a high level of convergence, even though the dialogues were all written by different scriptwriters. A comparison of columns 1 and 2 of Table 2 shows a striking similarity between the CS patterns of the interactions between young speakers in the two movies from the 1980s. Comparisons of columns 2 and 3, and of columns 4 and 5 also reveal near-identical distributions of turn types within the 1990s and the post-2000 periods. Such similarities could only have arisen if the scriptwriters were mindful of contemporary speech patterns (Haugen's norms), and were consciously trying to reproduce them. This was clearly articulated by Sabrina Dhawan, scriptwriter for *Monsoon Wedding*, who in an interview for the web magazine *rediff.com* commented:

The film has its characters flitting from English to Hindi with doses of Punjabi thrown in . . . that's the way we speak at home in Delhi, I wanted to stay very true to that. I borrowed big-time from life.

However, there are certain features of Bollywood CS patterns that one needs to be mindful of. One important phenomenon is the difference in the amount of English used between the post-2000 'commercial' movies *Dil Chahta Hai* and *Main Hoon Na*, on the one hand, and the 'art-house' *Monsoon Wedding*, on the other (compare columns 5 and 6 in Table 2 with data from *Monsoon Wedding*: H or P only, 0%; H or P with E, 0%; (H or P)/E, 7.4%; E with H or P, 12.9%; E only, 79.6%). The volume of English in the commercial movies could have been deliberately reduced as a concession towards their largely Hindi-speaking domestic audience; *Monsoon Wedding*, directed by Mira Nair, a New York-based director, was pitched towards a more international audience, and

Table 2. Comparison of the relative proportions of turn types in the dialogue of young speakers within pairs of movies from each decade (in percentages)

	Masoom (1982)	Saath Saath (1982)	Dil (1990)	Dil Ka Kya Kasoor (1992)	Dil Chahta Hai (2001)	Main Hoon Na (2004)
H or P	77.2	63.7	69.8	68.3	33.2	30.4
H or P with E	10.3	23.9	14.6	22.6	22.0	23.7
(H or P)/E	7.6	4.4	7.8	5.2	20.7	21.1
È with H or P	0.0	0.0	0.0	0.0	0.3	1.5
E	4.8	8.0	7.8	3.9	23.9	23.2

Notes: H or P - Hindi or Punjabi only; H or P with E - English insertions into Hindi or Punjabi; (H or P)/E - English alternation with Hindi or Punjabi; E with H or P - Hindi or Punjabi insertions into English; E - English-only. Chi-square tests indicate no significant differences between the movies in each decade.

was free from such considerations. Such 'language planning' in the commercial movies often leads to the creation of Hindi dialogue that seems to be at odds with the characters that are made to speak it. For instance, in *Dil Chahta Hai*, the three young central characters are clearly from very affluent backgrounds, either flying off to Sydney to manage family business interests, or driving from Mumbai to Goa in a sporty Mercedes. Yet, they switch to very eloquent 'pure' Hindi at various points in the movie, especially when discussing philosophical or emotional topics, as illustrated in the following dialogue:

Samir: kitnī khūbsūrat jagah hai.

'It's such a beautiful place.'

Akash: Hmm . . . it's beautiful.

Samir: You know what, hamĕ har sāl kam se kam ek hafte ke liye goa ānā cāhiye.

"... we should come to Goa for at least one week every year."

Akash: Definitely. kyŏ sid?

'Well, Sid?)'

Sid: Umm.

Akash: kyā soc rahe ho?

'What're you thinking?'

Sid: dekh rahā hŭ vo jahāz jo zarā der bād dikhāī nahĭ degā. jānte ho, ham tinŏ us jahāz kī

tarah haĭ, jo āj nahĭ to kal apnī manzilŏ ko ḍhŭḍte hue niklĕge. aur ho sakta hai ki hamāre manzilĕ alag alag hŏ.

'I'm just looking at that ship which will be over the horizon soon. You know, the three of us are like that ship, we'll soon head off for our own destinations, and it's possible that we'll end up in different places.'

Dil Chahta Hai (2001)

It seems very unlikely that a character of Siddharth's (abbreviated Sid) background would ever in real life launch into a monologue on the nature of life in Hindi-only. Even Samir's initial comment on the beauty of the location seems unnatural for someone who would soon be living in an apartment, in an upmarket part of Sydney, Australia, and with a view of the Harbour Bridge and the Opera House. A similar phenomenon was noted by Gardner-Chloros and Charles (2007), following an analysis of CS content in several different programme genres on the popular satellite TV channel Zee TV. While CS was present at 'normal' levels in programmes involving unscripted speech, it was underrepresented in scripted comedy and drama serials. The authors ask whether 'the portrayal of CS by fiction writers might be "tidied up" along with other "inconvenient" features such as overlaps, interruptions, repetitions etc' (Gardner-Chloros & Charles, 2007, p. 111) – the findings from my study indicate that this might well be the case.

Finally, sociolinguistic surveys of bilingual Hindi–English inhabitants of Delhi have revealed patterns of English use that support the findings reported in this study. Although Malhotra (1980) found that both Hindi and English were used within the home domain between 1972 and 1977 in upper-middle-class families, she mentions that 'in situations which are related to intimacy [i.e. home, family, kindred and close friends] there is a preference for the use of Hindi, and in those related to status a preference for English. . . . Most often a parent talking to his/her partner will switch to Hindi' (Malhotra, 1980, p. 42). Similarly, while reporting findings from her 1983 MPhil thesis, Sahgal (1994) notes that Hindi speakers from elite residential areas of Delhi used English in the family domain at a level lying between 'rarely' and 'occasionally'. English was used at a higher level ('occasionally') in the friendship domain, although this was still lower than the level of Hindi

use (between 'occasionally' and 'most of the time'). English use was highest in the institutional domain, equalling the level of Hindi use (between 'occasionally' and 'most of the time'). More recently, comparable sociolinguistic studies are harder to find, although Eilert (2006) reports some language use patterns in Delhi, using a survey-based study that did not control for socioeconomic status. Despite this drawback, it is interesting to note that 34 per cent and 41 per cent of respondents claimed to speak English with family and friends respectively.

Conclusions

This study has demonstrated that the manner and frequency of English use in Bollywood movies has changed in significant ways over the last three decades. I have argued in the latter half of this article that these changes are a reasonable indicator of English usage in upper-middle-class urban India. However, the small sample of movies used in the study, combined with the fact that the data were drawn from fictional sources, make it advisable to exercise a certain amount of caution when attempting to extrapolate these findings to real-world situations. The greatest utility of results such as these lies not in providing precise descriptions of complex sociolinguistic phenomena, but in indicating broad trends, and suggesting novel hypotheses, which can later be tested through the collection of more reliable field data. The results from this study have also suggested further questions regarding diachronic aspects of CS, that one may attempt to answer using Bollywood movie dialogues, or using real-life synchronic data. As children in affluent Indian families are exposed to large amounts of English from an early age, the volume and complexity of CS could be heavily dependent on variables such as age and socioeconomic status; such issues will be investigated in a future study. Children from affluent urban households probably also have a very limited ability to produce monolingual Hindi speech. Hindi proficiency could therefore be an important factor associated with CS patterns, particularly among the younger generation. This could also be easily tested by means of bidirectional translation tests on groups of age- and SES-matched children. A high volume of English in conversation should correlate well with low scores on the English-to-Hindi translation task, and vice versa.

The data indicate a strong 'turnover' in the language of choice among young people, but what might be driving this phenomenon (if indeed it exists in contemporary Indian society)? Two possible explanations are that either (a) the overall volume of English usage has increased dramatically, or (b) that new types of code switching have emerged. While nothing conclusive can be said on this topic on the basis of the current study, my data do indicate that both factors might be in operation, albeit to differing extents. Figures 1 and 2 show that the proportion of utterances with some English was significantly higher post-2000 than it was in the 1980s, indicating that English started to be used more frequently in recent years. Figure 3, on the other hand, shows that although Hindi insertions into English were present at a very low level in 1982, their frequency was far greater in 2001 – this indicates that the Hindi insertions into English strategy only recently started to be popular. Nevertheless, the answer to the above question must necessarily remain conjectural for the time being.

While I would suggest that the trends presented in this study have high predictive power in terms of overall levels of English, the fictional nature of the data places definite restrictions on one's ability to make generalizations about finer-scale events: predicting the locations of switch points, or estimating the probability of an insertion as opposed to an alternation in a particular turn, for instance. Nevertheless, Bollywood movies have proven a rich source of data, on which models concerning sociolinguistic issues can be based. The study of change in any spoken language over time presents considerable methodological challenges for any researcher, but this study has provided some insight into this issue through the use of an unconventional data source.

Acknowledgements

I am grateful to Professor Michael Clyne for friendly advice and valuable comments.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

- Note on transcriptions: Retroflex consonants in the Hindi parts of the dialogues are indicated by t and d. Long vowels are represented as ā, and ă is a long nasalised vowel.
- Abbreviations used in interlinear glosses are as follows: ACC = accusative; CAUS = causative; COND = conditional; CONV = converb; COP = copula; DIM = diminutive; F = feminine; IMP = imperative; INT = interjection; M = masculine; PRES = present; PST = past; PL = plural; POL = polite; SG = singular; 1, 2, 3 = 1st, 2nd, 3rd person.

References

Acharya, S. (2004). *Bollywood and globalization*. Unpublished master's thesis, San Francisco State University. Auer, P. (1998). Introduction: Bilingual conversation revisited. In P. Auer (Ed.), *Code-switching in conversation* (pp. 1–25). London: Routledge.

Bhatt, R. (1997). Code-switching, constraints, and optimal grammars. Lingua, 102, 223-251.

Clyne, M. (2003). Dynamics of language contact. Cambridge: Cambridge: Cambridge University Press.

Clyne, M., Eisikovits, E., & Tollfree, L. (2002). Ethnolects as in-group varieties. In A. Duszak (Ed.), *Us and others* (pp. 133–158). Philadelphia, PA: John Benjamins Publishing.

Dua, H. (1993). The national language and the ex-colonial language as rivals: The case of India. *International Political Science Review, 14*, 293–308.

Edwards, M., & Gardner-Chloros, P. (2007). Compound verbs in codeswitching: Bilinguals making do? *International Journal of Bilingualism*, 11, 73–91.

Eilert, R. (2006). *English in India, a study of native Hindi speakers in Delhi*. Unpublished master's thesis, Australian National University.

Eliashberg, J., Elberse, A., & Leenders, M. (2006). The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science*, 25, 638–661.

Gafaranga, J. (1997). Code-switching/code-mixing or Kinyarwanda for all practical purposes: The base language issue. *Actas do I Simposio Internacional sobre o Bilingüismo* (pp. 504–514).

Gafaranga, J., & Torras, M. (2002). Interactional otherness: Towards a redefinition of codeswitching. *International Journal of Bilingualism*, 6, 1–22.

Gardner-Chloros, P., & Charles, R. (2007). Subiko welcome (Welcome to everyone): Hindi/English codeswitching in the British–Asian media. *Birkbeck Studies in Applied Linguistics*, 2, 89–122.

Gumperz, J. (1961) Speech variation and the study of Indian civilization. American Anthropologist, 63, 976–988.

Gupta, R. (1991). English and Indian languages: Code mixing. In R. Gupta & K. Kapoor (Eds.), English in India: Issues and problems (pp. 207–215). Delhi: Academic Foundation.

Haugen, E. (1977). Norm and deviation in bilingual communities. In P. Hornby (Ed.), *Bilingualism: Psychological, social and educational implications* (pp. 91–102). London: Academic Press.

Kachru, B. (1978). Code-mixing as a communicative strategy in India. In J. Alatis (Ed.), *International dimensions of bilingual education* (pp. 107–124). Washington, DC: Georgetown University Press.

Kumar, A. (1986). Certain aspects of the form and functions of Hindi–English code-switching. *Anthropological Linguistics*, 28, 195–205.

Lipski, J. (2005). Code-switching or borrowing? No sé so no puedo decir, you know. In L. Sayahi & M. Westmoreland (Eds.), *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics* (pp. 1–15). Somerville, MA: Cascadilla Proceedings Project.

Malhotra, S. (1980). Hindi–English, code-switching and language choice in urban, upper-middle-class Indian families. In P. Hamel and R. Schafer (Eds.), *Kansas Working Papers in Linguistics*, *5*, 39–46 (reprinted 1992).

Malik, L. (1994). Sociolinguistics: A study of code-switching. New Delhi: Anmol Publications.

Muysken, P. (2000). *Bilingual speech: A typology of code mixing*. Cambridge: Cambridge University Press. Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in code-switching*. New York: Oxford University Press.

Parasher, S. (1991). Indian English functions and form. New Delhi: Bahri Publications.

Rao, S. (2006). The globalization of Bollywood: An ethnography of non-elite audiences in India. Paper presented at the International Association for Mass Communication Research, Cairo, Egypt.

Romaine, S. (1995). Bilingualism. Cambridge, MA: Blackwell Publishers.

Sahgal, A. (1994). Patterns of language use in a bilingual setting in India. In R. Agnihotri & A. Khanna (Eds.), Second language acquisition: Socio-cultural and linguistic aspects of English in India (pp. 299–306). New Delhi: Sage.

Singh, R. (1995). Hindi–English code-mixing: Some implications for a general theory. *Indian Linguistics*, 56, 95–105.

Singh, R. (1998). Linguistic theory, language contact and modern Hindustani: The three sides of a linguistic story. New Delhi: Munshiram Manoharlal Publishers.

Srivastava, R., & Sharma, V. (1991). Indian English today. In R. Gupta and K. Kapoor (Eds.), English in India: Issues and problems (pp. 189–206). Delhi: Academic Foundation. Copyright of International Journal of Bilingualism is the property of Sage Publications, Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.