Forms and functions of codeswitching to Hindi/Urdu in Indian English and Pakistani English

Henna Kortelainen

MA Thesis

English, Researcher Path

School of Languages and Translation Studies

Faculty of Humanities

University of Turku

May 2020

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

UNIVERSITY OF TURKU
School of Languages and Translation Studies/Faculty of Humanities
KORTELAINEN, HENNA: Forms and functions of codeswitching to Hindi/Urdu in Indian English and Pakistani English

MA Thesis, 110 p., 20 p. appendices
English, Researcher Path
May 2020

– –

This study examines the forms and functions of codeswitching to Hindi/Urdu in Indian English and Pakistani English. The aim is to find out how these local languages manifest and are used in the local non-native variety of English. Previous research on codeswitching to Hindi and Urdu in Indian English and Pakistani English has mostly concentrated only on the use of single lexical items and their impact on the lexicon of the local English leaving the more varied ways in which the local languages show relatively unresearched. The material was collected from the Corpus of Global Web-based English (GloWbE). This study served additionally as a methodological experiment using the most frequent Hindi and Urdu words to locate and collect codeswitches in the corpus.

The analysis of the structural patterns showed that Hindi and Urdu codeswitches manifest in a variety of different forms ranging from longer intersentential codeswitches for complete Hindi and Urdu sentences to interclausal switches, i.e. switches between main and dependent clauses, and to shorter intraclausal codeswitches like words and phrases. The structural analysis also revealed that the structural patterns appear to follow the same tendencies in both Indian and Pakistani English.

The Hindi/Urdu codeswitches also served diverse types of functions. The switches could roughly be divided into switches with a communicative function and cultural switches. Communicative functions included, among others, quotations, figurative language, conveying greetings and prayers, interjections, reiterations, and metalinguistic commentary. Cultural codeswitches expressed objects and concepts specific to the Indian and Pakistani culture. Cultural switches also functioned as references to the Indian and Pakistani culture implying the Indianness or Pakistaniness of something or someone.

Keywords: codeswitching, Indian English, Pakistani English, Hindi, Urdu

Table of contents

Appendices

List of tables

List of figures

List of abbreviations:

CS	codeswitching
IE	Indian English
PE	Pakistani English
GloWbE	Corpus of Global Web-based English
ML	matrix language
EL	embedded language

1 Introduction

When English arrived in the Indian subcontinent hundreds of years ago, it was introduced to the vastly multilingual and multicultural settings of the area. The roots of the presence of English are in the colonial era, which started in the early 17th century. As the British gradually gained more political power over the area, English started to be used in administrational system as well as in education, which were important developments in helping consolidate the position of English in the Indian subcontinent. After the end of the British rule and the declaration of independence in 1947, English was adopted as the official language of the state along with Hindi in India and along with Urdu in Pakistan. Today, English continues serving several functions in the society; it is the language of administration and higher education, and it is also used as the medium of instruction in English-medium schools. English also serves as an intranational lingua franca in both countries. In India and Pakistan, English is also used widely in the media, and both countries have their English-language literature.

India and Pakistan are both vastly multilingual societies where language contact with its different outcomes and manifestations is everyday life. English has become part of the linguistic repertoire and is one language more into the mix to influence and to become influenced by the regional, indigenous languages. The coexistence means linguistic influence in both directions; while English, both as an official language and as a global language, exerts influence on Hindi, Urdu, and other local languages, English is also influenced by the indigenous languages of the area. However, codeswitching to English in Hindi and Urdu appears to have received considerably more attention from researchers (for example Verma (1974), Kachru (1978a), Patel (2016), Rafi (2013), Parveen and Aslam (2013), Ehsan and Aziz (2014), Iqbal (2011)) than codeswitching to Hindi and Urdu in the local English. The previous research and accounts on the use of Hindi and Urdu in Indian and Pakistani English have concentrated more on the impact of the local languages on the lexicon of the local English (for example Sailaja (2009), Rahman (1990), Mahmood and Shah (2011), Ahmad and Ali (2014)) ignoring the more varied ways in which the local languages manifest and are used as part of the local English. Thus, the research into both the functions and structural patterns of codeswitching in both Indian English and Pakistani English has remined relatively fragmentary. While the lexical influence of Hindi and Urdu on the local variety of English has been studied somewhat, no comprehensive accounts on either the functions or forms of codeswitching seem to exist. The aim of this study is thus to find out what forms the Hindi/Urdu codeswitches take in Indian English and Pakistani English and what functions they serve. My research questions are:

-What are the structural patterns of codeswitching to Hindi/Urdu in Indian English and Pakistani English?

-What are the functions of codeswitching to Hindi/Urdu in Indian English and Pakistani English?

The motivation to look into the codeswitching practices in Indian English and Pakistani English simultaneously is to find out if in these two contexts, where the societies are overall characterized by extensive multilingualism and language contact, and where English plays similar roles in the society, display similar characteristics in codeswitching to the local languages – in this case Hindi and Urdu – both in terms of forms and functions in the local English. From the structural point of view, examining the use of Hindi in Indian English and the use of Urdu in Pakistani English side by side further enables the inspection of whether the forms of codeswitching converge or whether they diverge in these two scenarios where the languages involved in the switching process are in the same typological relationship with each other. Hindi and Urdu are linguistically very closely related, and in their everyday spoken forms the languages are mutually intelligible. Because of the degree of linguistic closeness, they could almost be thought of as dialects of the same language. Hindi and Urdu are in the same typological relationship with English and thus it could be expected that the structural patterns of codeswitching to Hindi and Urdu in Indian English and Pakistani English would show similar tendencies rather than diverge remarkably. As for the functions of codeswitches, at least one type of codeswitches that can be expected in light of the previous research and knowledge of the use of Hindi or Urdu in Indian English and Pakistani English are codeswitches that are related to the local culture (for example Sailaja (2009), Mahmood and Shah (2011), Ahmad and Ali (2014)). The material for this study was collected from the Corpus of Global Web-based English (GloWbE).

The terms Indian English and Pakistani English are generally used to refer to the non-native varieties of English in India and Pakistan that developed as a result of the introduction of English into the non-native contexts of Asia and Africa – the outer circle countries (Kachru 1998, 94) – in the colonial era and that display variation and deviations from the native norms on different levels of language such as lexicon, grammar, phonology as well as discourse and style, which form the idiosyncratic, characteristic features of these varieties (Kachru 1982, 45-48). However, I employ the terms Indian English and Pakistani English (henceforth IE and PE, respectively) to refer to – not a variety characterized by specific grammatical, lexical, and phonological features – but more generally to English used by Indians and Pakistanis.

I will begin this thesis by first describing the linguistic scenario and multilingualism in India and Pakistan in more general terms, and briefly introduce Hindi and Urdu as languages. In the same section on languages and multilingualism in India and Pakistan, I will discuss the presence and status of English in these two countries both in the past and in the present day as well as the position of English in relation to local languages. In the third section, I will discuss codeswitching, its definitions, and the different theories – social, interactional, and structural – developed to account for codeswitching. In the same section I will also discuss the relationship between codeswitching and borrowing. Previous studies on the use of Hindi and Urdu in IE and PE will also be looked into in the section on codeswitching. The theoretical part is followed by the methodology section in which I will present my material as well as explain and break down the process of material collection from GloWbE. Sections 5 and 6 comprise the analytical part of the study, where the structures are analysed quantitatively and functions qualitatively. The analysis is followed by discussion on the results and findings together with further considerations on the methodology used in the material collection. In the last section I will summarize the main findings of this study and give suggestions for possible future research.

2 Languages and multilingualism in India and Pakistan

2.1 Languages in India and Pakistan

India and Pakistan are both multilingual societies. The estimates of the number of languages in India range from a few hundred to over a thousand languages depending on the source. Reasons for the variation between sources are on the one hand differences in how the distinction between languages and dialects is made, and on the other hand the fact that languages with fewer speakers are not always documented. The census of India in 1961 identified 1652 mother tongues (Mallikarjun 2002, n.p.). According to the 2011 census of India, there are 121 major languages and 270 mother tongues spoken in India. A mother tongue refers here to a language or a dialect of a language. However, the census only documented those languages with more than 10,000 speakers. Thus, the number of languages with less than 10,000 speakers remains unspecified in the census. According to People's Linguistics Survey of India, there are 780 languages spoken in India. However, the researchers estimate the real number of languages to be about 100 more than what was documented in the survey because some languages may have so few speakers that they could not be documented (Lalmalsawma 2013, n.p.). In Pakistan, the information on the number of languages ranges from 61 (Rahman 2006, 73) to above 70 (Eberhard, Simons and Fennig 2020, n.p.) depending on the source.

The nationwide official languages in India are Hindi and English. Additionally, India has 22 officially recognized languages that are listed in the Eighth Schedule of the constitution (Government of India, n.d.). The languages in the Eighth Schedule have an official status, and states and union territories can choose their official language(s) out of these 22 languages. Despite not being listed in the Eighth schedule, English is the official language – or one of the official languages – in some of the states.

India can be divided linguistically into the Hindi-speaking north and the Dravidian-speaking south where Dravidian languages like Tamil, Telugu, Kannada, and Malayalam are spoken. According to the Indian 2011 census, 43.63% of the population speaks Hindi as their first language (Government of India 2011a, 15-16). Comparison of the census data from different decades reveals that the percentage of people returning Hindi as their mother tongue has steadily increased over time. Other languages among the ten most spoken mother tongues in India are Marathi, Telugu, Tamil, Gujarati, Urdu, Kannada, Odia, and Malayalam. English is spoken as a first language by a clear minority in India. In the 2011 census, English was returned as a mother tongue by 259 678 people, which makes up 0.02% of the population (Government of India 2011a, 8). However, as a second and third language, English is spoken

by 10.19% (128 279 412) of the population (Government of India 2011b), which makes English the second most spoken language in India after Hindi.

The official languages of Pakistan are the national language Urdu and English. Pakistan has six major languages: Urdu, Punjabi, Pashto, Sindhi, Saraiki, and Balochi (Rahman 2006, 73). The language with most mother-tongue speakers – 44.15% of the whole population – is Punjabi. Although Urdu is spoken as a mother tongue by only 7.57 % of the population (Government of Pakistan 2017a), which is around 15 million people, it is the national language of Pakistan and it serves as the nationwide lingua franca. According to Rahman, Urdu is the main language of communication and "the most widely understood language" especially in the urban areas (Rahman 2006, 74). There does not, however, appear to be updated information available on how many people actually speak Urdu as a second language in Pakistan. The number of people speaking Urdu as their second language was 94 million in 1999 (Eberhard, Simons, and Fennig 2020, n.p.), which is approximately 69 % of the Pakistani population in 1999 (Government of Pakistan 2006, 25). English is hardly spoken as a native language in Pakistan. The estimated percentage of people speaking English as a second language is 11.72% of the population (Crystal 2003, 64). Although this information is from the year 2001 and might no longer be completely accurate, it can still be regarded as illustrative of the approximate share of English speakers in Pakistan.

2.1.1 Hindi-Urdu

Hindi and Urdu are Indo-Aryan languages and members of the Indo-Iranian branch of the Indo-European language family. They both derive from *Khari Boli*, which was a dialect of Western Hindi spoken in the Delhi region (King 2001, 45), and which provides Hindi and Urdu their common linguistic base (Dua 1991, 382). The most easily observable difference between Hindi and Urdu is their script; Hindi is written in the Devanagari script from left to right whereas Urdu is written in the Perso-Arabic script from right to left (King 2001, 43). The other main difference can be found in vocabulary; Hindi takes more lexical influence from Sanskrit while Urdu borrows from Persian and Arabic (Kelkar 1968, 8). When it comes to the grammatical structure of the languages, Hindi and Urdu can be said to be almost identical. In their spoken forms, Hindi and Urdu are mutually intelligible. In fact, the spoken forms are so close to each other that they could even be regarded as dialects of the same language. However, differences between Hindi and Urdu become more pronounced in higher, more formal registers of the languages. The divergence of Hindi and Urdu from each other is cumulative; while they are mutually intelligible in their spoken forms and speakers usually have no difficulties

understanding each other, the differences become greater when moving from lower registers to higher registers – eventually reaching the point where the languages have become mutually incomprehensible. The differences have to do with the vocabulary which is highly Sanskritized in higher registers of Hindi and Persianized in higher registers of Urdu (Kelkar 1968, 6-8).

According to Shapiro (2003, 253), the term *Hindustani* is sometimes used to refer to the two languages collectively in their mutually intelligible varieties. Another umbrella term for Hindi and Urdu is *Hindi-Urdu-Hindustani* which, unlike *Hindustani*, also encompasses highly Sanskritized Hindi and highly Persianized Urdu. *Hindi-Urdu-Hindustani* can be regarded as highlighting the existence of a type of a continuum where Hindustani represents the point where the varieties converge and are mutually comprehensible while Hindi and Urdu are the two ends of the scale where the polarization reaches its extreme (Shapiro 2003, 253).

According to Dua (1991, 382), four stages can be distinguished in the development of Hindi and Urdu into separate varieties: 1) formative period, 2) emergence of different bases, 3) consolidation period, and 4) polarization and identity formation. These stages cover a time period of around nine hundred years. The formative period began approximately in 1100 when Muslims invaded India (Dua 1991, 382). The Muslim invasion of Delhi in 1192 was the beginning of the Muslim rule that gradually expanded to cover most of India and that lasted for several centuries until it finally collapsed in the 18th century. Among the Muslim rulers, Persian was the language of high prestige, and it was used as the literary language and as the language of administration by the Muslim elite (Shackle and Snell 1990, 3). Sometime in the 16th and 17th centuries, a tradition of writing poetry in an archaic form of Urdu known as Dakani Urdu arose in Deccan. When these Dakani poets and writers migrated from Deccan to Delhi, they brought the Dakani poetry with them. The introduction of the "new" vernacular poetry together with the decreasing popularity of Persian as the literary language caused by the Afghanistani invaders led to Urdu replacing Persian as the preferred literary language among the Muslim elite of Delhi (Shackle and Snell 1990, 5). The formative period can be regarded as having lasted all the way to the 17th century. A significant development that took place in the formative period was the emergence of *Khari Boli* – a dialect spoken in the Delhi region that came to form the linguistic base for Hindi and Urdu. Khari Boli itself evolved from a wide dialect base, and due to the influx of Persian loans into local languages caused by the prestige status of Persian, Khari Boli had also received notable Persian influence (Dua 1991, 383).

The different bases for Hindi and Urdu started to emerge towards the end of the 17th century and the beginning of the 18th century. One crucial development leading to the formation of the different bases was that prose started to be written in the local vernacular Khari Boli - in

Devanagari script by Hindus and in Perso-Arabic script by Muslims. In addition, the Hindu writers tended to use more Sanskrit words in their writing, whereas Muslims incorporated more Persian and Arabic vocabulary into their texts (Dua 1991, 383-384). Among other significant factors in the divergence of Hindi and Urdu was the British patronage of Urdu as the official language and court language in northern India in the 18th century, which led to increased popularity of the Hindi movement that, among other things, demanded the replacement of Urdu with Hindi written in the Devanagari script for the official purposes of the court. In addition, the emergence of Hindi poetry written in the Sanskritized Khari Boli in the late 19th century further deepened the gap between Hindi and Urdu (Dua 1991, 384-385).

After the emergence of the different bases, the consolidation period followed. That was the time when three different varieties of Hindi and Urdu arose: the colloquial form of Hindi-Urdu (also known as Hindustani), High Hindi and High Urdu. The colloquial Hindi-Urdu was the spoken variety and neutral in a sense that its vocabulary was not heavily influenced by either Persian, Arabic, or Sanskrit. The high varieties, High Hindi and High Urdu, were in turn characterized by heavy absorption of Sanskrit and Perso-Arabic vocabulary, respectively (Dua 1991, 385).

The fourth stage, polarization and identity formation, took place from the late 1920s onwards and lasted until the Indian independence in 1947. The associations of Hindi and Urdu with Hindu identity and Muslim identity, respectively, grew even stronger and the rivalry between the two languages as well as Hindus and Muslims intensified. Attempting to prevent any further escalation of the language debate and to find some common ground, Gandhi suggested the adoption of a language variety he called *Hindustani* as the national language of India. Hindustani was supposed to be a mixture of vernacular, Perso-Arabic and Sanskritic features, and it would be written in both the Devanagari script and the Perso-Arabic script. However, the attempts to find a compromise solution failed and even had the opposite effect leading to increased dissatisfaction on both sides (Dua 1991, 385-386). After the partition of India in 1947, the divergence of Hindi in India and Urdu in Pakistan has increased. However, in the case of Indian Urdu this process has been slower, and Indian Urdu takes more influence from Hindi (Dua 1991, 386).

The purpose of this subsection was to provide a brief introduction to Hindi and Urdu and to give an idea how closely related the two languages are – despite certain, mostly lexical, differences. The linguistic closeness and similarities between Hindi and Urdu often make the distinction between the two languages difficult. What this means for the current study is that in many instances strictly assigning a switch to either Hindi or Urdu is practically impossible. The

decision whether a codeswitch or a codeswitched passage should be regarded as Hindi or Urdu can be dependent on one single word that can be said to be a Hindi item or an Urdu item, while the rest of the codeswitch is shared material between Hindi and Urdu – and even then the possibility of codeswitching between Hindi and Urdu cannot always be excluded. Due to the obscurity of the boundaries and the arbitrariness of the distinction between Hindi and Urdu, distinguishing between the two languages is not attempted in this study – and not even regarded as necessary. Since the purpose of this study is to look into the codeswitching practices in IE and PE rather than into how Urdu in particular is used in PE compared with Hindi in particular in IE, the distinction between Hindi and Urdu is considered neither needed nor meaningful. I will thus refer to the language as Hindi/Urdu – although, occasionally, the codeswitched language will be referred to as either Hindi or Urdu if the codeswitch is clearly either one of them and if the language of the codeswitch being either Hindi or Urdu has some specific relevance for the analysis of the codeswitch.

2.2 English in the Indian subcontinent

2.2.1 Historical background

India and Pakistan share the colonial history all the way to the partition of the Indian subcontinent into India and Pakistan and the independence of the countries in 1947. The date that is regarded as marking the beginning of the English influence in the subcontinent is 31 December 1600 when Queen Elizabeth I granted the British trading company known as the East India Company the monopoly of the Indian trade (Kachru 1978b, 479). Merchants of the East India Company arrived in the area in the early 17th century and founded their first trading station in Surat in 1612 (Crystal 2003, 47). During the British rule over the Indian subcontinent, which started with the East India company gaining political power in the area in 1765 (Kachru 1978b, 481) and ended in the independence in 1947, English became the language of administration and education (Crystal 2003, 47). The foothold of English was further reinforced by the approval and implementation of Macaulay's Minute, the British educational policy, in 1835. In 1857, three universities were established in Bombay, Calcutta, and Madras, and towards the end of the 19th century numerous English-medium schools were founded (Parasher 1998, 93). The educational policy together with the founding of universities and other educational institutions with English as the medium of instruction were central in consolidating the position

of English as the language of education and civilization, which led to English becoming viewed as the language of prestige and high status (Parasher 1998, 94).

After the end of the British rule and the declaration of independence in 1947, English was adopted as the official language of the state along with Hindi in India and along with Urdu in Pakistan. Having English as the official langue was supposed to be a temporary solution to facilitate the transition from the English administrational system into a system run in local languages in both countries. In India, the transition from English to Hindi was supposed to take place within the timeframe of 15 years. However, 15 years was gradually discovered to be too short a time for a complete replacement of English with Hindi. Consequently, the time period was first extended until the directive for the replacement was completely abolished (Parasher 1998, 95-96). Like in India, English as official language was first meant to be a temporary solution in Pakistan as well. It was meant to continue serving as an official language until it would be replaced by Urdu. However, this transition has not happened to this date in Pakistan either (Rahman 2006, 77).

2.2.2 English in India and Pakistan today

Today, English serves as the official language of the state together with Hindi in India and together with Urdu in Pakistan. In both India and Pakistan, English is used as a language of education, both as a taught subject and as a medium of instruction. India has a policy for language teaching in school education called the *three-language formula*. The policy entails that the first language to be taught to children is either their mother tongue or the regional language if the regional language is other than the child's mother tongue. The second language to be taught is either one of the state's official languages, Hindi or English. The third language can be either a modern Indian language or a foreign language – foreign language meaning a European language (Government of India 1966, 192). English also serves as the medium of instruction in English-medium schools (Meganathan 2011, 75). Pakistan does not have as clearly stated a language education policy as India does. The National Education Policy of Pakistan, however, decrees that English be made a compulsory subject from the first grade onwards (Government of Pakistan 2017b, 49). English is also the medium of instruction in English-medium schools, which are usually private schools or schools run by the Pakistani armed forces (Coleman 2012, 15). Most of the state schools employ Urdu as the medium of education (Coleman 2012, 15). In both countries, English is the primary language of higher education and dominates as the language of universities and academia (Parasher 1998, 107; Coleman 2012, 15).

English is used in the public administration and especially in higher-level administrational matters. In India, the proceedings are conducted in English in the Supreme Court and High Courts, while lower courts generally use either Hindi or a regional language in their proceedings. English is always used in authoritative documents such as bills, amendments, acts, court orders and other legislative texts. English is also the main working language of the Parliament (Parasher 1998, 102). In Pakistan, English is the primary language in civil administration and bureaucracy as well as in the higher levels of the legal system. Proceedings in the Supreme court and the provincial High courts take place in English. Court orders are always in either English or Urdu (Abbas 1993, 148).

In India, English is used as an intranational lingua franca both on the state level as well as on interpersonal level. The communication between the Indian government and non-Hindi states as well as the communication between Hindi states and non-Hindi states oftentimes take place in English. The language policy called the Official Languages Act decrees that in communication between Hindi states and non-Hindi states English translations have to be provided with Hindi texts (Government of India 1963, n.p.). In the multilingual India, English also functions as the link language between people from diverse linguistic backgrounds and people not sharing a mother tongue in interpersonal communication (Parasher 1998, 104). In Pakistan, English is mostly used as the lingua franca of the national elite (McArthur 2003, 327). Otherwise, the nationwide lingua franca that serves as the link between different ethnic groups is the national language Urdu (Rahman 2006, 74).

English is also used in the mass media in broadcasting television and radio programmes. English is especially used in nationwide reporting. The national television of India, Doordarshan, broadcasts in Hindi, English, and Urdu on its national channels (Government of India 2017, 37). It also broadcasts the nationwide news bulletins in Hindi, English, and Urdu – with some news content also produced in Sanskrit (Government of India 2017, 108-110) – while programmes on regional channels are in regional languages. In Pakistan, the Pakistan Broadcasting Corporation broadcasts national programmes primarily in Urdu and English (Government of Pakistan n.d.), whereas the regional languages are used in local reporting in their respective areas (U Din, n.d.). India and Pakistan also have their English press. The English-language press is the second largest press in both India and Pakistan. In India, it is only outnumbered by the Hindi-language press (Government of India 2018, 17) and in Pakistan, the number of English newspapers and periodicals is only exceeded by those published in Urdu (Government of Pakistan 2017c, 384).

2.2.3 English vs. local languages

In India and Pakistan, English can be observed to have assumed functions and roles in people's lives that are complementary to Hindi, Urdu, and other local languages. In a society with prolonged bi- or multilingualism, languages or language varieties sometimes specialize in serving specific, distinct functions in people's lives. As the linguistic situation in the Indian subcontinent is characterized by centuries of multilingualism, it is possibly not surprising to find functional compartmentalization of languages in the area. A situation where the use of two languages or language varieties is heavily compartmentalized is called *diglossia* (Ferguson 1959, 325-326). Characteristically, one code is consistently used in formal situations and official life, for example administration, education, media, and religion, while the other code is used in informal, everyday communication, for example among family members and friends (Ferguson 1959, 328-329). There is a direct, one-to-one relationship between a communicative context and a code choice. The languages are commonly associated with different degrees of prestige; the code of formal communication enjoys higher prestige and is regarded as the superior and more sophisticated *high variety*, while the code of informal everyday communication is regarded as the inferior *low variety* (Ferguson 1959, 329). The coexistence of languages and their functional specialization in a society can take more complex patterns involving more than just two languages. Sociolinguistic situations where three languages are specialized in serving distinct purposes in the society can be described as triglossia (cf. Mkilif 1972), while situations involving the distribution of functions between more than three languages can be described in terms of polyglossia (cf. Platt 1977).

Based on the studies by Malhotra (1980) and Parasher (1980), English can be observed to be in a diglossic relationship with Hindi and other mother tongues in India. Malhotra (1980) analysed patterns of language choice among upper-middle class Hindi-English bilinguals in different domains of life in urban settings in North India. She discovered that Hindi was used in domains involving greater degree of informality and intimacy, such as family and friends, whereas English was the preferred language in more formal situations, such as profession and education. The domains she distinguished were family, friends, religion, education, employment, professional, and business. In family and friend circles the communication usually took place in Hindi – with codeswitching to English. In the religious domain, Hindi was generally the preferred language except for prayers, which were read in Sanskrit. In the educational domain, the language of teaching was English both in English-medium schools and at the university level. In the domain of employment, English was often preferred as the language of interaction, but the language choice was also dependent on who people were talking

to; among peers, the language could be either English or Hindi, but when talking to superiors, the interaction took place in English. In seller-buyer interactions at a marketplace, the interaction between the shopkeeper and the customer took place in Hindi, whereas in businesses where customers were mainly upper-middle class the language was English (Malhotra 1980, 41-44).

The study by Parasher (1980) corroborates the findings of Malhotra's study showing how English was similarly used in more formal situations while the mother tongue was used in more informal communication. English was the dominant language in the domains of education, government, and employment, whereas mother tongue was used with the family. English also dominated in the domain of transactions. Parasher (1980, 157-158) however points out that the dominance of English in transactions is very much influenced by the fact that the subjects were members of the well-educated urban elite and that Hindi would definitely be the preferred lingua franca in this domain in more rural areas and among less educated people. In Parasher's (1980, 156-159) findings, English was also the dominant language in friendships – although Hindi and other regional languages were also used. This was due to the fact that many of the subjects did not share a mother tongue with their friends, and therefore English served as the lingua franca. In the domain of neighbourhood, English, Hindi, and regional languages were all used without any of them being particularly dominant. Both of these two studies support the view of English as the higher language in relation to Hindi and other regional language as it is used in more formal interactional domains, whereas the local languages are used in more informal communication. However, both of these studies were conducted on the more affluent proportion of the society; the study by Parasher was conducted on educated bilinguals and the study by Malhotra examined upper-middle-class families in urban settings. The scenario can look very different for less educated and uneducated people in rural settings.

Quite similar studies do not appear to have been conducted on the functional specialization of languages in Pakistan. However, it appears that similar functional specification of languages can also be observed in Pakistan. Ud-Din, Khan and Iqbal (2011) argue for the existence of a triglossic situation in Pakistan involving regional languages at one end, Urdu in the middle, and English at the other end. In Pakistan, regional languages like Punjabi are low varieties when their status is compared with Urdu and English. Urdu and English are thus higher varieties in relation to regional languages. However, in relation to English, Urdu ranks lower on the prestige scale, and therefore English has the highest status of all the languages. In this scenario, Urdu is both a low language and a high language depending on what its status is compared with – the regional languages or English. According to Ud-Din, Khan and Iqbal

(2011, 16-17), aspects that support this analysis are for example that local languages or vernaculars are learnt and used in home settings, while Urdu is the language used in education and as the national lingua franca in nationwide communication. Further aspects that overall support the superior status of both Urdu and English to regional languages is the fact that they are the official languages of the country. Urdu and English are both used in education whereas the use of regional languages for educational purposes is very marginal – almost non-existent (Rahman 2002, 4557). The superiority of English to Urdu, in turn, shows especially in English being the lingua franca of the national elite as well as the language of elite schooling (Rahman 1996, 13).

In neither India nor Pakistan is diglossia restricted to English vs. local languages only but it exists between local languages as well. As was already seen in the discussion about triglossia in Pakistan, there is functional specialization between local languages as well – in this case between Urdu and other local languages. Diglossia is not only possible between different languages, but it can also be observed between varieties of the same language. This is for example the case with Urdu. In Urdu, one can distinguish between two varieties: the formal persianized Urdu and the informal conversational variety of Urdu, the "layman's Urdu" (Ud-Din, Khan & Iqbal 2011, 14). The persianized variety is used, for example, in religious sermons, whereas the layman's Urdu is used in conversations with family and friends etc. In Khalapur, India, Hindi has assumed different roles from the two local dialects of Hindi, *saaf boli* and *moti boli*, and is used as the language of formal situations, whereas the local dialects are used in situations characterized by greater degree of informality (Gumperz 1964, 144-145).

It should be kept in mind, however, that due to the vast linguistic diversity and variation in the multilingual settings, variation in patterns of language choice across domains can be expected to exist in both India and Pakistan. The roles of languages can to vary both between different regions and between individuals depending on the individuals' own linguistic repertoire. Therefore, no absolute country-wide generalizations can be made concerning the roles of languages.

3 Codeswitching

3.1 Defining codeswitching

In general terms, and to put it simply, codeswitching is usually defined as alternate use of two or more languages. Switching does not only occur between different languages but also takes place between varieties of the same language. However, codeswitching has been defined in various ways, which is why it is not automatically clear what type of alternate use is meant by the term. The confusion around and the lack of agreement on the exact meaning of codeswitching is on the one hand caused by the several different definitions and on the other hand by the existence of several closely related terms that partly overlap and partly differ slightly from each other.

Milroy and Muysken define codeswitching as "the alternative use by bilinguals of two or more languages in the same conversation" (Milroy and Muysken 1995, 7). This definition highlights the fact that codeswitching has generally been studied more in spoken language and has been regarded as a spoken phenomenon. Heller defines codeswitching as "the use of more than one language in the course of a single communicative episode" (Heller 1988, 1). Montes-Alcalá, in her study on written codeswitching, uses the definition "the alternative use of two languages within the discourse" (Montes-Alcalá 2000, 193). These two definitions can be interpreted as encompassing both spoken and written codeswitching.

All these three definitions are relatively broad and allow a relatively wide range of different language alternation phenomena to be counted as codeswitching. Some researchers, in turn, have a more restricted view on what counts as codeswitching. When Auer (1999) differentiates between codeswitching, language mixing and fused lects, he reserves the term codeswitching for those instances of alternate use of languages where the switch has a conversational function and is locally meaningful in the interaction. Instances where the alternation between languages is frequent and the individual switches do not serve any particular discourse purposes are called language mixing. Fused lects are mixed language varieties that show structural stabilization.

Some researchers, for example Kachru (1978a) and Sridhar and Sridhar (1980), differentiate between codeswitching and codemixing. They reserve codeswitching for intersentential switching, i.e. switching of language between sentences, and use codemixing to refer to switch of language occurring intrasententially, i.e. within a sentence. In his classification of different patterns of intrasentential switching into three categories – insertional,

alternational, and congruent lexicalization –, Muysken (2000) uses the term codeswitching for alternational type of shifting between languages and employs the term codemixing for the two other types, insertion and congruent lexicalization. Jacobson (2001) uses the term language alternation to describe instances where the languages are in an equal relationship and contribute equally much to the discourse to distinguish them from codeswitching where there is a clear dominance relationship between the languages involved with one language being clearly the main language. According to him, there is characteristically neither a clearly dominant language nor a subordinate language in language alternation. A more recent term is translanguaging. According to García, translanguaging refers to not just shifting between languages but rather emphasises how bilinguals draw from their entire linguistic repertoire or resources and use multiple discursive practices to convey their bilingual reality (Garcia 2009, 45). The concept translanguaging assumes that "bilinguals have *one linguistic repertoire* from which they select features *strategically* to communicate effectively" (García 2012, 1). The different terminology that has been used to refer to and to distinguish between different language alternation phenomena somewhat reflects the different perspectives that different researchers take on the phenomenon and what aspects they emphasise: structural or discourse functional. Yet another question, and one that has been disputed for a long time, is the distinction between codeswitching and borrowing. This topic, however, is discussed in its own subsection below.

Despite the terminological ambiguity and somewhat contradicting terms, codeswitching is still the most commonly used one – whatever is meant by it. I employ codeswitching (henceforth CS) in a broader sense as a kind of umbrella term to refer to various types of alternate uses of languages in order to allow a variety of multilingual phenomena to be examined under the same general term.

3.2 Social and conversational aspects

The concept of diglossia was already discussed in 2.2.3. To briefly recap, diglossia refers to a situation where two languages or varieties of the same language are consistently used for different purposes and in different functions in a society, and where one code is usually used in formal situations while the other code is used in more informal situations. In a diglossic context, the language of formal situations is associated with higher prestige and is the *high variety* while the code of informal communication is viewed as the *low variety* (Ferguson 1959, 325-329). The concept of diglossia is taken up here again because the idea of such compartmentalization is relevant from the point of view of certain CS theories. Another similarly relevant concept is Fishman's sociolinguistic domains. Fishman argues that different sociolinguistic domains,

which he defines as "institutional contexts and their congruent behavioural co-occurrences" (1972, 441), should be analysed in terms of various variables in order to be able to fully understand the language choices people make in different situations. These variables are, according to him, for example the participants involved in the interaction, the social groups the participants belong to, role relations between the participants, the overall setting of the interaction, and the topic under discussion (Fishman 1972, 437-444). Although alternating between languages across different situations is generally not regarded as CS proper as CS is usually defined as switching languages within the same discourse, the concepts of diglossia and sociolinguistic domains, however, lay a useful foundation for discussion about the impact of the broader social circumstances on CS practices.

Gumperz (1982, 60-61) distinguishes between *situational* and *metaphorical* codeswitching. *Situational* codeswitching refers to instances where the code choice is conditioned by situational factors such as the setting, activity type, or participants of the interaction. In situational switching, the codes are used in separate domains and with different interlocutors; for example, the language used at home can be different from the language of school or work place, activities such as public speeches or different ceremonials can be regularly performed in a specific code, or different codes are used in communication with family members and friends than with authorities. Situational switching is basically related to diglossia. *Metaphorical codeswitching*, also called *conversational codeswitching*, refers to CS that occurs when a speaker wishes to convey particular conversational meaning and bring about certain conversational effect. According to Gumperz, the metaphorical switches derive their communicative effect from their associations and social value that the codes have received because of their habitual contexts of use, i.e. situational switching.

In her *Markedness model*, Myers-Scotton (1988, 151) argues that CS is a strategy of interlocutors to arrive at certain social consequences. The model is based on the idea of *markedness* versus *unmarkedness* of code choices in particular situations or interactions. The concept of *markedness* refers to the degree of predictability of a certain linguistic form in a particular situation and the community's expectations of the appropriate code choice in a particular communicative exchange. The expected code choice that conforms to the norms of the society concerning the correct code choice in a particular interaction is called the *unmarked* choice, whereas the unexpected code choice that violates the social norms is called the *marked* choice (Myers-Scotton 1988, 152).

While choosing the unmarked code in a conventionalized exchange signals the acceptance of the prevailing rights and obligations, the use of the marked code of the interaction

can be interpreted as an attempt to alter and renegotiate the rights and obligations of the interaction (Myers-Scotton 1988, 156). Conventionalized exchanges are, as Myers-Scotton (1988, 152) defines them, interactions that people in the community have a "script" for, i.e. mental representation of what can be expected in the particular type of interaction including, for example, the status or the relationship between participants and the appropriate language choice. The conventionalized exchanges are linked with particular sets of rights and obligations. Rights and obligations are "a theoretical construct for referring to what participants can expect in any given interaction type in their community" (Myers-Scotton 1998, 23).

The conventional use of codes in a society leads to the codes having certain social and psychological associations as well as social meaning linked to them. These associations are exploited in the form of CS to arrive at certain social consequences. Through CS, interlocutors can signal disagreement with the current rights and obligations and convey their own, differing perception of the rights and obligations. Renegotiating the rights and obligations can mean an attempt to either increase or decrease the social distance between the participants of the interaction (Myers-Scotton 1988, 167). In non-conventionalized exchanges, on the other hand, CS can function as an exploratory choice which allows the speakers to signal multiple identities and negotiate the norms for the current exchange (Myers-Scotton 1988, 155-157). Myers-Scotton (1988, 154-155) also adds that markedness should not be regarded as two strict categories but rather as a continuum; some code choices can be more marked – or unmarked – for a particular type of interaction than others, but the correct code choice is not absolute. It is also worth noting that CS does not always occur as negotiation of the participants' roles and position, and it is by no means always the marked choice. Sometimes it is actually CS that is the unmarked code choice. In these instances, the switches themselves do not necessarily serve any particular discourse functions, but it is rather the overall pattern of CS that is the discourse function and that has social significance (Myers-Scotton 1988, 161-162).

Gumperz's situational vs. metaphorical codeswitching and Myers-Scotton's Markedness model carry the assumption that codes draw their communicative functions from the associations they have in a particular community because of their habitual contexts of use. However, sociolinguistic approaches like these two theories have been criticized by Auer (1998, 20) for basing their explanations of CS too much on the social roles of the codes, relying on external factors to explain CS and seeing CS too much as a result of the functions that the codes serve in the broader social context. Auer (1995, 118) argues that for CS to be explainable in terms of the codes' social functions, stable and established multilingual communities would be required in order for the codes to have such established functions. He also notes that the

appropriate code choice is not, by any means, so clearly defined in all situations that the participants could rely on their mutual understanding on the "right" code for the interaction.

Gumperz (1982, 66) suggests that codes become viewed as *we*-code and *they*-code; we-code is generally the "ethnically specific, minority language" and the language of in-group communication and informal activities, whereas they-code, the majority language, becomes associated with formal, out-group interaction. However, this is not always the case. It could be assumed that Hindi would function as the we-code for Indian Hindi-speakers and English would be the they-code. However, according to Singh (1983, 71-73), affluent educated Hindi-English bilinguals in India tend to use English as their we-code and Hindi, the ethnically specific language, as their they-code. In their conversations, Hindi was observed to be often used with ironical intent. Therefore, Singh argues that we/they-code should not be understood merely in terms of socioeconomic position but also with regard to the role expectations and social aspirations of the participants involved in the interaction. However, as Singh notes, this finding corresponds to Gumperz's proposition that it is the language of the minority that becomes viewed as the we-code: here, English is the language of the affluent educated minority of India. There can also be situational variation in what counts as a we-code and they-code. Sebba and Wootton (1998, 162-164) argue that which code functions as the we-code or they-code depends on situational factors, and the same code can be assigned the role of either the we-code or the they-code in different situations. Thus, we- and they-codes do not necessarily remain static and stable across all situations. On further problems of labelling languages as we-code and they-code, Sebba and Wootton remark that the distinction between a we-code and a they-code does not even necessarily apply in all bilingual or multilingual communities (1998, 162). Even Gumperz himself reminds that "association between communicative style and group identity is a symbolic one: it does not directly predict actual usage" (Gumperz 1982, 66).

Auer (1984, 5-6) argues that CS is most preferably to be analysed in terms of sequential position of the switch in the conversation and the local function that it serves within the interaction. Therefore, he regards conversation analysis as the best analytical tool to account for the functions of and reasons behind code choices. Rather than relying on the potential social value attached to the codes to explain code choices, CS should be interpreted locally with regard to its position within the discourse. The analysis should show how the switches are locally meaningful in the on-going interaction. As another proponent of the conversation analytical perspective on CS, Wei (1998, 286) argues that it is unlikely that interlocutors consciously draw from the associations of codes to construct meaning or evaluate each other's language choices in light of the broader social functions of the codes. According to Auer (1984, 6), the

conversation analytical approach reduces the analyst-centeredness in the interpretation of the switches, i.e. the dependence of the interpretation on what meanings the analyst gives to the switch, enabling the analysis of the switches from the speakers' perspective and showing how the switches are meaningful to the speakers themselves. While acknowledging the potential impact of the social factors on CS as well, Auer (1998, 20) and Wei (1998, 290) criticize the emphasis that is put on the social explanations of CS in previous research. On the other hand, the conversation analytical approach can be criticized for relying too much on the situated analysis of the codeswitches as part of the conversation and for assuming the conversation to hold all required knowledge needed for the explanation of CS completely disregarding the broader social setting and contextual factors.

Linked to the conversation analytical approach is CS functioning as a discourse strategy to signal shifts in footing. Footing is a concept originally introduced by Goffman (1979). It means the alignment participants take towards each other as well as themselves in the course of an interaction. A change in footing "implies a change in the alignment we take up to ourselves and the others present as expressed in the way we manage the production or reception of utterance" (Goffman 1979, 5). For example, in Zentella's (1997, 93-95) analysis of Spanish-English CS, CS served as a strategy to signal change in footing and realignments such as shift in topic, quoting or reporting someone's speech, shifting from narrating a story to evaluating it, shifting roles in the interaction and in attempts to appeal to or to control the addressee. Another central concept in the conversation analytical approach is the notion of CS as a contextualization cue. A contextualization cue is, as Gumperz defines it, "any feature of linguistic form that contributes to the signalling of contextual presuppositions" (Gumperz 1982, 131). In other words, contextualization cues signal how the interlocutor intends what they say and how they wish it to be interpreted by others. Such cues can be auditory cues like intonation and rhythm, or visual cues such as posture or gestures (Auer 1995, 123).

Codeswitching can function similarly to these audiovisual cues in the speech of bilingual people. The study on monolingual versus bilingual conversational strategies by Gardner-Chloros, Charles, and Chesire (2000) demonstrated that the same conversational functions that are achieved through contextualization cues such as intonation or rhythm in monolingual speech are achieved through CS in bilingual discourse. According to Gardner-Chloros, Charles, and Chesire (2000, 1335), the conversational effect achieved by bilingual means is often even more powerful and more effective than what can be achieved monolingually. CS was observed to coincide with other contextualization cues, in which case these instances were doubly marked, which reinforced their effect further.

Based on the analysis of codeswitched data from three different language pairs (Hindi-English, Spanish-English, Slovenian-German), Gumperz (1982, 75-84) developed a preliminary typology of the functions CS was observed to frequently serve. The functions Gumperz identified in his data were *quotation*, *addressee specification*, *message qualification*, *interjection*, *reiteration*, and *personalization vs. objectivization*. *Quotation* is relatively self-explanatory and covers both direct quotation and reported speech. *Addressee specification* means switching code to direct the message to a particular audience or person. The category *interjection* consists of switching for interjections and sentence fillers. *Reiteration* means repeating what was previously said for the sake of either clarification or emphasis of the message. Reiteration may involve a direct translation of the message or expressing it in a slightly modified form. The category *message qualification* consists of such instances of CS where CS is used to somehow qualify or add information about the topic. These qualifications often take the form of "sentence and verb complements or predicates following a copula" (Gumperz 1982, 79). The last category, *personalization* vs. *objectivization*, is, as Gumperz himself writes, slightly harder to determine. He describes this category in the following way:

> The code contrast here seems to relate to such things as: the distinction between talk about action and talk as action, the degree of speaker involvement in, or distance from, a message, whether a statement reflects personal opinion or knowledge, whether it refers to specific instances or has the authority of generally known fact.
>
> (Gumperz 1982, 80)

Gumperz notes that the category personalization vs. objectivization encompasses "a large class of stylistic and semantic phenomena" (Gumperz 1982, 83). Personalization vs. objectivization are related to we-code and they-code; we-code is used to express more personal statements, such as personal opinion, and to signal greater degree of the speaker's involvement, while they-code is used to convey more objective information and to reflect the speaker's distance from what is said. Gumperz (1982, 83) argues that codeswitched passages can be interpreted as either personalized or objectivized exactly because of the choice of we-code or they-code.

Other researches have composed similar lists of CS functions both augmenting one another and leaving out certain functions depending on the findings of their own study. Researchers also sometimes name the same functional categories differently. Auer (1995, 120-121), on the other hand, criticizes such typologies as they often confuse actual discourse functions, conversational structures, and linguistic forms. For instance, as he exemplifies,

reiteration is actually a conversational structure which can serve functions such as putting emphasis on or further clarifying a statement, but it is not a discourse function as such. The typologies can also give an impression that CS functions would always be unambiguous and categorizable into strict categories although, in fact, switches can serve more than one function simultaneously.

The theories and typologies created to account for CS have primarily been based on findings in spoken material which is caused by the fact that CS has been studied traditionally more in spoken language than in written language. Accounts on written CS include CS in literary texts or novels, in a variety of different historical texts, as well as in different genres of online writing. As the material of the present study consists of online texts, the focus in the following discussion on CS in written genres will be on studies investigating CS online.

CS has been examined on Facebook for example by Halim and Maros (2014) who looked into English-Malay CS, by Frick (2010) who studied the functions of Finnish-Estonian CS and by Parveen and Aslam (2013) who investigated the functions of Urdu-English CS. Montes-Alcalá (2007), in turn, studied English-Spanish CS in blogs, while Siebenhaar (2006) focused on CS and the code choice between German and Swiss-German in Internet Relay Chat rooms. Studies on more formal online discourse include investigation of CS in e-mail discourse, for example English-Jamaican Creole CS by Hinrichs (2006) and Greek-English switching by Georgakopoulou (1997).

To exemplify the functions of CS in a study on written CS, Halim and Maros (2014, 129-133) analysed Malay-English CS in Facebook posts of five Malay-English bilinguals and identified switching for *quotation*, *addressee specification*, *reiteration*, *message qualification*, *clarification*, *emphasis*, *checking*, *indicating emotions*, *availability*, *principle of economy*, and *free switching*. The first four categories are also found on Gumperz's list and have thus already been discussed. *Clarification* and *emphasis*, as the terms suggest, refer to switching for further clarification and elaboration on a message, and to switching to emphasise a point, respectively. The category *checking* means switching to ask for opinion or seeking for approval or confirmation from the other interlocutors. Checking often takes the form of tag-questions, although other strategies are also used, for example yes/no-questions and wh-questions. The subjects also switched to *indicate emotions*, i.e. conveyed their feelings in another language. Switching for *availability* refers to switching motivated by the existence of a more accurate expression in the other language. Trying to explain or translate the expression would be too cumbersome, and therefore the word or expression from the other language is used. Switching for *principle of economy* occurs when bilinguals switch code due to the existence of an easier

and less complicated word or expression in the other language. The category *free switching*[1] comprises instances where no specific discourse function could be attributed to the switch.

It has been a somewhat contested issue whether CS in spoken interaction and in written discourse are manifestations of the same phenomenon or whether they should be regarded as two separate phenomena. This disagreement has to do with the differences between the spoken and the written mode, the former being characterized through spontaneity, whereas the latter is subject to more planning, more conscious linguistic choices, and editing processes. As written mode involves more planning and enables later modification, the linguistic choices can be thought of as more conscious and carefully thought unlike in spoken mode where the language use is more spontaneous. It has therefore been questioned – for example by Sebba (2012, 98-99) – whether the language alternation in written texts even counts as CS and whether the theories on CS developed on the basis of spoken CS are applicable to written CS.

As for the applicability of the conversation analysis on written data, the fact that written mode lacks central aspects of face-to-face interaction such as turn-taking system and visual cues (Androutsopoulos 2013, 670) might question the suitability of the conversation analytical approach. However, Georgakopoulou (1997, 158) argues that this absence of the contextualization cues, such as visual and auditory cues, that are usually present in spoken, face-to-face interaction actually "results in an increased reliance on code-centered contextualization cueing" (Georgakopoulou 1997, 158) in written interaction. This means that the lack of other channels in written interaction increases the importance of linguistic choices and contextualization that would be signalled by other means in spoken interaction is signalled through the linguistic means in written interaction.

There are also differences between written genres in how and to what extent they differ from spoken interaction. It has been noted in studies on online CS that online writing actually exhibits similar features to spoken interaction. The resemblances include for example greater degree of informality and less planning and editing of the content (Georgakopoulou 1997, 142). Schendl (1997, 76-82) compared codeswitches in historical, written data and studies on modern, spoken CS showing that the differences in the switches are rather quantitative than qualitative.

[1] *Free switching* has also been used in another sense in the codeswitching research. Nurmi, Tyrkkö, Petäjäniemi and Pahta (2018) referred to switches that are free in their form – not function – as free switches to distinguish them from switches that are more established in their form (pre-fabricated and conventionalised switches).

He further noted that the differences in the codeswitches seem to be more related to the degree of formality and the text type rather than the medium of communication.

As can be seen from the preceding discussion on different CS theories, some of the theories emphasise the broader macro-level factors in explanations of the emerging CS patterns and linguistic behaviour, whereas others emphasise the importance of micro-level factors in understanding the functions of and motivations for CS. However, CS does not function independently of either of these factors but is the result of the interplay between both macro- and micro-level aspects. Therefore, in order to attain a deeper and more comprehensive understanding of the CS practices, both the macro-level and micro-level factors have to be taken into consideration. Relying only on either the broader social aspects or conversation-internal aspects would fail to provide a holistic account of the CS practices ignoring the contribution of certain essential factors.

3.3 Syntactic aspects

The syntactic patterns of CS can be characterized in terms of intersentential or intrasentential CS. However, these concepts, intersentential and intrasentential, have received different definitions and have been understood differently by different researchers. Myers-Scotton (1993, 3-4) defines intersentential CS as switching between sentences or at a major sentence boundary, and intrasentential CS as switching within a sentence. Romaine (1995, 122-123), in turn, defines intersentential CS as switching between sentences or clauses, and intrasentential as switching within a sentence or a clause. According to Romaine's definition, switches occurring at a clause boundary within a sentence – for example between an independent and a dependent clause – are to be treated as intersentential. A third category in syntactic categorization of switches is a category that has variously been called tag-switches, extrasentential switches, or emblematic switches (Poplack 1980). These switches are elements that are syntactically rather independent and "freely distributable within discourse" (Poplack 1980, 603), such as tag phrases and interjections.

The motivation for distinguishing between inter- and intrasentential CS is that in intrasentential CS the grammars of the participating languages come into contact, which leads to intrasentential switching involving structural convergence between the languages. Consequently, intrasentential switching has a greater risk of ungrammaticality and violation of syntactic rules associated with it. In intersentential switching, the languages remain structurally separate. Defining intersentential switching as switching between sentences or clauses entails the idea that codes remain separate in the same way in switching between clauses as they do in

switching between complete sentences, and that the real structural contact occurs when the code is switched within a clause. This is what Myers-Scotton (2006, 239-241) argues later on when she states that a better unit to measure the structural complexity of CS is rather a clause, not a sentence, suggesting that it is the *intraclausal* switches – instead of intrasentential – that are informative of the structural constraints of CS.

The differing definitions and applications of the terms "intersentential" and "intrasentential" possibly stem partly from the fact that the terms themselves can be thought of as somewhat confusing. Sentence, by definition, is made up of at least one independent clause in addition to which it can contain other independent clauses or dependent clauses (Myers-Scotton 2006, 238). If "sentential" is taken literally to refer to a sentence, then intersentential would refer to switching between complete sentences only and intrasentential would encompass CS between clauses within a sentence. Another possible interpretation would be to apply "intersentential" to switches between coordinated independent clauses as independent clauses can, unlike dependent clauses, form sentences on their own. All other types of sentence-internal switches – including switches between independent and dependent clauses – would be treated as intrasentential. This latter interpretation is how for example Schendl (1997, 56-57) employs the terms. His motivation for categorizing both sentences and independent clauses as intersentential is that they are "syntactically rather independent units, so that their integration into the text does not present any particular difficulties for the bilingual speaker" (Schendl 1997, 56) whereas intrasentential switches "involve the matching of the syntactic rules of two or more languages" (Schendl 1997, 57), which is why they are structurally more complex.

I will use the term intersentential to refer to codeswitches occurring between complete sentences and coordinated independent clauses. Coordinated independent clauses can be regarded as similar to two distinct sentences in their syntactic independence, as Schendl pointed out, and they can form complete sentences on their own. Intrasentential is used to refer to all other sentence-internal switches. I will further employ the terms *interclausal* and *intraclausal* to differentiate between switches that occur between independent clauses and dependent clauses within sentences, and switches occurring within a clause, respectively. Interclausal and intraclausal are basically treated as subcategories of intrasentential switching.

3.3.1 Constraints on codeswitching

Since intrasentential CS involves structural integration of the languages, it has been of particular interest for researchers aiming at finding structural constraints that govern CS. Poplack (1980, 585-586) has suggested CS to be conditioned by the *free morpheme constraint* and the

equivalence constraint. The *free morpheme* constraint states that "[c]odes may be switched after any constituent in discourse provided that constituent is not a bound morpheme" (Poplack 1980, 585-586). According to the *equivalence constraint* CS occurs "at points in discourse where juxtaposition of L_1 and L_2 elements does not violate a syntactic rule of either language" (Poplack 1980, 586). For example, in Spanish-English CS switching between a noun and its modifying adjective would be prohibited because that would violate the permissible structure of both the languages. However, counterexamples to both of these proposed constraints can be found, for example, in Finnish-English codeswitched data (Halmari 1997, 75-77). According to Halmari, embedded English nouns and verbs can be observed to receive Finnish inflectional endings in Finnish-English CS. Instances such as *lunchboxiin* and *pretendattiin* where Finnish bound morphemes are attached to an English noun and an English verb go against the free morpheme constraint. As for the equivalence constraint, Halmari shows that there is CS to English in a Finnish postpositional phrase which is a type of switch that would be prohibited according to the equivalence constraint; whereas English has prepositions, the adpositions used in Finnish are postpositions. Because of this clash in the syntactic rules between Finnish and English, switching in an adpositional phrase should not occur. However, switches such as *lunchin alla* can be found in Finnish-English CS. Therefore, the proposed constraints should rather be seen as tendencies than restrictive rules.

According to the Matrix Language Frame model by Myers-Scotton (1993), one language in CS forms the structural frame into which elements from the other language are inserted. The more dominant language of the interaction that creates the morphosyntactic frame is called the matrix language (ML), while the language from which elements are inserted is called the embedded language (EL) (Myers-Scotton 1993, 3). Myers-Scotton distinguishes between ML+EL constituents, ML Islands, and EL Islands. ML+EL constituents contain both ML and EL morphemes, and their structure follows the grammatical rules of the ML (1993, 77-78). ML islands consist solely of ML constituents, and they conform to the grammatical structure of the ML. EL islands, in turn, consist entirely of EL constituents and follow the grammar of the EL. The MLF model is primarily concerned with the structure of the ML+EL constituents. Myers-Scotton argues that the morphosyntactic structure of constituents containing both ML and EL elements comes from the ML (the ML hypothesis) (Myers-Scotton 1993, 82). She proposes the Morpheme-Order Principle and the System Morpheme Principle to account for the structure of and constraints for constituents consisting of both ML and EL material (1993, 83). The Morpheme-Order Principle holds that the order of morphemes in constituents containing elements from both ML and EL (ML+EL constituents) follows the

permissible morpheme order of the ML. The System Morpheme Principle maintains that all the system morphemes in constituents containing elements from the ML and the EL come from the ML. In other words, the ML system morphemes set the structural frame into which content morphemes from the EL are embedded. The distinction that the MLF model makes between content morphemes and system morphemes is almost equivalent to – although not identical with – the distinction between lexical words and function words. Myers-Scotton further argues that if EL content morphemes lack congruence with the ML, their occurrence is inhibited in constituents with elements from both ML and EL (the Blocking Hypothesis) (1993, 120). Congruence here refers to parallel functional properties of the ML and EL morphemes or their compatibility in terms of pragmatic or discourse function (1993, 120-121).

The Matrix Language Frame model has received criticism, for example, for its somewhat unclear distinction between content and system morphemes. Muysken (1995, 183) notes that what counts as a system morpheme or as a content morpheme can be different in different languages. Partly in response to the criticism it has received, the Matrix Language Frame model has been amended and refined later on, for example in Myers-Scotton and Jake (2000).

Rather than attempting to formulate grammatical constraints on intrasentential CS, Muysken (2000) aims at describing intrasentential CS in terms of different patterns. He distinguishes between three patterns: insertional, alternational, and congruent lexicalization. In insertional CS, elements like single words and phrases are inserted from one language into the other. The other language is clearly more dominant and provides the grammatical frame in insertional switching (Muysken 2000, 60-64). Alternational CS involves switching between structures of the languages (Muysken 2000, 96). For example, switching between main and subordinate clauses is alternational switching. Alternational switching tends to occur at points where there is linear equivalence in the word order between the languages (Muysken 2000, 114). The third type, congruent lexicalization, occurs between structurally similar languages. In congruent lexicalization, both languages are equally involved in supplying the lexical items (Muysken 2000, 122). In this CS pattern, alternating between languages is so frequent that it is not possible to pinpoint which language is the more dominant one.

3.3.2 Matrix language

The concept of matrix language originates from Myers-Scotton's Matrix Language Frame model. As was already explained above, ML is the language that provides the morphosyntactic frame – is so to speak the underlying language – into which elements from EL are inserted. ML

as a concept has been criticized as it automatically assumes that CS always involves a clearly more dominant language. First of all, the main language or the dominant language does not necessarily remain the same throughout the interaction, but it can change in the course of the interaction. Muysken (2000, 68) notes that the concept of ML is most suitable for insertional CS where elements of one language are clearly embedded into the structure of the other and one language dominates. However, it becomes problematic in alternational type of CS, which involves switching between structures of languages. In CS involving switches for clauses and sentences, the matrix frame also switches with the change of language causing the matrix to fluctuate. CS can also be so frequent that it is difficult or even impossible to pinpoint the more dominant language. In her comment on the data she examined, Callahan (2004, 58) notes that in instances where codeswitches were not that frequent and were primarily single words, the matrix could be determined easily. However, when switching was more frequent and the switches were clauses and sentences, the distinction between ML and EL language failed. Which language is the more dominant one in the interaction overall is thus a different question from which language forms the morphosyntactic frame for a clause or a sentence at any given point. Therefore, for example Moyer (1998) distinguishes between *base* language, which sets the grammatical frame for the sentence, and *main* language, which is the dominant language of the interaction.

Despite being somewhat controversial and having problems associated with it, the concept of ML is relevant from the perspective of the present study in certain aspects. Determining the ML is particularly central in the structural categorization of Hind/Urdu codeswitches for the reason that not only are there codeswitches to Hindi/Urdu in English, but the Hindi/Urdu codeswitches – sentences, clauses and even phrases – also contain English elements embedded in them. Therefore, one needs to be able to distinguish what is actually embedded where in sequences containing both English and Hindi/Urdu elements in order to assign the Hindi/Urdu switches into structural categories.

Different criteria have been formulated to answer the question of how to distinguish the matrix – some of them structural, others psychological. The criterion that Myers-Scotton (1993, 68) initially suggested for determining the ML was morpheme-count; the language that provides most morphemes is the ML. However, it is an unresolved matter how many morphemes exactly are enough for one language to qualify as the ML, as Myers-Scotton herself consents (1993, 68). This criterion is also problematic because, as Muysken points out, language-typological factors influence the number of morphemes that languages supply (2000, 66). Another possibility would be to determine which language is the provider of system

morphemes in a codeswitched discourse; as ML is the language that provides the grammatical frame, it could be assumed that it is also the language of most system morphemes. However, again, as Muysken (1995, 183) points out, languages can differ from one another in what they treat as system morphemes vs. content morphemes.

According to Muysken (2000, 67), the main verb has sometimes been suggested to determine the matrix of a sentence. However, this criterion is problematic especially in the case of Hindi/Urdu because Hindi and Urdu adapt foreign verbs with the verb *karna* ('to do') which functions in these instances as a dummy auxiliary. Applying the main verb criterion to Hindi/Urdu would mean that the matrix of the sentence *hum apnay deen to apni achi examples kay thru portray kerain* ('We have to portray our religion through our good examples') would be English as it has the English verb "portray" as the main verb. Quite visibly, this interpretation is not conceivable.

Psycholinguistic and sociolinguistic criteria have also been suggested to determine the ML. According to the psycholinguistic criterion, the ML is the language that the speaker is most proficient in. The sociolinguistic criterion proposes unmarkedness of the code in an interaction as the decisive factor; ML is the code that is most unmarked for the particular type of interaction in a society (Myers-Scotton 1993, 67). However, how to measure a speaker's proficiency and how to determine the unmarked code of an interaction are not easy questions to answer – and if these criteria are hard to apply in spoken interaction, they are even more so in written data.

When a sentence or a clause contains elements from both English and Hindi/Urdu, the ML of that sentence or clause will be determined on the basis of two criteria: the word order and the predicate verb/verb phrase. There is a major difference in the word orders of English and Hindi/Urdu: while English has the word order SVO, Hindi and Urdu have SOV. Thus, the first criterion is that if the sentence or clause containing elements from both English and Hindi/Urdu follows the SVO word order, the matrix is most likely English; if the word order is SOV, the matrix is Hindi/Urdu. The second criterion is concerned with the placement of the predicate verb/verb phrase in the sentence as well as its structure. Not only do English and Hindi/Urdu differ in the placement of the predicate verb/verb phrase in a sentence but there is also an essential difference in the structure of the verb phrase. The order of verbs in a Hindi/Urdu verb phrase is main verb + auxiliary unlike in English where the order is auxiliary + main verb. If the predicate verb phrase of a clause or a sentence is constructed following the rules of Hindi/Urdu, it is regarded as an indicator – together with the word order criterion – that the matrix is Hindi/Urdu. While Hindi/Urdu verb phrases can contain English verbs embedded in them, English verbs need to be adapted with the help of the Hindi/Urdu verb *karna* which

functions as a dummy auxiliary in these instances receiving the Hindi/Urdu inflectional endings because the Hindi/Urdu inflections cannot be added directly on English verbs. If the predicate verb phrase is a Hindi/Urdu verb phrase, but the clause or sentence follows the word order SVO, it is regarded as an instance of embedding a Hindi/Urdu verb phrase into an English clause/sentence – unless it can be regarded as an instance of variation in the Hindi/Urdu word order. Therefore, the word order criterion and the predicate verb/verb phrase criterion are considered together in determining the matrix language of a sentence or a clause.

3.4 Codeswitching vs. borrowing

One of the major unresolved issues in the field of contact linguistics is how to reliably distinguish between codeswitches and borrowings. This question arises especially in the case of single lexical items of foreign origin. Borrowing means, as defined by Thomason and Kaufman (1988, 37) "the incorporation of foreign features into a group's native language by speakers of that language". Winford (2010, 171) calls the language that is the original source of the lexical item a source language and the language into which the foreign item is borrowed a recipient language.

Several different criteria have been suggested to distinguish codeswitches from borrowings, such as the use by monolinguals, replacement of the native item, morphological and phonological adaptation, and frequency. The criterion concerning the use of the word by monolinguals (Winford 2003, 107) holds that if the foreign item appears in the use of monolinguals, it can be considered a borrowing. However, one problem with this criterion is how to define monolingualism and bilingualism; how little knowledge of a foreign language is enough for one to already qualify as bilingual, or the other way round, how much knowledge of a foreign language is one allowed to have in order to still be regarded as monolingual. CS is, theoretically, possible with relatively limited language skills (Auer 1999, 312).

Another criterion suggested to distinguish between borrowings and codeswitches is the morphological and phonological integration of the item into the recipient language. According to this criterion, if the foreign item is morphologically and phonologically adapted into the recipient language, it is a case of borrowing. Myers-Scotton (1993, 177-181), however, argues against this criterion stating that, firstly, not all borrowings are phonologically integrated into the recipient language, and secondly, single-lexeme codeswitches are integrated into the morphosyntax of the matrix language.

Myers-Scotton (1993, 193-195) argues for the frequency counts as the best determinant of whether an item qualifies as a codeswitch or a borrowing; the more frequently a foreign item

occurs, the more likely it is a borrowing. Codeswitched items, conversely, occur less frequently. This argument derives from Myers-Scotton's distinction between two types of borrowed forms, cultural borrowed forms and core borrowed forms, which differ in the way they enter the recipient language (1993, 169-170). Cultural borrowed forms are lexical items denoting objects or concepts introduced to the recipient language via contact with other cultures. Cultural borrowings do not have equivalents in the recipient language, and thus they fill lexical gaps (Myers-Scotton 1993, 169). Core borrowed forms, in turn, are borrowed items that have valid equivalents in the recipient language. Myers-Scotton argues that these two types of borrowings differ essentially in how they become part of the recipient language. Cultural borrowings enter the recipient language, so to speak, abruptly and suddenly with relatively high frequency of occurrence right from the beginning, whereas core borrowings enter the language with relatively low initial frequency which gradually increases over time (Myers-Scotton 1993, 171-172). She argues that core borrowed forms start out as codeswitches in the recipient language and become established borrowings with increasing frequency. Therefore, Myers-Scotton suggests the absolute and relative frequency of the lexical item as the most reliable determinant of whether it is a borrowing or a codeswitch (1993, 176). However, Winford (2010, 182) criticizes frequency counts for their inconclusiveness, which is why frequency fails to account for the distinction in a satisfactory way.

To solve the problem of identifying singly occurring lexical items as either codeswitches or borrowings, Sankoff, Poplack, and Vanniarjan (1990, 74) introduced the additional category of *nonce borrowings*. Nonce borrowings are single content words or compounds that are morphologically and syntactically integrated into the recipient language like established loans but have lower frequencies of occurrence than established borrowings. However, according to Myers-Scotton, this addition of an extra category still does not explain the distinction between borrowings and codeswitches, and it actually blurs the boundary between the two "*as end products*" (Myers-Scotton 1992, 32).

Gysels (1992, 53-54) proposes discourse function as the criterion to distinguish between codeswitches and borrowing. She argues that if the foreign item contributes to the structuring of the discourse, it is a codeswitch; if it does not fulfil any clear discourse function, it is a borrowing. Although this may be a useful and valid criterion in some instances, discourse function as a criterion is questionable especially in situations where CS is, in Myers-Scotton's terms, the unmarked code choice. In contexts where frequent CS is the communicative norm, individual codeswitches typically do not serve any specific functions, but it is rather the overall pattern of switching that is socially meaningful. In her study on Spanish-English CS in Puerto

Rican community, Poplack also noted how individual switches could not always "be attributed to stylistic or discourse functions" (1988, 230). Therefore, discourse function also fails to account for the distinction between codeswitches and borrowings.

The replacement of the native synonym by the foreign item, as mentioned by Poplack and Sankoff (1984, 103), is yet another suggested criterion. If the foreign item has displaced the native synonymous item, it has become an integrated borrowing. Replacement can be measured in terms of availability, as suggested by Mackey (1970, 203-208). Availability means the range of items that speakers of a certain language have at their disposal in their language to refer to certain concepts. Availability can be used as an indicator of whether the foreign item has started to displace or has already displaced the native synonymous item. However, Poplack and Sankoff (1984, 104) remark that a foreign item becoming established in the recipient language does not necessarily involve the replacement of the native alternative, but they can also coexist. Two further criteria are acceptability and translatability, as suggested by Hasselmo (1969, 59-69). Acceptability refers to the acceptability of the foreign lexical item in the recipient language; if speakers of the recipient language consider the foreign item an acceptable word choice in the recipient language, it can be regarded as an integrated part of the recipient language. Translatability has to do with a person's ability to provide an equivalent of the foreign item in the recipient language. If they are not able to provide an equivalent, the item can be assumed to be part of the recipient language. However, Poplack and Sankoff (1984, 104) criticize the acceptability criterion for being affected by language attitudes and the status of languages, which influence the judgements of the item's acceptability. The problem with translatability, as also stated by Hasselmo (1969, 65) himself, is the question of what actually constitutes a translation.

What can be gathered from the counterarguments to all the suggested criteria is that each of these criteria – at least when applied alone – fails to reliably distinguish between codeswitches and borrowings in one way or the other. Thus, it seems to be a better idea to rely on more than one criterion to make the distinction. One should also take into consideration that different criteria can work in different situations. Callahan (2004, 11) even raises the question if a universal method for separating borrowings from codeswitches even exists. This statement leads one to think of the potential language- or language pair-specificity of the adequate criteria as well as the possible impact of the type of language contact situation. Firstly, what distinguishes borrowings from codeswitches can vary across languages and language pairs. Secondly, what counts as a borrowing or a codeswitch in a mainly monolingual setting can be different from what the distinction is in a multilingual context characterized by intense language

contact. It is further contestable whether a strict division between borrowings and codeswitches is even meaningful and desirable. As Winford notes, "the distinction between a switch and a borrowing is not transparent to bilinguals" (2010, 182).

Myers-Scotton (1993,163) argues that CS and borrowing should be seen as a continuum rather than as two separate processes. According to her, borrowing and CS are ultimately part of the same phenomenon and subject to the same underlying morphosyntactic procedures. Therefore, they should be viewed as "part of the same developmental continuum" (1993, 163). On this continuum, codeswitches become borrowings through increasing frequency (Myers-Scotton 1993, 174-176). The two ends of the continuum are codeswitches with low frequency of occurrence, for example switches that occur only once and never again, and foreign items with high frequency of occurrence, such as cultural borrowings. Obviously, and as Myers-Scotton (1993, 182-183) herself notes, this is not meant to suggest that all codeswitches end up becoming established borrowings. Neither does this suggest that all borrowings enter the language via CS. It only shows one possibility of how codeswitches become established borrowings.

3.4.1 Borrowing vs. codeswitching in non-native varieties of English

There are roughly speaking two types of Hindi and Urdu loanwords in English: those that have spread worldwide into other varieties of English and are well-established as part of the English vocabulary, and those that are specific to IE and PE. Sailaja (2009, 69-72) calls these two different groups of Hindi/Urdu borrowings assimilated items and restricted items, respectively. Assimilated items have become an integral part of the native English lexicon, and in many instances their foreign origin would no longer even be recognized. Examples of assimilated items include, for example, *karma* or *guru*. Restricted items are limited to the local variety of English forming part of the local element in that variety's lexicon. As they are specific to that particular variety, there is a great chance that restricted items may not even be understood by other English speakers than the users of that particular variety.

Non-native varieties of English can bring new dimensions to the borrowing vs. CS dilemma, one being the speakers' own perception of the status of their native language items in the local English or in English overall. Dako (2002) takes up this aspect of the speakers' view in her discussion on the status of lexical items from local languages in Ghanaian English. She argues that people tend to be "fully aware of the local 'foreign' element […] no matter how integrated the item may be in daily discourse" (Dako 2002, 48). This can be applied in the context of IE and PE as well; even if and when Hindi/Urdu words are recognized as borrowings

in English and included in an English dictionary, it is questionable how "English" – or rather "un-English" – these borrowed Hindi and Urdu words are from the perspective of Hindi and Urdu speakers. This can be the case especially with restricted items specific to the local variety. For speakers of Hindi and Urdu, the local words can be associated strongly with their own language and culture, and would not be perceived as English words by them at all. Because the words are strongly bound to the Hindi/Urdu language and Indian/Pakistani culture, they can function as cultural references and devices of ethnic signalling in the use of Hindi/Urdu speakers. Despite their potentially frequent appearance in the local English, they would not be recognized as English items. Therefore, treating such words in the data as borrowings and excluding them form the analysis would leave out crucial aspects of how local languages manifest in the local variety of English. Because the goal of this study is to examine how local languages are used and how they show in the local variety of English, their potential establishedness in the local variety is not regarded as a relevant issue. Actually, if a Hindi/Urdu word or expression is in such frequent use in the Indian or Pakistani variety of English that it can be seen as established in it, the more its use is worth noting as it is an indicator of what local elements are often adopted and how local languages commonly show in the non-native variety.

For this reason, a strict division between Hindi/Urdu codeswitches and Hindi/Urdu borrowings is regarded as neither meaningful nor desirable. The differentiation made between Hindi/Urdu codeswitches and borrowed items here is the following: If an English speaker who is not familiar with the Indian/Pakistani culture and has no knowledge of any Indian/Pakistani languages can be expected to understand the meaning of the word or expression, it is regarded as a borrowing. Otherwise the Hindi/Urdu item is treated as a codeswitch. Obviously, the line between what requires specific knowledge about Indian and Pakistani culture and what would rather count as common or general knowledge is sometimes obscure and subject to subjective interpretation. This criterion will potentially also lead to the inclusion of Hindi/Urdu words and expressions that are accepted in an English dictionary and qualify as borrowings in English.

3.5 Previous studies on codeswitching in Indian English and Pakistani English

While CS between the non-native variety of English and the local languages is recognized and often mentioned in different works on World Englishes (for example Kachru and Nelson 2006, 255-265) in general, it does not appear to have been studied extensively in IE and PE. Research on CS in IE and PE has remained relatively fragmentary lacking more comprehensive accounts

and systematic studies on both functional aspects and structural patterns of CS to local languages. The previous research on the manifestation of local languages in IE and PE has mainly been concerned with the impact of the local languages on the lexicon of the local English and on how the local languages give the local English the local flavour and characteristics that identify the variety of English as essentially the English of that specific region. These studies have mostly paid attention to single-word uses of Hindi/Urdu in IE and PE.

Lexical items from Indian and Pakistani languages in the local English often appear to be related to culture-specific domains, as can possibly be expected because of the differences in the sociocultural contexts of India and Pakistan and the contexts where English is spoken natively. In Sailaja's account of different Hindi lexical items in IE (2009, 73-75), many lexical items can be observed to be related to such sociocultural domains as religion, food items and clothing. These domains can be observed to be influenced by Urdu in PE as well. Especially the influence of Urdu in the domain of religion is relatively well-attested in PE (Rahman (1990), Mahmood and Shah (2011), Ahmad and Ali (2014)), and words related to religion and Islamic culture are often taken from Urdu in PE. Another notable category of Urdu words in PE is words denoting items and concepts specific to the Pakistani society and culture (Mahmood and Shah (2011), Ahmad and Ali (2014)). These Urdu items related to the socio-cultural domain are for example garments, food items, and kinship terms. According to Mahmood and Shah (2011, 73), religious and sociocultural vocabulary often serves the purpose of filling lexical gaps since they do not have valid English equivalents. Obviously, the use of Urdu words in PE is not limited to these two domains only, and Urdu lexical items of various kinds and with available English equivalents can be observed. Mahmood and Shah (2011, 82-85) argue that sometimes the Urdu expression carries special meaning and associations that its English equivalent lacks and therefore the Urdu word is preferred over the corresponding English expression. They exemplify this, for instance, with the Urdu word *paindo*; *Paindo*, meaning 'a villager', is derogatory in tone and is strongly associated with concepts such as "ignorant, uncultured, and uncivilized", unlike its English counterpart. The Urdu forms of address, such as *ji*, *sahib*, *yaar* and *bhai*, have been attested to be used in PE (Talaat and Anwar 2010, 99). *Ji* and *sahib* are more formal and denoting respect and honour, whereas *yaar* and *bhai* are informal and used between people close to each other.

According to Bhatt (2008), different ideological associations with English and Hindi – and other local languages – manifest and are exploited in CS practices. Because of the different statuses and uses of English and local languages, English has become associated with "modernity and progress" while the local languages are associated with "tradition and cultural

values" (Annamalai 2004, 160). Annamalai argues that as a consequence of the economical value of English in the Indian society, local languages have become more limited to and associated with domains such as home and religion as well as cultural practice and ethnic identity. In his study on the use of Hindi in IE newspapers, Bhatt (2008) argues that CS between English and Hindi is a means to switch between and present different identities and to juxtapose old, local and traditional ideologies (Hindi) with new, global, and modern ideologies (English). According to him, this practice signals a new socio-ideological consciousness creating a *third space* where the "competing cultural collectives" such as traditional-modern, indigenous-foreign, and local-global meet and are negotiated. The interpretation of these codeswitches rests on the shared bicultural knowledge and experience, and their interpretation requires bilingual as well as bicultural competence.

Neither are structural aspects of CS in IE and PE researched extensively. Anwar (2009) studied the use and types of Urdu phrases and clauses in PE. Her findings concerning the different types of Urdu phrases included noun phrases, genitive phrases, postpositional phrases, adjective phrases, and verb phrases. On a clause level, the data contained both Urdu main clauses and subordinate clauses. The Urdu main clauses were coordinated main clauses as well as main clauses occurring with an English subordinate clause. Urdu subordinate clauses occurred as subjects of an English main clause, complements of a verb, and objects of a prepositional phrase and as adverb clauses. The data in Anwar's study were analysed qualitatively, and there is no information on the proportions of the types of phrases and clauses. Similar studies do not appear to have been conducted on the structural aspect of the use of Hindi in IE.

CS from Hindi to English and from Urdu to English appears to have received more attention than the use of Hindi and Urdu in English. CS between Hindi and English has attracted researchers' attention for several decades. Some of the older studies include Verma's (1974) study on different features of Hindi-English CS, Kachru's (1978a) investigation of structural aspects of Hindi-English CS, and Kumar's (1986) account on formal and functional aspects of Hindi-English CS. More recently, the phenomenon has been investigated for example in Bollywood movies by Si (2010), in tv-programmes in British-Asian media by Gardner-Chloros and Charles (2007), and in Hindi-English advertisements by Patel (2016). Another more recent study is Klingler's (2017) investigation of patterns of Hindi-English CS among female Hindi speakers from different age groups in North India. As for Urdu-English CS, especially Rafi (2013) has studied the use of Urdu-English CS in online discourse. Parveen and Aslam (2013) looked into the functions of Urdu-English CS on Facebook. Sociolinguistic aspects of Urdu-

English CS were examined by Khan (2010) in his doctoral dissertation in which he looked into the CS practices of people by gender, age group, and occupation as well as into code choice across different situations in Hyderabad, Delhi, and Lucknow in India. Ehsan and Aziz (2014) examined Urdu-English CS in Urdu news of a private tv channel. Urdu-English CS in class-room interaction has been studied for example by Iqbal (2011).

In multilingual countries like India and Pakistan, CS, or any other language contact phenomena for that matter, are no unique scenario. Because of the multitude of languages, language contact with its various outcomes and manifestations is part of everyday life and reality in India and Pakistan. English is only one language more into the mix to influence and to be influenced by the local languages. CS and other language contact phenomena do not only take place between English and the local languages but also between local languages. Therefore, CS between English and the local languages should not be seen as separate from the broader multilingual setting and language contact taking place between the local languages. CS not only occurs between English and Hindi or Urdu but between English and other local languages as well, for example English-Telugu (cf. Jitta 2018), English-Bengali (cf. Chanda, Das and Mazumdar 2018), English-Tamil (Kanthimathi 2009) and English-Kannada (cf. Shivaprasad 2015) – just to name a few for the sake of illustration. CS does not necessarily involve only two languages, but there can be several languages contributing to the mix, for example Hindi-Kannada-English CS (cf. Bhattacharjee, Rahman, and Chengappa 2009), Tamil-Hindi-English CS (cf. Prasad 2011), or Urdu-Punjabi-English CS (cf. Alam, 2011).

4 Material and method

4.1 Corpora and corpus linguistics

A corpus is a set of texts of naturally occurring language (Biber, Conrad, and Reppen 1998, 4). Corpora enable the study of a variety of different linguistic phenomena in people's actual use, and they help in handling large amounts of language data (Biber, Conrad, and Reppen 1998, 3-8). Corpora can be general corpora consisting of a variety of different types of texts from different text genres, or they can be specialized corpora consisting of texts or text samples that represent a specific genre (Meyer 2002, 36). General corpora represent the language or language variety as a whole providing a general picture of the language, whereas specialized corpora enable the study of features in specific genres. Corpora can consist of written language, spoken language, or both (Meyer 2002, 34-35).

Corpora can be used to investigate various features of language as they occur in people's actual language use. These features can be both lexical and grammatical, and corpora enable the examination of their frequencies and distribution across different genres, registers, dialects, and time periods. The choice of a corpus and the corpus design for a study depend on the types of research questions one wishes to answer (Biber, Conrad, and Reppen 1998, 246-247). If one wishes to find out about the occurrence or the use of a specific feature in the entire language variety, the corpus should be representative of and cover the language variety in question as well as possible. General corpora are suitable for studies aiming at answering research questions concerning the language as a whole and researching the occurrence of a feature or a phenomenon in the entire variety as they include samples from various genres and registers. If one, in turn, is interested in examining features in a specific genre, the texts that the corpus is compiled of should be representative of that particular genre, in which case one would choose a specialized corpus.

In this study, a corpus was used for material collection. The corpus that was chosen was the Corpus of Global Web-based English (GloWbE), which is a corpus containing texts from the different varieties of English around the world – including Indian English and Pakistani English. In GloWbE, the Hindi/Urdu codeswitches were located with the most frequent Hindi and Urdu words, which were determined with the help of the Hindi and Urdu corpora *Hindi Web 2013 (hiTenTen13)* and *Urdu Web (UrduWaC)* in Sketch Engine.

4.1.1 Corpus of Global Web-based English (GloWbE)

The data for this study were collected from the Corpus of Global Web-Based English (GloWbE). Compiled in 2012-2013, GloWbE is a 1.9 billion-word corpus consisting of texts gleaned from the Internet from 20 different English-speaking countries all over the world, for example the United States, Great Britain, Australia, India, Bangladesh, Malaysia, Nigeria, and Jamaica to name a few. GloWbE was accessed via english-corpora.org, which is a website containing several English-language corpora created by Mark Davies.

The corpus can be divided into subcorpora according to country and text type. The text type categories distinguished by the corpus are "general" and "blogs". "General" consists primarily of news articles, user comments, and the like with occasional blog texts, whereas "blogs" only contains texts from blogs. Table 1 shows the numbers of web sites, web pages, and words as well as their distribution across the two text type categories. The compilers of GloWbE aimed at a 60/40 ratio of more informal blog texts and more formal texts belonging to different genres. Although 60% of the corpus are texts belonging to the category "general" and 40% are texts belonging to the category "blogs", "general" also contains blog texts, which is how the ratio of 60% informal texts and 40% formal texts is achieved (Davies and Fuchs 2015, 4).

Due to its remarkable size and the functionalities provided by the english-corpora.org interface, GloWbE enables the study of various aspects and features of the different varieties of English. The english-corpora.org interface provides a variety of functions for a variety of different research purposes. GloWbE is annotated for part of speech with the CLAWS 7 tagger. GloWbE can be used to examine variation in frequencies of different lexical items across varieties of English as well as variation in their use and meaning. GloWbE can also be used to seek answers to research questions concerning structural aspects such as variation in the use of different grammatical constructions and morphological variation in the varieties of English.

Table 1 Size of GloWbE

	GloWbE		
	Web sites	Web pages	Words
General	220,405	1,140,741	1,300,348,146
Blog	170,224	651,304	583,923,681
Total	340,619	1,792,045	**1,885,632,973**

In this study, GloWbE served as a source of material. It was used to locate and collect Hindi/Urdu codeswitches that constituted the material of the study. GloWbE was chosen because, as a corpus consisting of texts from different varieties of English, it has both IE and

PE represented. Secondly, GloWbE is large in size containing a variety of different text genres, which is why it enables the examination of CS across different genres providing a more representative and comprehensive picture of the uses of CS.

For this study, I used the subcorpora of Indian English and Pakistani English. The IE subcorpus is 96,430,888 words in size. PE subcorpus is almost half the size smaller with 51,367,152 words. Table 2 shows the corpus sizes together with the number of web sites and web pages, and their share across the two text type categories. As can be seen, a bigger share of the texts in the corpora belong to the category "general", whereas the category "blogs" is smaller in size.

Table 2 Sizes of the Indian English and Pakistani English subcorpora

	Indian English			**Pakistani English**		
	Web sites	Web pages	Words	Websites	Web pages	Words
General	11,217	76,609	68,032,551	3,070	25,852	38,005,985
Blog	9,289	37,156	28,310,511	2,899	16,917	13,332,245
Total	18,618	113,765	**96,430,888**	4,955	42,769	**51,367,152**

4.1.2 Sketch Engine

The corpora that were used to determine the most frequent Hindi and Urdu words are part of Sketch Engine. Sketch Engine is a corpus tool containing 500 corpora and covering over 90 languages. Sketch Engine was developed by Lexical Computing CZ s.r.o in 2003. Sketch Engine has various different features and functions enabling one to explore various aspects of language. The most relevant function for this study was the word list tool, which was used to generate word lists of the most frequent Hindi and Urdu words. The word list tool generates a word list of the most frequent words based on all the words in the corpus. In the list, the items are organized by frequency in a descending order starting with the most frequent item.

The Hindi corpus that was used to determine the most frequent Hindi words was *Hindi Web 2013 (hiTenTen13)*. *Hindi Web 2013* is 350 million words in size, and it was compiled in 2013. The Urdu corpus was *Urdu Web (UrduWaC)*, which consists of more than 53 million words and was compiled in 2015. Both the Hindi corpus and the Urdu corpus are compiled of texts gleaned from the web.

4.2 The data and data collection

The amount of material collected for this study was altogether 6362 instances of Hindi/Urdu CS. Out of these 6362 instances 4183 were from IE and 2179 from PE. The analysed data itself consisted of 5449 Hindi/Urdu codeswitches – 3650 from IE and 1799 from PE – after the

elimination of duplicates and other Hindi/Urdu instances that were not accepted in the analysis (see subsection 4.2.3.). The amount of data was considered sufficient for a quantitative analysis and enough for qualitative analysis to provide a representative picture.

This study was also a methodological experiment using the most frequent Hindi and Urdu words to locate Hindi/Urdu codeswitched instances in GloWbE. The frequencies of Hindi and Urdu words were determined by generating word lists of the most frequent Hindi and Urdu words with the help of Hindi and Urdu corpora in Sketch Engine. The reason for using the most frequent Hindi/Urdu words to search for Hindi/Urdu codeswitches in GloWbE was that it was assumed that Hindi/Urdu words that have a high frequency of occurrence in Hindi and Urdu would also likely occur in Hindi/Urdu codeswitches and thus the high-frequency Hindi/Urdu items would be an effective way of locating Hindi/Urdu codeswitches. In the following subsections, the method of locating the Hindi/Urdu codeswitches in GloWbE will be explained in more detail.

4.2.1 The Hindi/Urdu frequency lists

To locate Hindi/Urdu codeswitches in GloWbE, I conducted the searches with the three most frequent Hindi and Urdu words: *ke*, *mein* and *ki* in Hindi, and *ke*, *ki* and *mein* in Urdu. All of these words are postpositions. As can be seen, the three most frequent words were the same in both Hindi and Urdu – only in a slightly different order. Extracts from the Hindi and Urdu frequency lists are provided in Appendix 3 and Appendix 4 with the 10 most frequent words of each list.

The first three items of the frequency lists were used because the amount of data arrived at with the three most frequent Hindi words and the three most frequent Urdu words was considered sufficient for the current study. Thus, there was no need to go further down on the frequency lists. The searches in the IE corpus were conducted with the Hindi words and the searches in the PE corpus were conducted with the Urdu words.

4.2.1.1 Romanization of Hindi and Urdu

In order to make the Hindi/Urdu words searchable in the GloWbE, the words had to be romanized, i.e. converted from the Devanagari and the Perso-Arabic script to the Roman script. However, the lack of any widely accepted standard for the romanization of either Hindi or Urdu and the great variation in the way the Hindi/Urdu users romanize Hindi/Urdu posed a challenge for searching and locating the Hindi/Urdu words in GloWbE. Although several transliteration and romanization systems have been created for Hindi/Urdu romanization, for example IAST (International Alphabet for Sanskrit Transliteration) for transliteration of Hindi or the ISO

15191:2001 standard created by the International Organisation for Standardization for the standard transliteration of Devanagari, Arabic and other related Indic scripts, there is no widely accepted, official romanization standard for either Hindi or Urdu.

Therefore, I devised a set of romanization principles based on the ALA-LC romanization tables for Hindi and Urdu, on the sounds that the Hindi/Urdu letters represent, and on my own observations on the romanization of certain letters. The ALA-LC romanization tables have been developed for the romanization of non-Roman scripts for bibliographical information and are approved by the Library of Congress and the American Library Association. The ALA-LC tables were chosen because of their availability for both languages and their accessibility. However, as the romanization system in question is created for romanization of bibliographical information, it does not necessarily fully correspond to how Hindi/Urdu users actually write the languages in the Roman script. Strictly following a system that does not take into account the variation in the spelling practices could be misleading and even distort the results. Therefore, the tables had to be adapted with slight modification and a few additions to better meet the needs of this study. The changes and additions were made based on the sounds that the letters represent and my own observations on the romanizations of certain letters and sounds. The romanization of Hindi and Urdu letters are presented in tables in Appendix 1 and Appendix 2, respectively.

The greatest variation in the Roman spellings of Hindi/Urdu words can be observed in the representation of vowel sounds. Therefore, vowels were replaced with wildcards in order to locate a wider variety of different spelling variants of the words. Wildcards are symbols that are used to replace characters in the corpus search to locate different forms of the word. The wildcard "?", which equals any one character, was used to replace the vowels in the Hindi/Urdu words as it enables locating various spelling variants of the same Hindi/Urdu word while keeping the search method as effective and systematic as possible. Vowels were substituted with the wildcard in the following way: both short and long vowels were replaced with one to two question marks and diphthongs with two to three question marks. For example, the first word on both frequency lists, Hindi के and Urdu کے /ke:/, is searched as *k?* and *k??*, i.e. two different searches were conducted. When the vowel in the word is a nasal vowel, the searches are formulated in the following way: if the nasal vowel is in the middle of the word, the vowel is replaced with one to two question marks and the letter *n* is added after them; if the nasal vowel is at the end of the word, it is 1) replaced with one to two question marks, 2) replaced with one to two question marks with the addition of the letter *n* after the question marks.

One of the modifications made to the ALA-LC romanization tables was the omission of diacritics added to the Roman letters. The modifications also involved a few omissions and replacements of certain Hindi and Urdu characters as well as a few additions of possible romanizations of certain letters. The changes and modifications are explained in more depth in Appendix 1 and Appendix 2.

4.2.2 Searches in the GloWbE

The searches in GloWbE were formulated by first mapping the consonants in the Hindi/Urdu words to their Roman equivalents and replacing the vowels with the wildcard "?" as explained in 4.2.1.1. The search forms of the three Hindi frequency list words and the three Urdu frequency list words are given in Table 3. When the search is conducted, GloWbE generates a list of words with all word forms that match the search (Figure 1). By clicking on a word form, the corpus shows a keyword in context (KWIC) display, which in turn shows the word used in context (Figure 2). In addition to the context provided in the KWIC lines, it is possible to view expanded context by clicking on the KWIC lines. The expanded context provided by GloWbE is approximately one paragraph.

Table 3 Search forms

Hindi		Urdu	
के	k?, k??	کے	k?, k??
में	m?, m??, m?n, m??n	کی	k?, k??
की	k?, k??	میں	m?, m??, m?n, m??n

All the word forms in the list that could potentially be spelling variants of the searched Hindi/Urdu word were checked. If the word appeared to be a spelling variant of the Hindi/Urdu word, and it did not yield more than 30 false hits per first 100 hits, it was considered an accepted spelling variant of the Hindi/Urdu word. If the number of false hits was higher than 30 per first 100, the spelling variant was regarded as a distraction and excluded. If the number of hits was less than one hundred, the acceptable number of false hits was 30% of the hits. The search forms and the word forms that were accepted spelling variants of the searched Hindi/Urdu frequency list words are given in a table in Appendix 5 together with the number of tokens. A false hit means here that the word was neither a Hindi/Urdu codeswitch nor part of a Hindi/Urdu

codeswitch. These false hits could be for example abbreviations, first names and last names, names of geographical areas, and switches to other languages.

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...] iWeb

		CONTEXT ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 14,292 | UNIQUE 40 +
1		KM	3470	
2		KI	1701	
3		K.	1591	
4		KG	1468	
5		KA	1390	
6		KE	1005	
7		KO	659	
8		KP	429	
9		KB	401	
10		KR	394	
11		KW	276	
12		KK	198	
13		KS	191	
14		KD	183	
15		KV	176	
16		KC	159	
17		KL	113	
18		KU	89	
19		KT	80	
20		KJ	65	
21		KN	45	
22		K1	39	
23		K2	35	

Figure 1 Word list in GloWbE

SECTION: India **(1,701)**
FIND SAMPLE: 100 200 500 1000
PAGE: << < 1 / 17 > >>

CLICK FOR MORE CONTEXT [?] SAVE LIST CHOOSE LIST ---------- CREATE NEW LIST [?] SHOW DUPLICATES

1	IN	G	india-forums.com	A	B	C	for you.. thank you.. for that one scene today.. # yeh dil ki baat apni dil main dabaake rakhna # That made me cry.. and that reminded
2	IN	G	india-forums.com	A	B	C	dance when her actions are a little seducing and Arnav's expression then... Dil ki baat scene... how he hugged her haq se in front of everyone... that
3	IN	G	...ibabaexperiences.org	A	B	C	if the said train went or not? The kooli replied me -- " Aane ki tayari main ha " (The train is about to come). It was
4	IN	G	delhiassembly.nic.in	A	B	C	proper attention to the grievances of the citizens, a Grievance Management System " Aap Ki Sunwai " has been set up and grievances of the public are being redr
5	IN	G	punjabnewsline.com	A	B	C	manner. # MUMBAI:Actor Imran Khan's rustic look in Vishal Bhardwaj's " Matru Ki Bijlee Ka Mandola " and Arshad Warsi's get up in upcoming film " Zila
6	IN	G	punjabnewsline.com	A	B	C	sports a beard, which is very similar to what Imran sports in " Matru ki Bijlee... ". " When I saw the first look of' Matru Ki
7	IN	G	punjabnewsline.com	A	B	C	ki Bijlee... ". " When I saw the first look of' Matru Ki Bijlee Ka Mandola',
8	IN	G	bolegaindia.com	A	B	C	A- list actress like Katrina Kaif via her chartbusting songs like Chikni Chameli and Sheila Ki Jawani. This is the reason why most of the B-town actresses want to sh
9	IN	G	bolegaindia.com	A	B	C	, the like of Vidya Balan is gearing up for an item number in Ferrari Ki Sawari. Now when someone akin to Vidya, who seems to have born with
10	IN	G	bolegaindia.com	A	B	C	A- list actress like Katrina Kaif via her chartbusting songs like Chikni Chameli and Sheila Ki Jawani. This is the reason why most of the B-town actresses want to sh
11	IN	G	bolegaindia.com	A	B	C	, the like of Vidya Balan is gearing up for an item number in Ferrari Ki Sawari. Now when someone akin to Vidya, who seems to have born with
12	IN	G	zeenews.india.com	A	B	C	Aam Aadmi. Yaar mera matlab Mango Man nahi jo iss par baithta hai visitor ki kursi per. Bechara! Nevertheless You did a good job madam.
13	IN	G	...tainment.oneindia.in	A	B	C	played great hosts to their guests. Mr Bachchan had even tweeted, " Deepavali ki anek anek shubhkamanayein... Sab ko... prosperity, success good health, and m
14	IN	G	...ami-krishnananda.org	A	B	C	about without touching the ground. The point here is is blessings. Sankara bhagavan ki jai! Here we have the central issue, practically, of the fourth skandha
15	IN	G	bollywoodlife.com	A	B	C	At least the director of the programme should have chosen singer anchor instead of mom ki gudiya, who should have complete knowledge of sur. This is not a fa
16	IN	G	bollywoodlife.com	A	B	C	. All are legendary and outstanding like Abidaji, Runaji and above all Indian sur ki devi Asha ji. # Ayesha Takia is the perfect host for the show --
17	IN	G	authorstream.com	A	B	C	Udaipur: - Lake Pichola, City Palace, Jagdish Temple, Museum, Saheliyon ki Bari, etc are the places that you would find quite amusing and interesting as
18	IN	G	thehindu.com	A	B	C	singer from Bangalore for an evening that couldn't be called anything else but Yaadon Ki Baarat. # Those of us who did not have loved ones sitting next to
19	IN	G	thehindu.com	A	B	C	. The bashful bridegroom and his sporting bride danced as Ganesh serenaded them with Bhanwre Ki Gunjan. (Kareena Kapoor's parents, Randhir Kapoor and Ba
20	IN	G	thehindu.com	A	B	C	Dil Lagta Nahin from " Baawre Nain " and Yeh Duniya yeh Mehfil Mere Kaam ki Nahin from " Heer Ranjha ", heartbreak seemed worth it too. Requests came
21	IN	G	hindustantimes.com	A	B	C	, except for two films. I thought Vijaypath (1994) and Hogi Pyar Ki Jeet (1999) wouldn't do well, but I was wrong. Other

Figure 2 KWIC lines in GloWbE

Some Hindi and Urdu words that have different spellings in the Devanagari or Perso-Arabic script can have the same romanized form. For example, the Urdu genitive postposition کے, which is the first word on the Urdu frequency list, and the conjunction کہ can both be romanized as *ke*. Therefore, *ke* appears in both meanings in the search results. Since the purpose of the Hindi/Urdu frequency lists was to help locate the Hindi/Urdu codeswitches in GloWbE

rather than to locate particular words with a particular meaning, the meaning was ignored as long as the element was an instance of Hindi/Urdu CS.

The spelling variants of the Hindi and Urdu frequency list words sometimes returned different numbers of false hits in the IE corpus and in the PE corpus. Consequently, a spelling variant that was regarded as a valid spelling variant of a Hindi or Urdu word in one of the corpora was not necessarily accepted in the other. Therefore, the frequency list words could have different numbers of accepted spelling variants in IE and PE. For example, the first item on the frequency lists (Hindi के, Urdu کے) had more accepted spelling variants in the PE corpus (*ke*, *kay*, *kee*, *keh*), while in Indian English the only accepted spelling variant was *ke*, as can be seen in Appendix 5.

4.2.3 Construction of Excel-files

When a word form was confirmed to be a valid spelling variant of a Hindi/Urdu frequency list word, the Hindi/Urdu codeswitches were collected into Excel-files together with their wider textual context provided by GloWbE in the expanded context-view to enable the structural and functional analysis of the switches. In addition to the Hindi/Urdu switch containing the spelling variant of a frequency list word, the expanded contexts often contained other switches to Hindi/Urdu as well. In order to allow a wider variety of switches to be represented in the data, some of these so-called additional switches were also included. The principle of inclusion was that switches that were part of the same sentence as the switch located with the search term were included in the analysis. Information such as which frequency list word and which search form the switch was located with was marked in the Excel-files. Additional switches were marked as additional material. The Excel-files also contain the source of the text, i.e. where GloWbE collected the text from.

In Excel, the data were subject to elimination of such Hindi/Urdu codeswitched instances that did not meet the criteria to be included in the analysis. Instances in which it was not possible to infer what context the Hindi/Urdu codeswitch was part of were left out of the analysis. These might have been for example links to articles or websites at the end of news articles or in the side bars of web pages collected into GloWbE. Codeswitches where part of the codeswitch was missing were also left out. The missing parts were most often caused by restrictions of the corpus; the corpus did not recognize the characters used in the text and replaced them with its own codes. A better illustration of the types of instances that were included in and excluded from the analysis is provided in Appendix 7 with explanations and examples.

The largest group of eliminated switches were duplicates. The data had two sources of duplicates: the corpus itself and the method. Corpus duplicates were caused by the corpus containing the same text or an extract of the same text multiple times. Duplicate texts or duplicate extracts of the same text result, for example, from the same text appearing on several different web sites, from users copying or citing each other's comments on the web pages, or from the same comment being posted twice on the web page. Some of the duplicate texts in the corpus are detected and eliminated by the corpus itself in which case the texts with several copies are marked with a number in parenthesis indicating how many copies of that text the corpus contains. For instance, the marking (1) indicates that the text in question has one duplicate, and (2) indicates that the text has three copies in the corpus. The duplicate texts detected by the corpus are eliminated from the KWIC view, i.e. each text shows only once. However, not all duplicate texts can be detected by the corpus – not least because of its large size – and some duplicate samples inevitably remain in the corpus. These remaining duplicates had to be eliminated manually from the data.

Another source of duplicates was the method of data collection. Method duplicates were caused by the searches with the Hindi/Urdu frequency list items and the criteria for the inclusion of additional switches. There were two ways in which the method led to duplicate switches: 1) a codeswitch contained more than one of the three frequency list items, and 2) more than one codeswitch in a sentence contained one of the three frequency list items. In the first scenario, the same codeswitch appeared more than once in the data when it 1) contained more than one accepted spelling variant of the same frequency list item (example (1)), 2) when it contained the same spelling variant multiple times (example (2)), or 3) when it contained accepted spelling variants of different frequency list items (example (3)).

(1) Imran the Rat **ke** courage yeah hay key American **kay** sath jaa raha hay.

(2) Afterall yeh pakistani family drama hai joh ek family **ke** saath beth **ke** dhekka ja sakta hai.

(3) doodh dene wali gaay **ki** laat kha kha **ke** besharm ho gaye hain

In the second scenario, duplicate switches resulted from the inclusion of additional switches, i.e. collecting all the codeswitches in a sentence if the sentence contained other codeswitches in addition to the one containing the frequency list item. If a sentence contained several codeswitches and more than one of those codeswitches contained an accepted spelling variant of a frequency list item, that led to duplicate occurrences of those codeswitches

containing a frequency list item that was searched later because they had already become collected as additional switches. This is illustrated with the following sentence:

(4) if a **kutta** calls others **randi ke bachay** then be ready to hear **randi kay bachay** in return from them.

This sentence contains three switches to Hindi/Urdu: *kutta*, *randi ke bachay*, and *randi kay bachay*. *Ke* and *kay* are two different spelling variants of the same frequency list item (Urdu کے). The two different searches formulated were *k?* and *k??* of which the search *k?* was conducted prior to *k??*. This means that the spelling variant *ke* was located before *kay* and the switches containing the spelling variant *ke* were collected – together with possible additional switches – before switches with *kay*. This in turn means that *randi kay bachay* was collected as an additional switch of *randi ke bachay* before the search *k??* and the spelling variant *kay*. *Randi ke bachay* has two switches as additional material: *randi kay bachay* and *kutta*. Therefore, when *randi kay bachay* was located with the search *k??* and collected in the data file, it was a duplicate occurrence of a switch because it had already been collected as additional material earlier. Consequently, it was marked as a method duplicate in the data file. Basically, *randi ke bachay* and *kutta* would also be additional material of *randi kay bachay*. However, in these instances the additional switches of a duplicate switch were not collected again.

5 Quantitative analysis: structural patterns

The analysed data consisted of 3650 instances of CS in IE and 1799 instances of CS in PE. As a general overall observation on the relative numbers of codeswitches in IE and PE, it can be noticed that when these raw numbers are adapted and viewed in proportion to the respective corpus sizes, i.e. when the numbers are normalized, it appears that the number of codeswitches is proportionately almost the same: 39 codeswitches per one million words in IE and 35 codeswitches per one million words in PE. These numbers do not by any means illustrate the overall frequency of CS in IE and PE, but they indicate that the commonness of the practice and the tendency to codeswitch are approximately the same. The conclusion that can be drawn is that Hindi/Urdu CS as a phenomenon appears to be equally common in both IE and PE.

Table 4 Normalized frequencies of Hindi/Urdu switches

	Corpus size	Codeswitches in data	Normalized (per one million words)
Indian English	96,430,888	3650	39
Pakistani English	51,367,152	1799	35

IE and PE displayed a wide range of different types of codeswitches to Hindi/Urdu varying from insertions of single Hindi/Urdu lexical items to longer stretches of Hindi/Urdu. The analysis of the forms of the Hindi/Urdu codeswitches consisted of three layers enabling a more profound examination of the structures. The structural analysis involved first categorizing the Hindi/Urdu switches into intersentential and intrasentential switches. Intrasentential switches were further subdivided into interclausal and intraclausal switches, and intraclausal switches were subcategorized according to word class and phrase type.

Figure 3 shows the share of intersentential and intrasentential codeswitches in IE and PE in percentages. These percentages are counted in relation to the total number of analysed switches in each variety – 3650 in IE and PE 1799. In IE, intersentential codeswitches constituted 14.55% (N = 531) and intrasentential codeswitches 85.45 % (N = 3119) of the entire number of codeswitches in the IE data. In PE, the percentages were 22.46 % (N = 404) for intersentential and 77.54% (N = 1395) for intrasentential codeswitches. It appears that intersentential switches were somewhat more frequent in Pakistani data and reversely, intrasentential switches occurred slightly more often in Indian data. These categories and the types of switches they contain are discussed in detail in the following subsections.

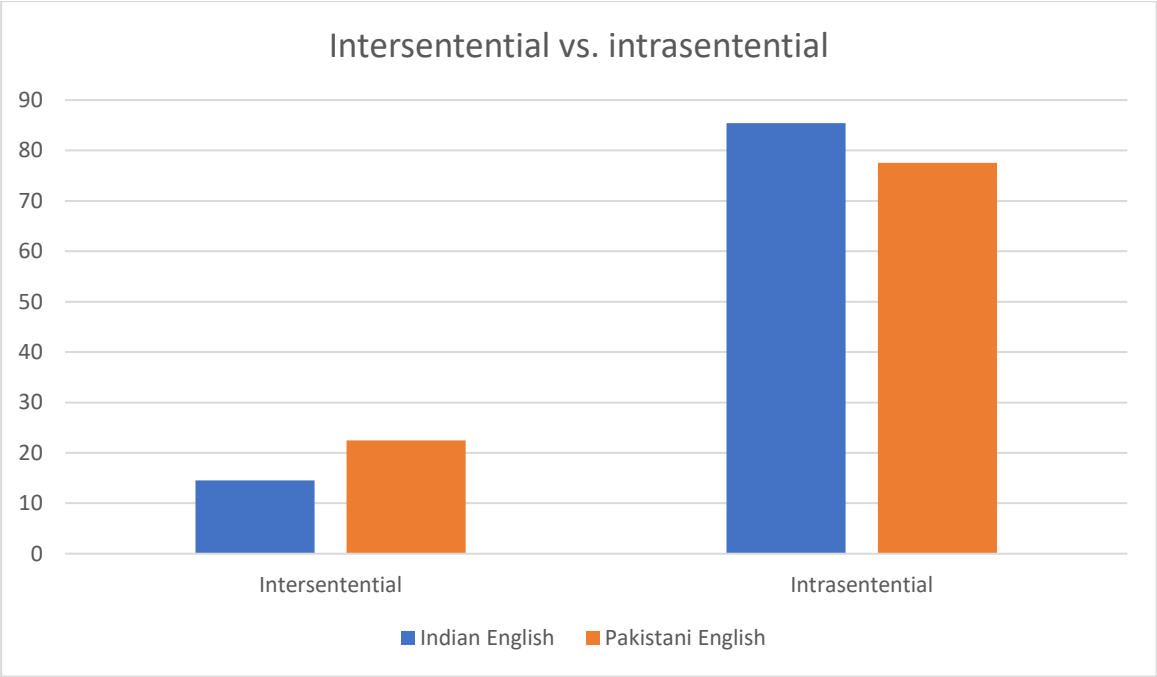

Figure 3 Intersentential and intrasentential codeswitches

5.1 Intersentential switches

Codeswitches between complete sentences as well as switching between coordinated independent clauses were analysed here as intersentential codeswitches. As shown in Figure 3, intersentential switches constituted 14.55% (N = 531) of the data in IE and 22.46 % (N = 404) of the data in PE. On the sentence level, code was switched for complete Hindi/Urdu sentences – either for one single sentence or longer sequences involving several consecutive Hindi/Urdu sentences. Example (5) illustrates switching for one complete Hindi/Urdu sentence. Intersentential switches sometimes involved a complete temporary switch of language to Hindi/Urdu. In (6), there is a switch of language to Hindi/Urdu for a longer stretch with multiple consecutive Hindi/Urdu sentences. The code is then switched back to English mid-sentence at a clause boundary. As a general note on the examples of the Hindi/Urdu codeswitches provided in these two analytical sections, the codeswitch under discussion is marked in bold in the example. The examples are also accompanied by translations[2] of the codeswitches in square brackets. A code indicating whether the example was from IE or PE and whether it was from

[2] Translations of the Hindi/Urdu codeswitches are provided in square brackets after the example. Codeswitches other than the one under discussion are not translated. Titles of movies, songs etc. and names of dishes are not translated. If a translation is already provided in the example itself, the codeswitch is not translated unless the translation lacks some of the content of the codeswitch.

the text type category “general” or “blogs” is provided after the translation, for example IB for India Blogs or PG for Pakistan General.

(5) yes now that he is no more, I would turn to Katrina because she seems to be a genius of Love, life and all the other intricacies of human relationships. **Anushka ke saath love ke bare mein baatein hongi, and Katrina ke saath Life ke bare mein!** But then again, I feel I have a learnt a lot about human emotions and relationships from Mr. Yash Chopra already! [I talk about love with Anushka and about life with Katrina.] (IG)

(6) or imagine.. What if you would go to Hajj and take alongwith you a translator gadget for your ears..:) Why don't you advise Christians of Gujranwala to tranlslate and read Bible in Punjabi during Sunday prayer service.. **yeh hamdardi becharay muslims kay saath hi kion.. Aray baba pehlay inhay prayers per to lay ao... baad main in ki language ka publishing idaara khol layna..:) tumhara predecessor Ata Turk bhi koshish ker gaya** but my sympathies are with all of you...:) [Why this sympathy only with poor Muslims…. Oh man at least bring them to prayers… Later, open a publishing house for their language. Your predecessor Ata Turk also tried] (PG)

Switches between independent clauses within a sentence were also classified as intersentential codeswitches. Coordinated independent clauses are syntactically independent and separate in the same way as individual sentences, and they can form sentences on their own. In sentences containing two independent clauses the independent clauses could be conjoined with a coordinating conjunction as in (7) and (8), or there could be a colon, a semicolon, or a comma linking the two independent clauses. In (7), the coordinating conjunction is the English *and* while in (8) it is the Hindi/Urdu *ya* ‘or’.

(7) I am Mr Bachchan's brand ambassador and **yeh mere contract mein hona chahiye**. Now I can't be Vijay so I am being his heroine in any film. [this should be in my contract] (IG)

(8) Honestly why do you call this stupid man Mushaid Ullah Khan, honestly if I would have been sitting there he deserved a few slaps, such a pathetic man who just got Es. 20 lac from Punjab Government not from Mian Shab's own pocket, **ya paisa in ke Bap ka ha**. [or the money is his father’s] (PG)

Not all the intersentential switches were complete, well-formed sentences, but they also included sentence fragments, i.e. sentences lacking the required features of a well-formed sentence like the predicate verb and the subject. They consisted of, for example, only one noun phrase etc. In (9), the sentence or sentence fragment consists of a postpositional phrase *mood ke hisaab se* ‘according to mood/depending on mood’.

(9) Since you have admitted to being a diehard romantic what's your definition of romance? **Mood ke hisaab se**. Since it's monsoon now, my idea of romance would be having chai and bhajiya on the street lined with trees and listening to Shubha Mudgal's Ab ke sawan.. [depending on mood] (IG)

5.2 Intrasentential switches

Intrasentential switches are further subcategorized into interclausal and intraclausal switches. Interclausal switches are switches between independent and dependent clauses as well as switches for other Hindi/Urdu clauses embedded in English main clauses. Intraclausal switches are switches within clauses such as switches for Hindi/Urdu words and phrases. Out of all the intrasentential switches in the Indian data, 3119 instances, 12.79% (N = 399) of the switches were interclausal and 87.21% (N = 2720) were intraclausal. In the Pakistani data, out of the 1395 intrasentential switches, 19.50% (N = 272) were interclausal and 80.50% (N = 1123) intraclausal. Here again, there were somewhat more interclausal switches in PE than in IE and conversely, IE had slightly more intraclausal switches than PE.

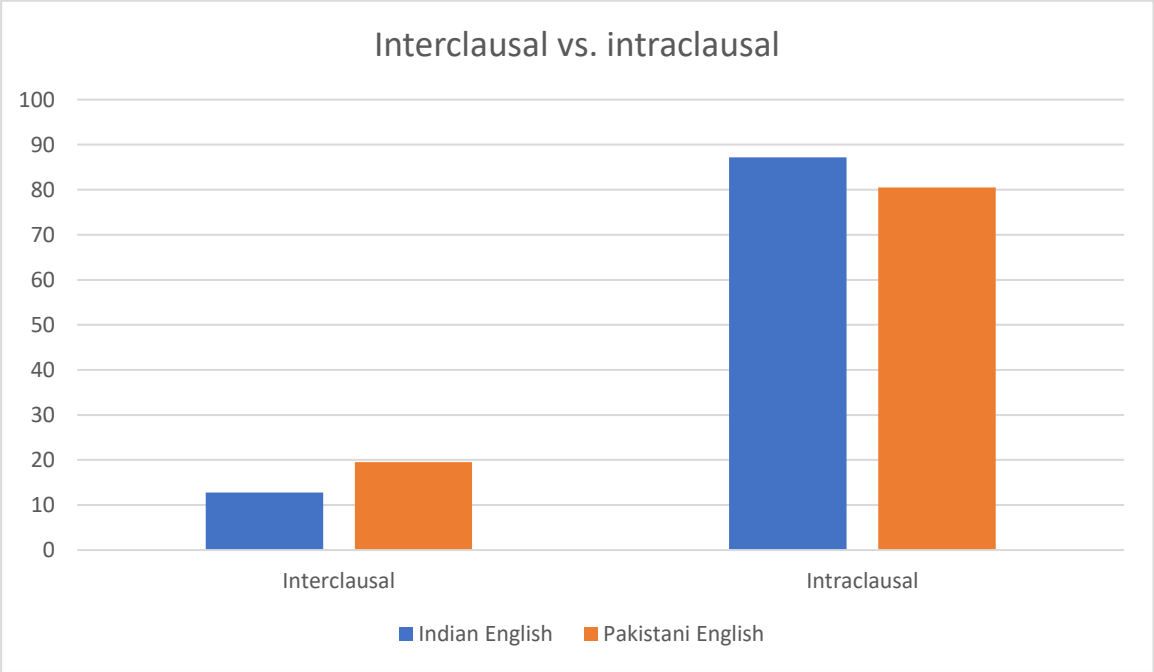

Figure 4 Interclausal and intraclausal switches

5.2.1 Interclausal switches

The term interclausal refers here to switches between independent and dependent clauses at a clause boundary as well as to switches for other Hindi/Urdu clauses embedded in English matrix clauses. The category of interclausal switches primarily consists of switches between English main clauses and Hindi/Urdu subordinate clauses, and switches for Hindi/Urdu clauses embedded in English matrix clauses. Included in this category are, however, also the few occasions of switching between a Hindi/Urdu main clause and an English subordinate clause (IE 19 instances, PE 8 instances).

5.2.1.1 Hindi/Urdu main clauses with an English dependent clause

Switching between a main clause and a subordinate clause occurred between an English main clause and a Hindi/Urdu subordinate clause as well as between a Hindi/Urdu main clause and an English subordinate clause, as in (10). In switches between a Hindi/Urdu main clause and an English subordinate clause, the subordinating conjunction could come from either English or Hindi/Urdu. In (10), the Hindi/Urdu main clause is followed by an English complement clause introduced by the Hindi(/Urdu) complementizer *ki*.

(10) As it's out of my budget, Heena will be gifting me one on my birthday! Then went to Cafe Mondegar (**pareshan kar ke rakha tha Heena ne ek mahine se, ki** I want to go to Cafe Mondegar). It was a great place and we had a great time eating pizza, Jain cheesetoast and a bottle of beer. [Heena had been complaining for one month that…] (IG)

5.2.1.2 Hindi/Urdu dependent clauses with English main clauses

The data exhibited several types of Hindi/Urdu dependent clauses embedded in English main or matrix clauses. Codeswitches for Hindi/Urdu subordinate clauses included, for example, complement clauses, adverbial clauses, relative clauses, and non-finite clauses. Hindi/Urdu clausal switches were also for example quotations introduced by an English introductory clause. The different types of Hindi/Urdu clauses could be observed to appear in various syntactic positions, for example as objects or subjects in an English main clause, or as complements in English prepositional phrases.

Several of the interclausal switches were switches between an English introductory clause and a Hindi/Urdu direct quotation. Although the quoted sequences are independent and autonomous units as such, they are objects of the reporting verb of the English main clause and that way dependent in relation to the main clause. Therefore, they are regarded as embedded in the English matrix clause. Examples (11) - (13) are illustrative of direct Hindi/Urdu quotations introduced by an English introductory clause.

(11) Kaushik secretly wrote to his family in India, telling them of the barbarism he was subjected to. In a letter, he asked: **" Kya Bharat jaise bade desh ke liye kurbani dene waalon ko yahi milta hai?** (Is this the reward a person gets for sacrificing his life for India?) " (IG)

(12) When asked knowledge about counting she said that she knows counting to 10 only. Looking at this I thing it's reasonable to understand the side or indifference from parents side for schooloing. Some parents says " **ab sarkaari mein padayi hoti hi nahin hai, mere bachhe aathvi (8th) mein aake ke bhi chaar line nahin padh sakta toh kyu school bhejna, es se ache ghar ka kaam kar, kheti jaaye "** # Few parents are sending their daughters to private schools for primary education.) [There is no studying

in the governmental school, when my child got to the 8th class they could not read even four lines then why send to school, they do good house work, farming] (IG)

(13) My father and mother were always there right from the start. They would often say,' **Agar issi field mein kaam karna hai toh achhe tareeke se Mumbai jaakar karo.'** It was in 2005 that I came to Mumbai. [If you have to work in this field, go to Mumbai and do it properly.] (IG)

Several of the subordinate clauses were Hindi/Urdu complement clauses in both IE and PE. The Hindi/Urdu complement clauses could be introduced by the Hindi complementizer *ki*, the Urdu complementizer *ke*, or the English complementizer *that*. The Hindi/Urdu complement clauses acted as complements of the preceding verbs as in (14) - (18). Like direct quotations, indirect quotations were also relatively common in the data and consequently many complement clauses were instances of reported speech, as in examples (16) and (17). In (18), the complement clause is an adjective complement.

(14) Now why was I shocked? Hmmm I really did not know **ke Pakistan itna " taraqi " kur gaya hai**! We are so modern, so westernized (I won't say enlightened for obvious reasons) mujhe tu pata hii nahi tha. [that Pakistan has progressed so much] (PG)

(15) He also adds that his three month old daughter is his darling and he loves cuddling her after a hard days work as it refreshes him completely! Hearing about this picture perfect family one can only wish and pray that " **inko kisi ki nazar na lage.** [that nobody lays their eyes on them] (IG)

(16) But he made a mockery out of it on a Youtube video, where he said that, **Police nahin bol sakte toh mein kaise bol sakta hoon?** [The police cannot talk so how could I talk] (IG)

(17) You would not believe I used to ask Yashji **ki kuchh toh action scenes dal dijiye**. [give some more in the action scenes] (IG)

(18) ahahaha it was good I agreed on some point people like girls like phuljarees... but its not necessary **ke woh punjabi girls men hi miley**... you never know when you will generate a capability to handle more then phuljari hahahah... [that only Punjabi girls are like that] (PG)

Hindi/Urdu clauses were also different adverbial clauses. The adverbial clauses in the examples below express reason ((19) and (20)), time (21) and condition (22). Here again, the subordinators could come from either English, as in (19) and (21), or Hindi/Urdu, as in (20) and (22). In (19), the Hindi/Urdu subordinate clause is introduced by the English conjunction *because* while in (20), the subordinate clause begins with its Hindi/Urdu equivalent *kyonki* (‘because’).

(19) i don't remember the exact words in the exact order but it was about whether non muslims would go to heaven if they do good all their lives and not believe in Allah and

in reply to that it was written that their good deeds are of no use because **iss se barra zulm kya hai agar Allah ke saath kisi aur ko shareek thehraya jaye** or maybe it was ke Allah per emaan na laya jaye. [what is a bigger wrong than giving someone else the same status as to Allah] (PG)

(20) Sorry for such a long post but it demanded the answer... I didn't post out side **kyonki thoda odd lagega ki** I am contradicting.. [because it feels a bit odd] (IB)

(21) Oh I love Eids and on such events and I love to talk to mye fans, friends, family, my nephew and niece. I keep my ears closed when **qurbani kay liye qasaab ata hai** and he is ready to slaughter it. [the butcher comes for the sacrifice] (PG)

(22) If a woman is not content and happy to share her husband, **agar us ki ijazat nahi hai**, he cant marry 4 times, 3 times or even 2 times. Times were different back then, i agree with Adnan on this.. ever heard of " Ijtihaad "? [if he does not have her permission] (PG)

The data contained a few Hindi/Urdu relative clauses, as in (23) and (24). In both examples, the Hindi/Urdu relative clause is introduced by the Hindi/Urdu relativizer *jo* – or *jin* as it is in its oblique case form in (24). In (24), the Hindi/Urdu relative clause *Jin ke pait mei dard ho rahi hai* ('who have fear in their stomach'; 'who are scared') has been placed before the English main clause. While a relative clause follows its antecedent in English, in Hindi/Urdu relative clauses can be placed right after the antecedent, or they can be detached from the antecedent and either follow or precede the entire main clause. Here the relative clause precedes its antecedent "they", the subject of the English main clause, i.e. its placement follows the permissible positioning of a relative clause in Hindi/Urdu.

(23) Exposure-wise, in every way, the Indian film industry is far better than the Pakistani film industry. But I hope and pray that the Pak industry gets the same prestige **jo ke aaj se 30 saal pehle thi**, [that it had 30 years ago] (PB)

(24) Javed Hashmi, whose reputation in politics and beyond is beyond reproach, is being perfectly consistent, ie; a patriot. **Jin ke pait mei dard ho rahi hai,** they can always pick up their bori bistra and move to india. [who are scared] (PG)

A few Hindi/Urdu non-finite clauses were also found embedded in English. In (25), a Hindi/Urdu non-finite participial clause is modifying the head noun in an otherwise English noun phrase. In Hindi/Urdu, participial relative clauses precede their nominal head unlike in English where the equivalent structure follows the head. Another example of a Hindi/Urdu non-finite clause is the adverbial participial clause "*bori bistar baandh ke*" in (26). This construction involves a non-finite form of the verb *bandhna* ('to close', here: 'to pack'), which is formed with the conjunctive participle *-ke* attached to the verb stem *bandh-*. Constructions involving a conjunctive participle somewhat equal the English 'having done something' (or 'as soon as').

In (27), there is a Hindi/Urdu infinitive clause with the infinitive form of the verb *phekna* 'to throw' functioning as the subject of the sentence.

(25) He darted back to his mother like I do when I go to receive a phone call and end up chit chatting forgetting about milk kept on the stove for boiling not realizing that he had not thought over in his mind about what and how was he going to ask his sari clad, bindi adorning, **haath mein puja ki thaali liye hue** mom about it. [with a prayer tray in her hand] (IG)

(26) It was, as they say in Hindi, people were running to Hyderabad **bori bistar baandh ke** [after packing up their stuff] (IG)

(27) And I totally agree, it is so sad that " **kisi haseen ki taraf ek gulaab phekna** " does not even enter the brain of today's men. Maybe because the haseena herself has somewhat changed? [throwing a rose to the direction of a beautiful woman] (IB)

Hindi/Urdu clauses could be found in various syntactic positions. Hindi/Urdu clauses were used as objects of verbs as in (28) and (29), as subjects as in (30) and (31), and as complements of prepositional phrases in English sentences like in (32), and (33). Hindi/Urdu clauses were also found embedded in English noun phrases as in (34) and (35).

(28) When I wasn't completely breastfeeding I was the center of discussion and auntys would discuss how my milk wasn't enough and when I started they would say **baby ko feeder ki adaat honi chai hai**. [a baby should not get used to the feeder.] (PB)

(29) He doesn't speak much English, but can read and write. I see him every morning reading the Times of India to know' **uske gaon mein kya ho raha hai**.' [what is happening in its villages] (IG)

(30) Even when alcohol was not banned legally, drinking was considered to be infra dig. **" Achchhe ghar ke log peete nahi hai "** was the norm. Soda was different. After the evening meal, people would flock to the Soda-wallah. [People from good homes do not drink] (IG)

(31) **" Zindagi aur maut uparwale ke haath hai jahanpanah, jise na aap badal sakte hai na main "** is one of his most famous dialogues from the film Anand. He died on April 12, 2011 of a cardiac arrest. He wrote around 90 films in his career, the last being Yuvraaj. [Life and death are in a higher hand, you cannot change it, neither can I.]

(32) When he insults her about' **paise ke liye kuch bhi karegi'**, it sounds extremely degrading and hurtful. # I wish they'd show something about him coming to know of her majboori and softening him up in this aspect! [doing something for money] (IG)

(33) It is strange that Imran's intentions of marching on to Waziristan did not wake TTP for any reaction, but all of a sudden TTP wakes up and threatened him of KILLING with lame excuse of " **Hum ek pro-western politician ki hamdardi ke muhtaj naheen hain. Imran ne long march kia to use qatal kar dia jaiga** ". [we don't need the sympathy of a pro-western politician. If Imran organizes a long march, then he will be killed.] (PB)

(34) however you can smoke if you want in public i dnt see the world coming down, but do prepare self for the **beta ache ghar ki larkiya yeh nhn karti** lecture!! [son, girls from a good home don't do that]

(35) The traditional' **takaluf ki koi zaroorat nahi hai**' announcement by your mother, and yet you are served with the samosas, chat, kebabs, drinks and everything they can afford and manage to get for you. [no need to take trouble] (PB)

5.2.2 Intraclausal switches

Intraclausal switches were Hindi/Urdu words, phrases, and other codeswitched sequences below the clause level embedded into English clauses. Intraclausal switches were further categorized according to their word class and phrase type. Switched elements or sequences that did not fit either the word class or phrase type categories were classified as "other intraclausal switches".

Figures 5 and 6 show the percentages of different categories of intraclausal switches in IE and PE, respectively. These percentages are counted out of the total number of intraclausal switches in the respective variety – 2720 in IE and 1123 in PE. Percentages of those word class and phrase type categories that exceeded one per cent are shown as their own categories in the figures. Word class and phrase type categories that constituted less than one per cent of the intraclausal switches are in turn grouped collectively as "other word classes and phrase types". The figures basically show that in IE and PE, 3-4% of the intraclausal switches were categories that constituted less than one per cent of the intraclausal switches as individual categories. The categories shown in the figures for IE and PE are partly different because of the slightly different percentages of certain categories in the Indian data and Pakistani data. The number of categories in the figures for IE and for PE is also different for the same reason; a different number of categories constituted more than one per cent in the IE data and the PE data.

As can be seen in the figures, proper names were the most prominent category making up roughly 70% of all the intraclausal switches in both IE (73.9%, N = 2010) and PE (69%, N = 775). The second largest and the third largest categories were genitive phrases and nouns. In IE, genitive phrases constituted 9.8% (N = 267) and nouns 7.1% (N = 194) of the intraclausal switches. In PE, the percentages were 11.6% (N = 130) and 10% (N = 112), respectively. In both IE and PE, "other word classes and phrase types" consists of diverse kinds of other intraclausal switches which were more marginal as individual categories, such as adjectives, postpositions, and acronyms. The word class and phrase type categories, that were over one per cent, will be discussed in their own subsections – regardless of whether they were recognized as their own categories in both the IE figure and the PE figure or just one of them.

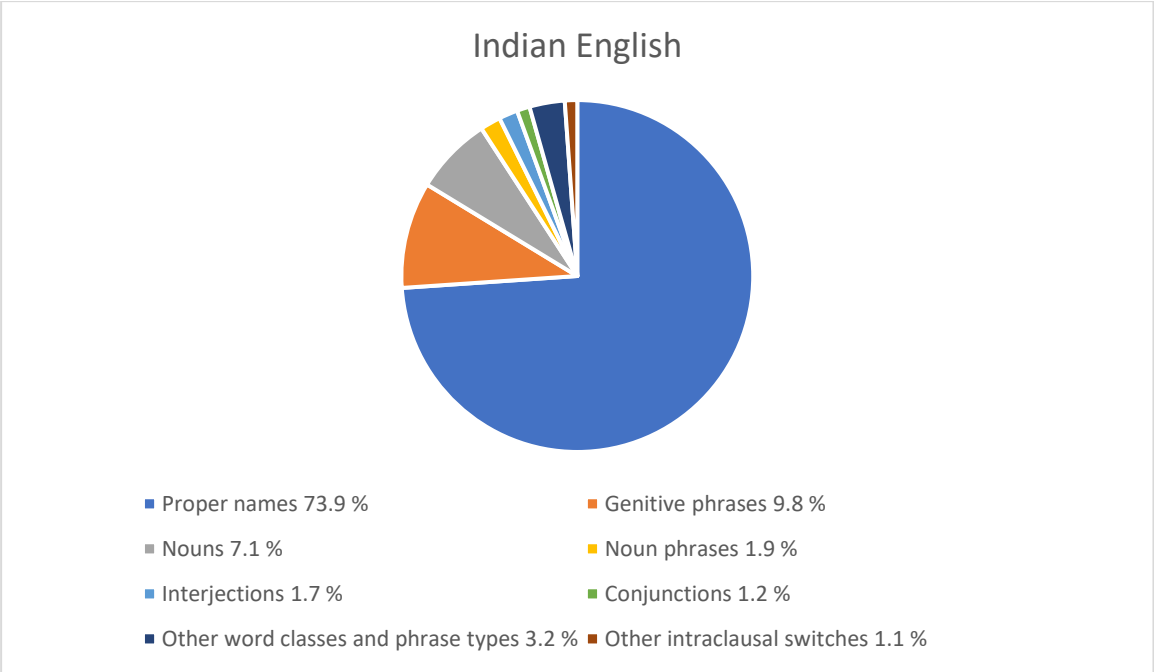

Figure 5 Intraclausal switches in Indian English

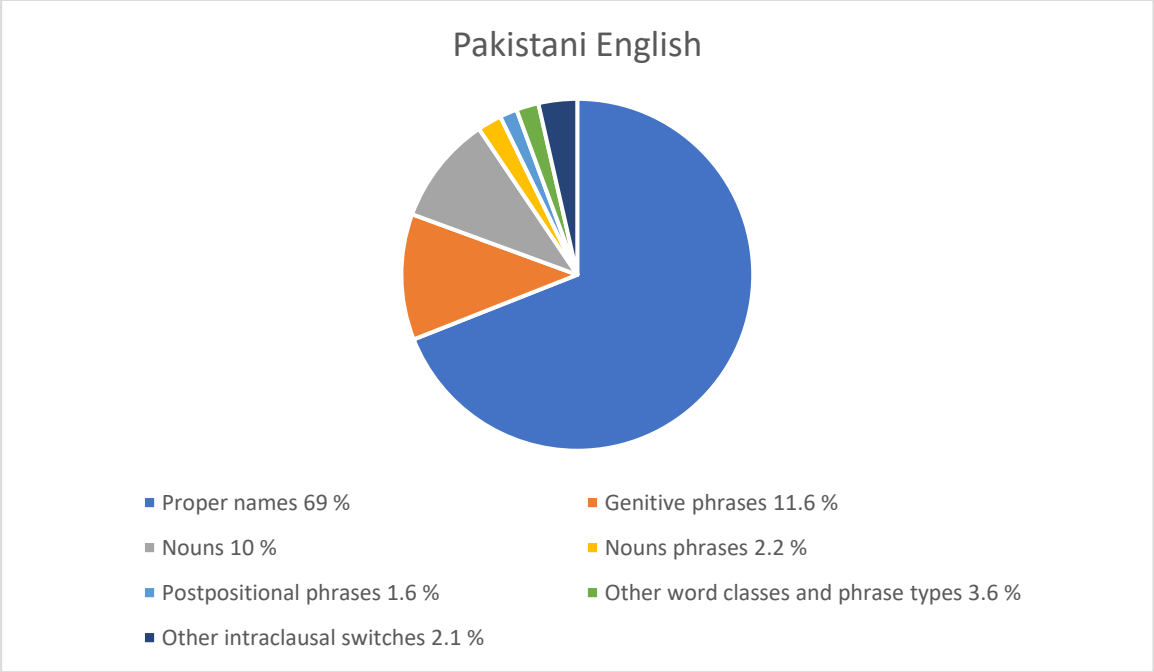

Figure 6 Intraclausal switches in Pakistani English

5.2.2.1 Proper names

Proper names were the largest category of intraclausal switches in both IE and PE. Proper names were mostly titles of movies, songs, tv-shows and books – with occasional place names and nicknames. The movie titles were names of Bollywood movies and movies produced by other local film industries. Song titles were names of Hindi/Urdu songs by Indian and Pakistani artists and songs appearing in Bollywood movies. The titles of tv-shows were names of Indian and

Pakistani tv-shows, and book titles were those of books written by local authors. Example (36) contains titles of Pakistani movies. Example (37) illustrates the use of titles of Bollywood songs. Example (38) contains Hindi/Urdu titles of Indian tv-shows. The proper name in (39) is a book title.

(36) **Khuda Ke Liye, Bol, Yeh Dil Aapka hua** and few other movies were technically sound and showcased strong scripts and some impressive performances. (PG)

(37) The songs**' Challa',' Ishq Shava'** and**' Saans Mein'** are a delight to the ears. The Master Musician's' Heer' is a beautiful song. The numbers are hummable, and nice ones; but just that. They aren't exactly exceptional. (IG)

(38) I started with a positive role, did a grey scale role (in **Kis Des Mein Hai Mera Dil**), did a haunted role (in **Aahat**) and a glamorous role (in **Baat Humari Pakki Hai**) so you see I have done a variety of roles. (IB)

(39) Upon completion of his research, in 1955, he published his first book, **Naye Ahd Ke Darwaze Par**, or' On the Threshold of a New Era'. (PB)

5.2.2.2 Genitive phrases

Another large category in the phrase-type classification was genitive phrases. This category consists of genitive constructions formed with the Hindi and Urdu genitive postpositions *ka* (masculine singular), *ki* (feminine singular and feminine plural), and *ke* (masculine plural). Genitive phrases are noun phrases with a genitive postpositional phrase premodifying the head noun. Therefore, they can also be seen as belonging to the category of noun phrases. However, in the previous research on Urdu CS in PE (Anwar 2009), genitive phrases have been discussed as their own category separately from other noun phrases, which is why I will also discuss them as a separate category.

Genitive phrases consist of a head noun and a genitive postpositional phrase premodifying the head noun. The genitive postpositional phrase in turn consists of a postpositional head and its nominal complement. The head noun and the nominal postpositional complement could come from either Hindi/Urdu or English. There were genitive constructions where both the head noun and the complement are Hindi/Urdu nouns, genitive constructions where one of them is a Hindi/Urdu noun and the other one is an English noun, and genitive constructions where both nouns are English. (40) - (43) are examples of different genitive phrases. They represent instances where the genitive construction is made up of Hindi/Urdu material (examples (40) and (41)) as well as instances where the genitive phrase also contains English material (examples (42) and (43)). However, although apparently English, the possible

status of these English nouns as potential borrowings in Hindi/Urdu cannot be completely ruled out, in which case the apparently English nouns would actually also count as Hindi/Urdu items.

(40) that **achy ghar ki larki** lecture was in fact the point of the article which u hypocrytically declared as non existant read ur comment again. Even in villages only elderly women indulge in Hukka n stuff young girls are not allowed there as well! [a girl from a good home] (PG)

(41) I want to make films that reach out to a wide audience. I will now make films that would exude the **mitti ki khushboo**. I won't pick up any and every film for remake [the smell of soil] (IG)

(42) But since after losing land for good money (ofcourse even a fortune wont last for life if you do nt have anything to work) and losing money on aiyyashi these jaats and bhaiyyas had to get into crime (chain snatching, robberies etc.) and since the bad habits come along when you have been doing aish at **baap ke cash**. [father's cash] (IG)

(43) At the same time, I didn't want to become a producer with PTV, because I was averse to the idea of a nine-to-five job. No **bank ki naukari**, no **denting-painting ka kaam**, and no **import-export ka business** for me, please! I've been a freelancer all my life. [work at bank; denting-painting job; import-export business] (PG)

Instances where the head noun of the genitive construction is an English noun are counted as Hindi/Urdu genitive phrases – although alternative analyses would also be possible. These could also be seen as instances of an inserted Hindi/Urdu postpositional phrase into an English noun phrase in which case the Hindi/Urdu element would be classified as a Hindi/Urdu postpositional phrase – also when the nominal complement of the Hindi/Urdu postposition is an English word. The Hindi/Urdu genitive postposition could, theoretically, also be analysed as an insertion of a Hindi/Urdu genitive postposition in instances where both the head noun and the postpositional complement are English. However, these genitive phrases are formed following the structure of a Hindi/Urdu genitive expression, which is why the genitive postposition would be hard to see as a mere insertion. Moreover, the choice of the genitive postposition depends on the gender and number of the head noun, which means that the English head noun is assigned a Hindi/Urdu gender when it is preceded by a premodifying Hindi/Urdu genitive postpositional phrase. This further raises the question concerning the status of the English word as a potential borrowing in Hindi/Urdu or potentially belonging to the grey area between not established and established English elements in Hindi/Urdu. Treating English elements, or "English-looking" elements, automatically as English in Hindi/Urdu genitive phrases would not always lead to the correct interpretation. Consider the following example:

(44) Calling the government " **zero-loss ke hero** " (the heroes of zero loss), the BJP's Mukhtar Abbas Naqvi said that government was looking for opportunity to target the CAG. " We will not allow this at all, " he said. (IG)

Here the phrase *zero-loss ke hero* would, at a first glance, appear to contain two English elements; *zero loss* and *hero*. However, the choice of the postposition *ke* indicates that the head noun *hero*, instead of being singular, is actually plural. This is further confirmed by the translation provided in brackets: *the heroes of zero loss*. This means that the word *hero* has received a Hindi/Urdu plural – which in this case means no explicit plural ending – instead of the English plural ending. *Hero* is basically treated here as a Hindi/Urdu word. Therefore, treating English words automatically as English in constituents containing both Hindi/Urdu and English elements and disregarding their status in Hindi/Urdu would not always lead to the correct analysis. Analysing *ke* here as a single inserted Hindi/Urdu element would not be correct either. In (45), the nominal complement of the genitive postpositional phrase is – at least apparently – an English noun; there is *note* in *noton ki gaddis*. However, *note* is actually in Hindi/Urdu plural form here and inflected for oblique case with the oblique inflectional ending (*-on*).

(45) Comparatively the more silent lot, these men seem to come well stocked, what with the endless amounts of **noton ki gaddis** coming out of their pockets. They're usually the last ones to disband for dinner, and the socialising and laughing only begins once they have left the table. [stacks of notes] (IG)

In Hindi/Urdu, genitive constructions are also used to express conceptual wholes or entities in a way that is similar to compound nouns – like for example *paint ki dukaan* in (46). Some genitive phrases were conventionalized and fixed as expressions referring to established concepts such as *khushi ke ansoon*, meaning ‘tears of joy’, in (47). Some established genitive constructions were also idiomatic, figurative expressions, like *lakeer ke faqeer* in (48), which were here categorized as genitive phrases as they are genitive constructions in form.

(46) When applying make up, don't make yourself like a **paint ki dukaan**, but don't miss the basics. Always apply a nice coat of lipstick or lip gloss - it attracts men like nothing else! A dash of mascara, a bit of gloss and you are done! [paint shop] (IB)

(47) It is a different kind of love story. It has not been seen before. I love it. All the macho men; the stud kind of guys are going to come out with **khushi ke ansoon**. [tears of joy] (IG)

(48) Actually trash coming out of Andhra and Tamil Nadu has lowered the image of Indian software developers. I fired few in last few months. By the way even Gora programmers are also in the same category.. Most of them are **Lakeer ke Faqeer**.. They

lack original ideas. [lit. line fakir: fig. a person who does everything by the book strictly abiding by the rules without applying common sense or their own thinking] (IB)

5.2.2.3 Nouns

Hindi/Urdu nouns were the third largest individual category of intraclausal switches after proper names and genitive phrases. As Figures 5 and 6 show, the number of single nouns is very much the same in both IE and PE. Examples (49) and (50) contain examples of inserted Hindi/Urdu nouns in English sentences.

(49) I was backstage when Pawar arrived, but when I arrived there and asked' Aap ke saath kaun kaun hai?' the ACP started hitting me with a **lathi**. [stick] (IB)

(50) our hypocritical society doesn't condemn the men who visit prostitutes.all the **besharmi**, **beghairti** is laid at the feet of the women while the men who visit prostitutes are called' studs',' mard kay bachay' while the prostitutes are considered the scum of society. [shamelessness; shamelessness] (PG)

In (51), there are two Hindi/Urdu nouns embedded in English: *kharchee* and *bachees*. The word *kharchee* ('expenses') is in Hindi/Urdu plural – singular *kharcha*. *Bachees* is interesting in the sense that it shows the use of a double plural; the word *bacha* ('a child') has first been inflected for Hindi/Urdu plural *bachee* and has then received the English plural ending -s. However, when Hindi/Urdu nouns were in plural, they most commonly received the English plural ending, like the word *haveli* ('mansion') in (52).

(51) ' Waderai Ka Beta' glorifies everything about the Saaen; their expensive jeeps (a must for long and frequent drives to the village), their **kharchee** (allowance from what they make off their dad's properties), their lack of burger-pun (who cares about a foreign degree when you got the money?) and their **bachees** (kyun kay Saeen ki bachee bhi Saeen). [expenses; children] (PG)

(52) One of the most famous **Havelis** in Udaipur is the Bagore Ki Haveli, built in the 18 th century and located on vantage point of the edge of Lake Pichola. [mansions] (IB)

The few instances where the Hindi/Urdu noun was part of a compound or a hybrid construction were included in the category of Hindi/Urdu nouns. In these compounds or hybrid constructions, one part of the item was English and the other part Hindi/Urdu, like in *saas-to-be* in (53). There were also instances where English derivational endings were attached to Hindi/Urdu nouns as in *chaiholics* in (54).

(53) More than men, saasu maas are obsessed with getting a chaand si bahu for their sons so that the projeny is also beautiful. Imagine the **saas**-to-be coming for a rishta and saying " zara larki ko inspection ke liye bhej dein ". Pathetic!!! [mother-in-law] (PG)

(54) i know how to cook desi khanna, just not chai (i do nt like to drink it). i know kind of sort of the dumm stuff. but can i get a recipe frm u **chai holics** for that PERFECT chai ki piyaa? [teaholics] (PG)

5.2.2.4 Noun phrases

The data contained a few Hindi/Urdu noun phrases. Classified as noun phrases were nouns with modifiers (constructions with a head noun and its modifiers) – except for noun phrases with a genitive modifier. Nouns that are modified by a genitive postpositional phrase are classified as genitive phrases, as was explained above. Singly occurring nouns are categorized as nouns.

Example (55) contains the Hind/Urdu noun phrase *aankhon hi aankhon mein isharas*, which consists of the head noun *isharas* and a premodifying postpositional phrase *aankhon hi aankhon mein*. If a noun is modified by a postpositional phrase in Hindi/Urdu, the postpositional phrase precedes the noun that it modifies. Instead of receiving the Hindi/Urdu plural *ishare*, the noun has received the English plural ending. The same sentence also contains the Hindi/Urdu echo compound *ishq-vishq* ('love and the like'), which is, however, counted as a noun here.

(55) Ishq-vishq is no more limited to the **aankhon hi aankhon mein isharas**... The virtual world is the place to be and networking websites have become the playground for all sorts of love games. [signs from eyes to eyes] (IG)

5.2.2.5 Postpositional phrases

Unlike English, Hindi and Urdu do not have prepositions but postpositions. In both IE and PE, there were a few Hindi/Urdu postpositional phrases embedded in English. Example (56) contains postpositional phrases *azan ke time par* and *namaz ke time par* both of which translate into 'at the time of prayer' and both of which have the postposition *par* as their postpositional head. The postpositional complements are genitive phrases *azan ke time* and *namaz ke time*. Example (57) in turn contains the postpositional phrase *shor dalnay kay baad* – 'after making noise' – which consists of the compound postposition *ke baad* and its complement *shor dalnay*.

Hindi/Urdu postpositions with an English noun or noun phrase as the postpositional complement are analysed as Hindi/Urdu postpositional phrases with English embedded in them, as in example (58). In these phrases it is the Hindi/Urdu postposition that determines the structure of the phrase. Thus, they would not be conveniently seen as mere insertions of a Hindi/Urdu postposition. Therefore, the phrase *near future mein* in (58) is analysed as an inserted Hindi/Urdu postpositional phrase, which contains an English noun phrase *near future* embedded in it as its postpositional complement.

(56) Nowadays my son keeps singing anything and everything, but I tell him **azan ke time par**, **namaz ke time par**, when someone is talking/sleeping he should not sing bc woh disturn hote hain. [at the time of prayer; at the time of prayer] (PG)

(57) it's time wives starting divorcing the husbands who cheat on them. **shor dalnay kay baad** if you still stick with him it just shows him that you are totally dependent on his income and you'll go nowhere no matter what he does. [after making noise] (PG)

(58) hahaha... i wish u told me earlier... i have already planned for Sikkim But then, i can surely visit andamans **near future mein**, i already have it in my mind..... thanx, this DP had been for a week in fb and now, its here too [in near future] (IG)

5.2.2.6 Conjunctions

There were also single-word switches involving the Hindi(/Urdu) conjunction *ki*. *Ki* is a complementizer introducing complement clauses. These singly switched Hindi complementizers were used to link an English main clause and an English subordinate clause as can be seen in examples (59) and (60). The use of *ki* as a single-word switch could especially be found in Indian English where its usage exceeded one per cent.

(59) For people like us, this self-censoring is so much in the DNA **ki** nowadays even when there is no one around and there is a saucy scene on TV, automatically you feel like changing the channel! [that] (IG)

(60) So only instead of asking for these things directly, they are putting all the blockade and bandh and doing all the other nonsense things. But government is thinking **ki** we will happily pass one bill and then go and tell **ki**,' We have met the demands of the people, please vote please vote'. [that; that] (IG)

5.2.2.7 Interjections

The category of interjections consists of switches for Hindi/Urdu interjections like exclamations, greetings, and discourse particles. Interjections like the exclamation *khuda kay liay* in (61), the greeting *Assalam Alaikum* in (62) and discourse particles *haan* in (63), could be attested in both IE and PE even though the category is provided only in the figure of IE. In IE, interjections constituted almost 2 % of the intraclausal switches whereas in PE the number of interjections remained significantly under one per cent (only five occurrences). Therefore, they are part of the combined category “other word classes and phrase types” in PE. Although we are discussing relatively small percentages and a relatively small share of the entire data in both cases, a noteworthy aspect about the interjections in IE is, however, that the majority of the interjections were interjections ending with *ki jay*. *Ki jay* roughly equals the English expression “long live (st)” and it is more of a feature of Hindi than that of Urdu. The difference between Hindi and Urdu in the use of this particular type of interjection is reflected in the

occurrence of the interjection in the local English, which somewhat explains the slight difference in the number of interjections found in IE and PE. The occurrence of several *ki jay*-interjections in the Indian data itself is caused by *ki jay* containing the frequency list word *ki*. Examples of interjections with *ki jay* are *bharat matha ke jai* in (64) and *google baba ki jay* in (65) meaning 'victory to/long live mother India' and 'long live father google' respectively.

(61) To the educated elite: **Khuda kay liay**, spend your time and energies fixing the huge pile of REAL issues in this country.. do nt just start blindly aping the latest imported psuedo-intellectual debate. [for God's sake] (PG)

(62) **Assalam Alaikum**, With reference to all the replies for the topic, " Hissay wali qurbani mein shave kar sakte Hai ya nhi ". [Hi] (IG)

(63) ok khushiji i dnt knw ki yeh sab kya but **haan** till the time I handle I wud like to not see u around.. [yes] (IG)

(64) We are proud to be an indian, **bharat matha ke jai**. [long live mother India] (IG)

(65) I have heard my mom saying having milk with honey at night increases weight and warm water with honey in the morning reduces weight # Later' **google baba ki Jay**'..., reading about honey helped me understand what she meant. [long live father Google] (IG)

5.2.2.8 Other word classes and phrase types

Word classes and phrase types that constituted less than one per cent of the intraclausal switches as individual categories were grouped collectively in the category "other word classes and phrase types". Among the more marginal intraclausal switches in the data were for example adjectives and postpositions. In IE, this category also contains for example postpositional phrases, and in PE, interjections and conjunctions are included in this category. The use of Hindi/Urdu adjectives in IE and PE is exemplified in (66) and (67). In these examples, Hindi/Urdu adjectives *haram*, *punjoo*, and *gori* are used as attributes in English noun phrases.

(66) rather than frothing at Muslim actor's marriage to non Muslims and their idol worship you people should protest their working in a **haraam** act(acting) and yourself remain away from this sinful act of watching such stuff which will take you to Dozakh ki aag. [prohibited] (PG)

(67) as for karachi kay larkay choosing **punjoo** girls, i think there's just one major reason: **gori** skin. [Punjabi; fair] (PG)

Acronyms are also counted in the category of "other word classes and phrase types". The acronyms were abbreviations of movie titles, like *MBKD* for *Mere brother ki Dulhan* in (68), names of political parties as well as shortened forms of other terms or expressions like *JKG* in (69). *JKG* is abbreviated from the preceding *Joru ke Gulam*, which means 'the wife's

slave'. In some occasions the acronym was followed by the entire name or phrase that it was an abbreviation of set in brackets as a sort of clarification, and sometimes it was the acronym that was given in brackets after the whole name or phrase, as in (68).

(68) Brimming with confidence and wit, Ali Zafar speaks to Ankur Pathak. Your role was appreciated in Mere Brother Ki Dulhan (**MBKD**) and the film is doing well too. Has the success sunk in yet? (PG)

(69) Some Joru Ke Gulam, **JKG** I am proud of:) Edited to add: In India it takes a lot of guts for a man to take a stand against customs that oppress women, it's worst if he is fighting in support of his wife, or his Joru! (IG)

The data contained a few instances of singly occurring Hindi/Urdu postpositions. All of these postpositions occurred in discussion about Hindi/Urdu grammar and the use of Hindi/Urdu postpositions. These postpositions did not have postpositional complements; had they had complements, they would have been classified as postpositional phrases.

(70) All the rest are compound, mostly with **ke**, like **ke ba ((d** (after) and **ke uupar** (above). When you encounter a promising new one, cherish it: give it a special place in your notebook and your mind. (IG)

Sometimes the absence of a feature is as noteworthy as the presence of one. Categories that are conspicuous by absence are finite Hindi/Urdu verbs and verb phrases. Whereas the method used in this study can potentially have played a role, the absence can also be caused by typological differences between English and Hindi/Urdu – namely the structure of verb phrases (main verb + auxiliary in Hindi/Urdu, auxiliary + main verb in English) and the placement of the verb phrase in a sentence (SOV in Hindi/Urdu, SVO in English). In Hindi/Urdu, verbs are inflected by adding inflectional endings to the verb stem, which is received by omitting the infinitival suffix *-na*, for example *karna* becomes *kar-*, and while the English inflectional endings could be attached to the Hindi/Urdu verb stems theoretically, it is another matter how likely that is to happen. These typological differences can make the insertion of Hindi/Urdu verbs and verb phrases into English potentially more challenging and decrease the likelihood of Hindi/Urdu verbs or verb phrases being switched in English.

5.2.2.9 Other intraclausal switches

The categorization according to word class and phrase type could be done with Hindi/Urdu switches that were well-defined, single constituents. There were also instances where the Hindi/Urdu codeswitches were consecutive Hindi/Urdu elements that did not form one unique constituent. When the successive Hindi/Urdu constituents could not be analysed conveniently

as consecutive insertions, they were included in the category "other intraclausal switches". Instances of such switches are exemplified in (71) and (72).

(71) It was only after 1983 that I started playing with a cricket ball. Before that, everywhere I played, **sarkoun pay, chhatoun pay ya school mein** (be it on the roads, roof-tops or in school), it was with the tennis ball. In Ramazan, of course, we used to have a tournament every night. [on the roads, on the rooftops or at school] (PG)

(72) I also belive that jan lokpal bill is not a solution. **Kyunki baad mein** if it get fail to achieve its aim... I am damn sure people will think " india ka kuch nahi ho sakta ". [because later] (IG)

The category of "other intraclausal switches" also includes idiomatic expressions that did not fit into the word class and phrase type categories. Fixed expressions that fitted in phrase type categories were included in their respective categories. In (73), the switched sequence is a Hindi/Urdu idiom – in a somewhat shortened, modified form, though. In its entirety, the idiom goes *dhobi ka kutta na ghar ka hai, na ghat ka hai* ('a washer's dog is neither at home nor at riverbank') and signifies that someone does not belong in either place or does not fit anywhere. An idiomatic expression forms one fixed unit that can be inserted into the other language as such and it is to be regarded as one whole. Although idiomatic, fixed expressions form unique, single constituents on their own, they are here included in the category of other intraclausal switches if they do not fall into any of the word class or phrase type categories.

(73) I think we 1st generation immigrants are " **dhobi ke kutte.. na ghar ke... na...** ". We look at our land of birth with nostalgia and powerful memories of our childhood, and when we do come back, we cant wait to get back to our adopted land. [dogs of a washer, not at riverbank, not…] (PG)

6 Qualitative analysis: functions

In this section, I will analyse and discuss the functions that codeswitches to Hindi and Urdu served in IE and PE. Codeswitches in the data could be divided into two broad categories; codeswitches with a communicative function and codeswitches related to culture. Communicative functions that codeswitches to Hindi/Urdu were observed to serve were, among others, quotations, figurative language, interjections, and switches to convey for example greetings and blessings. Culture-related switches were objects and concept related to Indian or Pakistani culture, or switches that could be interpreted as references to the Indian/Pakistani culture. These were for instance titles of movies and songs, food items, religious terminology, and kinship terms. Despite the distinction between switches with a communicative function and cultural switches, this division does not mean to suggest that culture-related uses of Hindi/Urdu would be completely devoid of communicative intent or referential meaning, or the other way round, that switches with a communicative function could not have cultural implications.

Functions of the codeswitches are analysed qualitatively. Quantitative analysis would require the different functions to be fitted into categories to enable their quantitative examination. However, CS functions are ambiguous at times and one switch can serve several functions simultaneously. For example, when switching to quote a literary work, the function of the switch is not necessarily just quotation, but the quotation serves a further function of creating intertextuality. Further, the role and functions of the codes on a broader social level in the community can be reflected in the micro-level functions that the codes serve in a discourse; in addition to the local, conversational function that the switch serves in the discourse, the codeswitch can also receive further communicative and interpretive effect from the broader macro-level functions of the codes in the society. The functions of switches can also overlap, and the same switch can be assigned to more than one functional category. Furthermore, functional analysis is also subjective, and one codeswitch can receive different interpretations depending on the analyst. Therefore, it would even be impossible to assign the switches conveniently into categories and on the other hand, forcing switches into strict categories that they would not even fit perfectly would give a somewhat distorted and too one-sided a picture of their functions.

6.1 Communicative functions

Codeswitches to Hindi/Urdu served a variety of different communicative functions in IE and PE. These functions included, among others, quotations and reported speech, reiteration, conveying greetings/blessings/prayers, side comments, and insulting to name just a few. In the

following subsubsections, the more prominent types of functions in the data – both IE and PE – will be presented and discussed in their own sections, while the somewhat more marginal categories will be presented in "other communicative functions" in 6.1.7.

6.1.1 Quotations and reported speech

One of the most prominent functional categories in both IE and PE was switching to quote. Quotations ranged from quoting famous works such as poems, books, movie dialogues and song lyrics to reproducing the speech and presenting the voices of Indian and Pakistani people, for example family members, celebrities, as well as other local people. In addition to quoting spoken interaction, the quotes were also sometimes reproductions of written discourse. Quotations not only illustrated utterances and speech in real-life situations, but they also represented people's hypothetical speech.

Sometimes CS to quote can simply serve to frame a quotation. CS can also sometimes be motivated by willingness to retain the original language of the utterance in some instances. Through the switch of language in quotations and representations of speech language users can create polyphony, present different voices, and characterize different actors in a narrated sequence. Quotations of other works, for example books or poems, also serve to create intertextuality. In the following three subsections the different types of quotations are discussed in more detail and illustrated with examples.

6.1.1.1 Quotations of works of art and intertextuality

Hindi/Urdu quotations were quoted sequences of and references to famous works of art such as poems, song lyrics, and movie dialogues functioning as devices to create intertextuality. Quoted sequences could be identified as quotations by an introductory clause and quotation marks – in which case they were easily identifiable as quoted text – or they could be more implicit allusions to a particular work of art. Sometimes the original source of the quote was identified, for example with a mention of the author or the work itself.

In (74) and (75) the Urdu(/Hindi) codeswitches are quotations of poems. In (74), the person exemplifies the use of the word *kafir* in context by quoting an Urdu poem where the word in question is used. In (74), the quotation is accompanied by an English translation. In (75), the quote of the lines of the poem by Allama Iqbal in a way reinforces or supports the point the person attempts to make concerning their opinion on how people should themselves take action to change the situation and circumstances in their country – in this case Pakistan. In (74), the poet is identified directly as Mirza Ghalib whereas in (75), the person behind the

original text is referred to as “our great poet”, which in this case refers to the famous Pakistani poet Allama Iqbal – a reference that would be understood by Pakistanis.

(74) In Urdu poetry the word kafir is used to describe the beloved, usually a beloved who spurns the poet's advances. The beloved can be haqiqi or mjazi, meaning spiritual or earthly, the object of love being either God or an earthling. Urdu's greatest mystic poet Mirza Ghalib says: **" Mohabbat mein nahin hai farq jeene aur marne ka; Usi ko dekh kar Jeete hain jis kafir pe dam nikle. Khuda ke waste parda na Kaabe se utha Zaalim; Kahin aisa na ho yan bh di wohi kafir sanam nikle.** # (In love, there is no difference between life and death, We live gazing at the same kafir beloved on whom we die. For God's sake, do not lift the veil from the face of Kaaba, Who knows, maybe the idol of the same kafir is installed there.) (PG)

(75) We all just want a miracle to happen that changes the destination of our beloved country. # We all know that a miracle happens only if we make it happen. If we just want a golden spoon than sorry to say this is impossible. As our great poet said **Khuda ne aj tak us qom ki halat nahi badli Na ho jisko khyal ap apni halat k badalnay ka** [Until today, God has not changed the conditions of such a nation that does not want to change its conditions itself] (PB)

Quotations of other literary works – other than poems – were quotes from religious texts. Quotes from religious sources involved especially quotes from the Quran, as in (76), where the person quotes an excerpt from the Quran in Urdu. Switching for religious quotations can be regarded as further relating to switching to Urdu in religious contexts more generally. Urdu is regarded as the language of Islam and Muslims in India and Pakistan, and these associations can play a role in the use of Urdu in religious contexts in IE and PE more generally. This more general aspect of using Urdu in religious contexts, which likely relates to the associations of Urdu with Islam and Muslim identity, will be discussed later on.

(76) i think he n his parents should take a stand for u n he should marry u..... if he is not loyal.... then pary to Allah to help u:)..... quran say..... **Allah ki rahmaat say na umeed mut ho**:)..... n life for a new muslim is tough coz u have to take a stand..... but Allah will be with u in every step u take [Don’t be hopeless with Allah’s blessings] (PG)

Examples (77) and (78) contain quotations of movie lines and dialogues. In both cases the quotes are explicitly identified as being excerpt from movies – in the former from the movie Paan Singh Tomar, and in the latter lines uttered by the actor Boman Irani in a movie that in this context remains unidentified. The quotation in (79) is a quoted excerpt of song lyrics.

(77) In Paan Singh Tomar, it was a story of an athlete, who becomes a rebel because of injustice and ill treatment caused to him by the society and ignites a crusade. With dialogues like **" Beehad mein to sirf baagi hotey hain.. chor to parliament mein hotey hain. "** and **" Sarkar toh chor hai sahab! "** don't expect our

govt to select such films. [In jungle there are only rebels, in parliament there are only thieves; Government is just a thief!] (IG)

(78) Irani also gets to deliver, in the half-muttered tones of a crabby old man, the film's most cracklingly sharp lines: from "**Yeh cricketer log nahi hai, yeh salesman log hai, tel-sabun bechte rehte hain** (These guys aren't cricketers, they're salesmen -- go around selling oil and soap)" to "**Jab safed log ke desh mein recession hota hai toh aisa scheme nikalta hai, camp-vamp ka** (When white people have a recession in their countries, they come up with these schemes: camps and suchlike.)" (IG)

(79) A glance at the roof top from the courtyard, you almost feel as if someone would peep down singing **hawa mein urta gaai, mera laal dupatta malmal ke**; all this prepares you to enter a similar environment indoors where you will actually taste the true essence of the eating place. [my red muslin scarf flew in the wind] (PG)

Slogans and advertisements were also among the different quoted texts. In (80), the person is making references to advertisements of different companies or brands by using their slogans – two of which are in Hindi/Urdu and one is in English. The Hindi/Urdu slogans are that of Pepsi (*yeh dil mange more*) and that of the Life Insurance Company of India (LIC) (*zindagi ke saath bhi zindagi ke baad bhi*). (81) involves a quotation of the slogan of the tv-channel SAB TV. The slogan itself contains a word play with the word *sab* which, on the one hand, is the name of the tv channel and, on the other hand, means 'everyone, all'. Thus, the slogan has a double interpretation as "real fun comes with everyone" vs. "real fun comes with SAB (TV)".

(80) Our friendship is not like Pepsi- **yeh dil mange more**, not like wills- made for each other, it is like LIC- **zindagi ke saath bhi zindagi ke baad bhi**... [this heart wants more; with life and after life] (IG)

(81) SAB is India's only family comedy entertainment channel with a core brand promise of' **Asli Mazaa SAB ke Saath Hai**'. SAB TV with its impressive lineup of innovative programs and light-hearted content ranging from daily family comedy to path breaking concepts like silent comedy is dedicated to promoting an enjoyable " family-viewing " experience. [Real fun comes with everyone/SAB] (IB)

Intertextuality not only showed as direct quotations of and direct references to famous works but manifested also as more indirect allusions. In (82) and (83), the Urdu passages are excerpts from Urdu poems. These quoted passages are not marked as a quotation by either an introductory clause or quotation marks but are a more implicit or indirect intertextual references. As can be noticed, the same lines of the same poem that were quoted directly in (75) are used as an indirect allusion in (83).

(82) Islamic Republic of Pakistan is absolutely contrary to the country the forefathers had envisioned. The disaster has led to outcome that **naa khuda hee milaa naa wisaal-e-sanam naa idhar kae rahe naa udhar ke rahay** # Pakistan has only desecrated the

name of Islam and has decimated itself trying to become a Median-like state, which it never can in the 21st century ecosystem. [I got neither God nor the company of my beloved, I don't belong here, I don't belong there] (PG)

(83) Every body is busy in looting the wealth of our beloved country and once they complete there tenure. They will comfortably move to foreign countries where they have all the money saved. **KHUDA NE AAJ TAK US QAUM KI HALAT NAHI BADLI NA HO JISKO KHAYAL AAP APNI HALAT KE BADALNE KA.** [Until today, God has not changed the conditions of such a nation that does not want to change its conditions itself] (PG)

The use of the same lines as direct quotation at one instance and as indirect reference at the other instance leads the discussion to the occasional blurriness of the line between quotations and proverbs. If a certain quotation is frequently used, it can gradually lose its ties with the original text and start to be treated as an established saying or proverb. An example of a quotation – originating from the Quran – that can be thought of as qualifying as an established proverb is *maa kay pairoan kay neeche jannat hai*.

(84) We are taught to respect women for **"maa kay pairoan kay neeche jannat hai"** (the paradise lies beneath the feet of the mother). However, the cases of sexual harassment faced by women in our society are one of the highest in the world. Gang rapes (we are now labelled as "Rape Friendly Country"), killing in the name of honour, abuse of women on the streets with one's eyes and physical assaults are not uncommon. (PB)

In addition to longer quoted sequences, intertextual references and allusions could also be shorter elements or extracts. For example, *Jai Ho* and *neele aasmon ke taale* are both references to the Hindi version of the theme song of the movie *Slumdog Millionaire*. The exclamation *Jai ho* is the song title and *neele aasmon ke taale* is a reference to the line "*aaja zariwale neele asman ke tale*" in the song.

(85) Slumdog Millionaire has won! Yeahhh! Every Indian can now leap out of human excreta and go **Jai Ho** under the **neele aasmon ke taale**. ["Jai Ho"; under the blue skies] (IB)

6.1.1.2 Reported speech of Indian and Pakistani people

Hindi/Urdu was used to reproduce and to illustrate the speech of different individuals, for example family members and friends, celebrities, as well as other local Indian and Pakistani people. People also represented their own speech by switching to Hindi/Urdu. Quotations and reproductions of speech involved reconstructions of past conversations and dialogues as well as presentation and creation of different voices in the course of the narration of a story.

A common denominator for these Hindi/Urdu quotations was that they represented – not just anyone's speech – but the speech of Indian and Pakistani people in particular. This is relevant because switching language to Hindi/Urdu to quote someone was thus not just a strategy to highlight or signal a quote, but to represent local voices. Therefore, the language in which the quote is delivered (Hindi/Urdu) is an essential part of the function of the switch. The use of Hindi/Urdu in presentation of voices of local Indian and Pakistani people creates the sense of localness and adds to the authenticity of both the utterance and the character whose voice it represents. The use of Hindi/Urdu is part of the characterization of these people as local. In examples (86)-(88), the voices of a teacher (86), a coach (87), and rickshaw driver (88) are presented in the Hindi/Urdu quotations.

(86) I lost faith in' Prestigious Universities' since their graduates are the one who messed up World's finance, so if is has to end like this, I prefer learning from my school teacher who always said' **Chadar dekh ke paon phelao**'. [lit. spread your legs as much as is the area of the bed sheet; fig. do not spend more money than you earn] (PG)

(87) Like with soccer, you and I don't know what is to be done at the soccer ground. But whatever the coach says, " **Aapke bete ko coaching ke liya Wales bhejna chhaiya** (You should send your son to Wales for coaching). (IG)

(88) At a very core level this is so true. I remember a ricksahawaalah telling me when HKL Baghat died, **aise log na aise hi saad saad ke marenge**. But, the flip of it is true too. [such people do not die like this] (IB)

Quotations of celebrities included, for example, reported sequences from interviews. Example (89) contains a reproduced question-answer sequence between an interviewer and Shah Rukh Khan, a Bollywood movie star. In this question-answer sequence, both the voice of the interviewer as well as that of the Bollywood movie star are represented in Hindi/Urdu. Other quotations of famous Indian/Pakistani people include a Pakistani a cricketer in (90) and Malala in her speech in (91).

(89) When a mediaperson asked him, " **Salman Khan ne ek interview mein aapki taarif ki hai, aap kya kahenge;** " Shah Rukh replied, " **Please aap aise sawaal mujhse na poocha karein, main sharminda ho jaata hoon.** " SRK seemed to be in good humour despite the early hour. [Salman Khan praised you in an interview, what do you say?; Please do not ask me such questions, I start feeling shy.] (IG)

(90) Kapil sidestepped queries on the issue of whether Sachin Tendulkar should retire from international cricket. # " **Bade player ke baare mein na hi baat karein to accha hai** (It is better not to talk about the decisions of big players), " he said. (IG)

(91) Also, CM recalled the time when Malala was invited as a guest of honour at the Prize Distribution Ceremony of Inter toppers at Lahore and she started her speech by saying that, **" Swat ki bachiyan bahadur hain aur kisi se nahee darteen ".** At the moment, CM announced naming the Danish School Attock after Malala Yousafzai and

also announced a quota for students from Swat in this Danish School " [Children from Swat are brave and not afraid of anything.] (PG)

In reconstructed family conversations the speech of family members and relatives was illustrated through CS to Hindi/Urdu. Code could also be switched to illustrate the speech of a friend or a discussion between friends. In instances where the writer has quoted their family members or others in their close inner circle using Hindi/Urdu in Indian and Pakistani context, the language of the quotation can be a relevant part of the function of the switch. Using Hindi/Urdu when representing the speech of family members and friends can be seen as reflecting the existence of diglossia in people's lives and a situation where Hindi/Urdu is the code used in close, informal relationships. Quoting family members and friends itself does not imply diglossia, but because the existence of diglossia is known, these codeswitches for quotation can be interpreted in that light. The associations of Hindi/Urdu with home conditions and friendships thus manifest in the language choice in representation of voices of family members and friends. Instead of functioning merely as a device to frame or signal a quote, the switch to Hindi/Urdu draws part of its communicative function from the situational switching in the community and is an essential part of the presentation of the voice of a family member or friend. These quotations can be seen as demonstrations of how the macrolevel functions of codes have impact on the functions that the codes have on the microlevel. It can further be assumed that the quoted conversations can actually have taken place in Hindi/Urdu, and so the original language of the utterance has been retained in the quotation.

The following extracts exemplify the Hindi/Urdu switching for quotations of family members and family conversation. In (92), the person writing the text presents the speech of their aunt (*khala*) in Hindi/Urdu, whereas in (93) the words of a grandfather are quoted. In (94), the dialogue between the mother and the person writing the text is reproduced in Hindi/Urdu. In (95), the writer illustrates her own speech in Hindi/Urdu when reproducing a discussion with her son.

(92) My khala trotted up to me on my birthday 13/08/2012, I was out with my friends and all she had to ask my mom was " **Aur iss ke result ka kya hua?** " I was like SERIOUSLY!? Pakistani's are SICK. Families are all about bringing eachother down. Its about upstaging eachother. [And what about her the result?] (IG)

(93) I recall a confession my late grandfather once made while puffing on his infamous Hukka: **" Beeta: Meeri shaadi aap ki daadi se, pain'sat saal ki hai. Leikin mujhe aab tack patta nahin ye kya cheez hai LOVE "**. Stick jealously to your marital roles and Allah (SWT) will SURELY keep your bond secure. [Son, I have been married to your grandmother for 65 years. But I still don't know what thing love is.] (PG)

(94) Well... for a number of years, Ammi has been wanting to go for Haj. And of course, I have often thought about it. But when Ammi said a couple of times **ki humko jaana hai**, I said **chaliye hum aapko le chalte hain, hamara bhi Haj ho jayega.** This is rather far out in time, but as Azad grows up, he will be looking to go for pilgrimages to sites of which faith [that we have to go; let's go, we have to go, we also go on pilgrimage] (PB)

(95) If he wants to sing he can go to his room and sing, so he is almost getting it but still forgets so remind him politely **mama ke kia bataya hai singing ka**?? he realizes and immediately follows.. [what did mom say about singing] (PG)

As can be seen in examples (94) and (95) above, people also presented their own speech through a switch to Hindi/Urdu. In these particular examples, the writers illustrated their speech in a conversation with a family member. An example of a person presenting their own speech in Hindi/Urdu in some other context than a family conversation can be found in (96) where the switch is part of a narrated past interaction with a ticket agent.

(96) When we reached the theatre the person told me there was only one ticket available. Or there were two seats in separate places. I pleaded with the ticket agent. " **Mujhe apni beti ke saath baithna hai. Aur agla show bohat late ho jayega. Kal school hai aur usko film dekhni thi.** (I need to sit with my kid. The next show will be too late since she has school tomorrow. She really wanted to watch the film.) (IB)

Hindi/Urdu was also used to present voices of different characters in a fictive narrative. In narrating the plot or contents of a movie or tv-series, the voices of the characters were presented in Hindi/Urdu. Examples (97) and (98) are accounts of the story lines or scenes from episodes of Indian/Pakistani tv-series. In these summarised reviews of the contents of the episodes, the speech of the character – likely their actual line – is quoted in Hindi/Urdu.

(97) Ram and Priya smile at their cute little chota packet taken back by her choice of words and the wonder how long they've been standing there for. Both walk over to pihu quickly comforting her Ram says "**aree app toh hum dono ki gudiya rani hai**" as he bends down on his knees. [come on, you are our little princess] (IG)

(98) He along with Jaman gears up to visit neighbor village to meet minister with their plea related to Sassi. His determination has not shaken by the conspiracies, and while encouraging Jaman he quips, **" Pathar marne wale se na daro bulke khud ko itna uncha karlo ke koi pathar tum tak pohnch hi na sakay**(Don't be afraid of the ones who throw stones at you, rather raise your caliber so that no stone reaches you). " (PG)

The above presented examples of quotations and reported speech are all reproductions of spoken interaction. Quoted sequences also included extracts of discourse from written

sources, not only from spoken communication. (99) is an example of an instance where the quote is an excerpt from a written discourse.

(99) Kaushik secretly wrote to his family in India, telling them of the barbarism he was subjected to. In a letter, he asked: **"Kya Bharat jaise bade desh ke liye kurbani dene waalon ko yahi milta hai?** (Is this the reward a person gets for sacrificing his life for India?)" (IG)

6.1.1.3 Hypothetical quotations

Some of the quotations and representations of speech were not anyone's actual speech but were rather representations of hypothetical speech. These hypothetical or imaginary quotes neither originated from any "real" source nor were they something that anyone had necessarily really said. They were rather illustrations of people's hypothetical or imagined speech in fictive situations. The people themselves were also generally fictive and imagined or the quotations were more generic representing the speech of people in general or that of members of a specific group of people.

Like the non-hypothetical quotes discussed in the previous subsection, hypothetical quotations also depicted voices of Indian and Pakistani people, or those of people with affiliation or connection to India or Pakistan, for example people with Indian or Pakistani family background. Hypothetical quotations also represented hypothetical speech of people in general or that of a particular group of people in general functioning as more generic quotations. Example (100) involves representation of fictive speech of the local "literate" people – in this case Pakistani – with a codeswitch to Urdu. With the hypothetical speech, the person describes biased attitudes that, according to the them, "even the literate people" have. In (101) and (102) the imaginary quotes are supposed utterances by "those people who stand up after the prayer" and "third-generation Indians in England", respectively.

(100) Even the literate people here talk like.. **Qadianio ke shakal per lanat hoti hai? Christians say boo ati hai?** [Curse on the face of heretics. Christians smell.] (PG)

(101) I think it is those people who standup after the prayer and say' **Namaz kay baad deen kee baat hogee**' and then they take the group in the corner. I think those are the people who have bigger influence. [after prayer we talk about religion] (PG)

(102) I don't want third-generation Indians in England to say, **Indian film ke liye theek hai** (it's not bad for an Indian movie). (IG)

The source of a hypothetical quote could be identified as one specific, yet hypothetical person, as in (103) and (104). Example (103) actually contains two instances of hypothetical speech. The first two quotes (*"Aray, yahan rape hua hai."* and *"Ek aurat ka Saddar mein murder*

hua hai, story chahiye?") are hypothesized speech of a crime reporter. The same context contains a third hypothetical quote representing the speech of people in general – a group of people which the writer is also part of ("We don't even stop and think…"). Example (104) involves hypothetical speech of a Pakistani mother-in-law (*saas*) in a situation where the imaginary mother-in-law comes to see a potential daughter-in-law for marriage arrangement and is primarily interested in the daughter-in-law having the desired skin tone.

(103) Every other day our crime reporter says, **"Aray, yahan rape hua hai."** (Hey, there's been a rape at this place.) or, " **Ek aurat ka Saddar mein murder hua hai, story chahiye?** " (A woman's been murdered in Saddar. Want a story on that?) We don't even stop and think twice before asking, " **Nahi, is mein naya kya hai? Mazay ka murder tha? Koi weird detail pata chali?** " (No, what's new in that? Was there something cool about the murder? Did you find out any weird details?) (PG)

(104) More than men, saasu maas are obsessed with getting a chaand si bahu for their sons so that the projeny is also beautiful. Imagine the saas-to-be coming for a rishta and saying " **zara larki ko inspection ke liye bhej dein** ". Pathetic!! [just bring the girl in for inspection] (PG)

6.1.2 Figurative language

Another relatively large category was switches for Hindi/Urdu figurative language which involved the use of Hindi/Urdu idiomatic expressions and proverbs. Example (105) contains the expression *bhens ke samne been bajana*, which literally means 'to play flute in front of a buffalo', and which is used to express that something is completely futile. In this particular context, *bhens ke samne been bajana* refers to the futility of trying to reason with safety mobs – or *ghairatmobs* as the writer calls them – on the subject of wearing a hijab. *Laaton ke bhoot baaton se nahin manta* (lit. 'ghosts of feet don't listen to talks') in (106) expresses that you have to be strict and harsh with someone in order to make them obey because they do not listen to normal talk and orders.

(105) Loneliberal PK: I guess, there are many ways to respond to " Ghairat'mob, one is redicule, which you do well, is it effective? I have very other mode to to respond, mainly reason and logic, true it is like' **BESS ke samane been bajana**', but some time it works some time not.But whether I agree or not, I get amused, and it bring smiles to my face, which is good at my age.Keeps me young and in good spirit.Cheers. [lit. to play flute in front of a buffalo; fig. something is completely futile] (PG)

(106) combination of both gandhiji &; subhash ch bose work properly today. Depend on people you deal, first apply gandhi rule if it not works then apply subhash rule because "**latho ke bhoot baaton se nhi manta**". [lit. ghosts of feet don't listen to talks; fig. you have to be strict with someone to make them obey] (IG)

Neki kar darya mein daal in (107) is an example of the use of a Hindi/Urdu proverb. This proverb implies that if one does something good or a favour for someone, one should not boast about it or talk about it any further. One should just rather forget about it and not expect anything in return. Sometimes the line between a proverb and a quote was somewhat obscure and not that easy to draw as was discussed earlier. However, in this particular instance the expression could be classified quite clearly as an established proverb.

(107) salman bhai is really to day's haatim taai like **neki kar dariya mein daal**. in the beginning of skr's carrier salman help to skr 7 support him as much as possible. in those days srk stay in salman's home. but salman not show never that he helped to srk and then what did srk. [do good, throw in the river] (IG)

The Hindi/Urdu figurative language also involved shorter idiomatic expressions. *Doodh ki dhuli*, literally 'washed in milk' or 'milk washed', in (108) is a Hindi/Urdu idiomatic expression meaning that something or someone is utterly flawless and totally beyond reproach. This expression is usually not complimentary but has a negative tone to it like in the current example where it is used in a sarcastic manner. In discussing idiomatic expressions, it is worth noting that this same extract happens to also contain the Hindi/Urdu proverb *100 chuhe khaake billi haj ko chali*. This proverb translates literally into 'after eating 100 mice, the cat goes on a pilgrimage'. Its figurative meaning is that after causing damage or harm to others for someone's own gain or profit, the person tries to cover up their wrongdoing with a good deed. Although this particular instance was not included in the actual analysis as it neither contains any of the search items nor does it meet the requirements to be collected as an additional switch, it is another example of the use of Hindi/Urdu figurative expressions. On the other hand, this proverb did, however, appear a couple of times in the analysed data as well when it was used in other contexts.

(108) This are the best ways of washing out her image in public and become as **doodh ki dhuli** as ever.... its like 100 chuhe khaake billi (in her case Jungli billi) haj ko chali... before jungli billi, there are so many actresses who have been associated with social causes. [lit. washed in milk/milk washed: fig. innocent, flawless, beyond reproach] (IG)

6.1.3 Greetings

Hindi/Urdu were used in greetings and well-wishes as well as to convey blessings and prayers. Greetings and well-wishes were mostly found in the beginning of a text as opening formulas whereas blessings and prayers were more typically found in the end of a text as closings.

Example (109) contains the greeting *Assalam Alaikum* in the beginning of a user comment. In (110), the person uses Hindi to wish happy Diwali.

(109) **Assalam Alaikum**, With reference to all the replies for the topic, " Hissay wali qurbani mein shave kar sakte Hai ya nhi ". [Hi] (IG)

(110) **Deepavali ki anek anek shubhkamanayein... Sab ko...** prosperity, success good health, and may you win the worth of it [Happy Diwali for everyone] (IG)

Switching for blessings and prayers, as in (111) and (112), appeared to occur especially in the context of Islam. In India and Pakistan, Urdu is associated with Islam and regarded as a marker of Muslim identity – just like Hindi is linked to Hinduism and Hindu identity. These associations between a religion and a language can be regarded as being reflected in switching to Urdu in blessings and prayers; as Urdu is the language of Islam and Muslims, it is the appropriate language choice for conveying blessings and prayers. Switching to Urdu in other religion-related contexts could be attested elsewhere in the data as well, for example in quotations of religious sources, as was shown above in 6.1.1.1.

(111) so, after performing Hajj, one must change his or her lifestyle completely and make it to conform to the teachings of Islam... one must NOT revert back to old ways because the chance that one gets to purify himself/herself may NOT get another chance again... so, one must cash in the benefits and not loose it again... **Allah ham ssab ko hidaayat de aur siraat ul mustaqeem par chalne kii taufeeq' ataa farmaaye... aameen** [May Allah guide us to the straight path] (PG)

(112) God Bless Our dear daughter Malala, **Jald sehat aabi ki Dua ke saath.** [may she get better with prayers] (PG)

6.1.4 Interjections

Hindi/Urdu interjections were used in both IE and PE. As could already be seen in the structural analysis, interjections were more common in IE than in PE. However, the percentages of interjections in the structural analysis do not correspond to the number of interjections as a functional category in either IE or PE because several of the interjections were intersentential, which is why they were not included in the intraclausal category “interjections”. Although functions are not analysed quantitatively, the difference between IE and PE in the number of interjections was still visible. This difference was caused, like was already explained in the structural analysis, by the fact that majority of the interjections in IE were interjections with *ki jay*, which are more of a feature of Hindi than that of Urdu, which explains why they occurred more in IE than in PE. The occurrence of several interjections with *ki jay* in the data, in turn, was a consequence of *ki jay* containing a search term. Therefore, one cannot really generalize that interjections would be overall more common in IE than PE. However, it shows that these

particular types of interjections are more common in IE than in PE, which is a consequence of the difference between Hindi and Urdu in the use of these interjections.

Some examples of Hindi/Urdu interjections were already seen in 5.2.2.7. Example (113) contains two interjections: *India ki jai ho* ('Long live India') and *Pakistan zindabad* ('Long live Pakistan'). *Zindabad* is basically the Urdu equivalent of the Hindi *ki jay*. Yet another example of a Hindi/Urdu interjection is *Allah ki marzi* in (114).

(113) you need a break. I recommend you 3 Idiots. Get well soon. Aal Izz Well. **Pakistan Zindabad. India ki jai ho.** [Long live Pakistan. Long live India] (PG)

(114) Lack of education, no big deal, nobody in the family went to college. Minor and major health crises- **Allah ki marzi**! People have been encouraged to adopt cultural ways, underpinned by pseudo-religious sanction, of reconciling to usurpation of their rights. [God's will] (PG)

6.1.5 Reiterations

Hindi/Urdu were used to reiterate what was previously expressed in English. Reiteration as a CS function usually serves further communicative purposes; it is not only the act of repeating itself but rather what is achieved through the act of repeating something that is significant. Reiterations can serve as a strategy to add emphasis on a statement, or to clarify or to elaborate on a point. In IE and PE, Hindi/Urdu reiterated sequences were also translations. Reiterations in the data involved both more and less accurate or precise repetitions of the previous sentence or statement; sometimes the repetition was an almost direct translation, sometimes the contents were repeated in the other language in a modified form.

In (115)-(117) the codeswitches are reiterations for emphasis where the emphatic effect is achieved through the repetition of the preceding sentence or phrase in another language. In (115) and (116) the contents of the preceding English sentence are repeated in Hindi/Urdu: *Paisey kay liey maan bhi bhaich day* ('also sells mother for money') in (115) and *purrriii mannn bharrr ke marna* ('beat to their heart's content') in (116). Reiteration involved repeating longer stretches such as sentences as well as shorter expressions such as single words and phrases, like in (117). *Mard ka bacha* literally means 'child/son of a man' denoting "a real, masculine man" which in this particular context is treated as synonymous to "stud".

(115) Naji is a known mother seller for money. **Paisey kay liey maan bhi bhaich day.** He wrote MQM walay apnay jalsay mein hazreen ki tadad per israr nahi kartey. Is that really so?. Because I have seen some MQM tattos claiming lakhoon ka thathen marta sumanadar. [Also sells mother for money.] (PB)

(116) They can heat &; beat each other.. as much as they want, **purrriii mannn bharrr ke marna**: [beat to their heart's content] (IG)

(117) all the besharmi, beghairti is laid at the feet of the women while the men who visit prostitutes are called' studs',**' mard kay bachay**' while the prostitutes are considered the scum of society. [children/sons of a man] (PG)

In (118) the codeswitch can be analysed as having a clarificatory function. Here the writer explains or clarifies what they mean by "bad intentions" or "seeing a girl with bad intentions" with the Hindi/Urdu expression *kisi ladki ko nazar utha ke nahi dekha*. The Hindi/Urdu expression translates into 'not lifting his gaze to look at a girl', which is a somewhat euphemistic way of insinuating "not to look at a woman in any sexual way/with sexual intentions" in Hindi/Urdu. This instance of reiteration can be regarded as somewhat indirect repetition where the preceding phrase is repeated in a somewhat modified form in the other language. In (119), "flaxseeds" is repeated in Hindi/Urdu by providing its Hindi/Urdu equivalent as a clarification to its meaning for those who are not familiar with the English word.

(118) he is not that type of person... even once he had said whole kannauj know guddu shukla has not seen any girl with bad intentions (**kisi ladki ko nazar utha ke nahi dekha**) [does not lift his gaze to look at a girl] (IG)

(119) While you may nibble on a fist full of walnuts and almonds, try staying away from too much cashew nuts and pistachios. You could also include flaxseeds (**alsi ke beej**) in your cooking. They come from a fibre crop and aids digestion. [flaxseeds] (IG)

Reiterations were also translations conveying the meaning of the English sentence or stretch in Hindi/Urdu. In (120), the Hindi/Urdu sentence placed in brackets is an instance of translation of the preceding English sentence. Some reiterations that would appear to be translations at a first glance were instances where the writer had seemingly provided the original utterance in brackets after quoting it first in English. Thus, the sequence in brackets is not so much of a translation but an illustration of the exact words of the person saying the original utterance. It is actually the English quote that is a translation or interpretation of the original utterance. In (121) and (122) the codeswitches to Hindi/Urdu in brackets are likely to be the original utterances provided after the English quotations.

(120) The Hindus in Pakistan were treated very nicely when they were migrating as opposed to the inhuman treatment meted out to the Muslim migrants from India. (**Musalmanon nein Pakistan se janay walay Hinduon ko her qissam ki sahulatein deen, lekin Baharat ke logon nein Musalmnon per bohat Zulm kiyay**). (PB)

(121) " One should think about one's own situation (**Bihar par tippani karne walon ko apne halaat ke bare me sochana chahiye**) ", he had told reporters in reply to a question when asked for his reaction to Modi's remarks targeting Bihar leaders [People who comment on Bihar should think about their own situation] (IG)

(122) Manek wryly maintains that a good love story should be uplifting to society (**" acchi prem kahaani samaaj ke liye kalyaankari honi chahiye "**) and that stories like Devdas are " sentimental junk " because they lack a " moral ", but his own actions in his narratives are less than edifying; he portrays himself as limp-wristed, responsibility-shirking and cowardly (IG)

In (123) and (124) the Hindi/Urdu sequences are direct repetitions of the preceding English sentence; that of "Last year she used to go to his house" in (123) and of "If you are my friend, you will come with me" in (124). Here the function of the Hindi/Urdu sentences overall is to illustrate and exemplify Hindi/Urdu grammatical rules and linguistic structures in practice with Hindi/Urdu sentences containing the structures in question. The English sentence, that the Hindi/Urdu sentence is a repetition of, is actually the translation or the explanation of the Hindi/Urdu sentence – just preceding the Hindi/Urdu sentence; the order of the Hindi/Urdu sentence and its English translation or explanation is in a way reverse.

(123) My old friend and former colleague in the language program at Columbia, David Rubin, coined the perfect name for such cases: "ghostpositions," Thus for example consider "Last year she used to go to his house": **pichhle saal vuh us ke ghar jaatii thii**. Here pichhlaa saal, "last year," and "his/her/its house," us kaa ghar, have each gone oblique before an invisible but effective ko. The ko is a "ghostposition" because its hovering presence is established by its obliquifying effect (and its grammatical necessity). (IG)

(124) If you are my friend, you will come with me. **Agar ap mere dost hai to mujh ke sath aunge** (same mistake: use' mere' in place of' mujh' esp. for' with me') (PG)

6.1.6 Metalinguistic commentary

Hindi/Urdu was used in explanation of and discussion about Hindi/Urdu grammar. In this kind of a context, switching of language happens naturally when the meaning and use of words and structures are discussed and exemplified with examples from the other language. For example, in (125) compound postpositions and their typical forms are illustrated through switches to Hindi/Urdu for the genitive postposition *ke* and compound postpositions *ke baad* and *ke uupar*. In (126), discussion about possessive forms and the use of genitive constructions in a sentence involves switching for Hindi/Urdu genitive postpositions *ka/ke/ki* as well as for a few Hindi/Urdu sentences, which are provided as examples of the use of possessives in a sentence. As a side note on the potentially peculiar looking forms like *ba ((d* or *be; Tiyaa; N*, these are apparently romanization conventions that are used to romanize Urdu in Urdu studies.

(125) All the rest are compound, mostly with **ke**, like **ke ba ((d** (after) and **ke uupar** (above). When you encounter a promising new one, cherish it: give it a special place in your notebook and your mind. (IG)

(126) Possession of humans, etc., with **kaa / ke / kii:** The way one " has " relatives, children, friends, loved ones, etc., is with the possessives **kaa / ke / kii**. Usually the resulting constructions are quite straightforward: **un kaa ek potaa hai** (" of them is one grandson "), **us ke do be; Te the** (" of him/her were two sons "), **us kii tiin be; Tiyaa; N hai; N** (" of him/her are three daughters ") (IG)

6.1.7 Other communicative functions

Hindi/Urdu served a variety of other communicative functions as well. Among the other, more marginal functions were for example switching to convey personal opinion, side comments, word play, and insults. In (127)-(130), Hindi/Urdu was used to convey personal statements such as personal opinion, point of view, or personal impression of something. All of these statements are explicitly identified as the writer's opinion by the beginning phrases *meray khayal say*/*mere khyaal mein* ('in my opinion'), *mujhe to lagta hai* ('it seems to me') and *I think*. Switching to Hindi/Urdu for personalized statements relates to Gumperz's concepts personalization vs. objectivization and how CS is used to create contrast between personal statements and more objective descriptions on the one hand, and to the notion that the ethnically specific we-code is used to convey the more personal statements on the other hand.

(127) If we at this point of time start mud slinging then what will happen when you will be running the country in future. " Unity is Discipline ", " Unity is strength ", I don't know how many times we have been taught this in our schools but we have never learnt that by heart. **Meray khayal say Pakistan ko aagay barhaanay kay liyay Unity sab say important cheez hai.** Please guys this is our Pakistan and it takes a hell lot of effort to form a country, don't just let it get away like that. [In my opinion, for Pakistan to grow forward unity is the most important thing.] (PG)

(128) Dr Rizvi has dedicated the whole of his life to serving humanity without thinking of reward. That's the way he is; he can't help himself. I think **hamari bakhshish un ki waja se ho gi** (we will be accepted in the afterlife because of Dr Rizvi). (PG)

(129) My talks are not sensible to you but some find it fine. I don't mind it may be subjective. **Mujhe to lagta hai is dushman-e-iman ke bagair tumhara dil nahi lagta.** Love to you my friend By mohd yunus – [it seems to me that you cannot be without the enemy of faith] (IG)

(130) **Mere khyaal mein taaj mahal duniya ke ajaibon mein se aik ajaib hai...** It is certainly the most magnificent example of indo-muslim architecture in the world and I would love to visit it oneday. [In my opinion, Taj Mahal is the most wonderful of the world's wonders.] (PG)

In (131), the codeswitch serves the function of addressee specification. In switching for addressee specification, the switch of language is on the one hand triggered by the change of

addressee and on the other hand the change of the recipient of the message is signalled through the change of language. The codeswitch to Hindi/Urdu occurs here when the writer addresses and directs the message to a particular person - in this case Saeed Anwar, a Pakistani Islamic scholar.

(131) I agree nobody is farishta in this world but we need honest man like quid azam and in the present situation only I can seen those qualities in Imran khan.... **Saeed anwar bhai... tum to bayan sun kar khana kha kar so jate ho jabke Imran din raat mehnat kar ke ghareeb logon ke khidmat karta hey... to tum batao koun acha he tum ya Imran... sure Imran.... lanat ho tum jase moulvioun par** [Saeed Anwar bhai…. after listening to a sermon, you eat and go to bed while Imran works hard day and night serving poor people. You tell me who the good one is, you or Imran. Surely Imran. Curse on maulvis like you] (PG)

Hindi/Urdu was used to address the reader more directly and personally and to signal directing the message to the reader. In (132), the writer discusses Pakistani politicians and their own opinion on Pakistani politicians in English and switches to Hindi/Urdu to address the reader while in (133), the writer is first explaining the state of the knowledge of and attitudes towards Urdu among Muslims in India and then shifts to addressing the reader more personally in the codeswitched sentence. In both instances, the change in the position that the writer takes towards the reader is accompanied by a switch of code to Hindi/Urdu. In these two instances, it appears that the writer is trying to appeal to the reader and trying to convince the reader of the validity of their argument in the codeswitched sentence. With both of these instances involving the use of imperatives (*ap moje yeh bataen* and *dil par haath rakhiye aur bataiye*), the codeswitched part could be interpreted as if the person was insinuating "think yourself" and trying to appeal to the reader's common sense. However, one cannot be entirely sure whether the person being addressed is one particular person or the wider audience in general. Both of these two texts appear to be user comments, but the corpus does not reveal whether they are so to speak separate comments or whether these are replies to another comment, which would potentially reveal who exactly the addressee is.

(132) nawaz sharif and zardari both courpt and chor imran khan is one who is not courpt. i m not in pti but i should to say there is only one hope of pakistan its imran khan he is clean i like this program very hard qustion but imran khan proved that he is leader. **ap moje yeh bataen ke phele kon sa leader awam ke samne a kar beta hai or un ke qustion ka jawab deya ho** imran do that pakistan zindabad. [Now tell me which leader has previously come and sat in front of the public and has answered their questions.] (PG)

(133) Muslim Living in North india, does not read or speak URDU as we perceived or wish. Once upon a time it was the case. Today we do not want our children to study urdu. **Dil par haath rakhiye aur bataiye ki India ke kis kone me Indian Muslims ka**

middle class apne bachhe ko Urdu padha raha hai. So if we want to communicate with Indian Muslims and other communities, we have to communicate in their language. [Put your hand on your heart and tell in which corner of India the middle class of Indian Muslims teach their children Urdu.] (IG)

In (134), the codeswitch to Urdu is an attempt to appeal to the reader, and in this instance more specifically to other Muslims. The effect of the appeal is reinforced through the switch of language to Urdu, the language of Muslims and Islam. The language choice can be interpreted as serving as a reference to the common Muslim background and the expected shared Islamic values, which make the appeal of the writer more forceful giving it more authority through the reference to religion. The quoting of the words of the Prophet *("Ek Musalman Dosre Musalman ki Dhaal Hai"*, 'a Muslim is other Muslim's shield') can also be analysed as an attempt by the writer to gain more authority for their statements via referencing to and quoting a religious source.

(134) Please stop fighting our beloved Prophet Mohammed (PBUH) said "Ek Musalman Dosre Musalman ki Dhaal Hai" so please **ek dosre ke hamdard bano, dushman nahi, aur sahi mai Musalman bano, ek dosrey se hamdardi rakho, dil saaf rakho, niyat aur sounch achi rakho... Insha'Allah Allah hum sub ko kamiyab karega... Allah hum sub ko Deen ka samjh ne wala banaye... Ameen...** [become compassionate towards each other, not enemies, and become truly Muslim, have compassion for each other, keep heart clean, keep intentions and thoughts clean. Inshallah Allah will make us all successful. May Allah make us people who understand religion] (IG)

Hindi/Urdu was used in parenthetical remarks and side comments. In the side remark in (135) the writer shifts from narrating the course of events in English to commenting on it, in this case commenting on why they decided to go to Café Mondegar. Example (136) is an objective, factual account of the mathematical system of the Quran in English involving a Hindi/Urdu switch in the writer's own remark placed in parenthesis. In the parenthetical remark the writer shifts from more objectively explaining the mathematical system to pointing out the importance of the number 19.

(135) As it's out of my budget, Heena will be gifting me one on my birthday! Then went to Cafe Mondegar (**pareshan kar ke rakha tha Heena ne ek mahine se, ki** I want to go to Cafe Mondegar). It was a great place and we had a great time eating pizza, Jain cheesetoast and a bottle of beer. [Heena had been complaining for one month that...] (IG)

(136) There are two major facets of the Quran's mathematical system: (1) The mathematical literary composition, and (2) The mathematical structure involving the numbers of suras and verses. Because of this comprehensive mathematical coding, the slightest distortion of the Quran's text or physical arrangement is immediately exposed. # **(yaad rahe ke 19 ke numeric bohat aham hai)** SIMPLE TO UNDERSTAND

IMPOSSIBLE TO IMITATE # For the first time in history we have a scripture with built-in proof of divine authorship -- a superhuman mathematical composition. [remember that the number 19 is very important] (PB)

Hindi/Urdu was also used in insults and statements that were meant to be offensive, as in (137) and (138). In (137), the Hindi/Urdu insult is part of a context where the purpose of the whole comment is to diss and insult Pakistanis. Here, it is not merely Hindi/Urdu that is used to insult. However, in this instance, there appears to be a slight difference in the offences conveyed in English as opposed to the offence expressed in Hindi/Urdu; the offences expressed in English are more implicit and indirect mocking with a somewhat sarcastic tone, whereas the Hindi/Urdu insult is a more direct attack and more overtly offensive and abusive – involving for example the pejorative *suvar ke bacho* ('children of a pig'). The Hindi/Urdu insults could also be singly occurring pejorative terms like *rundee ke bachoo* ('children of a whore') in (138).

(137) you paki's don't have balls so why talk abt us, for us any blast in the country is new but for u it's usual news go ask in u r country itself people say india is moving in a right way... and by the way mughal empire did not belong to u they came from afghan and c it's state now hahaha.. i heard pak is famous for porn what a go guys u r atleast famous for that keep it up... **apne maa, bahen, patni ko bejte raho suvvar ke bachho** [Keep seeding your mother, sister, wife, you children of a pig] (PB)

(138) You need to grow up freaks. You guise have been repeatdly raped by US and thats a fact. NaPak army is only good to kill their own people. I was hoping to get an answer on the points I raised instead I get personal abuse. Its easy to label people with Indian Hindu agents, its norm in Pakistan. Now both of you crawl back to same hole you came from **rundee ke bachoo**. [children of a whore] (PB)

In (139), Hindi/Urdu is used to convey a word play that would not work in English. In (139), the word play involves the Hindi/Urdu expression *ullu ka patha* (lit. 'son of an owl') – or in plural *ullu ke pathe* (lit. 'sons of an owl') – which means a fool or an idiot. As an expression that would be an equivalent of *ullu ka patha* does not exist in English, the word play could not be conveyed in English. Therefore, the writer has preserved the Hindi/Urdu expression to convey the word play in their narrative of a past conversation that the word play was part of.

(139) The story of how I quit my second film company is similar. It was proposed that we name a new film **Ullu ke do pathe**. I objected, as the title should have been **Do ullu ke pathe**. But it stuck and they said to me, " Who're you anyway? We're paying for it so if we want it call it **Pathe ke do ullu** then we'll go ahead and call it that. " So the shooting began and I quit. [Two sons of an owl; Sons of two owls; Two owls of a son] (IG)

As was discussed above, reiterations can function to elaborate on or to clarify a message that the person is trying to convey, or they can serve to add emphasis on a statement. However,

reiterations are not the only way to accomplish clarificatory or emphatic effect. Codeswitches can serve purposes of clarification, elaboration, or emphasis without being reiterations as these functions can be achieved through other means as well. In (140), the person elaborates on why they think of Farman as a hypocrite by describing him and his actions in the codeswitch. In (141), the codeswitch can be analysed as putting emphasis on the previous words of the writer.

(140) I totaly agree with you Azk, Farman is an hypocrate, **bivi kay saamnay kuttay kee tarah dum hilata hay aur peechay chup ker shaadi ker lee... is say bara dhoka kia ho ga jo us nay zubaida ko dia?** I am not supporting zubaida's behaviour at all but if Farman was not happy with his wife, he should have told her that he is not happy and he wants to marry another woman... [wagging his tail like a dog in front of the wife but goes and marries behind back. It's such a big betrayal he did to Zubaida.] (PB)

(141) The poison and racist malaise within you is evident for all to see. The hatred you harbour for Pakistan will turn round and haunt you one day, mark my words. **Hameyn tbah karne ki koshish main khud mit jao ge.....** [if you try to destroy us you will perish yourself] (PG)

There were also Hindi/Urdu codeswitches without any apparent communicative function. Not always can a codeswitch be identified as having a clear function or a motivation, and on the other hand, not all codeswitches even have a specific communicative intent or are communicatively purposeful. For example, in (142)-(144) no reason that would explain why the switch to Hindi/Urdu occurred could be deduced either by the switch itself or its context.

(142) it's time wives starting divorcing the husbands who cheat on them.**shor dalnay kay baad** if you still stick with him it just shows him that you are totally dependent on his income and you'll go nowhere no matter what he does. [after making noise] (PG)

(143) as he realises the extent of the damage he and the family members have done Anjali by keeping her cocooned. Khushi helps him sort out Anjali, by getting out of her **Di Ke Liye** mode and helping Di face facts, rather than hiding them. [for Di] (IG)

(144) Exposure-wise, in every way, the Indian film industry is far better than the Pakistani film industry. But I hope and pray that the Pak industry gets the same prestige **jo ke aaj se 30 saal pehle thi**, [that it had 30 years ago] (PB)

6.2 Cultural codeswitches

Cultural switches is the other broad category that could be distinguished in Hindi/Urdu CS. Hindi/Urdu was used to express cultural objects and concepts that are part of and specific to the Indian or Pakistani culture or the meaning and connotations that they evoke are otherwise heavily culture-bound. In addition to cultural objects and concepts, the Hindi/Urdu expressions also served as references signalling the relatedness of something to the Indian or Pakistani

culture and eliciting the interpretation of something in the context of the Indian or Pakistani culture.

Cultural switches included, among others, titles of movies or songs, food, concepts related to religion, and names of celebrations and festivals. In the following subsections, the culture-related switches will be first discussed in relation to the sociocultural domains that they represented and then with regard to the functions of and the motivations for the use of the culture-related switches.

6.2.1 Sociocultural domains of cultural switches

Several of the cultural switches were categorizable into sociocultural domains. Table 5 shows cultural switches categorized according to the domains they belonged to. The lists in the table are not meant to be exhaustive containing all the items belonging to each category in the data. Some of the categories, especially titles of movies, songs, and tv-shows as well as cuisine were so large in the data that it would be impossible to include all the items of those categories in this table. The purpose of the table and the listed items is rather to provide a general idea of and to illustrate the types of Hindi/Urdu switches in different cultural domains.

Table 5 Sociocultural domains of cultural switches

Titles of movies, songs, tv-shows etc.	**Indian/Pakistani cuisine**	**Religion**	**Kinship**	**Marriage and wedding**
Aahat, Kahaani, Bol, Ek Tha Tiger, Pyaar ke Side Effects, Mrs Kausik Ke Paanch Bahuein, Zindagi Na Milegi Dobara, Sheila ki Jawani, Aap ki Adalat Lakhon mein Ek	naan roti ghee lassi kulfi dosa murgh cholay aloo ke parathey biryani Punjabi pulao laban ke kofte raita	dua namaz azaan fatwa jannat ummah sunnah qiyamah puja bhajan rakah	khala mamu beta saas daadi ammi didi bhai chacha	rishta nikah valima dulhe raja shadi ke din jahaiz moon dikhai dulha ki salami
Indian/Pakistani music and poetry	**Celebrations and festivals**	**Clothing**	**Forms of address**	**Miscellaneous**
raga qawwali ghazal nazm mushaira sur dholki swarmandal	Diwali Deepawali Eid Dusshera Navaratri	sherwani saree shalwar kameez	ji bhai sahib yaar	aate ke diye bindi agarbatti gori

As Table 5 shows, the cultural switches were categorizable into the following sociocultural domains: titles of movies, songs, and tv-shows, Indian/Pakistani cuisine, religion, kinship, marriage and wedding, Indian/Pakistani music and poetry, celebrations and festivals, clothing and forms of address. The category "miscellaneous" consists of a miscellaneous group of culture-related objects and concepts that did not belong to any of the distinguished nine domains. The meanings of words and expressions in the table will be explained in the following discussion on the motivations for the use of cultural switches.

6.2.2 Motivations for culture-related switches

The functions of and motivations for the use of cultural switches were primarily of two types; either they were motivated by a lexical gap between English and Hindi/Urdu and the absence of a valid equivalent, or they functioned as references to the Indian and Pakistani culture. As cultural references, the Hindi/Urdu codeswitches implied the Indianness or Pakistaniness of either an object, a concept, or a person evoking cultural or culture-specific connotations and encouraging the interpretation of something in the context of the Indian or Pakistani culture. The lexical motivations and Hindi/Urdu codeswitches functioning as cultural references are discussed in the following subsections.

6.2.2.1 Lexical motivations

There were several occasions where the Hindi/Urdu item or expression either completely lacked an English equivalent or did not have a direct English equivalent. In the absence of an English equivalent, the Hindi/Urdu item was needed to fill a lexical gap. Such was the case particularly with objects and concepts specific to the Indian and Pakistani culture. These were for example names of festivals and celebrations, like the Hindu festivals *Diwali* and *Dusshera*, or the Islamic celebration *Eid*, traditional Indian and Pakistani garments like *saree* or *shalwar kameez* and names of food items and dishes. For example, *roti* and *naan* are both types of flat round bread typical of both India and Pakistan. *Parathey* are also a kind of flat round bread somewhat similar to roti and naan, and *aloo ke parathey* are parathas filled with mashed potatoes (*aloo* 'potato'). Dishes included *biryani* and (*Punjabi*) *pulao*, which are both different types of rice dishes, and *murgh cholay*, which is a dish made of chicken and chickpeas. *Lassi* is a yoghurt-based drink made by mixing water, yoghurt, and usually sugar. Although some of the names of dishes could basically also be expressed in English, the concept, however, does not exist in English as such. Some of the religion-related terms, like *rakah,* which denotes movements that are conducted

during prayers, and *fatwa*, which is a term related to Islamic law, also lack English equivalents. Hindi/Urdu words related to Indian and Pakistani music and poetry, like *raga* ('a genre of India classical music'), *qawwali* ('a type of Islamic devotional religious music'), or *ghazal* ('a genre of Urdu poetry') or the musical instruments *dholki* and *swarmandal* do not have equivalent expressions in English either. Cultural objects filling lexical gaps were also for example *agarbatti* ('incense'), *bindi* ('a coloured dot on the forehead') and *aate ke diye* ('a type of an oil lamp made of wheat flour') – just to give a few more examples. Titles of movies, songs, books, tv-shows etc. can also be counted as Hindi/Urdu items lacking English equivalents.

The lack of an exact English equivalent can also have acted as a motivation for the preference of the Hindi/Urdu item over the English one. The Hindi/Urdu item can sometimes be semantically more specific conveying a more exact meaning than the English alternative. The availability of the more exact Hindi/Urdu term or expression can potentially motivate the choice of the Hindi/Urdu item. The closest English equivalent would fail to reach the same specificity and exactness of expression, and in order to convey as specific a meaning, it would require further explanation. Therefore, the Hindi/Urdu expressions are preferred because they are simpler and more economical to use.

Examples of Hindi/Urdu items that lack a direct English equivalent are, among others, *nikah*, *shadi* and *valima* which are all names for different parts of an Islamic wedding. *Nikah* is the part of the wedding where the marriage contract is signed, *shadi* is the main wedding reception dinner arranged by the bride's family, and *valima* is the wedding party usually organized by the husband's side of the family. *Shadi*, on the other hand, can also be used as an umbrella term to refer to the entire wedding encompassing all its different parts and ceremonies. Although all of these could basically be expressed in English as wedding, due to the different structure of wedding and different wedding traditions, these concepts do not have direct, convenient English equivalents. Another example of a concept lacking a direct English equivalent is *rishta* which indirectly corresponds to English "marriage proposal". In an arranged marriage system, *rishta* means a proposal for marriage from either the man's family or the woman's family, and thus it does not equal English concept of marriage proposal.

Another good example of an instance where the Hindi/Urdu words are more specific and more precise in meaning than their English alternatives are kinship terms. Hindi and Urdu have a more elaborate kinship vocabulary making more precise distinctions between family members and relatives like siblings, aunts, uncles, and cousins. The more elaborate vocabulary of Hindi and Urdu in the domain of kinship can have acted as a partial motivation for the preference of the Hindi/Urdu items. The choice of the word for, for instance, uncle or aunt

depends on whether they are maternal or paternal aunts and uncles as well as on whether they are the parent's siblings or spouses of the parent's siblings. For example, *mamu* 'uncle' means mother's brother, while *chacha* 'uncle' is father's younger brother. *Khala* is the word for 'aunt' when referring to mother's sister and *daadi* the word for paternal grandmother. Further, it is customary – and polite – to address relatives using the kinship term or adding the kinship term after the first name of the relative. Thus, the kinship term is in a way part of the name of the relative. Other kinship terms in the data were for example, *ammi* ('mother'), *bhai* ('brother'), *didi* ('elder sister').

6.2.2.2 Hindi/Urdu switches as cultural references

Apart from lexical motivations, Hindi/Urdu switches were used as cultural references to evoke cultural connotations, express Indianness/Pakistaniness, and to add local flavour. Further, despite the potential lexical motivations, it is unlikely that even in instances of a lexical gap, lack of a suitable English equivalent, or greater semantic specificity of the Hindi/Urdu item the motivations for the use of the Hindi/Urdu terms would be purely linguistic; these Hindi/Urdu switches can also convey cultural signalling and cultural connotations. And, on the other hand, in instances where a valid English equivalent exists, the English item or expression can still fail to capture the same meaning or content as the Hindi/Urdu word or phrase.

Hindi/Urdu were used to express Indianness or Pakistaniness of something or someone, like the switches *chai* in (145), *dulhe raaja* in (146), and *shaadi* in (147). Whereas *bhaji* is the name of an Indian savoury snack, *chai* is the Hindi/Urdu word for tea which in this instance can be interpreted as referring to tea that has been prepared in Indian style. In example (146), the Hindi/Urdu word for groom, *dulhe raaja*, which is here used in the context of an Indian wedding, serves as a specific reference to an Indian bridegroom in particular. In (147), *shaadi* ('wedding') brings about the idea of an Indian-style wedding. In (148), *jahaiz* is used for 'dowry' in the context of Pakistani/Indian wedding as a reference to the local wedding traditions. (148) additionally contains *dulha ki salami* and *moon dikhai*, which are the traditional wedding gifts given at the Indian/Pakistani wedding. *Dulhe ki salami* is a gift, for example money, given by the bride's family and relatives to the groom, while *moon dikhai* is a valuable gift – often jewellery – from the groom to the bride. It is relatively unlikely that the writers would have used these Hindi/Urdu words to refer to tea, bridegroom, wedding and dowry in any other occasion than when referring to them in the Indian/Pakistani context. These would most likely not have been used as neutral terms to refer to for example western wedding

or bridegroom, or to refer to these concepts in general. The choice of the Hindi/Urdu expressions highlights their Indianness/Pakistaniness.

(145) Since you have admitted to being a diehard romantic what's your definition of romance? Mood ke hisaab se. Since it's monsoon now, my idea of romance would be having **chai** and **bhajiya** on the street lined with trees and listening to Shubha Mudgal's Ab ke sawan.. [tea] (IG)

(146) Finally Shadi Ke Din Invitations clearly mention that nikaah is at 7 pm but **dulhe raaja** arrives at 8:30, 9:00. 9:30. Soon after nikah, people are desperately waiting for some one to announce " aaiye "?? (call for dinner) because on his way to the function hall he will stop at a cool drink store. [bridegroom] (IG)

(147) There are elements of your vision in Rowdy Rathore. The **shaadi** sequence reminded me of' Aankhon Ki Gustakhiyan' from Hum... Dil De Chuke Sanam? No. I was no part of the creative process of Rowdy Rathore. [wedding] (IG)

(148) material assets such as the size of the house, the number and brands of their cars, the price tag of the watch given as dulha ki salami, the value of the jewel given as moon dikhai, the brand name and number of crockery pieces in the dinner set of the **jahaiz**, and the number of meats at the wedding reception of their adult offspring -- to just name a few. [dowry] (PG)

Hindi/Urdu terms related to religion and religious concepts can also be interpreted as cultural references. While some of the religious terms do not have valid equivalents in English, others would basically have available English alternatives which, however, are likely not used because they would not convey the same content and same associations with Islam or Hinduism. For example, *haraam* not only means ‘prohibited’ or ‘forbidden’, but it coveys that something is prohibited by Islam. *Dozakh ki aag* and *jannat* are religious concepts translating into English as ‘purgatory’ and as ‘heaven’ or ‘paradise’, respectively. Despite the potential English expressions, the Urdu terms can be thought of as having stronger associations with Islam and therefore preferred.

Hindi/Urdu expressions were also used as references to local people to further add to their characterization as local. In (149), the Hindi/Urdu phrase *karachi kay larkay* (‘karachi boys’) is used to denote men from the Pakistani city Karachi, and the adjective *punjoo* is used in the phrase *punjoo girls* for women from the Punjab region. This example also contains a third switch that can be analysed as cultural: the word *gori* in “gori skin”. *Gori* means ‘fair’ or ‘white’ and is used especially as a reference to fair or white skin tone, which is the local beauty ideal in India and Pakistan. Therefore, it can also be interpreted as having cultural implications.

(149) punjoo aunties always say they are looking for urdu-speaking boys for their daughters because they're tameezdar. as for **karachi kay larkay** choosing **punjoo** girls, i think there's just one major reason: gori skin. [Karachi boys; Punjabi] (PG)

Some of the Hindi/Urdu codeswitches were expressions that evoke cultural stereotypes. In (150) , *ache ghar ki larki* ('a girl from a good home') denotes the Pakistani stereotype of an ideal, good, well-raised girl who, among other things, does not smoke or drink, does not hang out with friends late, and does not have affairs with men etc. *Ache ghar ki larki* is here juxtaposed with *awara larki* ('wandering girl') which refers to a woman displaying completely opposite characteristics to those of an *ache ghar ki larki*, which are considered highly condemnable traits for a woman to possess in the Pakistani culture. An *awara larki* is a woman with a low, condemnable character, and the expression raises highly negative connotations. *Ache ghar ane ki* ('coming from a good home') in this context refers to an *ache ghar ki larki* meaning the same thing.

(150) that **achy ghar ki larki** lecture was in fact the point of the article which u hypocrytically declared as non existant read ur comment again. Even in villages only elderly women indulge in Hukka n stuff young girls are not allowed there as well! These stigmas that you say ET is crazy enough to allow are further substantiated by your own comments! Stereotyping ppl as you say " **awara larky** " and gender biasing " **Achy gharany ki** or not " are plain WRONG! Every one as an adult shud hav the right to choose [a girl from a good home; a wandering girl; coming from a good home] (PG)

The Urdu word for mother-in-law, *saas*, evokes the connotations of a stereotypical Pakistani mother-in-law who is overly critical and controlling of their daughter-in-law and seeks to exercise utmost control and dominance over their daughter-in-law. It creates a contrast between mother-in-law (*saas*) and daughter-in-law (*bahu*), which is further enforced by the use of the word *bahu* in the previous sentence, bringing about the culturally stereotypical rivalries between these two family members.

(151) More than men, saasu maas are obsessed with getting a chaand si bahu for their sons so that the projeny is also beautiful. Imagine the **saas**-to-be coming for a rishta and saying " zara larki ko inspection ke liye bhej dein ". [mother-in-law] (PG)

7 Discussion

In this section, the results and findings of the preceding structural and functional analyses will be discussed further. In the discussion on the forms of the Hindi/Urdu codeswitches, the results of the structural analysis and their implications will be discussed in 7.1 together with a few comments on the potential impact of the method on the results and the generalizability of the results. In discussion on the functions of the codeswitches in 7.2, the findings of the functional analysis will be viewed from a broader perspective and links will be drawn between different functions to create a bigger and more general picture of the tendencies that arose in the analysis. In 7.3, I will make a few general comments and observations on the Hindi/Urdu codeswitches which are unrelated to the research questions as such, but which are, however, points worth mentioning. In 7.4, I will further evaluate the method of data collection and discuss the impact of the method on the results in more depth.

7.1 Structural patterns

In the analysis of the CS forms, the aim was to investigate what types of structural patterns CS manifests in and what different forms Hindi/Urdu CS takes in IE and PE. Considering IE and PE together further allowed the examination of whether the structural patterns follow similar tendencies in these two contexts where the languages involved are in the same typological relation with each other (English-Hindi, English-Urdu) or whether the structural patterns somehow differ.

As a general remark on the commonness of CS as a phenomenon, it could be stated that the practice of Hindi/Urdu CS itself appears to be equally common in both IE and PE. Overall, both IE and PE displayed a variety of structurally different codeswitches ranging from intersentential switches involving sequences of several consecutive Hindi/Urdu sentences to intraclausal, single-word switches. The structural analysis of the Hindi/Urdu codeswitches showed that the switches followed largely similar tendencies in IE and PE. The similar trends could be observed in the share of intersentential, intrasentential, interclausal as well as intraclausal switches in both IE and PE. However, IE had somewhat higher percentages of intrasentential and intraclausal switches, whereas PE had more intersentential and interclausal switches. Language typological aspects can be ruled out as potential explanations for these differences in the share of different structures, as Hindi and Urdu are typologically in the same relation to English. A possibility that could explain the differences in the percentages of the structures would be relatedness of certain structural forms to certain text types the balance

between which would be different in the Indian subcorpus and the Pakistani subcorpus. This possibility remains, however, purely speculative here.

Which structural patterns and categories appeared and were prevalent in the structural analysis was affected by the method and the search items to a certain extent. The fact that all the three most frequent Hindi/Urdu words that were used to conduct the searches were grammatical words – all postpositions – had impact on the types of switches that were located with them. For example, the searches with the genitive postpositions explain the prevalence of genitive phrases among intrasentential switches. The impact of the method is discussed in more detail in 7.4. Because of the impact of the search method on the appearance and prevalence of certain structural patterns in the data on the one hand and the marginal appearance of some other structural patterns on the other hand, no generalizations concerning the relative frequencies of structural patterns can be made on the basis of these results; the data do not allow the investigation of, for example, whether single-word switches and other shorter switches are more common than, for example, sentences. However, the generalization that can be made is that Hindi/Urdu codeswitches in IE and PE appear to follow similar structural patterns – with IE having somewhat more of intrasentential and intraclausal switches and PE having more of intersentential and interclausal switches.

In both IE and PE, intersentential uses of Hindi/Urdu included switches for complete Hindi/Urdu sentences, coordinated independent clauses, as well as longer Hindi/Urdu stretches. The category of interclausal switches contained a variety of Hindi/Urdu subordinate clauses and other embedded Hindi/Urdu clauses in different syntactic positions in both IE and PE. The subcategorization of intraclausal switches according to word class and phrase type revealed parallel patterns in the share of the three largest word class and phrase type categories in IE and PE: proper names, genitive phrases, and nouns. Although the numbers and percentages of the smaller word class and phrase type categories were so minimal that no generalizations can be made based on them, it can be said that IE and PE follow the same tendencies in share of the three largest categories of the intraclausal switches.

Single Hindi/Urdu nouns emerging as the third largest category of the intraclausal switches is interesting for the reason that none of the nouns was located with a frequency list word directly but became included in the data as additional material. While it was just stated that the data and the results do not enable generalizations to be made or conclusions to be drawn concerning the relative frequencies of the different structures, the fact that single nouns were the third largest category of intraclausal switches with all of them being collected as additional switches can, on the other hand, be interpreted as a hint of their relatively frequent overall

occurrence and them being more frequently switched than other word classes, for example adjectives.

7.2 Functions

In the analysis of the functions, the purpose was to find out what types of functions the codeswitches to Hindi/Urdu serve and what motivations possibly lie behind the Hindi/Urdu switches. The functions were analysed qualitatively. The large data provided a representative and varied picture of the functions Hindi/Urdu codeswitches serve in IE and PE. In terms of functions, the Hindi/Urdu codeswitches could be categorized broadly into switches with a communicative function and switches related to culture. Switches with a communicative function were codeswitches that could be identified as serving discourse functions such as quotation, reiteration, conveying greetings/well-wishes/blessings/prayers, figurative language, side comments, word play, insulting etc. Cultural switches were switches for objects and concepts related to the Indian and Pakistani culture and switches serving as cultural references.

This division into switches with a communicative function and cultural switches is not meant to suggest that cultural switches would be devoid of communicative intent or that switches with communicative functions would be devoid of cultural implications. Cultural switches also have referential function that serves as a cue for the interpretation of the switch in the context of Indian/Pakistani culture. Switches determined to have a communicative function are not detached from the culture or devoid of cultural motivations either. For example, Hindi/Urdu figurative language and idioms can also be seen as culture-specific expressions and their use can be motivated by the absence of an English equivalent. Further, the local, communicative functions could sometimes have implications to and draw part of their communicative function from the broader cultural context and be in that sense culture-related as well. An example of such instances is the use of Urdu in conveying blessings and prayers; the use of Urdu in these religion-related switches is also linked to the associations of Urdu with Islam and Muslim identity, which puts these switches into a broader cultural context making these switches also culture-related.

Some of the switches with a communicative function can be interpreted as deriving part of their communicative function from the macro-level functions of and associations with Hindi/Urdu bringing another dimension to the functional analysis of the switches and illustrating the multifunctionality of codeswitches. For example, while the act of switching code to quote can serve as a means of simply framing a quote without the language of the switch itself having greater significance, in instances where the voices of local people are represented

in Hindi/Urdu, the local language can actually serve as an important tool of characterizing the people as local Indian and Pakistani. Quoting family members and friends can also be interpreted against the backdrop of diglossia. As was discussed in section 2.2.3, languages can be observed to be in a diglossic relationship in Hindi-English bilinguals' lives in India. While English is the language of work, education and other more formal contexts, Hindi and other local languages are the languages used in more informal interactions with family members and friends. Quoting family members and friends and reproducing family dialogues in Hindi/Urdu can be interpreted as relating to the habitual language of communication in inner circles and can thus be seen as a demonstration of how macro-level functions of codes can become part of the functions that the codes serve on micro-level.

As a more general note on the themes and contexts of CS, religion as a theme surfaced on several occasions especially in PE – either in the switches themselves or in the more general contexts that the codeswitch was part of. The Urdu codeswitches could be either directly religion-related – like for example Urdu religious terminology, switching to convey blessings and prayers and quoting religious sources – or they occurred in otherwise religious, Islamic contexts. In addition to the local functions that Urdu codeswitches served in religious contexts, switching to Urdu in the contexts of religion can further be interpreted as having a more general, overall function of expressing Muslim identity. In India and Pakistan, Urdu is the language associated with Islam and Muslims and regarded as part of Muslim identity. Looking at the bigger picture of CS to Urdu in religious contexts, the use of Urdu can be thought of as a manifestation – or at least a partial one – of this link between the Urdu language and religion. It may also be possible that a religious context itself prompts switching to Urdu.

Cultural switches were switches for cultural objects and concepts particular to the Indian and Pakistani culture as well as Hindi/Urdu switches functioning as cultural references. Cultural switches often belonged to sociocultural domains such as Indian/Pakistani cuisine, clothing, festivals/celebrations, religion, kinship etc. Several of the cultural domains that that the Hindi/Urdu codeswitches represented correspond to the findings of previous studies and accounts on the Hindi/Urdu lexical items in IE and PE. As was already discussed in 3.5, the Hindi/Urdu items used in the local English have also previously been observed to come from domains such as food and drink, clothing, kinship, religion etc. The use of address forms *bhai*, *yaar*, *ji*, and *sahib* in PE has also been attested for example by Talaat and Anwar (2010). Although different studies have used different ways of categorizing the lexical items, the same themes manifest.

The borderline cases between borrowings and codeswitches could be found in the category of cultural switches. As a matter of fact, some of the items that were here classified as cultural switches are recognized as established borrowings in the *Oxford English Dictionary*. These were words for Indian/Pakistani cultural objects and concepts, for example food items and dishes such as *roti* and *ghee*, celebrations and festivals like *Diwali*, garments like *saree* and *shalwar kameez* as well as a miscellaneous group of other cultural items like *agarbatti* or *bindi* etc. Among the Hindi/Urdu borrowings in the *Oxford English Dictionary*, there are also some forms of address: *sahib*, *ji*, and *yaar*. As was explained in 3.4.1, the criterion used in this study to distinguish between borrowings and codeswitches was that if an English speaker, who does not possess any particular familiarity with or knowledge of the Indian and Pakistani culture or any Indian or Pakistani language, can be expected to understand the meaning of the Hindi/Urdu word, the item is regarded as a borrowing – if not, the item is thought of as a codeswitch and included in the analysis. This criterion has led to the inclusion of both restricted items, i.e. items the use of which is limited to the local variety of English, and assimilated items, i.e. items that have spread to other varieties of English. Even if the Hindi/Urdu word has spread to other varieties, it is no guarantee that it is understood by everyone and without knowledge of Indian/Pakistani culture. As was also explained in 3.4.1, the reason for not attempting any stricter distinction between codeswitches and borrowings in this study is that it is questionable how "un-English" the Hindi/Urdu words and expressions recognized as borrowings by *OED* would be from the perspective of a Hindi/Urdu speaker. For Hindi/Urdu speakers, the Hindi/Urdu items can be strongly associated with their own language and culture, and they would not be perceived as English at all. Therefore, a distinction between borrowings and codeswitches was not even regarded as meaningful.

7.3 General comments on the Hindi/Urdu codeswitches

As a general observation on the treatment of the Hindi/Urdu switches it can be noted that the Hindi/Urdu codeswitches were sometimes accompanied by English translations or other types of English explanations on the meaning of the Hindi/Urdu item or sequence. Writers either translated the Hindi/Urdu codeswitch directly into English or explained its meaning by other means such as paraphrasing or providing a more elaborate account on the meaning of the switch. This implies that the readership is not automatically expected to be Hindi/Urdu-speaking. The translations are a means of inclusion of non-Hindi/Urdu speakers and keeping the text or the message accessible and understandable to broader audience, not just to Hindi/Urdu speakers.

(152) Jundal said the common words after the attack came to an end were' **Mubarak ho! Hamne teen din tak Hindustan ki security forces ke saath ladai ki, jo wo hamesha yaad karenge** (we fought with Indian security forces for three days, which they are not going to forget) (IG)

Hindi/Urdu codeswitches were also sometimes marked as foreign language – or non-English – elements, i.e. flagged (Poplack 1988, 238). Flagging was both visual and verbal. Visual flagging involved the switch being framed and highlighted with the use of punctuation, for example quotation marks or brackets. In verbal, flagging the language of the switch was explicitly identified in the text. For example, in (153) the language of the quoted sequence "*Allah aap kay aibon pe parda dal day ga*" is identified as Urdu in the introductory clause "but in Urdu it goes as". In (154), the Hindi codeswitch is preceded by the phrase "the old hindi saying" which identifies the language as Hindi.

(153) "Allah is merciful, i am sure if at the moment you are honest and faithful with him, then i do nt know how to say it in english, but in urdu it goes as **"Allah aap kay aibon pe parda dal day ga".** [Allah covers your flaws with a curtain] (PG)

(154) The old hindi saying of' **Jis thaali mein khaata hai, usi mein he thookta hai'** came to our mind but we ignored him yet again. It was not worth it! [spit on the plate you eat from] (IG)

Although the focus of this study was on the use of Hindi/Urdu in English, an aspect worth addressing is English embedded in the Hindi/Urdu codeswitches. As could possibly be noticed in several examples provided in the analytical part of this study, many of the Hindi/Urdu codeswitches contained English elements embedded in the Hindi/Urdu frame. English elements being embedded in Hindi/Urdu was by no means a rare phenomenon. English could be found embedded in Hindi/Urdu sentences, clauses as well as phrases, and the English elements ranged from open-class items such as nouns, adjectives, and verbs to closed-class items such as conjunctions. There are even instances where the grammatical frame of a sentence is provided by Hindi/Urdu and all the content words in that sentence are supplied by English as in example (155). CS could also be observed in movie titles and song titles.

(155) why IK, s dad Musharf cases in these 14 months as he registered a case of plane hijacking in this time..... starnge..... very strange that **Musharf ne plane hijacking ka case register kiya but Corruption ka koi case register nia kiya inn 14 months mein**..... [Musharf registered a case of plane hijacking but did not register a case of corruption in those 14 months] (PG)

Although the surrounding English context can partly have had an impact on the appearance of English elements embedded in the Hindi/Urdu switches, I, however, argue that

the embedded English elements are primarily reflections of the commonness of the practice of switching to English in Hindi and Urdu. Hindi/Urdu-English CS is even the social, interactional norm of everyday communication among Hindi and Urdu speakers, and the people's tendency to use English as part of their Hindi/Urdu also shows in their CS to Hindi/Urdu in the local English.

7.4 Methodological considerations

This study also served as a methodological experiment where the data were collected using the most frequent Hindi and Urdu words to locate Hindi/Urdu codeswitches in GloWbE. The motivation for using Hindi/Urdu words with a high frequency in Hindi and Urdu was the assumption that these high-frequency Hindi/Urdu words would also occur frequently in Hindi/Urdu codeswitches. Overall, judging by the amount of data that was gathered, the method proved to be an effective way of extracting data and locating Hindi/Urdu switches in GloWbE.

The first step in searching for Hindi/Urdu codeswitches with the frequency list words was to convert the Hindi/Urdu words from the Devanagari and the Perso-Arabic script into the Roman script in order to make them searchable. This was made challenging by the absence of any widely accepted standard romanization system for Hindi and Urdu, and the large variation in the romanization practices of Hindi and Urdu users when writing the languages in the Roman script posed a challenge for the formulation of the searches. In order to locate different spelling variants of the frequency list items as effectively and as systematically as possible, I formulated the searches by replacing the vowels in the frequency list words by wildcards and romanizing the consonants following the conventions of the ALA-LC romanization tables. Inevitably, some spelling variants were still left out, and not all of them could be located. However, this strategy to formulate the searches was the most effective and most systematic out of the available options while still taking into account the large variation in the romanization strategies.

The fact that the frequency list words used to conduct the searches were function words (postpositions) had impact on the types of switches that were located with them, which in turn affected the results. The appearance of clauses and sentences is explainable by Hindi/Urdu clauses and sentences containing grammatical structures that the frequency list words were part of, and thus they led to the location of clauses and sentences. As sentences and clauses are longer units, they also hold a greater chance of containing one of the frequency list words than shorter elements like phrases do. The searches with the two genitive postpositions *ke* and *ki* explain the relatively high rate of occurrence of genitive phrases in the data, which shows in the classification of intraclausal switches. The items in the largest intraclausal category, proper

names, often contained a frequency list word (genitive postpositions *ke* or *ki*, or the postposition *mein*) and were therefore frequent in the data. Many of the proper names have also became collected as additional material because a sentence containing a movie title or a song title relatively often also included mentions of other movies and songs, which were then collected as additional switches. The method also affected the results in the way that – because of the searches with grammatical words – the occurrence of single-word switches remained more marginal. Single-word switches such as nouns and adjectives only occurred in the data when they were collected as additional material appearing in the same sentence as the switch containing the search item. Because of the impact of the method, the results are not generalizable, and one cannot draw conclusions concerning the overall frequencies of different types of codeswitches. However, searches with these grammatical words resulted in a wider range of different types of switches than searches with lexical words most likely would have.

I attempted to allow a wider variety of switches – not only switches containing the search item – to be represented in the data in two ways. First was not to be too restricted by the meaning of the frequency list item; if the word form in the corpus searches could be regarded as a spelling variant of the frequency list item, it did not matter in what other meanings it also appeared among the search results as long as they were codeswitches to Hindi/Urdu. Some of the romanized forms of the Hindi/Urdu frequency list words happen to be identical with the Roman spellings of some other Hindi/Urdu lexical items. For example, the searches with the genitive postpositions *ke* and *ki* had subordinators *ki* (Hindi) and *ke* (Urdu) among the results, or in the searches with the postposition *mein*, the word form also appeared in the meaning of the first person singular pronoun *mein*. Second was the criterion of collecting all switches in a sentence – not only the one containing the search item. This criterion was formulated to be able to include more different types of switches. Without this criterion, for example single nouns and adjectives would be practically non-existent in the data.

The method of locating codeswitches with the Hindi/Urdu words led to the existence of method duplicates in the material. Because the search items are high-frequency grammatical items, there is a strong possibility that a switch contains more than one of the frequency list items, more than one spelling variant of the same frequency list item, or the same spelling variant of the same frequency list item multiple times. This is especially true with longer switches such as clauses and sentences where the likelihood of the switch containing more than one of the frequency list words or spelling variants of the same frequency list word increases with the length of the switch. The number of method duplicates is likely to grow gradually the further down one goes on the frequency list. If the numbers of method duplicates in IE and PE,

which are shown in Appendix 6, are viewed in proportion to the number of switches collected from each variety, it can be observed that the PE material contained more method duplicates than the IE material. This is most likely caused by PE having higher percentages of longer switches, i.e. intersentential and interclausal switches, than IE. Thus, the number of method duplicates appears to depend on the types of switches that emerge in a particular context of CS.

The study of multilingualism sets certain requirements for the corpus as well. For a corpus to better enable the study of multilingualism, it should be annotated for foreign elements. That way the foreign elements could be located more directly and more effectively with a search for the tag. Taking multilingualism into consideration in corpus annotation would be particularly relevant in corpora dealing with global varieties of English where multilingualism can be expected to be the norm rather than the exception. That way the corpus would enable the study of a central aspect of non-native varieties of English – namely the presence and the use of the local languages in the local English.

However, especially when dealing with corpora of countries like India and Pakistan, the occurrence of a multitude of languages can be expected. Therefore, even if the corpus was annotated for foreign features in general, one would be faced with extracting the sought-after languages from other foreign languages – that is, if the focus of the research was on particular languages. The corpus would have to be annotated not only for foreign elements in general but preferably also for languages in order to make it possible to limit the searches for foreign elements to elements from a certain language or certain languages. However, how all this could be done in a large-sized corpus is another question. A second question would be the automatic recognition of languages; the variation in romanization practices could be a challenge for automatic detection of foreign languages in Indian and Pakistani context. Variation in romanization practices not only concerns Hindi and Urdu but can be expected to concern romanization of other Indian/Pakistani languages written in non-Roman scripts as well.

Overall, the method proved to be effective in locating Hindi/Urdu codeswitches and providing a large amount of data. This method would be worth trying with languages that have standardized spellings. It could be assumed that the method of locating codeswitches with the most frequent words could be even more effective and systematic when applied in the context of languages with standardized spellings or languages written in the Roman script.

8 Conclusion

The aim of this study was to examine the forms and functions of Hindi/Urdu CS in IE and PE. The data consisted of altogether 5449 instances of CS out of which 3650 were from IE and 1799 from PE. The material was collected from the Corpus of Global Web-based English. This study also served as a methodological experiment using the most frequent Hindi and Urdu words to conduct the searches and locate codeswitches to Hindi and Urdu in the corpus.

The structural analysis of the switches showed that the tendency to switch to Hindi/Urdu is largely the same in both IE and PE. The analysis of the forms showed that the Hindi/Urdu codeswitches in IE and PE display a variety of structural patterns. The structural patterns ranged from intersentential switches, i.e. switches for complete Hindi/Urdu sentences and switches between independent clauses, to intrasentential switches, which in turn varied from interclausal switches, i.e. switches between main and dependent clauses, to intraclausal switches like insertions of Hindi/Urdu words and phrases. The analysis further revealed that the different structural patterns tend to occur to approximately the same proportions in both IE and PE – with PE having slightly more of longer intersentential and interclausal switches and IE having slightly more of shorter intraclausal switches.

Based on their functions, the Hindi/Urdu codeswitches could be divided roughly into switches with a communicative function and cultural switches. Communicative functions were for example quotations, figurative language, conveying greetings and blessings, interjections, reiterations, metalinguistic commentary as well as other, more marginal functions, such as personalization, word play, and insults. The Hindi/Urdu cultural switches encompassed the use of culture-related expressions, which could be either switches for cultural objects and concepts, or switches functioning as references to the Indian and Pakistani culture.

The present study filled a gap in research on Hindi/Urdu CS practices in IE and PE by providing an account on the variety of different forms and functions of the Hindi/Urdu codeswitches. The previous research on CS in IE and PE has remained relatively fragmentary lacking more comprehensive accounts on both forms and functions of CS to local languages. The previous studies on the use of Hindi and Urdu in IE and PE have focused more on the use of single Hindi/Urdu lexical items and the lexical influence of the local language on the local variety of English not providing insight into the more varied structures or functional aspects.

The structures of Hindi/Urdu CS in IE and PE could be studied further using a different methodology that would not have such a great impact on the structures that appear in order to have more generalizable results showing in what proportions the different structures appear and what structures are more frequent than others. The topic could also be researched further by

relating the results and findings to similar accounts on the forms and functions of CS to local languages in other non-native varieties of English to see if similarities – or differences – emerge in terms of functions, for example presenting local voices and cultural switches. This would also enable the inspection of structural patterns of CS in typologically different languages and language pairs where the typological relationship between the languages involved in the switching process is different.

References

Primary sources

Davies, Mark. 2013. Corpus of Global Web-based English (GloWbE). Available online at https://www.english-corpora.org/.

Secondary sources

Abbas, Shemeem. 1993. "The Power of English in Pakistan". *World Englishes* 12., no. 2: 147-156.

Ahmad, Sajid and Ali, Sajid. 2014. "Impact of Urduised English on Pakistani English Fiction". *Journal of Research (Humanities)*: 62-75.

Alam, Suraya. 2011. "Integration of Various Approaches towards the Functions of Code-Switching between Punjabi, Urdu and English". *International Journal of Linguistics* 3, no .1: 1-18.

Anwar, Behzad. 2009. "Urdu-English Code Switching: The Use of Urdu Phrases and Clauses in Pakistani English (A Non-native Variety)". *International Journal of Language Studies (IJLS)* 3, no. 4: 409-424.

Annamalai, E. 2004. "Nativization of English in India and its Effect on Multilingualism". *Journal of Language and Politics* 3, no. 1: 151-162.

Auer, Peter. 1984. *Bilingual Conversation*. Amsterdam: John Benjamins Publishing Company.

———. 1998. "Introduction: Bilingual Conversation Revisited". In *Code-switching in Conversation: Language, Interaction and Identity*, edited by Peter Auer, 16-56. Taylor & Francis Group, ProQuest Ebook Central.

———. 1999: "From Code-switching via Language Mixing to Fused Lects: Toward a dynamic typology of bilingual speech". *International Journal of Bilingualism* 3, no. 4: 309-332.

Bhatt, Rakesh M. 2008. "In Other Words: Language Mixing, Identity Representations, and Third Space". *Journal of Sociolinguistics* 12, no. 2: 177-200.

Bhattacharjee, Mandira, Natasha Rahman, and Shyamala K. Chengappa. 2009. "Socio-linguistic Constraints of Code-switching in Hindi-English- Kannada Multilinguals". *Language in India* 9, no. 12: 75-89

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Callahan, Laura. 2004. *Spanish-English Codeswitching in a Written Corpus*. John Benjamins Publishing Company, Amsterdam.

Chanda, Arunavha, Dipankar Das, and Chandan Mazumdar. 2016. "Unraveling the English-Bengali Code-Mixing Phenomenon". *Proceedings of the Second Workshop on Computational Approaches to Code Switching*: 80-89.

Coleman, Hywel. 2012. “Profile of Pakistan”. In *Language in Education in Pakistan: Recommendations for Policy and Practice*, edited by Hywel Coleman and Tony Capstick, 13-17. Islamabad: British Council.

Crystal, David. 2003. *English as a Global Language*. 2nd edition. Cambridge: Cambridge University Press.

Dako, Kari. 2002. “Code-switching and Lexical Borrowing: Which is What in Ghanaian English”. *English Today* 18, no. 3: 48-54.

Davies, Mark, and Robert Fuchs. 2015. “Expanding Horizons in the Study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE)”. *English World-Wide* 36, no. 1: 1–28.

Dua, Hans R. 1991. “Hindi – Urdu as a Pluricentric Language”. In *Pluricentirc Languages: Differing Norms in Different Nations*, edited by Michael Clyne, 381-400. Berlin: De Gruyter.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2020. *Ethnologue: Languages of the World*. 23rd edition. Dallas, Texas: SIL International. Accessed 6 April 2020. https://www.ethnologue.com.

Ehsan, Andleeb, and Syed Abdul Aziz. 2014. “Code-mixing in Urdu News of a Private Pakistani Channel: a case study”. *Academic Research International* 5, no. 1: 160-169.

Ferguson, Charles A. 1959. “Diglossia”. *WORD* 15, no. 2: 325-340.

Fishman, Joshua A. 1972. “Domains and the Relationship between Micro- and Macrosociolinguistics”. *Directions in Sociolinguistics: the ethnography of communication*, edited by John J. Gumperz and Dell Hymes, 435-453. Oxford: Basil Blackwell.

Frick, Maria. 2010. “Suomi-viro koodinvaihto Facebookissa”. *Lähiõrdlusi. Lähivertailuja*, no. 20: 49-67.

Garcia, Ofelia. 2009. *Bilingual Education in the 21st Century: A Global Perspective*. Malden, MA: Blackwell.

———. 2012. “Theorizing Translanguaging for Educators”. In *Translanguaging: a CUNY-NYSIEB Guide for Educators*, edited by Christina Celic and Kate Seltzer, 1-6. New York: City University of New York.

Gardner-Chloros, Penelope, Reeva Charles, and Jenny Chesire. 2000. “Parallel patterns? A Comparison of Monolingual Speech and Bilingual Codeswitching Discourse”. *Journal of Pragmatics* 32, no. 9: 1305-1341.

Gardner-Chloros, Penelope, and Reeva Charles. 2007. “Subiko Welcome (welcome to everyone): Hindi/English Code-switching in the British-Asian Media”. *Birkbeck Studies in Applied Linguistics* 2, no. 2007: 64-97.

Georgakopoulou, Alexandra. 1997. “Self-presentation and Interactional Alliances in E-mail Discourse: the Style- and Code-switches of Greek Messages”. *International Journal of Applied Linguistics* 7, no. 2: 141-164.

Goffman, Erving. 1979. “Footing”. *Semiotica* 25, no. 1-2, 1-29.

Government of India. 1963. “Official Languages Act 1963”. Accessed 13 January 2020. https://rajbhasha.gov.in/en/official-languages-act-1963.

Government of India. 1966. *Report of the Education Commission 1964-1966*. Government of India, New Delhi. Accessed 13 January 2020. https://archive.org/stream/ReportOfTheEducationCommission1964-66D.S.KothariReport/48.JpReportOfTheEducationCommission 1964-66d.s.kothari_djvu.txt.

Government of India. 2011a. *Census of India 2011*. Government of India, New Delhi. Accessed 13 January 2020. http://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf.

Government of India. 2011b. "Population by Bilingualism, Trilingualism, Age and Sex". Accessed 13 January 2020. https://censusindia.gov.in/2011Census/Language_MTs.html.

Government of India. 2017. *The annual report of Prasar Bharati 2017-2018*. Accessed 13 January 2020. http://prasarbharati.gov.in/Annual_ReportPDF/2017-18-EnglishVersion.pdf.

Government of India. 2018. *Press in India*. Accessed 15 January 2020. http://rni.nic.in/all_page/press_india.aspx.

Government of India. n.d. "Languages added in the eighth schedule of the Indian constitution". Accessed 28 December 2018. https://rajbhasha.gov.in/en/languages-included-eighth-schedule-indian-constitution.

Government of Pakistan. 2006. *Pakistan Statistical Pocket Book 2006*. Government of Pakistan, Islamabad. Accessed 14 January 2020. http://www.pbs.gov.pk/content/pakistan-statistical-pocket-book-2006.

Government of Pakistan. 2017a. "Population by Mother Tongue". Accessed 14 January 2020. http://www.pbs.gov.pk/sites/default/files//tables/POPULATION%20BY%20MOTHER%20TONGUE.pdf.

Government of Pakistan. 2017b. *National Education Policy 2017*. Government of Pakistan, Islamabad. Accessed 14 January 2020. https://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/pakistan_national_education_policy_2017-2025.pdf.

Government of Pakistan. 2017c. *Pakistan Statistical Yearbook 2017*. Government of Pakistan, Islamabad. Accessed 14 January 2020. http://www.pbs.gov.pk/content/pakistan-statistical-year-book-2017.

Government of Pakistan. n.d. "PBC Family". Accessed 15 January 2020. http://www.radio.gov.pk/pbc-family#Channels123.

Gumperz, John J. 1964. "Linguistic and Social Interaction in Two Communities". *American Anthropologist* 66, no. 6: 137-153.

———. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.

Gysels, Marjolein. 1992. "French in Urban Lubumbashi Swahili: Codeswitching, Borrowing, or Both?". *Journal of Multilingual & Multicultural Development* 13, no. 1-2: 41-55.

Halim, Nur Syazwani, and Marlyna Maros. 2014. "The Functions of Codeswitching in Facebook Interactions". *Procedia - Social and Behavioral Sciences* 118: 126 – 133.

Halmari, Helena. 1997. *Government and Codeswitching: Explaining American Finnish*. Amsterdam: John Benjamins Publishing Company.

Hasselmo, Nils. 1969. “On Diversity in American Swedish”. *Svenska landsmål och svenskt folkliv* 92, no. 293: 53-72.

Heller, Monica. 1988. “Introduction”. In *Codeswitching: Anthropological and Sociolinguistic Perspectives*, edited by Monica Heller, 1-24. Berlin: De Gruyter.

Hinrichs, Lars. 2006. *Codeswitching on the web: English and Jamaican Creole in e-mail communication*. Amsterdam: John Benjamins Publishing Company.

Iqbal, Liaqat. 2011. “Linguistic Features of Code-Switching: A Study of Urdu/English Bilingual Teachers’ Classroom Interactions”. *International Journal of Humanities and Social Science* 1, no. 14: 188-194.

Jacobson, Rodolfo. 2001. “Language alternation: The Third Kind of Codeswitching Mechanism.” In *Codeswitching worldwide II*, edited by Rodolfo Jacobson, 59-72. Berlin: De Gruyter.

Jitta, Divya Jai. 2018. “Towards Understanding Code-mixed Telugu-English Data”. MA thesis, International Institute of Information Technology, Hyderabad.

Lalmalsawma, David. 2013. “India Speaks 780 Languages, 220 Lost in Last 50 Years – Survey”. Reuters, 7 September. Accessed 2 May 2020. http://blogs.reuters.com/india/2013/09/07/india-speaks-780-languages-220-lost-in-last-50-years-survey/.

Kachru, Braj B. 1978a. “Toward Structuring Code-mixing: an Indian Perspective”. *International Journal of the Sociology of Language* 1978, no. 16: 27-46.

———. 1978b. “English in South Asia”. In *Advances in the Study of Societal Multilingualism*, edited by Joshua A. Fishman, 477-552. The Netherlands: De Gruyter.

———. 1982. “Models for Non-native Englishes”. In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru. University of Illinois Press, Urbana.

———. 1998. “English as an Asian language”. *Links & Letters* 5: 89-108.

Kachru, Yamuna, and Cecil L. Nelson. 2006. *World Englishes in Asian Contexts*. Hong Kong: Hong Kong University Press.

Kanthimathi, K. 2009. “Tamil-English Mixed Language Used in Tamilnadu”. *The International Journal of Language Society and Culture*, no. 27: 48-53.

Kelkar, Ashok R. 1968. *Studies in Hindi-Urdu I: Introduction and Word Phonology*. Poona: Deccan College Postgraduate and Research Institute.

Khan, Waquar Ahmad. 2010. “Sociolinguistics of Code switching: an Investigation of Urdu-English Code Switching in Indian Metropolis”. Doctoral thesis, Aligarh Muslim University.

King, Robert D. 2001. “Poisonous Potency of Script: Hindi and Urdu”. *International Journal of Sociology of Language*, no. 150: 43-59.

Klingler, Anita. 2017. “Changes in Code-Switching Patterns among Hindi-English Bilinguals in Northern India”. *Lifespans and Styles* 3, no. 1: 40-50.

Kumar, Ashok. 1986. “Certain Aspects of the Form and Functions of Hindi-English Code-switching”. *Anthropological Linguistics* 28, no. 2: 195-205.

Library of Congress. [1997] 2017. “ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts”. Accessed 8 July 2018. https://www.loc.gov/catdir/cpso/roman.html.

Mackey, William F. 1970. “Interference, Integration and the Synchronic Fallacy”. In *Georgetown University Round Table on Languages and Lingusitics 1970: Bilingualis and Language Contact*, edited by James E. Alatis, 195-227. Washington D.C.: Georgetown University Press.

Mahmood, Rashid, and Sayed Kazim Shah. 2011. “Culture-specific Vocabulary: A Study of Urdu Nouns in Pakistani English”. *Komunikacija i Kultura Online* 2, no. 2, 72-86.

Malhotra, Sunita. 1980. “Hindi-English, Code-switching and Language Choice in Urban Upper-middle-class Indian Families”. *Kansas Working Papers in Linguistics* 5, no 2: 39-46.

Mallikarjun, B. 2002. “Mother Tongues of India according to the 1961 Census”. *Language in India* 2, no. 5. n.p.

McArthur, Tom. 2003. *Oxford Guide to World English*. Oxford: Oxford University Press.

Meganathan, Ramanujam. 2011. “Language Policy in Education and the Role of English in India: from Library Language to Language of Empowerment”. In *Dreams and Realities: Developing Countries and the English Language*, edited by Hywel Coleman, 57-85. London: British Council.

Meyer, Charles F. 2002. *English Corpus Linguistics: an Introduction*. Cambridge: Cambridge University Press.

Milroy, Lesley, and Pieter Muysken. 1995. “Introduction: Code-switching and Bilingualism Research”. In *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*, edited by Lesley Milroy and Pieter Muysken, 1-14. Cambridge: Cambridge University Press.

Mkilifi, M.H. Abdulaziz. 1972. “Triglossia and Swahili-English Trilingualism in Tanzania”. *Language in Society* 1, no. 2: 197-213.

Montes-Alcalá, Cecilia. 2000. “Written Codeswitching: Powerful Bilingual Images”. In *Codeswitching Worldwide II*, edited by Rodolfo Jacobson, 193-219. Berlin: De Gruyter.

———. 2007. “Blogging in Two Languages: Code-Switching in Bilingual Blogs”. In *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*, edited by Jonathan Holmquist, Augusto Lorenzino and Lotfi Sayahi, 162-170. Somerville, MA: Cascadilla Proceedings Project.

Moyer, Melissa G. 1998. “Bilingual Conversation Strategies in Gibraltar”. In *Code-switching in Conversation: Language, Interaction and Identity*, edited by Peter Auer, 378-414. Taylor & Francis Group, ProQuest Ebook Central.

Muysken, Pieter. 1995. “Code-switching and Grammatical Theory”. In *One speaker, two languages: cross-disciplinary perspectives on code-switching*, edited by Lesley Milroy and Pieter Muysken, 177-198. Cambridge University Press, Cambridge.

———. 2000. *Bilingual Speech: a Typology of Code-Mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1988. “Code Switching as Indexical of Social Negotiations”. In *Codeswitching: Anthropological and Sociolinguistic Perspectives*, edited by Monica Heller, 151-186. Berlin: De Gruyter.

———. 1998. “A Theoretical Introduction to the Markedness Model”. In *Codes and Consequences*, edited by Myers-Scotton, 18-38. Oxford: Oxford University Press.

———. 1992. “Comparing Codeswitching and Borrowing”. *Journal of Multilingual and Multicultural Development* 13, no. 1-2: 19-39.

———. 1993. *Duelling Languages*. Oxford: Clarendon Press.

———. 2006. *Multiple Voices: an Introduction to Bilingualism*. Malden, MA: Blackwell Publishing.

Myers-Scotton, Carol and Janice L. Jake. 2000. “Testing the 4-M Model: An Introduction”. *The International Journal of Bilingualism* 4, no. 1: 1-8.

Nurmi, Arja, Jukka Tyrkkö, Anna Petäjäniemi, and Päivi Pahta. 2018. “The Social and Textual Embedding of Multilingual Practices in Late Modern English: A Corpus-based Analysis”. In *Multilingual Practices in Language History: English and beyond*, edited by Päivi Pahta, Janne Skaffari and Laura Wright. Berlin: De Gruyter.

Parasher, S.V. 1980. “Mother Tongue – English Diglossia: a Case Study of Educated Bilinguals’ Language Use”. *Anthropological Linguistics* 22, no. 4: 151-162.

———.1998. “Language Policy in a Multilingual Setting: the Indian Scenario”. *Asian Englishes* 1, no. 1: 92-116.

Parveen, Sabahat, and Shazma Aslam. 2013. “A Study on Reasons for Code-switching in Facebook by Pakistani Urdu English Bilinguals”. In *Language in India* 13, no. 11: 564-590.

Patel, Jinali. 2016. “Bilingual Creativity of English Language in Indian Advertisements”. *Englishes Today* 2, no 2.: n.p.

Platt, John T. 1977. “A Model for Polyglossia and Multilingualism (with Special Reference to Singapore and Malaysia)”. *Language in Society* 6, no. 3: 361-378.

Poplack, Shana 1980: “Sometimes I’ll Start a Sentence in Spanish y Termino en Español”. *Linguistics* 18., no. 7-8, p. 581-618.

———. 1988. “Contrasting Patterns of Codeswitching in two Communities”. In *Codeswitching: Anthropological and Sociolinguistic Perspectives*, edited by Monica Heller, 215-244. Berlin: De Gruyter.

Poplack, Shana, and David Sankoff. 1984. “Borrowing: the Synchrony of Integration”. *Linguistics* 22, no. 1: 99-135.

Prasad, G. J. V. 2011. “Tamil, Hindi, English: the New Ménage à Trois”. In *Chutnefying English: the Phenomenon of Hinglish*, edited by Rita Kothari and Rupert Snell, 141-160. New Delhi: Penguin Books.

Rafi, Muhammad Shaban. 2013. “Urdu and English Contact in an E-discourse: Changes and Implications”. *Gomal University Journal of Research* 29, no. 2: 78-86.

Rahman, Tariq. 1990. *Pakistani English*. Islamabad: National Institute of Pakistan Studies.

———. 1996. *Language and Politics in Pakistan*. Karachi: Oxford University Press.

———. 2002. “Language, Power and Ideology”. *Economic and Political Weekly* 37, no. 44-45: 4556-4560.

———. 2005. “Passports to Privilege: The English-medium Schools in Pakistan”. *Peace and Democracy in South Asia* 1, no. 1: 24-44.

———. 2006. “Language Policy, Multilingualism and Language Vitality in Pakistan”. In *Lesser-known languages of South Asia: Status and Policies, Case Studies and Applications of Information Technology*, edited by Anju Saxena and Lars Borin, 73-104. Berlin: De Gruyter.

Romaine, Suzanne. 1995. *Bilingualism*. 2nd ed. Malden: Blackwell Publishing.

Sailaja, Pingali. 2009. *Indian English*. Edinburgh: Edinburgh University Press.

Sankoff, David, Shana Poplack, and Swathi Vanniarajan. 1990. “The Case of the Nonce Loan in Tamil”. *Language Variation and Change* 2, no. 1: 71-101.

Schendl, Herbert. 1997. “‘To London fro Kent / Sunt predia depopulantes’: Code-switching and Medieval English Macaronic Poems”. *VIEWS: Vienna English Working Papers* 6, no 1: 52-66.

Sebba, Mark. 2012. “Multilingualism in Written Discourse: An Approach to the Analysis of Multilingual Texts”. *International Journal of Bilingualism* 17, no.1: 97-118.

Sebba, Mark, and Tony Wootton. 1998. “We, They and Identity: Sequential versus Identity-related Explanation in Code-switching”. In *Code-switching in conversation: language, interaction and identity*, edited by Peter Auer, 460-500. Taylor & Francis Group, ProQuest Ebook Central.

Shackle, Christopher, and Rupert Snell. 1990. *Hindi and Urdu since 1800: A Common Reader*. New Delhi: Heritage Publishers.

Shapiro, Michael C. 2003. “Hindi”. In *The Indo-Aryan Languages*, edited by Cardona, George and Dhanesh Jain, 250-285. Oxfordshire: Routledge.

Shivaprasad, Chaitra. 2015. “The Optimal Grammar of Code-switching between Kannada and English”. MA thesis, University of Illinois.

Si, Aung. 2010. “A Diachronic Investigation of Hindi–English Code-switching, Using Bollywood Film Scripts”. *International Journal of Bilingualism* 15, no. 3: 388-407.

Siebenhaar, Beat. 2006. “Code Choice and Code-switching in Swiss-German Internet Relay Chat Rooms”. *Journal of Sociolinguistics* 10, no. 4: 481-506.

Singh, Rajendra. 1983. “We, They, and Us: a Note on Code-switching and Stratification in North India”. *Language in Society* 12, no. 1: 71-73.

Sridhar, S. N., and Kamal K. Sridhar, 1980. “The Syntax and Psycholinguistics of Bilingual Code mixing. *Canadian Journal of Psychology/Revue canadienne de psychologie* 34, no. 4: 407-416.

Talaat, Mubina, and Behzad Anwar. 2010. “The Impact of Urdu-English Code-Switching on Pakistani English”. *Kashmir Journal of Language Research*: 95-119.

The Office of Registrar of Newspapers for India, n.d. Accessed 13 January 2020. http://rni.nic.in/all_page/press_india.aspx.

Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkley: University of California Press.

Verma, S.K. 1976. “Code-switching: Hindi-English”. *Lingua* 38, no. 2: 153-165.

Wei, Li. 1998. “The ‘Why’ and ‘How’ Questions in the Analysis of Conversational Code-switching”. In: *Code-switching in Conversation: Language, Interaction and Identity*, edited by Peter Auer, 280-316. Taylor & Francis Group, ProQuest Ebook Central.

U Din, Najam. n.d. “Pakistan”. Accessed 14 January 2020. https://medialandscapes.org/country/pakistan.

Ud-Din, Umar, Muhammad Kamal Khan, and Muhammad Iqbal. 2011. “Multiple Nested Triglossic Situation in Pakistan”. *Language in India* 11, no. 3: 9-18.

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden, MA: Blackwell Publishing.

———. 2010. “Contact and Borrowing”. In *The Handbook of Language Contact*. Hickey, Raymond (ed.), Malden, MA: Blackwell Publishing.

Zentella, Ana Celia. 1997. *Growing up Bilingual*. Malden, MA: Blackwell Publishers Inc.

Appendix 1 Hindi romanization

Hindi character	ALA-LC table	Romanized	Hindi character	ALA-LC table	Romanized	Hindi character	ALA-LC table	Romanized
क	ka	k	फ	pha	ph	ष	sha	sh
ख	kha	kh	ब	ba	b	स	sa	s
ग	ga	g	भ	bha	bh	ह	ha	h
घ	gha	gh	म	ma	m	क़	qa	q
ङ	ṅa	n	य	ya	y	ख़	kha	kh
च	ca	ch*	ढ	ḍha	d	ग़	gha	gh
छ	cha	ch	ढ़	ṛha	rh, dh *	ज़	za	z
ज	ja	j	ण	ṇa	n	फ़	fa	f
झ	jha	jh	त	ta	t			
ञ	ña	n	थ	tha	th			
ट	ṭa	t	द	da	d			
ठ	ṭha	th	ध	dha	dh			
ड	ḍa	d	र	ra	r			
ड़	ṛa	r, d*	ल	la	l			
न	na	n	व	va	v, w*			
प	pa	p	श	śha	sh*			

*The following additions or changes were made to the ALA-LC Urdu table:

- ड़ is romanized as both *r* and *d*

- ढ़ is romanized as both *rh* and *dh*

- व is romanized as both *v* and *w*

- च are romanized as *ch* instead of *c*

- श romanized as *sh* instead of *s*

Further, the following rules concerning the romanization of the Hindi diacritics anusvāra ◌ं and anunāsika ◌ँ are followed (in word-medial position):

-Anusvāra ◌ं is romanized as *n* before gutturals, palatals, cerebrals, and dentals, and as *m* before labials.

-Anunāsika ◌ँ is romanized as *n* before gutturals, palatals, cerebrals, and dentals, and as *m* before labials, sibilants, semivowels, and aspirates.

-In word-final position anusvāra and anunāsika are romanized as *n* or not included in the romainzed spelling of the word. Because they signal the nasalization of the preceding vowel, the searches are formulated as was explained concerning the search forms of vowels and nasal vowels.

Appendix 2 Urdu romanization

Urdu character	ALA-LC table	Romanized	Urdu chrarcter	ALA-LC table	Romanized
ب	b	b	ص	ṣ	s
پ	p	p	ض	ẓ	z
ت	t	t	ط	ṭ	t
ٹ	ṭ	t	ظ	ẓ	z
ث	s	s	ع	ʻ	-
ج	j	j	غ	gh	gh
چ	c	ch*	ف	f	f
ح	ḥ	h	ق	q	q
خ	kh	kh	ک	k	k
د	d	d	گ	g	g
ڈ	ḍ	d	ل	l	l
ذ	z	z	م	m	m
ر	r	r	ن	n	n
ڑ	ṛ	r	ں	n	-, n*
ز	z	z	و	v	v, w*
ژ	zh	z	ہ	h	h
س	s	s	ی	y	y*
ش	sh	sh			

*The following additions or changes were made to the ALA-LC Urdu table:

- ڑ is romanized as both r and d

- و is romanized as both v and w

- چ is romanized as ch instead of c

- The letter ں known as *noon ghunna* is either romanized as *n* or not included in the romanized spelling of the word. *Noon ghunna* marks the nasalization of the preceding vowel in word-final position and therefore the romanization rules of vowels and nasal vowels are followed.

- و and ی are used to represent both consonants and vowels. When they represent consonants, they are romanized as shown in the romanization table. As vowels they are substituted with wildcards.

- The Arabic letter ك "*kaaf*" was replaced by another type of "*kaaf*" ک that is more commonly found in Urdu.

- The Arabic letter ة (*ta' marbutah*) was excluded from the Urdu table as it is not generally used in the Urdu script.

Appendix 3 Hindi frequency list

	Hindi word	Romanized form	Translation/Meaning
1.	के	ke	Genitive postposition (masc. and masc. pl.)
2.	में	mein	Postposition, ”in”
3.	की	ki	Genitive postposition (fem. and fem. pl.)
4.	है	hai	3rd person singular form of *hona* (to be), present tense
5.	से	se	Postposition, ”from”
6.	और	aur	coordinating conjunction ”and”
7.	को	ko	Postposition, ”to”
8.	का	ka	Genitive postposition (masc.)
9.	पर	par	Postposition, ”on”
10.	है।	hai |	3rd person singular form of *hona* (to be) with *danda*, the Hindi full stop (vertical line)

Appendix 4 Urdu frequency list

	Urdu word	Romanized form	Translation/meaning
1.	کے	ke	Genitive postposition (masc. and masc. pl.)
2.	کی	ki	Genitive postposition (fem. and fem. pl.)
3.	میں	mein	Postposition, ”in”
4.	اور	aur	coordinating conjunction ”and”
5.	سے	se	Postposition, ”from”
6.	ہے	hai	3rd person singular form of *hona* (to be), present tense
7.	کا	ka	Genitive postposition (masc.)
8.	کو	ko	Postposition, ”to”
9.	اس	is	oblique form of 3rd person singular pronoun “yeh”
10.	کہ	ke	subordinate conjunction, ”that”

Appendix 5 Frequency list words, search forms and spelling variants

Indian English

Frequency list item	Search form	Spelling variants	Tokens (in GloWbE)	Codeswitches*
के	k?	ke	1005	897
में	m??n	mein	617	547
की	k?	ki	1701	1482
	k??	kii	32	30

*codeswitches out of the tokens in GloWbE, no additional switches included

Pakistani English

Frequency list item	Search form	Spelling variants	Tokens (in GloWbE)	Codeswitches*
کے	k?	ke	482	430
	k??	kay	270	203
		kee	45	41
		keh	38	33
کی	k?	ki	757	688
	k??	kii	10	10
میں	m??n	mein	334	303

*codeswitches out of the tokens in GloWbE, no additional switches included

Appendix 6 Frequency list words, spelling variants, and numbers of codeswitches

Hindi

Hindi word	Spelling variant	Codeswitches
के	ke	897
	ke additional	402
	ke duplicate	96
	corpus duplicates	40
	method duplicates	56
	excluded	22
	total in analysis	1159
में	mein	547
	mein additional	178
	mein duplicate	106
	corpus duplicates	22
	method duplicates	84
	excluded	6
	total in analysis	619
की	ki	1482
	ki additional	656
	ki duplicate	267
	corpus duplicates	72
	method duplicates	195
	excluded	24
	total in analysis	1837
	kii	30
	kii additional	13
	kii duplicate	8

	corpus duplicates	0
	method duplicates	8
	excluded	0
	total in analysis	35

Urdu

Urdu word	Spelling variant	Codeswitches
کے	ke	430
	ke additional	170
	ke duplicate	78
	corpus duplicates	32
	method duplicates	46
	excluded	9
	total in analysis	504
	kay	203
	kay additional	40
	kay duplicate	25
	corpus duplicate	8
	method duplicate	17
	excluded	2
	total in analysis	215
	kee	41
	kee additional	9
	kee duplicate	8
	corpus duplicate	1
	method duplicate	7
	excluded	0
	total in analysis	42
	keh	33

	keh additional	5
	keh duplicate	8
	corpus duplicate	1
	method duplicate	7
	excluded	0
	total in analysis	30
کی	ki	688
	ki additional	204
	ki duplicate	125
	corpus duplicates	31
	method duplicates	94
	excluded	13
	total in analysis	749
	kii	10
	kii additional	0
	kii duplicate	1
	corpus duplicates	0
	method duplicates	1
	excluded	0
	total in analysis	9
میں	mein	303
	mein additional	65
	mein duplicate	110
	corpus duplicates	13
	method duplicates	97
	excluded	1
	total in analysis	250

Appendix 7 Accepted and not accepted codeswitches in the analysis

Accepted	**Explanation**
i agree with the comment on falcon on ET. sweetheart, you need to cool down a bit. **itna ghussa sehet ke liye achha nahin hota.** you can't fight all the battles. choose your battles and leave the others.	The intersentential switch contains an accepted spelling variant of a frequency list word (ke).
you suggest to give more importance to your mother tongue over the language of Rasool (S.A.W).. **Yeh Hadith to suni ho gee. (mafhoom) kay koi us waqt tek musalman nahi ho sakta jab tek main us kay baap, maan or her shay say ziada muaqaddam na ho jaoun.. Ji say mohabbat hoti hay us ki her shay say muhabbat ka izhaar houta hay.. (yeh diloun kay souday hain serkaar)..** and another thing history of muslim scholars show that there have been more religious and arabic scholars in Ajjam rather than in Arabs	The intersentential switch contains an accepted spelling variant of a frequency list word (kay). Note: The same sequence contains the spelling variant *kay* not only once but three times. In addition, the switch also contains an accepted spelling variant of another frequency list item (ki). This means, that this switched sequence occurs as a duplicate altogether three times in the material.
hi friends before impress a girl look at her face impresion when you see one another then she look at yoou or not check also that there should be not any person if she smile there may be two reason 1. is that **ki wo apko dekh ke hans rhhee hi ya hans ke dekh hans rhi hi** then conclusin your mind thanks	The interclausal switch contains an accepted spelling variant of a frequency list word (ke). Note: Again, the same switch contains the spelling variant *ke* twice and additionally, the switch contains an accepted spelling variant of another frequency list item (ki). Thus, this switch occurs twice as a duplicate in the material.
# Anonymous said... # pakistanis are just frog in the wells or **lakir ke fakirs**. its not fair to equate them with islam in general and some posts are in real bad taste.	The intraclausal switch contains an accepted spelling variant of a frequency list word (ke).
More than men, saasu maas are obsessed with getting a chaand si bahu for their sons so that the projeny is also beautiful. Imagine the **saas**-to-be coming for a **rishta** and saying **" zara larki ko inspection ke liye bhej dein "**. Pathetic!!	The interclausal switch "zara larki ko inspection ke liye bhej dein" contains an accepted spelling variant of a frequency list word (ke). Note: The codeswitches *saas* and *rishta* are also collected for analysis as well because

	of the criterion "if the sentence that contains a switch containing a frequency list item includes other switches to Hindi/Urdu, those switches will be collected additional material.
Not accepted	
English and Urdu pehle us ne shalwar utari phir us ne kameez utari phir us ne dupata bi utar dia Recent Info Of Your Interest # English If you are ever in doubt as to whether or not to kiss a pretty girl, always give her the benefit of the doubt English Benefits of Kissing Benefits of Kissing; Changes taste; Burns Calories; Lips never go dry; Relieves Stress; Makes face muscles strong; So KEEP KISSING: **Urdu Dil ker raha hai, ke aaj tum ko** (In Urdu)... ##3500131 Top 8 Furnishing Must Have for Your Home # Four walls and one roof make a building but inhabitants turn this building into a home.	The switch appears separate from the surrounding discourse, and it cannot be determined what the switch is part of.
Since, First Impression may be the last impression, the aesthetic appeal of any solution will need to be given its due value. Just after all, only the good looking ones get the second glance and awe! Everybody is gearing up for another holiday purchasing season ##3859497 # **Aaj kamran khan ke saath** on Geo news	The switch is a separate element from the surrounding discourse, potentially a link to an article or a video on a web page.
Habib Jalib is said to have visited her in hospital after she attempted suicide following the humiliation at the hands of Khan. He composed the verses: # **Tu ke **27;3339;TOOLONG hai abhi # Raks zanjeer pehn ker bhi kiya jata hai** # (You are still unaware of the etiquette of being a slave # One dances even when chained...) # The verses were later sung by Mehdi Hassan in the film Zarka (1969) and became the metaphor to remember the late governor with.	Part of the switch is missing due to restrictions of the corpus.
# This Topic and Thread ended right here # " Lets not dwell into matters which have already been dealt with 1400 years ago. Go read for yourself, and understand it yourself. " Sh. Hamza Yusuf # found this page by using these recent terms: **nikah ke baad**, nikah sex, sex after nikah, nikah without rukhsati, first night after nikah, nikah and sex, sex after nikah before rukhsati, can you have sex after nikah, can poeple who have had sex do Nikah, after nikah first night, sex before rukhsati, sex after nikah before rukhsati islam, islamic sex nikah, rukhsati in islam, sex nikah, can i have sex after nikah, nikah and rukhsati, sex without nikah, intercourse after nikah, what does rukhsati mean,	This switch is part of a list of search terms that does not belong to any particular whole.

Appendix 8 Finnish summary

Johdanto

Tässä tutkimuksessa tutkitaan koodinvaihtoa hindiin ja urduun intianenglannissa ja pakistaninenglannissa. Tarkoituksena on kartoittaa sekä koodinvaihtojen rakenteellisia muotoja että niiden käyttöfunktioita. Saapuessaan siirtomaa-aikana Intian niemimaalle – alueelle, jota leimaa vahvasti laaja monikielisyys – englannin kieli tuli kontaktiin alueen monien paikallisten kielten kanssa. Siinä missä englanti on vaikuttanut paikallisiin kieliin, myös paikallisten kielten vaikutus näkyy paikallisessa englannissa. Vaikka koodinvaihtoa paikalliskieliin tiedetään tapahtuvan, sitä ei kuitenkaan ole tutkittu paljoa intian- ja pakistaninenglannin kontekstissa. Aiempi tutkimus hindin ja urdun käytöstä intian- ja pakistaninenglannissa on lähinnä keskittynyt yksittäisten hindin ja urdun sanojen käyttöön ja siihen, kuinka paikalliset kielet vaikuttavat paikallisen englannin sanastoon. Näin ollen aiemman tutkimuksen ulkopuolelle on jäänyt, miten monenlaisin eri tavoin paikalliset kielet näkyvät paikallisessa englannissa. Motivaationa tutkia koodinvaihtoa hindiin ja urduun intianenglannissa ja pakistaninenglannissa yhtä aikaa on, että tarkastelemalla näitä rinnakkain voidaan selvittää, näkyvätkö paikalliskielet paikallisessa englannissa samoilla tavoin näissä konteksteissa, joissa englanti palvelee samanlaisia funktioita ja on pitkälti samanlaisessa asemassa yhteiskunnassa. Lisäksi voidaan tarkastella, ovatko koodinvaihtojen rakenteet samanlaisia näissä kahdessa kontekstissa, joissa koodinvaihdossa mukana olevat kielet ovat samanlaisessa typologisessa suhteessa toisiinsa (Englanti-Hindi, Englanti-Urdu), vai eroavatko ne toisistaan. Hindi ja urdu ovat kielellisesti hyvin lähellä toisiaan ja puhutussa muodossaan keskenään täysin ymmärrettävissä; niitä voisi lähestulkoon pitää saman kielen eri murteina.

Intian ja Pakistanin kielet ja monikielisyys

Sekä Intia että Pakistan ovat monikielisiä maita. Arviot Intiassa ja Pakistanissa puhutuista kielistä vaihtelevat välillä suurestikin eri lähteiden välillä. Valtion viralliset kielet ovat Intiassa hindi ja englanti. Pakistanissa puolestaan ne ovat urdu ja englanti. Intiassa ylivoimaisesti eniten äidinkielenä puhuttu kieli on hindi. Englantia äidinkielenään puhuvien ihmisten määrä on Intiassa hyvin pieni, mutta jos englantia toisena kielenä puhuvat ihmiset lasketaan, englanti on Intian toiseksi puhutuin kieli. Pakistanin väestöstä lähes puolet puhuu punjabia äidinkielenä, mutta urdu toimii valtionlaajuisena lingua francana erityisesti urbaaneilla alueilla. Englantia puhuu toisena kielenä arviolta noin 10% väestöstä – suurin piirtein sama prosenttiosuus väestöstä kuin Intiassa.

Sen lisäksi, että englanti on molemmissa maissa valtion toinen virallinen kieli, englantia käytetään Intiassa ja Pakistanissa yhteiskunnan monilla eri tasoilla. Englantia käytetään hallinnon kielenä erityisesti hallinnon korkeammilla asteilla. Kouluissa englanti on sekä opetettava kieli että opetuksen kieli. Englanti dominoi opetuksen kielenä erityisesti ylemmillä koulutusasteilla. Erityisesti Intiassa englantia käytetään ihmisten välisenä lingua francana tilanteissa, joissa keskustelukumppaneilla ei ole yhteistä äidinkieltä. Pakistanissa englanti toimii lingua francana lähinnä eliitin keskuudessa. Intiassa ja Pakistanissa englanti on myös median kieli, ja sitä käytetään niin televisiossa kuin radiossakin. Lisäksi molemmissa maissa on englanninkielinen lehdistö.

Intiassa ja Pakistanissa englannin voidaan havaita olevan diglossisessa suhteessa hindin, urdun sekä muiden alueellisten kielten kanssa ihmisten elämässä. Diglossialla tarkoitetaan tilannetta, jossa kaksi kieltä on yhteiskunnassa erikoistunut toimimaan kommunikaation kielenä eri elämän aloilla, ja kielet palvelevat eri funktioita yhteiskunnassa. Englantia voidaan havaita käytettävän virallisemmissa konteksteissa, kuten töissä ja koulutuksessa, kun taas hindiä, urdua ja muita paikallisia kieliä käytetään epävirallisemmissa yhteyksissä kuten perheen ja ystävien kanssa. On hyvä kuitenkin huomata, ettei diglossia-ilmiötä esiinny ainoastaan englannin ja paikalliskielten välillä, vaan myös paikalliskielten kesken.

Koodinvaihto

Yksinkertaisimmillaan koodinvaihdolla voidaan katsoa tarkoitettavan vaihtelua kielten tai kielen eri varianttien välillä diskurssin sisällä. Koodinvaihto on kuitenkin saanut erilaisia määritelmiä, jotka tavalla tai toisella hieman eroavat toisistaan, ja toisaalta koodinvaihto-termin rinnalle on myös syntynyt muita, läheisesti toisiinsa liittyviä termejä. Esimerkiksi termi koodien sekoitus (*code-mixing*) tekee eron lauseensisäisen ja lauseiden välisen koodinvaihdon välille, joista edellistä kutsutaan koodien sekoitukseksi ja jälkimmäistä koodinvaihdoksi. Useista eri määritelmistä johtuen aina ei ole täysin selvää, mitä koodinvaihdolla kulloinkin tarkoitetaan. Tästä huolimatta koodinvaihto-termiä käytetään kaikista yleisimmin, kun puhutaan kielten välillä tapahtuvasta vaihtelusta.

Koodinvaihtotutkimuksessa voidaan karkeasti erottaa kolme suuntausta: sosiolingvistinen, interaktionaalinen ja kieliopillinen. Sosiolingvistiset teoriat selittävät koodinvaihtoa kielten yhteiskunnallisen aseman ja roolin pohjalta. Yksi keskeisiä konsepteja on jako tilanteiseen ja metaforiseen koodinvaihtoon. Tilanteisella koodinvaihdolla tarkoitetaan tilanteesta johtuvaa kieleltä toiselle vaihtamista jonkin tilanteeseen liittyvän seikan vuoksi, kuten ympäristö/puitteet, aktiviteettityyppi tai osallistujat. Metaforisella koodinvaihdolla tarkoitetaan

puolestaan sitä, kun koodia vaihdetaan jonkin kommunikatiivisen efektin aikaansaamiseksi. Interaktionaalisten teorioiden mukaan koodinvaihdot tulee analysoida itse keskustelun sisällä ja täytyy osoittaa, mikä merkitys koodinvaihdolla on itse keskustelussa. Koodinvaihtojen on myös osoitettu toimivan kontekstivihjeinä, joiden avulla puhuja osoittaa kuulijalle, miten hän tarkoittaa jonkin sanomansa. Koodinvaihto toimii samanlaisena kontekstivihjeenä kuin esimerkiksi äänen paino, intonaatio tai eleet ja ilmeet. Koodinvaihdolla on havaittu olevan erilaisia käyttöfunktioita, joiden pohjalta on koottu erinäisiä listauksia ja typologioita. Havaittuja käyttöfunktioita ovat muun muassa puheen referointi/lainaaminen, puhuteltavan spesifiointi, toisto, sanotun modifiointi, huudahtukset ja persoonallisen ja objektiivisen näkökulman osoittaminen. Eri tutkijat ovat koonneet omien tutkimustulostensa pohjalta omia typologioitaan, jotka osittain täydentävät toisiaan ja osittain jättävät joitakin funktioita pois.

Kieliopilliset teoriat pyrkivät selvittämään, kuinka rakenteelliset ja kielitypologiset seikat vaikuttavat koodinvaihtoon, ja millaisia kieliopillisia rajoitteita koodinvaihdolle on. Kieltä, joka dominoi interaktion kielenä ja luo kieliopilliset raamit, kutsutaan matriisikieleksi. Kieli, josta elementtejä on integroitu matriisikieleen, kutsutaan puolestaan upotetuksi kieleksi. Syntaktisesti voidaan puhua lauseiden välisestä ja lauseensisäisestä koodinvaihdosta, joista edellinen viittaa koodinvaihtoon lauseiden välillä ja jälkimmäinen koodinvaihtoon lauseen sisällä. Kolmantena tyyppinä voidaan erottaa lauseenulkoinen koodinvaihto. Lauseenulkoiset koodinvaihdot ovat elementtejä, jotka ovat irrallisia itse lauseen rakenteesta.

Koodinvaihdolle on esitetty erilaisia kieliopillisia rajoitteita, kuten vapaan morfeemin rajoitus, jonka mukaan koodinvaihtoa ei tapahdu sidonnaisissa morfeemeissa, ja ekvivalenssirajoitus, jonka mukaan koodinvaihtoa tapahtuu vain kohdissa, joissa kielet ovat rakenteeltaan samanlaisia. Rajoitukset eivät kuitenkaan ole absoluuttisia, ja ne tulee ennemmin nähdä tendensseinä kuin sääntöinä.

Koodinvaihtojen rakenteellisia ominaisuuksia on myös pyritty kuvailemaan jakamalla koodinvaihdot rakenteiden perusteella kolmeen kategoriaan: insertio, alternaatio ja kongruoiva leksikalisaatio. Insertio tarkoittaa yksittäisten konstituenttien, kuten sanat ja lausekkeet, insertiota toiseen kieleen. Alternaatio puolestaan on koodinvaihtoa rakenteiden välillä. Kongruentti leksikalisaatio merkitsee tilannetta, jossa koodinvaihdossa sanoja otetaan yhtä lailla ja yhtä paljon molemmista kielistä, ja kielten välinen vaihtelu on erittäin frekventtiä.

Yksi kielikontaktitutkimuksen ikuisuuskysymyksiä on, kuinka tehdä ero koodinvaihtojen ja vakiintuneiden lainasanojen välille. Kriteereiksi on ehdotettu muun muassa sanan käyttöä

yksikielisten keskuudessa, sanan yleisyys, ja sanan morfologinen ja fonologinen integraatio. Kaikkiin kriteereihin liittyy kuitenkin ongelmia tavalla tai toisella. Tässä tutkimuksessa ei pyritty tekemään tiukkaa eroa koodinvaihtojen ja lainojen välille. Syy tähän oli, että osa hindistä ja urdusta englantiin lainautuneista sanoista voi assosioitua hindin ja urdun puhujilla hyvin vahvasti omaan kieleen ja kulttuuriin, eivätkä he pitäisi hindistä tai urdusta englantiin lainautunutta sanaa englannin sanana.

Aineisto ja metodi

Tutkimuksessa käytetty materiaali kerättiin Corpus of Global Web-based English (GloWbE)-korpuksesta. GloWbE koostuu eri puolilla maailmaa puhuttavia englanteja edustavista netistä kerätyistä teksteistä. Korpus on jaettu alakorpuksiin maan ja tekstityypin mukaan. Tutkimuksessa käytettiin Intian ja Pakistanin alakorpuksia. Kerätty aineisto koostui kaiken kaikkiaan 6362 koodinvaihdosta. Analysoidun aineiston määrä oli puolestaan 5449 koodinvaihtoa – 3650 intianenglannissa ja 1799 pakistaninenglannissa.

Koodinvaihtoja haettiin GloWbE:sta hindin ja urdun kielen yleisimmillä sanoilla. Syynä hindin ja urdun yleisimpien sanojen käytölle oli oletus, että hindissä ja urdussa usein esiintyvät sanat esiintyisivät usein myös koodinvaihdoissa hindiin ja urduun, ja näin niillä voitaisiin paikallistaa koodinvaihtoja hindiin ja urduun englannin kielen korpuksessa. Hindin ja urdun yleisimmät sanat määritettiin Sketch Enginessä olevilla hindin ja urdun korpuksilla. Haut suoritettiin listojen kolmella ensimmäisellä sanalla, jotka olivat sekä hindissä että urdussa postpositioita. Jotta hindin ja urdun sanoilla voitiin suorittaa hakuja GloWbE:ssa, ne täytyi romanisoida. Johtuen suuresta variaatiosta ihmisten romanisointistrategioissa hauissa pyrittiin paikallistamaan sanojen erilaisia kirjoitusvariantteja. Tämä tapahtui korvaamalla sanojen vokaalit wildcard-symboleilla. Konsonantit romanisoitiin ALA-LC-romanisointitaulukoita mukaillen.

Koodinvaihdot kerättiin Excel-tiedostoihin analyysia varten. Kerättyjen koodinvaihtojen ympäröivä tekstikonteksti sisälsi usein muitakin koodinvaihtoja kuin vain sen, joka sisälsi itse hakutermin. Saman kontekstin sisältämiä muita koodinvaihtoja kerättiin mukaan analyysiin sillä periaatteella, että jos virke, jossa hakutermin sisältävä koodinvaihto esiintyi, sisälsi tämän lisäksi muita koodinvaihtoja, virkkeen muut koodinvaihdot sisällytettiin analyysiin.

Kvantitatiivinen analyysi: rakenteet

Hindin ja urdun koodinvaihdot jaettiin rakenteellisesti ensin virkkeiden välisiin ja virkkeensisäisiin koodinvaihtoihin. Virkkeensisäiset koodinvaihdot jaettiin edelleen lauseiden

välisiin ja lauseensisäisiin koodinvaihtoihin. Lauseensisäiset koodinvaihdot kategorisoitiin vielä sanaluokan ja lauseketyypin mukaan.

Virkkeiden välisiä koodinvaihtoja oli intianenglannissa 14,55% kaikista intianenglannin koodinvaihdoista. Pakistaninenglannissa vastaava luku oli 22,46% pakistaninenglannin koodinvaihdoista. Virkkeensisäisiä koodinvaihtoja oli puolestaan intianenglannin koodinvaihdoista 85,45 % ja pakistaninenglannin koodinvaihdoista 77,54%. Intianenglanti ja pakistaninenglanti seurasivat virkkeiden välisten ja virkkeensisäisten koodinvaihtojen määrissä jokseenkin samoja tendenssejä, joskin pakistaninenglannissa oli jonkin verran enemmän virkkeiden välisiä ja intianenglannissa virkkeensisäisiä koodinvaihtoja. Virkkeiden väliset koodinvaihdot olivat joko yksittäisiä kokonaisia virkkeitä tai ne olivat pidempiä, useammasta perättäisestä virkkeestä koostuvia koodinvaihtoja hindiin ja urduun. Virkkeiden välisiksi koodinvaihdoiksi luettiin myös koodinvaihdot kahden rinnasteisen päälauseen välillä.

Virkkeensisäiset koodinvaihdot jaettiin lauseiden välisiin ja lauseensisäisiin koodinvaihtoihin. Virkkeensisäisistä koodinvaihdoista lauseiden välisiä koodinvaihtoja oli intianenglannissa 12,79% ja pakistaninenglannissa 19,50%. Lauseensisäisiä koodinvaihtoja oli intianenglannissa 87,21% ja pakistaninenglannissa 80,50%. Jälleen kerran pakistaninenglannissa oli hieman enemmän pidempiä, lauseiden välisiä koodinvaihtoja, kun taas intianenglannissa oli enemmän lyhyempiä, lauseensisäisiä koodinvaihtoja.

Lauseiden välisiksi koodinvaihdoiksi kategorisoitiin koodinvaihdot pää- ja sivulauseiden välillä sekä muut englannin päälauseisiin upotetut hindin- ja urdunkieliset lauseet. Koodinvaihtoa pää- ja sivulauseiden välillä tapahtui lähinnä englanninkielisten päälauseiden ja hindin- tai urdunkielisten sivulauseiden välillä, vaikkakin joukkoon mahtui muutama tapaus, jossa päälause oli hindin- tai urdunkielinen ja sivulause englanninkielinen.

Lauseensisäiset koodinvaihdot jaettiin sanaluokan ja lauseketyypin mukaan. Lauseensisäisten koodinvaihtojen kategorisointi osoitti, että kolme laajinta kategoriaa olivat samat sekä intian- että pakistaninenglannissa: erisnimet, genetiivilausekkeet ja substantiivit. Erisnimiä oli lauseensisäisistä koodinvaihdoista 73,9% intianenglannissa ja 69% pakistaninenglannissa. Genetiivilausekkeita oli intianenglannissa 9,8% ja pakistaninenglannissa 11,6%. Substantiiveja oli intianenglannissa 7,1% ja pakistaninenglannissa 10%. Muita kategorioita olivat muun muassa nominilausekkeet, postpositiolausekkeet, konjunktiot ja interjektiot. Muiden kategorioiden prosenttiosuudet olivat sen verran pieniä lauseensisäisten koodinvaihtojen kokonaismäärästä, ettei niiden perusteella kykene tekemään vertailua.

Kvalitatiivinen analyysi: funktiot

Koodinvaihtojen käyttöfunktiot analysoitiin kvalitatiivisesti. Koodinvaihdot voitiin jakaa funktioiden perusteella karkeasti kahteen kategoriaan: koodinvaihtoihin, joilla on kommunikatiivinen funktio, ja kulttuurisiin koodinvaihtoihin. Merkittävä koodinvaihtojen kommunikatiivisten funktioiden kategoria olivat lainaukset ja puheen referointi. Lainaukset olivat sekä lainattuja otteita tunnetuista teoksista kuten runot, elokuvien dialogit tai laulujen sanat, että intialaisten ja pakistanilaisten ihmisten puheen esittämistä. Ihmiset saattoivat esimerkiksi lainata intialaisia/pakistanilaisia julkisuuden henkilöitä tai omia perheenjäseniään ja ystäviään vaihtamalla koodia hindiin ja urduun. Myös fiktiivisten henkilöiden hypoteettista puhetta voitiin esittää hindiksi ja urduksi. Lainaukset voivat olla myös geneerisiä, jolloin henkilöä, jolle ääni kuuluu, ei ollut tarkemmin määritelty, vaan ”puhuja” oli jokin ihmisryhmä tai joukko ihmisiä yleisemmin ottaen. Ihmiset esittivät myös omaa puhettaan hindiksi ja urduksi. Yhteinen nimittäjä kaikille lainauksille, joissa esitettiin ihmisten puhetta, oli, että henkilö tai ihmiset, joiden puhetta esitettiin, olivat intialaisia ja pakistanilaisia. Koodinvaihto hindiin ja urduun toimi välineenä paikallisten äänien esittämisessä ja esiin tuomisessa intianenglannissa ja pakistaninenglannissa.

Koodinvaihtoa esiintyi figuratiivisessa kielessä. Figuratiivinen kieli piti sisällään hindin- ja urdunkielisten idiomien ja sanontojen käyttöä. Koodia vaihtamalla ilmaistiin myös tervehdyksiä ja toivotuksia sekä välitettiin siunauksia ja rukouksia. Koodinvaihtoa siunauksia ja rukouksia välitettäessä esiintyi erityisesti islaminuskon kontekstissa. Tämän voidaan nähdä heijastelevan urdun kielen yhteyttä ja assosiaatioita islaminuskoon ja muslimi-identiteettiin Intiassa ja Pakistanissa. Koodinvaihtoa esiintyi myös interjektioissa, jotka olivat esimerkiksi huudahduksia ja diskurssipartikkeleita. Toistoa esitettiin myös koodia vaihtamalla. Toistoa käytettiin sekä selventämään että täsmentämään aiemmin ilmaistua. Toistot voivat olla myös hindin- ja urdunkielisiä käännöksiä siitä, mitä edellä oli englanniksi ilmaistu. Hindiä ja urdua käytettiin myös metakielisissä kommenteissa keskusteltaessa hindin ja urdun kieliopista ja rakenteista. Muita funktioita olivat muun muassa persoonallisen näkökulman ilmaisu, puhuteltavan spesifiointi, sivukommentit, loukkaukset ja sanaleikin välittäminen. Jotkin koodinvaihdot olivat myös selventäviä tai täsmentäviä koodinvaihtoja ilman, että ne olivat toistoa, vaan selventävä tai täsmentävä efekti saavutettiin toisin keinoin.

Kulttuuriset koodinvaihdot ilmaisivat objekteja ja konsepteja, jotka kuuluvat intialaiseen ja pakistanilaiseen kulttuuriin. Kulttuuriset koodinvaihdot toimivat myös viittauksina intialaiseen ja pakistanilaiseen kulttuuriin vihjaten jonkin asian tulkintaa intialaisen ja pakistanilaisen

kulttuurin kontekstissa. Monet kulttuurisista koodinvaihdoista oli kategorisoitavissa sosiolingvistisiin domeeneihin kuten elokuvien, laulujen ja tv-sarjojen nimet, ruoka ja juoma, uskonto, sukulaisuussanat, häät ja avioliitto, intialainen/pakistanilainen musiikki ja runous, juhlat, vaatteet, ja puhuttelumuodot. Osa kulttuurisista koodinvaihdoista kuului sekalaiseen joukkoon erilaisia kulttuurispesifejä ilmaisuja.

Kulttuurisia koodinvaihtoja voitiin käyttää siitä syystä, ettei englannin kielestä löydy hindin/urdun sanalle sopivaa vastinetta – joko niin, ettei vastaavaa sanaa tai ilmaisua löydy englannista ollenkaan tai englannin ja hindin/urdun sanojen välillä on hienoisia merkityseroja, eikä täysin suoraa vastinetta hindin/urdun sanalle löydy englannin kielestä. Saattoi esimerkiksi olla, että hindin/urdun sana oli semanttisesti tarkempi kuin englannin vastaava ilmaus, minkä vuoksi hindi/urdun sanaa käytettiin. Leksikaalisten motivaatioiden lisäksi kulttuurisia koodinvaihtoja käytettiin viittauksina intialaiseen ja pakistanilaiseen kulttuuriin. Hindin/urdun sanat tai käsitteet joko pitivät sisällään kulttuuriin vahvasti sidonnaisia konnotaatioita tai niillä saatiin aikaan mielleyhtymiä intialaiseen ja pakistanilaiseen kulttuuriin. Hindin/urdun koodinvaihdoilla voitiin vihjata jonkin asian tulkintaa Intian/Pakistanin kontekstissa. Koodinvaihdolla tuotiin myös esille jonkin asian – joko esineen tai ihmisen – intialaisuutta tai pakistanilaisuutta.

Loppuyhteenveto

Tiivistäen voidaan todeta, että hindin ja urdun koodinvaihtoja esiintyi rakenteellisesti monenlaisissa eri muodoissa. Koodinvaihtojen muodot vaihtelivat pidemmistä virkkeiden välisistä koodinvaihdoista virkkeensisäisiin koodinvaihtoihin, jotka puolestaan voivat olla sekä lauseiden välisiä koodinvaihtoja, eli koodinvaihtoja pää- ja sivulauseiden välillä, että lauseensisäisiä koodinvaihtoja, kuten yksittäisiä sanoja tai lausekkeita. Eri rakenteiden osuudet koodinvaihtojen kokonaismäärästä olivat pitkälti samassa suhteessa toisiinsa sekä intian- että pakistaninenglannissa.

Kuten jo edellä mainittiin, funktioiden perusteella koodinvaihdot voitiin karkeasti jakaa kommunikatiivisiin ja kulttuurisiin koodinvaihtoihin. Jako koodinvaihtoihin, joilla on kommunikatiivinen funktio, ja kulttuurisiin koodinvaihtoihin ei ole missään nimessä absoluuttinen, eikä tarkoita, etteikö kulttuuriset koodinvaihdot olisi kommunikatiivisesti merkittäviä, tai etteikö kommunikatiivisen funktion omaavat koodinvaihdot olisi toisinaan myös kulttuurisidonnaisia. Kulttuuristen koodinvaihtojen funktiona voidaan katsoa olevan toimiminen viittauksina intialaiseen ja pakistanilaiseen kulttuuriin, ja toisaalta esimerkiksi koodinvaihto urduun rukouksia ja siunauksia välitettäessä voidaan nähdä heijastelevan urdun

yhteyttä ja assosiaatioita islaminuskoon ja muslimi-identiteettiin Intiassa ja Pakistanissa, minkä vuoksi kyseisten koodinvaihtojen voidaan nähdä olevan yhteydessä kulttuuriin. Tämä jako kommunikatiivisiin ja kulttuurisiin koodinvaihtoihin ainoastaan heijastelee niitä suuria linjoja, jotka aineistossa voitiin erottaa. Koodinvaihtojen funktioiden analyysissä voitiin toisinaan havaita, kuinka kielten funktiot makrotasolla näkyivät niiden funktioissa mikrotasolla. Esimerkiksi perheenjäsenten äänten referointi voitiin analysoida linkittyvän hindin ja urdun käyttöön kommunikaation kielenä perheenjäsenten kesken. Koodinvaihtoa käytettiin myös ilmaisemaan paikallisuutta sekä paikallisten intialaisten ja pakistanilaisten ihmisten äänten esittämisessä että monissa kulttuurisissa koodinvaihdoissa. Löydökset kulttuuristen koodinvaihtojen kuulumisesta tiettyihin sosiokulttuurisiin kategorioihin tuki aiempien tutkimusten löydöksiä hindin ja urdun koodinvaihtojen sidonnaisuudesta tiettyihin sosiokulttuurisiin kategorioihin.

Tämä tutkimus toimi myös metodologisena kokeiluna, jossa hindin ja urdun koodinvaihtoja etsittiin GloWbE:sta käyttäen hindin ja urdun frekventeimpiä sanoja. Yleisesti metodista voi sanoa, että hindin ja urdun yleisimpien sanojen käyttö osoittautui tehokkaaksi menetelmäksi koodinvaihtojen paikallistamiseen, vaikkakin yleisesti käytössä olevien romanisointistandardien puuttuminen loi omat haasteensa. Hakutermeinä käytetyt hindin ja urdun sanat vaikuttivat osaltaan tuloksiin. Esimerkiksi se, että kaksi kolmesta hakuterminä käytetystä frekvenssilistan sanasta oli genetiivipostpositioita, näkyi genetiivilausekkeiden määrässä, ja haut kielioppisanoilla johtivat yksittäisten sanojen, kuten substantiivien ja adjektiivien, verrattain marginaaliseen esiintymiseen aineistossa. Kuitenkin kielioppisanoilla hakeminen johti todennäköisesti paljon laajempaan skaalaan erilaisia koodinvaihtoja kuin leksikaalisilla sanoilla hakujen suorittaminen johtaisi.

Tämä tutkimus täytti aukon aiemmassa tutkimuksessa tarkastelemalla hindin ja urdun koodinvaihtoja intian- ja pakistaninenglannissa laajemmin kuin vain keskittyen yksittäisiin sanoihin. Koodinvaihtoa hindiin ja urduun intian- ja pakistaninenglannissa voitaisiin tutkia lisää esimerkiksi erilaisella metodilla, jolla ei olisi samalla tavalla vaikutusta esiintyvien koodinvaihtojen rakenteisiin, jolloin saataisiin tietoa eri koodinvaihtorakenteiden keskinäisistä suhteista.