e-ISSN 2785-5953

Bhasha
Vol. 4 — Num. 1 — April 2025

Code-Switching in South Asia: Comparing the Equivalence Constraint and Matrix Language Frame Models with Hinglish

Anmol Sahni
Emory University, USA

Abstract This article examines code-switching in South Asia by applying Shana Poplack's Equivalence Constraint Model and Carol Myers-Scotton's Matrix Language Frame Model to Hinglish, a hybrid of Hindi and English. Through syntactic and sociolinguistic analysis, it argues that the Matrix Language Frame Model better accounts for South Asian patterns of intra- and inter-sentential switching. The article also advocates translanguaging in English classrooms, linking structural theory to pedagogical practice in India's multilingual context.

Keywords Code-switching. Hinglish. South Asia. Equivalence constraint model (ECM). Matrix language frame model (MLFM). Transformational generative grammar. Bilingualism.

Summary 1 Introduction. – 2 Aims and Objectives. – 3 Literature Review of CS with an Illustrative Hinglish Example. – 4 Sociolinguistic and Structural Approaches to Code-Switching. – 5 Structural Dimensions and Constraint Approaches to Code-Switching. – 6 The Equivalence Constraint Model. – 7 Matrix Language Frame Model. – 8 Conclusion.

Edizioni Ca'Foscari

Peer review

Submitted 2025-03-25
Accepted 2025-05-07
Published 2025-07-01

Citation Sahni, A. (2025). "Code-Switching in South Asia: Comparing the Equivalence Constraint and Matrix Language Frame Models with Hinglish". *Bhasha*, 4(1), 51-76.

DOI 10.30687/bhasha/2785-5953/2025/01/003

1 Introduction

Over the past three and a half decades (1990-2024), scholarly interest in code-switching (CS) has grown significantly, challenging earlier negative perceptions of language mixing. This phenomenon is particularly salient in the Indian context, where Hindi, the dominant language of northern India, coexists with English – a language shaped by its colonial past and its global significance as the language of trade and commerce. Hindi is primarily spoken across the northern Indian states collectively referred to as the 'Hindi Belt', which includes Uttar Pradesh, Bihar, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, and Rajasthan. Within this region, Hindi encompasses various dialects that reflect its linguistic diversity. For example, the Awadhi dialect spoken in Lucknow, the capital of Uttar Pradesh, is distinct from the Bhojpuri dialect prevalent in Bihar.

The 2001 Census of India reported that Hindi was spoken by approximately 422 million individuals, including speakers of various dialects classified under Hindi (Koul 2008, 1). By the 2011 Census, this figure had risen to approximately 528 million, reflecting a significant increase in the demographic representation of Hindi speakers. Beyond India, the online journal *Ethnologue*[1] identifies countries with substantial Hindi-speaking populations due to immigration, including Nepal, Australia, Bhutan, Canada, Mauritius, Germany, and the United States, among others.

Hindi is traditionally classified into two main groups: Eastern Hindi and Western Hindi. Eastern Hindi includes dialects such as Awadhi, Bagheli, and Chhattisgarhi, while Western Hindi encompasses dialects like Haryanvi, Brajbhasha, Bundeli, Kanauji, and Khariboli. Additional dialects such as Maithili, Bhojpuri, and Magahi in Bihar, and Marwari, Jaipuri, and Malvi in Rajasthan, contribute to Hindi's linguistic richness. Some dialects from northwestern Uttar Pradesh and Himachal Pradesh, previously excluded from earlier classifications, are now recognized under the broader category of 'Hindi'. As one of India's official languages, Hindi serves as a vital medium for communication among diverse communities. Its historical roots in the Indo-Aryan language family are intertwined with the socio-cultural and political developments of the region. Over centuries, Hindi has been shaped by languages such as Sanskrit, Persian, Arabic, and, more recently, English, reflecting its adaptability to the dynamic landscape of Indian society.

Hindi's literary tradition is equally diverse, spanning religious poetry, prose, dramas, and novels written in dialects like Awadhi, Braj Bhasha, and Khariboli. Foundational figures like Kabir (fifteenth

1 *Ethnologue*: https://www.ethnologue.com/.

century) and Tulsidas (sixteenth century), deeply influenced by the Bhakti movement, laid the groundwork for a rich literary heritage. Kabir's *dohas* (couplets) and Tulsidas's *Ramcharitmanas* in Awadhi not only exemplified devotional literature but also elevated vernacular Hindi as a medium for artistic and spiritual expression. This tradition evolved further in modern Hindi literature, marked by luminaries such as Munshi Premchand, known for his social-realist novels (*Godaan*), Jaishankar Prasad, celebrated for his prose and dramas (*Chandragupta*), and Ramdhari Singh Dinkar, revered for his patriotic and philosophical poetry. During the colonial era, Hindi literature emerged as a powerful medium for resisting colonial hegemony, with nationalist themes prominently featured in the works of Bharatendu Harishchandra and Maithili Sharan Gupt.

The history of English in India parallels this complex and rich evolution of Hindi, originating during British colonization when it became a language of administration and education. Over time, English developed into a distinct variety known as Indian English, marked by unique phonetic, grammatical, and lexical features shaped by local languages, particularly Hindi and other regional tongues. This blend of native elements with English reflects India's diverse cultural and social fabric. In contemporary India, Hindi functions as a lingua franca for many speakers, while English, often used in formal contexts, education, and media, complements this linguistic landscape.

The interplay between Hindi and English has given rise to Hinglish – a hybrid language that fuses elements of both languages. The linguistic tapestry of Hinglish not only facilitates communication among diverse groups but embodies the speakers' cultural identities and social affiliations. Despite Hinglish's rich character and the insights it provides into code-switching and sociolinguistic dynamics, research on Hinglish remains limited. This article aims to deepen the understanding of Hinglish by exploring the structural principles underlying code-switching through the frameworks of Shana Poplack's Equivalence Constraint Model (ECM) and Carol Myers-Scotton's Matrix Language Frame Model (MLFM). By examining Hinglish, this study seeks to highlight its significance within the broader context of bilingualism and the sociolinguistic complexities of contemporary India.

2 Aims and Objectives

Research on code-switching has evolved into two primary approaches: the sociolinguistic approach, which examines code-switching at the discourse level, and the structural or syntactic approach, which focuses on code-switching below the sentence level. This article provides a comprehensive literature review of generativist approaches to code-switching, with a particular focus on two influential theories: Shana Poplack's Equivalence Constraint Model (ECM), developed from a Spanish-English corpus (Poplack 1980), and Carol Myers-Scotton's Matrix Language Frame Model (MLFM), based on Swahili-English CS data (Myers-Scotton 1993).

The article in part three begins by exploring key debates and challenges that animate CS research, including the difficulty of establishing a singular, definitive definition of CS. Part four of the article outlines the sociolinguistic context of CS research, tracing how the increasing body of work in this field has led to the development of structural approaches to CS. Part five then contextualizes these structural approaches – specifically, constraint-based theories of CS – within Noam Chomsky's influential theory of transformational generative grammar. Parts six and seven apply the ECM and MLFM to Hinglish, providing insights into the structural dynamics of code-switching and a comparative evaluation of the models' effectiveness in analyzing Hinglish. Finally, the article concludes by pointing to future directions for the development of CS research.

The primary objectives of my analysis are twofold: (a) to analyze Hinglish through the lenses of both ECM and MLFM; and (b) to demonstrate how ECM and MLFM contribute to the scholarship on bilingualism and code-switching, particularly in South Asia. Overall, the article seeks to address the underexplored application of the ECM and the MLFM to Hinglish within CS research, thus bridging a significant gap in the field.

3 Literature Review of CS with an Illustrative Hinglish Example

3.1 Challenges in Defining Code-Switching

Despite the significant contributions of scholars such as Ad Backus (1992), Susan Berk-Seligson (1986), Carol Myers-Scotton (1993a), and Shana Poplack (1980), there remains a lack of consensus regarding the definition and governing principles of CS. Poplack (1980) characterizes CS as the alternation between two languages within a discourse, sentence, or constituent, resulting in a hybrid utterance. Poplack describes CS as the

> alternation of two languages within a discourse, sentence or constituent […] [CS] is categorized according to the degree of integration of items from one language (L1) into the phonological, morphological, and syntactic patterns of the other language (L2). (Poplack 1980, 583)

The challenge of providing a singular, definitive explanation of CS is compounded by the ongoing debate among linguists regarding its distinction from lexical borrowing. Two primary viewpoints emerge: one group, including Poplack and her colleagues at the University of Ottawa Sociolinguistics Laboratory, advocates for differentiating between isolated foreign-language items and more extensive instances of code-switching. Conversely, Myers-Scotton and her supporters argue against treating CS and lexical borrowing as separate phenomena within bilingualism.

In her work *Borrowing* (2018), Poplack meticulously delineates code-switching, loanwords, and nonce borrowings as distinct phenomena of language mixing, based on their synchronic or diachronic manifestation in a speech community. Nonce borrowing refers to the temporary use of a word from another language on an ad-hoc basis to express a particular idea, without it becoming a permanent part of the borrowing language. For example: "I start my day with *chai* (tea) and *samachar* (news)". Here, *chai* and *samachar* are Hindi words temporarily borrowed to be used in an otherwise English sentence to convey specific meanings in the context of the sentence. In contrast, loanwords are more permanent lexical borrowings that result from sustained cultural contact. A good example is the anglicized term cummerbund, which can be traced back to its origins in Hindustani and Persian. Over time, cummerbund has been fully adopted into English and is commonly used without reference to its foreign roots. Furthermore, Poplack categorizes established loanwords – such as the English word 'country', which etymologically originates from French, according to

the Oxford English Dictionary – as a diachronic phenomenon. Poplack differentiates loanwords from CS by emphasizing that loanwords undergo a process of diachronic integration, wherein they lose their original linguistic attributes and assimilate into the morphology and, often, the phonology of the recipient language.

3.2 Distinction Between Code-Switching and Lexical Borrowing

Poplack argues that the fundamental distinction between CS and lexical borrowing lies in their structural integration. CS primarily involves intra-sentential mixing at the constituent level, where the grammar of the donor language is operational. Conversely, lexical borrowing involves the integration of words from a donor language into a recipient language, where the grammar of the recipient language is adhered to, maintaining grammaticality (Poplack 1980).

Poplack provides a framework for categorizing these phenomena, as illustrated in Figure 1 of her 1980 publication titled "Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL" **[fig. 1]**.

Type	Levels of Integration Into Base Language			Code–Switching?
	Phonological	Morphological	Syntactic	
1	√	√	√	No
2	×	×	√	Yes
3	√	×	×	Yes
4	×	×	×	Yes

Figure 1 Poplack's (1980) analysis of code-switching using threefold criteria of lexicon integration into the base language

Poplack evaluates the integration of a foreign word into the base language according to three criteria: phonological, morphological, and syntactic integration. In Figure 1, she identifies 'Type 1' cases, which do not qualify as CS but as loanwords. These instances involve foreign lexicon from (L2) that have been fully integrated into the phonology, morphology, and syntax of the recipient language (L1), shedding their donor language properties **[fig. 1]**. In contrast, cases classified as 'Type 2' to 'Type 4' are considered instances of CS due to their partial integration into the base language. In other words, CS occurs when a speaker alternates between two or more languages within the same conversation or sentence, but the switch remains consistent with the grammatical rules of each language.

For example, speaker code-switching between Spanish and English is demonstrated as follows:

I was talking to María, y ella me dijo que v. a la tienda.
'I was talking to María, and she told me she's going to the store.'

In this sentence, the speaker switches from English to Spanish without blending the grammar of the two languages. The two languages – namely, Spanish and English – are clearly distinguished. In contrast, Poplack would consider an utterance a case of borrowing when a word from one language is adopted into another language and is often integrated into the borrowing language's grammar and pronunciation. The borrowed word might be slightly modified to fit the phonological or grammatical rules of the new language, as evidenced in the following example:

Voy a parquear el coche.
'I am going to park the car.'

In this second example, the word 'parquear' is a borrowing from the English verb 'to park', but it has been adapted to Spanish grammar by taking on the Spanish verb ending '-ar'. Over time, such borrowings become fully integrated into the borrowing language.

Figure 2, as adopted from Poplack's coauthored work (1989), further distinguishes between established loanwords and nonce borrowings by examining parameters such as the degree of phonological integration and the role of the lexicon as a content word **[fig. 2]**. This nuanced approach highlights the complexity of language-mixing phenomena and underscores the importance of context in understanding their manifestation and classification.

Figure 2 The continuum for levels of borrowing in code-switching utterances (Poplack, Wheeler, Westwood 1989)

3.3 Divergent Perspectives on Code-Switching

Poplack's conceptualization of code-switching, characterized by the partial integration of phonological, morphological, and syntactic elements, contrasts sharply with Myers-Scotton's perspective, which challenges the differentiation between CS and lexical borrowing as distinct phenomena within bilingualism. Myers-Scotton has described

code-switching as "the selection by bilinguals or multilinguals of forms from an embedded language (or languages) in utterances of a matrix language during the same conversation" (1993b, 4).

Myers-Scotton's model posits that the languages or codes involved in CS hold equal status but fulfill distinct roles within the process, specifically within the framework of the Matrix Language (ML) and the Embedded Language (EL). According to this model, the ML is responsible for establishing the morpho-syntactic structure of the clause and supplying both the system morphemes and the majority of the content morphemes. In contrast, the EL contributes by inserting constituents and individual content morphemes into the ML's structural framework. Myers-Scotton thus asserts that "there is no need to make the borrowing vs. code-switching distinction" (2002, 153). This position of Myers-Scotton is reaffirmed in a later work, where she argues that both CS and borrowing "undergo the same morphosyntactic procedures during language production" (Myers-Scotton 2003, 8). This foundational divergence in the definition of CS is critical, as it shapes the subsequent discussion of the ECM and the MLFM, developed by Poplack and Myers-Scotton, respectively.

3.4 Code-Switching in Hinglish

An example of CS, as defined by both Poplack and Myers-Scotton, is demonstrated in 1(a), where the speaker alternates fluidly between two typologically distinct languages: Hindi, which follows an SOV word order, and English, characterized by an SVO word order. This type of language mixing is commonly referred to in the literature as Hinglish, denoting the blending of Hindi and English during conversation. Similar to other contact-blend languages such as Franglais (a mixture of French and English, prevalent in Canada), Taglish (a blend of Tagalog and English, common in the Philippines), and Portunhol (a mixture of Portuguese and Spanish), Hinglish incorporates both intra-sentential CS and borrowed loanwords from Hindi and English.

1 (a) *Bhai yeh tera hard - earned money hai isliye isko udana mat.*
*1 (b) Bhai this your *hard - earned money isliye ise udana mat.*
'Brother this is your hard-earned money therefore don't squander it.'

While (1a) illustrates a typical example of Hinglish, (1b) is deemed ungrammatical due to its violation of certain implicit rules or constraints governing CS, which will be discussed in greater detail later in the article. This example highlights that, despite appearing spontaneous or arbitrary, code-switched sentences are not random

assemblages of words from different languages. Instead, CS follows systematic patterns, occurring at intra-sentential sites – specific points within a sentence, such as major constituency breaks – and at inter-sentential sites, where switching occurs at the discourse level. Although bilingual speakers often engage in inter-sentential code-switching, this type of switching has received relatively less attention from syntacticians compared to intra-sentential mixing, as syntactic analysis typically focuses on grammatical relationships within the sentence.

4 Sociolinguistic and Structural Approaches to Code-Switching

This section provides a historical overview of code-switching research by examining the development of two parallel approaches: sociolinguistic and structural. The structural approach focuses on the morphosyntactic constraints governing CS, addressing issues related to the grammaticality of code-switched utterances. In contrast, the sociolinguistic approach explores the functional aspects of CS at the discourse level, elucidating how meaning is constructed when two languages are juxtaposed within a conversation. Historically, both approaches have evolved significantly; whereas early studies of bilingualism often viewed CS as indicative of a speaker's linguistic inadequacy, contemporary perspectives recognize it as a demonstration of proficiency in the grammars of the involved languages.

4.1 Historical Perspectives on Code-Switching

In the early stages of bilingualism research, beginning around the 1950s with foundational works such as Weinreich's *Languages in Contact* (1953), CS was largely viewed as detrimental to the grammar of mixed languages. This perspective was endorsed by prominent linguists like Leonard Bloomfield and Uriel Weinreich. During this early period, Weinreich distinguished between the "ideal bilingual" and the "imperfect bilingual", reflecting the prevailing erroneous attitudes towards grammaticality concerning research on CS. According to Weinreich, an ideal bilingual was a speaker who seamlessly switched languages based on changes in the speech context (e.g., interlocutors, topics) without disrupting the flow of conversation (1953). In contrast, the imperfect bilingual was perceived as someone with uneven proficiency across the languages being switched.

Weinreich's dichotomy between ideal and imperfect bilingualism, arising from the lack of a theoretical framework for CS grammar, contributed to the development of pejorative attitudes towards CS. Notable among these were the concepts of semilingualism, the deficit hypothesis, and prescriptivism. Semilingualism posited that CS indicated a bilingual's lack of proficiency in the languages being used, suggesting that mid-sentence switches reflected linguistic inadequacy. The deficit hypothesis, closely related to prescriptivism, asserted that speakers of non-standard dialects, such as African American English (AAE), were linguistically deficient and incapable of conveying complex ideas. In tandem, prescriptivism endorsed the supremacy of standard dialects, like Standard American English (SAE), and promoted linguistic uniformity within the education system, marginalizing non-standard dialects and reinforcing hierarchical views of language competence.

4.2 Code-Switching in Postcolonial and Multilingual India: The Case of Hinglish

From the 1950s to the 1990s, the absence of a robust theoretical framework for grammar in CS led many linguists to perceive CS as a threat to the linguistic integrity and grammatical structures of blended languages. This view, grounded in the erroneous belief that CS negatively impacts the grammars of the involved languages, prompted some prescriptive grammarians to caution against CS, fearing it would lead to language convergence and the bastardization of grammatical norms. This negative perception of CS was compounded by widespread linguistic insecurity concerning the lingua franca of newly formed nation-states during decolonization. For instance, post-1947, after India achieved independence from British colonial rule, there were a series of public and intellectual debates that attempted to renegotiate the hegemonic status of English in India. Notable among these debates was the controversy surrounding the 'three-language formula' introduced by the Indian central government under the 1968 National Policy of Education. This policy mandated that English be taught alongside Hindi and a modern Indian language to bridge the gap between the educated elite and the broader population (Krishnaswamy 2006).

Krishnaswamy observes that

> recognizing the role of English in globalization, India has shed its colonial complexes towards English and has come to terms with the language; Indians have separated the English language from the English. (2006, 153)

Consequently, Indian English has been adopted as one of India's official languages alongside Hindi. Recent research (Kothari 2011) indicates that code-switching between Hindi and English, exemplified by Hinglish, is a natural linguistic phenomenon facilitated by the bilingual nature of the Indian speech community. There is no empirical evidence to suggest that the use of Hinglish has negatively impacted the grammars of English and Hindi. Instead, Hinglish has enriched the linguistic repertoire of its speakers while adhering to the constraints of grammaticality associated with CS, as this paper demonstrates.

4.3 The Ebonics Debate's Influence on Sociolinguistic Views of Code-Switching

In the 1990s, while linguistic prescriptivism was increasingly being discredited as a valid theoretical framework, it continued to maintain marginal influence, particularly among linguists who prioritized the notion of linguistic 'purity'. From a contemporary perspective, this insistence on purity can be understood as a form of linguistic racism. As prescriptivism's credibility waned, descriptivism gained widespread acceptance within the linguistic community. This transition signified a shift from prescribing rigid linguistic norms to objectively identifying and documenting the intra-sentential sites of CS within speech communities.

The paradigmatic shift towards descriptivism can be traced back to the 1960s, with the pioneering work of linguists such as William Labov. Labov and others challenged the prevailing attitudes toward AAE, which had previously been treated as a deviation from SAE. Instead, AAE came to be recognized as a legitimate dialect with its own grammatical rules. The descriptivist approach in U.S. linguistics fostered research on code-switching between AAE and SAE, further establishing the legitimacy of AAE as a dialect.

A key moment in this shift was the Ebonics debate, sparked by the 1996 resolution passed by the Oakland, California Board of Education (CBE). This resolution called for schools across the U.S. to recognize "Ebonics" as the "primary language" of African American students, with the goal of enhancing their proficiency in SAE (Young 2014, 40). In response to this resolution, a wave of opposition emerged, with critics dismissing AAE as an improper or 'slang' version of English. This ignited a national debate in which both linguists and legislators grappled with the question of whether AAE constituted a separate dialect or merely a corrupted form of standard English. Central to the argument for recognizing AAE as a distinct dialect was Labov's seminal 1972 work Language in the Inner City. Labov testified before a Senate subcommittee in January 1997, at the height of the Ebonics

controversy, where he presented evidence that AAE is indeed a distinct dialect of English. Labov's research demonstrated that AAE possesses unique phonological and grammatical features, such as negative concord and contractions and deletions in the copula (e.g., "Sam here" instead of "Sam is here") (Labov 1972, 65). These findings provided a robust defense of AAE's systematic and rule-governed nature, solidifying its status as a legitimate variety of English. From the perspective of research into the social and structural role of CS, Labov was one of the first to suggest that CS between AAE and SAE should not be viewed as a linguistic deficit but rather as a manifestation of bilingualism. In his testimony, Labov advocated for educators to support and cultivate their students' proficiency in AAE, rather than stigmatizing the natural alternation between AAE and SAE. Labov's stance marked an important shift in recognizing the legitimacy of language mixing as part of a broader sociolinguistic competence.

In subsequent years, U.S. linguists further developed Labov's ideas, with sociolinguist Suresh Canagarajah playing a prominent role in popularizing CS and other forms of language mixing, such as translanguaging and code-meshing, as effective tools for rhetorical education (Canagarajah 2013). In his work, Canagarajah has conceptualized translanguaging as the ability of a speaker to seamlessly transition between languages. According to Canagarajah (2011), translanguaging can serve as a pedagogical approach that encourages and supports this linguistic fluidity in educational settings. Through translanguaging, students are able to engage in cognitive processes across multiple languages concurrently, utilizing their home language as a tool for acquiring and mastering academic English. Canagarajah's work builds upon the legacy of linguists like Labov, emphasizing the pedagogical value of harnessing multilingual students' full linguistic repertoire.

4.4 Hinglish as a Catalyst for Translanguaging in the Multilingual Context of India

As code-switching between Hindi and English, Hinglish naturally promotes translanguaging by enabling speakers to draw from their entire linguistic repertoire for communication and learning. By combining vocabulary, syntax, and idiomatic expressions from both languages, Hinglish allows speakers to convey nuanced meanings and cultural references that may be inaccessible in either language alone. If fostered in pedagogical settings and classrooms, Hinglish can promote inclusivity by creating a shared communicative space in the multilingual context of India, bridging gaps between speakers with varying proficiencies in Hindi, English, and other marginalized

regional languages. Hinglish can further facilitate learning in classrooms by integrating familiar linguistic elements with complex or challenging content, fostering active engagement and deeper understanding. Since Hinglish challenges linguistic hierarchies that privilege English over regional languages like Hindi, the use of code-switching in the form of Hinglish can ultimately promote a more equitable linguistic environment central to translanguaging pedagogy. By embracing Hinglish, educators and learners can create flexible, inclusive, and culturally relevant frameworks for communication and education.

4.5 From Sociolinguistic Foundations to Generative Approaches in Code-Switching

In addition to Labov's contributions, the pioneering research of Myers-Scotton significantly advanced the public and scholarly understanding of CS. Through works such as Social Motivations for Code-Switching (1993b), which focused on multilingualism in African contexts, Myers-Scotton demonstrated that CS is a skilled and purposeful linguistic practice in which bilinguals strategically draw upon their linguistic resources from different speech communities. Myers-Scotton's research highlighted the social and communicative motivations behind CS, positioning it as a sophisticated form of language use rather than a sign of linguistic deficiency. The sustained growth of sociolinguistic scholarship on CS led to a parallel expansion of research into its grammatical structures. Early studies by sociolinguists, particularly between the 1950s and 1970s, sought to challenge the prevailing negative attitudes toward CS, which often regarded language mixing as indicative of a lack of proficiency in either language. From the 1990s onward, researchers such as Poplack (1980) and Backus (1992) contributed to the development of structural approaches to CS. Their work helped to neutralize linguistic prejudices against code-switching by showing that it follows systematic grammatical rules and reflects a high degree of linguistic competence.

4.6 Generative Grammar's Influence on Structural Studies of Code-Switching

In light of the significant influence of generative syntax, recent scholarship has proposed conflicting constraint models of CS as part of the expanding body of structural approaches to CS. Before applying the ECM and the MLFM to analyze Hinglish data, it is essential to address three key inquiries that have driven much of

the research in this area. These questions, outlined by Bullock (2009), are central to the generative approach to CS: (a) How should CS be defined, and should nonce borrowing be treated as distinct from CS? (b) What rules, if any, constrain CS? (c) How can a constraint model for CS be formulated in alignment with generativist grammar theory? The prevailing trend in CS scholarship is to develop a constraint model capable of accounting for the spontaneous and dynamic nature of bilingual language use. Among these models, Poplack's ECM initially gained prominence. However, it was later challenged by Myers-Scotton, whose research on Swahili-English CS identified instances that could not be explained by ECM. In response, Myers-Scotton introduced the MLFM as an alternative framework to account for these linguistic phenomena. The following section defines both the ECM and MLFM models, beginning with a detailed explanation of ECM. Subsequently, the section applies both models to the analysis of a representative Hinglish sentence, contrasting their explanatory power in this context.

5 Structural Dimensions and Constraint Approaches to Code-Switching

5.1 Historical Foundations of the ECM

The publication of the ECM by Poplack in 1978 marked a pivotal development in the study of CS. The ECM's major contribution was its explanation of CS within the framework of universal grammar, applying the same principles that govern monolingual syntax to bilingual speech. This represented a significant departure from earlier explanations of CS, which focused primarily on surface-level grammatical structures where the syntax of the two languages coincided, thereby permitting a switch. Such approaches lacked a theoretical basis grounded in the universal principles of generative grammar, as advanced by Chomsky (1965).

Prior to the ECM, most CS research analyzed language alternation at the discourse level, providing a macro-level understanding of how blending languages could create meaning but offering little insight into the micro-level grammatical constraints that governed permissible switch sites within a sentence. For instance, in Discourse Strategies, Gumperz (1982) proposed that CS was impossible in specific syntactic environments, such as between verbs and pronominal subjects or between conjunctions and conjuncts. While Gumperz's work was instrumental in offering a discourse-based analysis of CS, his treatment of the grammatical aspects of CS remained limited and lacked the formal rigor found in generative approaches. Poplack's ECM filled this gap by aligning CS analysis with generativist syntactic

theory, allowing for a more nuanced and theoretically grounded understanding of the constraints governing bilingual speech. This shift enabled researchers to move beyond descriptive accounts of language alternation toward a more systematic explanation of CS rooted in universal grammar.

5.2 Generativist Foundations of the ECM

Building on Chomsky's transformational-generative theory, first introduced in *Aspects of the Theory of Syntax* (Chomsky 1965), Poplack's ECM pioneered a constraint-driven approach to studying CS. Two key insights from Chomsky's work have become foundational in contemporary CS scholarship (Newmeyer 1986): (a) all human languages are governed by constraints, and (b) syntax analysis consists of two components: base structure and transformational rules. Figure 3, reproduced from Aspects (Chomsky 1965), illustrates Chomsky's framework for syntax analysis **[fig. 3]**.

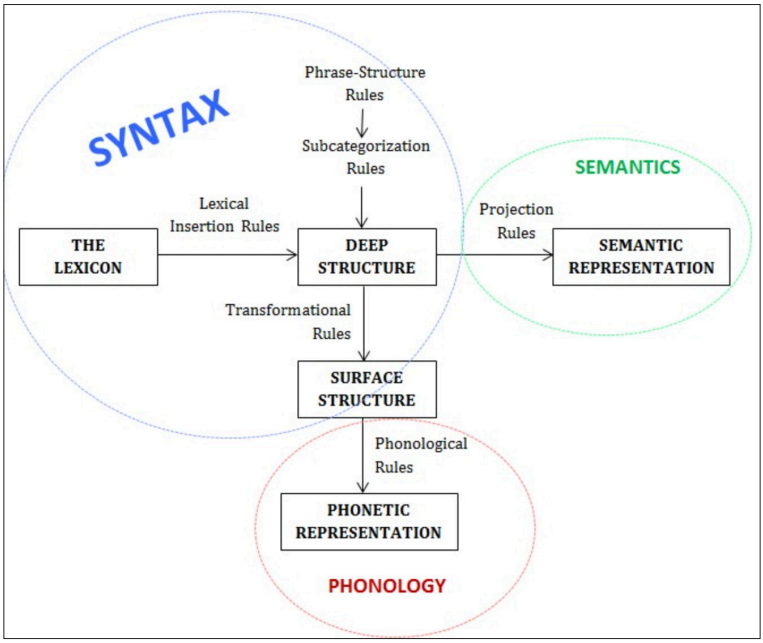

Figure 3 The generativist-transformational model presented in Chomsky 1965

Explaining this framework, Chomsky wrote:

> The syntactic component consists of a base and a transformational component. The base, in turn, consists of a categorial subcomponent and a lexicon. The base generates deep structures. A deep structure enters the semantic component and receives a semantic interpretation; it is mapped by transformational rules into a surface structure, which is then given a phonetic interpretation by the rules of the phonological component. (1964, 151)

The impact of Chomsky's transformational model on CS research is notable. Chomsky's generative-transformational model introduced the possibility of analyzing bilingual speech through finite syntactic rules and constraints, applicable across all languages, regardless of their typological differences or surface structures. The emergence of all constraint models of CS, including Poplack's ECM and Myers-Scotton's MLFM, can be traced back to the Chomskyan revolution in linguistics. Despite their differences in methodology, both models are predicated on the linear and hierarchical analysis of sentence structures. The following section first discusses ECM with an illustrative example and then elucidates the workings of MLFM.

6 The Equivalence Constraint Model

Poplack defines the ECM and its related Free Morpheme Constraint as follows:

- The equivalence constraint states that code-switched sentences are made up of concatenated fragments of alternating languages, each of which is grammatical in the language of its provenance. The boundary between adjacent fragments occurs between two constituents that are ordered in the same way in both languages, ensuring the linear coherence of sentence structure without omitting or duplicating lexical content (Poplack 2001, 2062).
- The Free Morpheme Constraint: Codes may be switched after any constituent in discourse provided that the constituent is not a bound morpheme (Poplack 1980, 585).

For example, code-switching between 'eat', an English verb stem, and '-iendo', the Spanish present progressive, is impossible unless the former is phonologically integrated into Spanish (Sankoff Poplack 1981). In summary, ECM suggests that CS only occurs at those sites where the surface order of constituents surrounding the switch point aligns with the order in both participating languages. Poplack

provides an example of a sentence that is considered ungrammatical due to its violation of ECM.

Example	**Gloss**	**Translation**
*told *le*	told to-him	"(I) told him".
le told	to-him I-told	
him *dije*	him I-told	
dije him	I-told him	(Poplack 1981, 176)

6.1 Application of ECM to Hinglish

The article now applies ECM to analyze the Hinglish sentence 2(a), which comprises two clauses conjoined by the coordinating conjunction 'and'. The sentence 2(a) reads as follows:

2 a Ram highly qualified *hai* and he can find a job *poori duniya mai.*
'Ram is highly qualified, and he can find a job anywhere in the world'.

This sentence contains two instances of CS, each occurring within one of the clauses. To analyze these CS instances, we can represent the sentence using two distinct phrasal trees: Tree 1 and Tree 2. Each tree separately depicts the grammatical structure of the clauses surrounding the CS points, allowing us to evaluate whether the switches conform to ECM's stipulations. This approach enables a detailed examination of how the CS instances align with the syntactic rules of the participating languages and whether they maintain the linear coherence required by ECM.

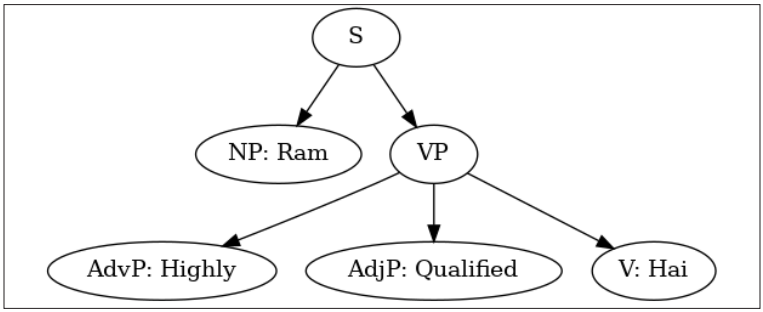

Figure 4 Tree 1

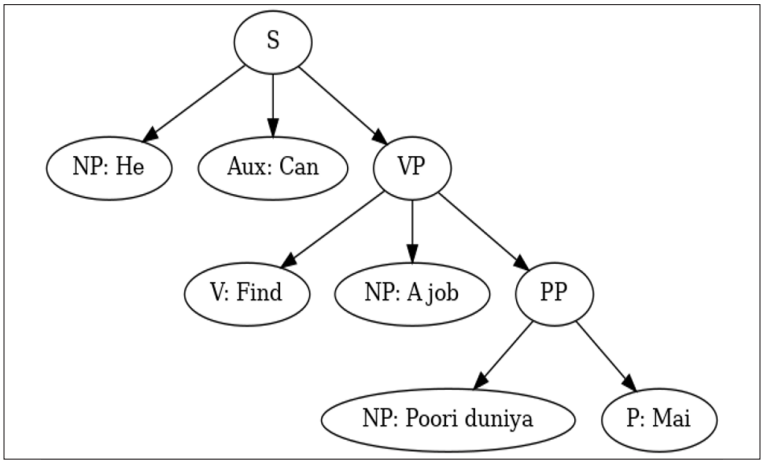

Figure 5 Tree 2

6.1.1 Analysis of the First Clause (Tree 1)

The sentence "Ram highly qualified hai" provides a compelling example of CS within the framework of ECM. In this instance, the speaker incorporates the English phrase "highly qualified" into an otherwise Hindi sentence. This switch occurs at the level of the Adjective Phrase (ADJP), where the integration of English elements is seamless due to the structural compatibility between Hindi and English. Both languages share a similar pattern for adjectival modification, where an adverb modifies an adjective (ADVP → ADV ADJ). This structural equivalence aligns with the ECM's principle that code-switching is permissible at nodes where the grammars of the participating languages are compatible. The ADJP node satisfies this criterion, facilitating the smooth incorporation of English into the Hindi sentence.

As the sentence progresses, the speaker transitions back to Hindi with the verb 'hai', illustrating the morpho-syntactic rules specific to Hindi. The verb phrase (VP) node adheres to Hindi's syntactic constraints, requiring the use of Hindi lexical items to maintain grammaticality. This transition exemplifies another key tenet of the ECM: while code-switching is allowed at structurally compatible nodes, constraints arise at points dictated by the morpho-syntactic rules of each language. The VP node, for instance, enforces these constraints, ensuring that the Hindi verb 'hai' occupies its expected position at the end of the clause.

This example highlights the broader implications of CS in Hinglish and other bilingual contexts. Adjective phrases often serve as frequent

sites of CS, as their structural alignment across languages enables smooth transitions. By demonstrating how the ECM applies to this example, the analysis underscores the model's utility in explaining the grammatical boundaries and possibilities of code-switching in bilingual speech.

6.1.2 Analysis of the Second Clause (Tree 2)

In the second clause, the speaker employs code-switching, transitioning from Hindi to English within the structures represented by the (S), (VP), and (NP) nodes. This phenomenon is facilitated by the shared phrase-structure rules governing these constituents across both languages, allowing for seamless integration of English elements without disrupting the hierarchical organization of the sentence. For example, the NP node contains the English pronoun 'He', while the VP node includes the auxiliary 'can' and the verb 'find', demonstrating a coherent English predicate.

However, a notable divergence arises in the treatment of postpositional phrases (PPs). Hindi typically utilizes postpositions – such as 'mai' (in) – which occur after the noun phrase, as seen in the Hindi phrase 'poori duniya mai'. In contrast, English generally employs prepositions, leading to structural incompatibilities when integrating PPs from both languages. While English can use postpositions in certain contexts (e.g., 'miles away'), these cases do not apply to the current example. Consequently, the speaker strategically reverts to Hindi for the phrasal structure below the PP node, maintaining adherence to Hindi's morpho-syntactic conventions.

This bilingual construction allows for a nuanced expression of meaning while preserving the grammatical integrity of the sentence. The linear arrangement of constituents reflects the hierarchical structure, where the English elements align with their corresponding Hindi counterparts. The combined analysis of the first and second clauses illustrates how the ECM can be effectively applied to Hinglish sentences. Figure 6 visualizes this ECM analysis by mapping the grammatical constraints and code-switching sites, highlighting the shared and language-specific structures identified within the clauses **[fig. 6]**.

Through this analysis, we observe that code-switching in Hinglish is not merely a random blending of languages; rather, it follows a systematic approach governed by the syntactic frameworks of both Hindi and English. By carefully navigating these frameworks, speakers can create complex sentences that reflect their bilingual competence, ultimately enriching the communicative potential of Hinglish.

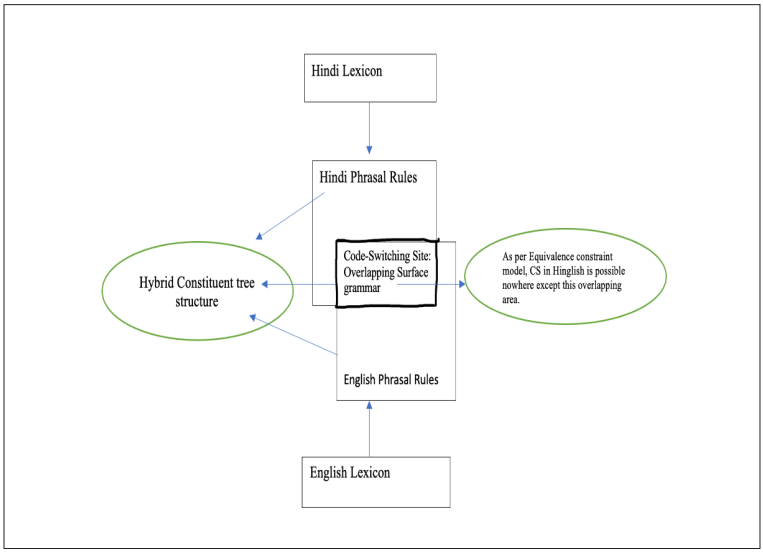

Figure 6 A graphical representation of Poplack's Equivalence Constraint Model of CS used to analyze the Hinglish example (2a)

7 Matrix Language Frame Model

7.1 Critique of the Equivalent Constraint Model

Myers-Scotton challenged Poplack's ECM on both empirical and theoretical grounds. Empirically, Myers-Scotton argued that while ECM showed promise for studying CS between typologically distinct languages, Poplack's ECM failed to account for CS data from the Swahili-English corpus, thus lacking empirical validation. Theoretically, Myers-Scotton contended that ECM did not address the concept of asymmetry, which is prevalent in most language contact scenarios. Joshi (1985) was among the first to highlight the disproportionate distribution of lexicon between the languages in a code-switched sentence, referring to this as the asymmetry of language involvement in bilingual speech. Joshi observed that "speakers and hearers generally agree on which language the mixed sentence is coming from. We can call this language the *matrix language* and the other language the *embedded language*" (Joshi 1985, 198; emphasis in original). Myers-Scotton's MLFM emerged as an alternative to ECM, incorporating Joshi's observation by distinguishing between the matrix and embedded languages. Unlike ECM, which treats loanwords and code-switching similarly, MLFM offers a framework that recognizes the asymmetry in language usage.

7.2 Influence of Psycholinguistic Theories

The development of the MLFM was also influenced by psycholinguistic theories. MLFM presupposes that speakers acquire their linguistic repertoire through a process of lemmatization, involving both conceptual and grammatical functions. This approach parallels cognitive theories of concept acquisition, such as those proposed by philosophers of mind like Jerry Fodor (1998, 40-9). The lemma-based approach in linguistics posits that language acquisition involves cognitive processes such as prototypicality and categorization, providing a foundation for understanding how speakers navigate and utilize multiple languages in code-switching contexts.

7.3 Definitions of Code-Switching

To illustrate Myers-Scotton's critique of ECM's inability to account for asymmetry in CS, it is important to contrast the definitions of CS proposed by Poplack and Myers-Scotton. Unlike Poplack, who treats CS and lexical borrowing as similar phenomena, Myers-Scotton argues that ECM fails to address the critical aspect of asymmetry in lexicon distribution within CS. Myers-Scotton defines CS as follows: "CS is the selection by bilinguals or multilinguals of forms from an embedded language (or languages) in utterances of a matrix language during the same conversation" (1993b, 11). This definition underscores asymmetry as a central feature of CS.

7.4 Key Principles of MLFM

To address the shortcomings of ECM, Myers-Scotton introduced the MLFM. The MLFM posits that in intra-sentential CS, speakers utilize the two languages asymmetrically. The first language – termed the matrix language (ML) – serves as the primary language of discourse, characterized by a higher frequency of morphemes and lexical items from that language. Conversely, the other language, known as the embedded language (EL), contributes fewer morphemes and lexical items relative to the ML. In essence, the MLFM captures the asymmetrical nature of language use in CS by distinguishing between the ML and EL.

According to the premise that the ML has a higher frequency of morphemes and lexicon distribution in code-switched sentences, Myers-Scotton proposed three key principles for conducting an analysis using the MLFM:

1. the Morpheme Order Principle: in ML+EL constituents consisting of singly occurring EL lexemes and any number of

ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML.

2. the System Morpheme Principle: in ML+EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e., which participate in the sentence's thematic role grid) will come from the ML.
3. the Blocking Hypothesis: in ML+EL constituents, a blocking filter blocks any EL content morpheme which is not congruent with the ML with respect to three levels of abstraction regarding subcategorization (Myers-Scotton 1993a, 83-120).

Simply put, according to the first principle of the MLFM, the ML dictates the ordering of elements within ML+EL constituents. MLFM differentiates between content morphemes (e.g., nouns and verbs) and system morphemes (e.g., articles and inflections), with the second principle asserting that function morphemes must originate from the ML. The third principle restricts the EL to contributing only specific content morphemes within the mixed constituents.

7.5 MLFM Analysis of Hinglish Sentence

To illustrate these principles, the article reanalyzes the Hinglish sentence 2(a), this time employing the MLFM. This analysis provides a comparative framework to evaluate the strengths and limitations of both the ECM and MLFM approaches. To reiterate, Hinglish sentence 2(a) is as follows:

2 a Ram highly qualified *hai* and he can find a job *poori duniya mai.*
'Ram is highly qualified, and he can find a job anywhere in the world'.

As the following analysis will demonstrate, the ECM insufficiently accounts for the presence of asymmetry in code-switched Hinglish sentences, whereas the MLFM demonstrates relatively effective performance in this context. Additionally, while the ECM aids in understanding the grammaticality of the code-switched sentence, it fails to address whether the speaker is inserting Hindi words into an English sentence or vice versa. The MLFM, on the other hand, can address this ambiguity concerning the predominance of one language over the other within a code-switched utterance. Myers-Scotton (1993a) proposes a discourse-oriented approach to identify the ML, suggesting that it is the language containing more morphemes in each utterance. However, identifying the ML can be complicated as it can shift within a single conversation. Alternative methods for determining the ML from a structural perspective have also been proposed by Joshi (1985) and Treffers-Daller (1994). Structural

analysis of the Hinglish sentence 2(a) reveals that Hindi provides the grammatical framework (such as the syntactical scaffolding of verbs and prepositional phrases), whereas English contributes key lexical elements (such as adjectives, pronouns, and verb phrases). Because the core grammatical framework of sentence 2(a) is provided by Hindi, one can infer Hindi to be the ML. This analysis can be explained as follows: First, 'Ram', a proper noun, is used consistently in both Hindi and English. Second, the verb 'hai' in Hindi, which roughly translates as 'is' in English, shows the ML structure. Lastly, the Hindi phrase "poori duniya mai", translating as "anywhere in the world", indicates location. In short, with grammatical elements like verbs and prepositional phrases being furnished from Hindi in Hinglish sentence 2(a), Hindi is inferred to be the ML in this case.

On the other hand, English serves as the EL inserted into the Hindi structure. The English phrase 'highly qualified' in the sentence describes Ram's qualifications. The coordinating conjunction 'and' in English connects the two clauses. The subject pronoun 'he' in English refers to Ram. Another English phrase 'can find a job' is used by the speaker to describe Ram's ability. Lastly, an English noun phrase, 'a job', is used to refer to the high chances of Ram's employability. With the contribution of the ML (Hindi) and EL (English) identified, the speaker's use of code-switching in the sentence can be further parsed as follows:

Matrix Language Frame (Hindi Structure)

Component	Description
Subject	"Ram" (remains constant as it's a proper noun used universally)
Predicate	"highly qualified hai" (mix of English adjective and Hindi verb)
Conjunction	"and" (English conjunction used to link clauses)
Additional Clause	"he can find a job poori duniya mai" (mixed structure with English embedded phrase)

Embedded Language Usage (English)

Component	Description
Clause	"highly qualified" and "can find a job" are English phrases embedded within the Hindi framework.
Pronoun	"he" is an English subject pronoun embedded within the sentence.
Prepositional Phrase	"poori duniya mai" (anywhere in the world) uses Hindi for the locative context.

7.6 Summary of Analysis

Matrix Clause 1: Ram highly qualified hai

Component	Description
Matrix Language	"Ram" (noun) + "hai" (verb)
Embedded Language	"highly qualified" (adjective phrase)

Matrix Clause 2: and he can find a job poori duniya mai:

Component	Description
Matrix Language	"and" (conjunction) + "poori duniya mai" (prepositional phrase)
Embedded Language	"he can find a job" (subject-verb-object phrase)

As demonstrated above, in this Hinglish sentence, Hindi supplies the grammatical framework through its use of verbs and prepositional phrases. In contrast, English introduces essential lexical components, including adjectives, pronouns, and verb phrases. This blending reflects a common pattern in Hinglish where English vocabulary is inserted into the grammatical framework of Hindi.

8 Conclusion

The preceding analysis compared two prominent constraint models of CS: the Equivalent Constraint Model (ECM) and the Matrix Language Frame Model (MLFM). As noted by scholars such as Nishimura (1985; 1986) and Chan (2009), ECM has been notably influential in the study of CS between typologically similar language pairs, such as English and Spanish. ECM's symmetrical approach to CS, which distinguishes between loanwords and code-switching, has garnered popularity among linguists focusing on such language pairs. In contrast, MLFM is favored for analyzing typologically distinct language pairs, such as English-Hindi (Pandit 1986) and Japanese-English (Nishimura 1985, 1986). The popularity of MLFM in these contexts can be attributed to its emphasis on asymmetry and its application of Occam's Razor to distinguish between borrowing and code-switching.

Recent scholarship has expanded the understanding of code-switching by incorporating diverse perspectives and methodologies, significantly contributing to the study of Hinglish. For instance, research by scholars like Kachru (2006) and Kothari (2011) emphasizes the socio-cultural dimensions of code-switching, highlighting how Hinglish serves as a marker of identity and cultural hybridity among speakers. Furthermore, the integration of

psycholinguistic theories into CS studies, as suggested by Wei (2009), enhances our understanding of how bilingual speakers navigate between languages in real-time, revealing cognitive processes that underpin code-switching behavior.

Emerging approaches that explore the neuro-linguistic dimensions of bilingual cognition (Kutas 2009) also provide valuable insights into how language processing occurs in bilingual individuals, further enriching the discourse on Hinglish. These developments signal a vibrant and evolving field, offering new insights into the complexities of bilingual language usage, such as Hinglish, and their broader implications for linguistic theory. By drawing from interdisciplinary research, scholars can deepen the analysis of code-switching phenomena, shedding light on the intricate relationships between language, identity, and cognition in bilingual contexts.

Bibliography

Backus, A. (1992). *Patterns of Language Mixing: A Study in Turkish-Dutch Bilingualism*. Wiesbaden: Harrassowitz.

Berk-Seligson, S. (1986). "Linguistic Constraints on Intrasentential Code-Switching: A Study of Spanish/Hebrew Bilingualism". *Language in Society*, 15(3), 313-48. Available at https://doi.org/10.1017/S0047404500011799

Bhatia, T.K.; Ritchie, W.C. (eds) (2013). *The Handbook of Bilingualism and Multilingualism*. 2nd ed. Hoboken: Wiley-Blackwell.

Bullock, B.E.; Toribio, A.J. (eds) (2009). *The Cambridge Handbook of Linguistic Code Switching*. Cambridge: Cambridge University Press.

Canagarajah, S. (2011). "Translanguaging in the Classroom: Emerging Issues for Research and Pedagogy". *Applied Linguistics Review*, 2, 1-28. Available at https://doi.org/10.1515/9783110239331.1

Canagarajah, S. (2013). *Translingual Practice: Global Englishes and Cosmopolitan Relations*. London: Routledge.

Chan, B. (2009). "Code-Switching Between Typologically Distinct Languages". Bullock, Toribio 2009, 182-98.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge (MA): MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge (MA): MIT Press.

Fodor, J.A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Gumperz, J.J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

Joshi, A. (1985). "Processing of Sentences with Intrasentential Code-Switching". Dowty, D.R.; Kattunen, L.; Zwicky, A. (eds), *Natural Language Parsing: Psychological,Computational and Theoretical Perspectives*. Cambridge: Cambridge University Press, 90-205.

Kachru, Y. (2006). *Hindi*. Amsterdam: John Benjamins Publishing Company.

Kothari, R. (2011). *Chutnefying English: The Phenomenon of Hinglish*. Gurgaon: Penguin Books India.

Koul, O. (2008). *Modern Hindi Grammar*. Hyatsville: Dunwoody Press.

Krishnaswamy, N.; Krishnaswamy, L. (2006). *The Story of English in India*. New Delhi: Foundation Books.

Kutas, M.; Moreno, E.; Wicha, N. (2009). "Code-Switching and the Brain". Bullock, Toribio 2009, 289-306.

Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.

Myers-Scotton, C. (1993a). *Duelling Languages: Grammatical Structure in Code-Switching*. Oxford: Oxford University Press.

Myers-Scotton, C. (1993b). *Social Motivations for Code-Switching: Evidence from Africa*. Oxford: Oxford University Press.

Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.

Myers-Scotton, C. (2003). "What Lies Beneath: Split (Mixed) Languages as Contact Phenomena". Matras, Y.; Bakker, P. (eds), *The Mixed Language Debate: Theoretical and Empirical Advance*. Berlin; New York: De Gruyter Mouton, 73-106.
Available at https://doi.org/10.1515/9783110197242.73

Newmeyer, F.J. (1986). *Linguistic Theory in America*. London; Oxford; Boston; New York; San Diego: Academic Press.

Nishimura, M. (1985). *Intrasentential Code-Switching in Japanese and English* [PhD Dissertation]. Philadelphia: University of Pennsylvania.

Nishimura, M. (1986). "Intra-Sentential Code-Switching: The Case of Language Assignment". Vaid, J. (ed.), *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*. Mahwah: Lawrence Erlbaum, 123-43.

Pandit, I. (1986). *Hindi–English Code-Switching: Mixed Hindi-English*. New Delhi: Datta Book Centre.

Poplack, S. (1978). *Quantitative Analysis of Constraints on Code-Switching*. New York: Centro de Estudios Puertorriqueños.

Poplack, S. (1980). "Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a Typology of Code-Switching". *Linguistics*, 18(7-8), 581-618.

Poplack, S. (2001). "Code-switching (Linguistics)". Smelser, N.; Batles, P. (eds), *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Elsevier Science Ltd, 2062-5.

Poplack, S. (2018). *Borrowing: Loanwords in the Speech Community and in the Grammar*. Oxford: Oxford University Press.

Poplack, S.; Wheeler, S.; Westwood, A. (1989). "Distinguishing Language Contact Phenomena: Evidence from Finnish-English Bilingualism". *World Englishes*, 8(4), 389-406.
Available at https://doi.org/10.1111/j.1467-971X.1989.tb00677.x

Sankoff, D.; Poplack, S. (1981). "A Formal Grammar for Code-Switching". *Papers in Linguistics*, 14(1), 3-45.

Treffers-Daller, J. (1994). *Topics in Sociolinguistics*. Vol. 9, *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin: Walter de Gruyter.

Wei, L. (2009). "Code-Switching and the Bilingual Mental Lexicon". Bullock, Toribio 2009, 270-88.

Weinreich, U. (1953). *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.

Young, V.A.; Lovejoy, K.B.; Barrett, R.; Young-Rivera, Y. (2014). *Other People's English: Code-Meshing, Code-Switching, and African American Literacy*. New York: Teachers College Press.