Chapter - III

Materials and Methodology

The chapter presents the purpose and the details regarding the topic undertaken for the study, its primary aim, its objectives and the scope. It also deals with, in detail, the source data, i.e. the International Corpus of English-Indian Component (ICE-IND) and the methodology adopted for the analysis and the assessment, along with the hypotheses.

3.1 Selection of the Topic

Indian English has now acquired the status of a standard non-native variety of British English. Its Indianness lies in its syntactic, lexico-semantic, phonemic and stylistic features. It has also been influenced by the bilingual and the multilingual phenomena that exist in this country. Code Mixing and Code Switching, as discussed in the previous chapter, are the inevitable consequences of the bilingual and the multilingual situations. The occurrences of Code Mixing and Code Switching in Indian English play important role in the Indianization of English in India. The studies on the language varieties, bilingualism, multilingualism and the role of Code Switching and Code Mixing come under the purview of Sociolinguistics. As Labru (1984:32) mentions, “The sociolinguistics of the sixties showed that language planning and the socio-dynamics of language in relation to social change and political independence are genuinely, and ought to be studied as, part of linguistic studies not necessarily as studies in language structure but positively, as studies in language function.” He, further adds, “India’s material culture and social institutions are rapidly changing, though not all at the same. When a multicultural society passes through a technological change, the overall cultural configuration undergoes a change. This change may be studied along and evaluated in terms of polarities like the following: 1. Literacy–Illiteracy 2. Rural–Urban 3. Isolation–Inculturation 4. National–Parochial.” With the social and the technological changes, he notes, the linguistic configuration of a society also changes. Hence, sociolinguistics is the discipline that considers such changes.
At present, there has been a wide-spread interest in Sociolinguistics, particular in regard to the realization of such kinds of mixed varieties. There has been a number of individual papers and research works dealing with CM, CS, Loan-words and Borrowing in Indian English.

Apart from such works, the compilation of the Indian English Corpora and the studies based upon them, have contributed considerably in the realization of Indian English as a distinct variety. Indian English, being the outcome of the bilingual and multilingual phenomena, also manifests the interferences of Indian languages.

The researcher proposes to study CM and CS in order to realize the sociolinguistic manifestations and constraints which distinguish Indian English as a standard variety. The prestigious characteristic of Indian English is that it is used in both forms speech and writing. Indian Writing in English already established itself in almost all literary and the critical forms. Many Indian Novelists and poets discovered their personal voice in this variety of English. It is being used widely also for the non-literary purposes like Education, Information, Administration, etc.

It is proposed to concentrate on the ‘spoken’ form, for two reasons. First, the spoken form is used more in India than the written one. The number of speakers of English is larger than that of those who write in English. Moreover, the peculiar Indian socio-cultural aspects of Indian English, the major concern of the study, are available in a profound number in the spoken form rather than in the written one. And above all, the spoken form, as we experience, is very much prone to CM and CS. CM and CS have less scope in the written form. The second reason for the preference to the spoken form is that, the research scholars such as Kachru (1978, 1983), Shastri (1988) whose works on CM and CS are mainly based on the written data, have asserted the need to study CM/CS occurrences in the spoken form of Indian English.

The other striking reason that motivates the researcher is: the availability of the International Corpus of English: Indian Component (ICE-IND). This Corpus is the recognized, authentic data which includes as many as 15 sub-
categories of the spoken form of IE. The details of this Corpus are given in the next sub-section.

3.2 Aims and Objectives of the Study

The aims of this study:

Code Mixing and Code Switching have been the inevitable consequences of language contact situation in the bilingual and multilingual countries like India. Hence, there is a need to focus on the socio-cultural and pragmatic factors of CM/CS in Indian English. To fulfill this need, with the help of the recognized database- Indian Component of International Corpus of English (ICE-IND), the researcher proposes to study Code Mixing and Code Switching in IE speech from the Sociolinguistic point of view. The researcher intends to establish IE as a widely recognized standard non-native variety of English on the basis of the peculiar use of CM and CS.

The objectives of this study are:

1) To explore the socio-cultural and the socio-semantic features of Code Mixing and Code Switching contributing to the process of Indianization of English,
2) To evaluate how the CM/CS choices are situationally determined and how shifts in situation are closely related to shifts in the language varieties,
3) To point out the pragmatic aspects of involved in the uses and shifts of Code Mixing and Code Switching,
4) To find out the frequencies of the CM and CS occurrences in the Spoken texts of the ICE-IND in order to determine the more common use.

3.3 Scope and Limitations

The scope of the study extends to 1) the register-wise socio-semantic analysis of the code-mixed items (at intrasentential level) and 2) the situation-wise analysis of the code-switched items (at inter- and intra- sentential levels) of
IE speech and 3) the assessment of the socio-linguistic, pragmatic factors involved in CM and CS, which equally contribute to the IE identity.

The syntactic constraints and the pronunciation peculiarities of CM and CS, however, remain out of its scope. The scope of the study is limited only to the Indian Component of ICE; the other components of ICE have not been referred to.

3.4 Materials

This sub-section deals with the views on the Corpus-based studies with respect to the choice of source material, the details of the Indian English Corpora and the texts selected for this study.

3.4.1 Choice of Source Material

There is a historical dichotomy over the choice of the source material for the language study. One view believes that the intuition of the native speaker is the only source of evidence, whereas the other view argues that the Corpus is the only source of linguistic evidence.

Chomsky (1964) and his colleagues support the first view, i.e. ‘intuition’. They argue that the Corpus source material is not of great value. The limitations of Corpus material pointed out by them are: a) The Corpus is bound to be biased. b) It is a record of language performance, not of the language competence. c) There is no method for deriving linguistic descriptions from a Corpus. They assert that those who use a Corpus for linguistic research must rely on ‘intuition’ as a means of interpreting the Corpus. Hence, Corpus cannot be the alternative for the intuitive evidence.

Those who hold the second view i.e. Corpus as the valid source material, are the post-Bloomfieldian structural linguists such as Fries, Hill and Harris and also the sociolinguists such as Labov. For them, the Corpus evidence is the fundamental reality which linguists must describe. Leech also prefers the Corpus data, though he does not dismiss ‘intuition’ altogetherly. He links the speakers’ intuition to Corpus by saying that the intuition performs two tasks: first, to recover the data from memory and second to interpret the data. The Corpus
material helps to extend the linguist’s ability of recollecting data. It does not however help to increase the ability of interpretation of the data. In Leech’s (1990:17) words, “The claim is not, then, that a Corpus replaces the need for introspection, rather than a native speaker’s intuition is rendered more powerful, particularly in the area of data collection, with the availability of the Corpus”. He also observes, “As soon as we become interested in generalizing descriptions beyond the individual’s idiolect, we have to become aware of the need for Corpus evidence”. Leech’s observation emphasizes the indispensability of the Corpus-based material for the sociolinguistic studies.

The drawback, that finite Corpora are skewed, can be eliminated by building up the more representative and larger Corpora. But, there is a biased phenomenon in ‘intuition’ method as well. The intuition unsupported by textual data may select most central, typical instances and neglect the vast mass of intermediate cases. In the Corpus analysis, however, such instances attract our attention. In order to overcome this limitation and to achieve the positive results, one has to maintain the balanced view towards ‘intuition’ and the Corpus data.

Biber (1994:169) illustrates two major advantages of the use of the text Corpora for Linguistic analysis: “First, they (Corpora) provide a large empirical database of natural discourse, so that analyses are based on the naturally occurring structures and the patterns of use rather than intuitions and perceptions, which often do not accurately represent the actual use. Second, they enable analyses of a scope not feasible otherwise, allowing researchers to address issues that were previously intractable. This is particularly true of the Computer-based text Corpora, which can be analyzed using semi-automatic techniques. Such analyses can examine much more language data than otherwise possible, including more texts, longer texts, a wider range of variation (text from different language varieties), a wider range of linguistics characteristics and the systematic co-occurrences patterns among linguistic features. In addition to quantitative analyses previously not possible, the corpus based approaches thus allow investigation of issues such as register variation and the discourse factors influencing the choice among structurally related variants (e.g. adverb
placement, or active vs. passive construction). He mentions the areas where this approach suits well: the cross-linguistic discourse comparison, the diachronic studies and the identification of the patterns of co-occurring linguistic features, for comparing the range of spoken and written registers. It is useful for comparing intuition against the actual patterns of language use and for analyzing the complex issues, such as the range of register variation in English.

The corpus-based studies are also well suited to the Applied Linguistics. Particularly, areas such as English grammar, Lexicography, Register variation can be dealt with by the Corpus-based approaches. Sinclair (1991) illustrates the ‘intuition’ point in the Corpus-based lexicographic research. In the analysis of the word *back*, he notes, “Our intuitions about a word often do not match the actual patterns of use. Most dictionaries list the human body part as the first meaning of *back*, and many people identify this as the core meaning. From the analysis of the COBUILD Corpus, however, this meaning is seen to be relatively rare”. He studies the lexical use by adopting the register perspective. He observes that in the social sciences texts, the word *back* is most commonly used in adverbial sense (e.g. *went back, came back*) and that in fiction, the body part meaning (e.g. *my back*) is much more common than in social science.

The Corpus-based approach, thus, facilitates investigations of wider scope, enabling macroscopic analyses of the inter-relations among several different parameters of variation.

The recent achievements in this regard are the building up of various text Corpora. Among them, the best known English Corpora are Brown Corpus, LOB Corpus, London–Lund Corpus, Kolhapur Corpus of Indian English and International Corpus of English (ICE). They observe the requirements. They remain bias-free, and are larger than the finite Corpus. A number of empirical researches have been carried out and are still being carried out on various linguistic aspects, based on these Corpora. Their observations and findings have provided new insights into many areas of language structure and use.

The present study deals with the behaviour of CM and CS in Indian English Speech. As per the views discussed above, the Corpus material suits
well to reach empirical results for such a kind of study. Since this study is based on Indian English Corpus, we may consider the Indian English Corpora in some detail.

3.4.2 Indian English Corpora

At present, there are two corpora available on Indian English, which can be used as sources of data for a study of Indian English. They are:

1) The Kolhapur Corpus of Indian English

2) International Corpus of English – Indian Component (hereafter referred to as ICE-IND)

3.4.2.1 Kolhapur Corpus of Indian English

It is a Corpus of written Indian English compiled by Shastri et al (1986). It is intended to serve as ‘source material for comparative studies of American, British and Indian English which in its turn is expected to lead to comprehensive description of Indian English.’ It is a representative corpus of the sample texts printed and published in 1978. It is built up on the same lines as Brown (American) and LOB (British) corpora.

Like the other two, it is a million-word corpus for use on digital computers. It consists of 500 texts of 2,000 words each. The texts are distributed over fifteen different categories of material, which represent a whole variety of written styles. Basically, there are two types of prose included in it: 1) Informative prose (374 texts) and 2) Imaginative prose (126 texts). The main and obvious difference is that LOB and Brown Corpora draw their samples from materials published in 1961, while Kolhapur Corpus of Indian English draws its samples from material published in 1978. The sampling procedures, the types of different styles of material represented and the weightage given to different categories of materials have all been kept very close to the others.

The texts selected for the corpus represent various registers. The distribution of the texts is as follows:

<table>
<thead>
<tr>
<th></th>
<th>Press: Reportage</th>
<th>44 texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Press: Editorial</td>
<td>27 texts</td>
</tr>
<tr>
<td>C</td>
<td>Press: Reviews</td>
<td>17 texts</td>
</tr>
</tbody>
</table>
As soon as this corpus was made available through several studies on different aspects on Indian English have been and still are being carried out. The major studies that have been carried out are: Katikar (1984), Babar (1996), Shinde (1991), Patil (1986), Shinde (1995) and the restricted studies carried out are: Ramtirthkar (1987), Shingate (1986), Salunkhe (1986), Gokhale (2008) and Kawale (2009). These studies have considerably contributed to the syntactic and semantic description of certain areas of Indian English. A brief summary of their studies appears in the Chapter I in the sub-section ‘Contribution of Corpus-based studies in Indian English.’

3.4.2.2 International Corpus of English- Indian Component (ICE-IND)

As the title suggests, ICE-IND is one component of International Corpus of English. The main International Corpus of English is an ambitious project. Initially, eighteen countries all over the world participated in the task. Among them, Great Britain, India, Australia, East Africa, New Zealand, Singapore, Hongkong and Ireland have completed the Corpus work. The Indian Component of International Corpus of English is compiled by Shastri and Leitner (2002). There is a common structure for all components of International Corpus of English, using the same criteria for text selection and the same frame time. The
design of ICE ensures the compatibility of various national and regional varieties with one other. The level of standardization also ensures the Corpus can be used for the direct comparative studies of varieties of English throughout the world. It is mainly intended to serve as source material for the comprehensive studies of language. Its computer based facility provides (semi-) automatic technique for the precise analysis of the texts.

This corpus is available on website http://www.ucl.ac.uk/english-usage/ice/index.htm.

3.4.2.2.1 Details of ICE-IND Data

ICE-IND consists of 500 texts. Each text contains 2000 words approximately. The texts are divided into two major categories: Spoken and Written. The Spoken category contains 300 texts and the Written category contains 200 texts. The Spoken category is further divided into two parts: Dialogue and Monologue.

The Dialogue category is further divided into two parts: Public and Private. Private Dialogues include Direct Conversations (90 texts) and Distance Conversation of (10 texts). Total Private Dialogues consist of 100 texts. Public Dialogues contain Class Lessons (20 texts), Broadcasts Discussions (20 texts), Broadcast Interviews (10 texts); Parliamentary Debates (10 texts), Legal Cross Examinations (10 texts) and Business Transactions (10 texts). Thus, total Public Dialogues consist of (80 texts) and the total number of Private and Public Dialogues is 180 texts.

The Monologue category is divided into two: Unscripted and Scripted. Unscripted Monologue consists of Spontaneous Commentaries (20 texts), Unscripted Speeches (30 texts), Demonstrations (10 texts), Legal Presentations (10 texts). Thus they amount to 70 texts. Scripted Monologues comprise of Broadcast News (20 texts), Broadcast Talks (20 texts) and Speeches (not Broadcast) (10 texts). Thus, Scripted Monologue texts are 50. The total number of the texts in the Spoken category thus, amounts to 300 texts, which are distributed over 15 sub-categories.
The Written category is broadly divided into two parts: Non-Printed (50 texts) and Printed (150 texts). The Non-Printed category is further divided into Non-professional writing (20 texts) and Correspondence (30 texts). The Printed category is further classified as Informational (Academic- 40 texts), Informational (Non-academic- 40 texts), Informational (Reportage- 20 texts), Instructional (20 texts), Persuasive (10 texts) and Creative (20 texts). Thus the 200 texts of the Written category are divided into 17 sub-categories. The following are the ICE-IND Text Categories and their Codes:

**SPOKEN (300)**

**DIALOGUE (180)**

Private (100) S1

- Direct Conversations (90) S1A
- Distanced Conversations (10) S1A-091 to S1A-100

Public (80) S1B

- Class Lessons (20) S1B-001 to S1B-020
- Broadcast Discussions (20) S1B-021 to S1B-040
- Broadcast Interviews (10) S1B-041 to S1B-050
- Parliamentary Debates (10) S1B-051 to S1B-060
- Legal Cross-examination (10) S1B-061 to S1B-070
- Business Transactions (10) S1B-071 to S1B-080

**MONOLOGUE (120)** S2

Unscripted (70) S2A

- Spontaneous Commentaries (20) S2A-001 to S2A-020
- Unscripted Speeches (30) S2A-021 to S2A-050
- Demonstrations (10) S2A-051 to S2A-060
- Legal Presentations (10) S2A-061 to S2A-070

Scripted (50) S2B

- Broadcast News (20) S2B-001 to S2B-020
- Broadcast Talks (20) S2B-021 to S2B-040
- Speeches (not Broadcast) (10) S2B-041 to S2B-050
### WRITTEN
- **NON-PRINTED (50)**
  - Non-Professional Writing (20)
  - Student Essays: W1A-001 to W1A-010
  - Examination Scripts: W1A-011 to W1A-020

- **Correspondence (30)**
  - Social Letters: W1B-001 to W1B-015
  - Business Letters: W1A-016 to W1B-030

### PRINTED (80)
- **Academic Writing (40)**
  - Humanities: W2A-001 to W2A-010
  - Social Sciences: W2A-011 to W2A-020
  - Natural Sciences: W2A-021 to W2A-030
  - Technology: W2A-031 to W2A-040

- **Non-academic Writing (40)**
  - Humanities: W2B-001 to W2B-010
  - Social Sciences: W2B-011 to W2B-020
  - Natural Sciences: W2B-021 to W2B-030
  - Technology: W2B-031 to W2B-040

- **REPORTAGE (20)**
  - Press News Reports: W2C-001 to W2C-020

- **INSTRUCTIONAL WRITING (20)**
  - Administrative Writing: W2D-001 to W2D-010
  - Skills & Hobbies: W2D-011 to W2D-020

- **PERSUASIVE WRITING (10)**
  - Press Editorials: W2E-001 to W2E-010

- **CREATIVE WRITING (20)**
  - Novels & Stories: W2F-001 to W2F-020
3.4.2.2 Details of the Texts for Study

As the researcher intends to study Code Mixing and Code Switching in Indian English Speech, she has selected the Spoken category of the ICE-IND data. The following table shows the Spoken Text Categories selected for the study:

Table - 4: ICE-IND Spoken Text-categories and Codes

SPOKEN (300)

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>Description</th>
<th>Sub-categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIALOGUE (180)</td>
<td>S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private (100)</td>
<td>S1A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Direct Conversations (90)</td>
<td>S1A-001 to S1A-090</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>2 Distanced Conversations (10)</td>
<td>S1A-091 to S1A-100</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>Public (80)</td>
<td>S1B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Class Lessons (20)</td>
<td>S1B-001 to S1B-020</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>4 Broadcast Discussions (20)</td>
<td>S1B-021 to S1B-040</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>5 Broadcast Interviews (10)</td>
<td>S1B-041 to S1B-050</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>6 Parliamentary Debates (10)</td>
<td>S1B-051 to S1B-060</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>7 Legal Cross-examination (10)</td>
<td>S1B-061 to S1B-070</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>8 Business Transactions (10)</td>
<td>S1B-071 to S1B-080</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>MONOLOGUE (120)</td>
<td>S2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unscripted (70)</td>
<td>S2A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Spontaneous Commentaries (20)</td>
<td>S2A-001 to S2A-020</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>10 Unscripted Speeches (30)</td>
<td>S2A-021 to S2A-050</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>11 Demonstrations (10)</td>
<td>S2A-051 to S2A-060</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>12 Legal Presentations (10)</td>
<td>S2A-061 to S2A-070</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>Scripted (50)</td>
<td>S2B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Broadcast News (20)</td>
<td>S2B-001 to S2B-020</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>14 Broadcast Talks (20)</td>
<td>S2B-021 to S2B-040</td>
<td>(all chosen)</td>
<td></td>
</tr>
<tr>
<td>15 Speeches (not Broadcast) (10)</td>
<td>S2B-041 to S2B-050</td>
<td>(all chosen)</td>
<td></td>
</tr>
</tbody>
</table>

Total texts 300 (all chosen)
3.4.2.2.3 Categories of the Selected Texts

Shastri (2002) in *User’s Manual of Indian Component of International Corpus of English (ICE-IND)* records the scope and limitations of ICE-IND corpus, the role of English in India, the kind of speakers and writers, the design of International corpus of English and the annotations in ICE-IND corpus. As our study is based on the Spoken category, it would be proper to give the details of the selected texts here. With regard to the Spoken category Shastri (2002) in his manual notes, “The main difference between the native language texts and that of second language/official language texts is that the former tends to be more natural and the later tends to be less so, particularly in the spoken texts.”

With respect to the kind of speakers he notes, “Our corpus represents the speakers of the major Indian languages included in the Indian constitution and there are a few minor languages such as Kokani, Tulu and so on.”

The category-wise details of the Spoken texts are as follows:

i) **Private Direct Conversations:**

Private Direct Conversations comprise three distinct subcategories. They are: a) genuine private conversations albeit often preplanned, b) discussions and c) mixed.

ii) **(Public) Class lessons:**

The Class Lessons range from the high school classes to the graduate level, with one exception (Text No. SIB-009) which happens to be at the post-graduate level and it is more of a lecture than a class lesson.

iii) **(Public) Broadcast Discussion and Interviews:**

All these texts were recorded by mass media; all of them are from T.V. (Door Darshan) channel. Most of them are full texts often exceeding the prescribed length of 2000 words. The participants are mostly politicians, journalists and other public figures.

iv) **(Public) Parliamentary Debates:**

As it was not possible to obtain the permission to record the face-to-face parliamentary debates, these texts were drawn in the form of
broadcast ones viz. “Question Hour”, “Joint sitting of the two Houses” and so on which are a tame affair, but no robust Parliamentary Debates as such.

v) **Legal Cross-examinations (Dialogues):**

Both these categories- Legal Cross-examination and Legal Representation- were drawn from only one place and one court i.e. Session Court at Bombay. About the difficulties faced while recoding these text categories, he notes, “There were a number of constraints: they were recorded in the presence of a judge and we had to adhere to the rules and with higher court, i.e. High Court and with its permission. For example, we had to change names of persons and the number/s of case/s etc. All the cases pertain to criminal cases and there are no cases of civil suits at all.”

**Legal Representation** includes three judgments from the same speaker. Also, it often tends to be dialogues and/or interruptions. This is probably because of the peculiar Indian ethos.

vi) **Business Transactions :**

These texts have been drawn from several places and several sources. 1) Pune (Poona): Maharashtra Board of Secondary and Higher Secondary School; 2) Kolhapur: certain minutes of the academic bodies of our own University; 3) Mumbai (Bombay) minutes of the bodies of the Educational Institutions; 4) Dombivali (a suburb of Bombay) minutes of the General body meeting.

vii) **Spontaneous Commentaries :**

All the twenty texts have been drawn from the broadcast ones. The ICE design does not specify the source. One could have drawn some of them ‘live’. But we found it easy to record broadcast commentaries. All these texts were taken from T.V. (Door Darshan) except two: One from All India Radio and another (a composite text) from International Film Festival-Zee channel-Filmfare Awards.
viii) **Unscripted Speeches:**

All the twenty texts of unscripted ‘speeches’ were drawn from several places. The speakers are from all over India. However, some of the ‘speeches’ are by definition and certain others are ‘lectures’ and yet others ‘presentations’ at Conference.

ix) **Demonstrations:**

Out of the ten texts in this category only three of them were drawn from live demonstrations. The rest of them were once again taken from broadcast ones-T.V.(Door Darshan).

x) **Broadcast News:**

Out of the twenty texts of Broadcast News, thirteen of them were drawn from T.V. (Door Darshan) and the remaining seven from All India Radio (AIR). The news readers are assorted. Hardly any news reader was repeated. Moreover, there was the practice of two different readers reading a full text. There were some composite texts but read by the same news reader. Most of the Door Darshan News readers were casual news readers except one who was (in the year 1990-91) and even now is a staffer. Similarly, most of the AIR news readers are casual readers, except the one who happens to be the chief news reader. There is an exception to the rule that this category comprises full texts not 2000+ words.

xi) **Broadcast Talks:**

Out of the twenty Broadcast Talks, fifteen were T.V. (Door Darshan) texts and the remaining five from All India Radio (AIR). Further, the Door Darshan Texts were taken from various EMRCs (Educational Media Research Centres). There are about 30 centres all over India.

xii) **Speeches (not broadcast):**

The ten texts of Speeches were drawn from scripted speeches as required. The sources were from various places and they were taken mainly from the Conferences. They were scholarly papers on assorted topics.
All 300 texts of 2000 words each (approximately) from 15 sub-categories are selected for this study. In other words, the selected part of the ICE-IND Corpus is approximately of six lakhs words.

3.5 Methodology

Several different methodologies and techniques have been used for analyzing the behaviour of Code Mixing and the Code Switching. They vary as per the thrust areas of the studies. The methodology used for studying syntactic constraints in CM/CS, for example, is different from the methodology used in analyzing the contextual CM/CS. Again the techniques used for classifying Code Mixing from Code Switching differ from one restricted definition to the other. There is not, it seems, an ideal methodology that can be used for studying all the aspects of the huge and complicated phenomena of CM/CS.

In the context of the observations of the studies and the theoretical framework built up in the foregone chapter, the researcher proposes to pursue the methodology concentrating on the following points:

1) To maintain the distinction between Code Mixing and Code Switching for the sake of description and propose the working definitions.
2) To provide the register-wise classification of Code Mixing,
3) To provide the category-wise classification for Code Switching,
4) To provide the Socio-linguistic and/or the pragmatic impressions for the occurrences.

The discussion of these points, offer the concrete statements about the methodology employed in the present study.

1) Distinction between Code Mixing and Code Switching: In the earlier chapter, the nature of CM and CS has been discussed the help of various definitions. It has also been observed that there is a controversy over the demarcation between CM and CS. In certain cases, the distinction between CM and CS is ambiguous, vague and overlapping in nature. The studies that maintain CM and CS as the manifestations of the same linguistic phenomenon, interpret them as ‘interference’ (Weinreich 1963), or ‘transfers’ (Haugen 1952),
or ‘embedding units’ (Myers Scotton 1983) or ‘They code’ (Gumperz 1982). These works are more theoretical in nature and their prime focus is on bilingualism and languages in contact. Some scholars have used Code-switching as the cover term for CM and CS whereas some others did exactly the reverse, i.e. they have used CM as the cover term for both manifestations. It is true that, it is difficult to provide precise definitions of CM and CS. This point has been discussed in detail in the previous chapter.

It has also been pointed out that the scholars, who have maintained distinction between these two phenomena have provided the definitions with reasoning for doing so (Kachru 1983, Gumperz 1978). In some cases their descriptions are limited as per the nature and the aspects of CM and CS they are dealing with (e.g., Myers Scotton 1992, Poplack 1980). In order to study the socio-cultural and the pragmatic factors involved in the interference of L1 into L2, i.e. Indian Language into Indian English, this study, for the purpose of description maintains the distinction between Code Mixing and Code Switching.

The distinction between CM and CS manifestations has been treated as a research tool for classifying the data. The following are the working definitions adopted for the purposes of our study:

*Code Mixing, is the transferring or mixing of the register-bound expressions, the content words, from the Indian language (L1) into the system of Indian English (L2).*

*Code Switching, is the successive use of clauses, sentences, discourse markers of Indian English (L2) and other Indian Language (L1) at intersentential and intrasentential levels.*

Both Code Mixing and Code Switching are influenced by the socio-cultural and the extra-linguistic factors such as topic, setting, participants, etc.

2) **Register-wise classification:** All 300 texts are read on computer one by one and occurrences of L1 interference have been located. One interesting noticed by this kind of procedure is that some texts did not have even a single
instance of L1 interference. Such texts are separated and labeled as ‘texts without L1 interference’.

In all, there are 55 texts which do not exhibit any interference of L1. It means the remaining 245 texts exhibit L1 interference. This division is important for us, because, with the help, of this we can trace out those domains or categories which are free from this distinctive phenomenon of IE.

The texts with L1 interference are read again carefully and by using the computer technique, examples are derived from the texts. The proper care has been taken in deriving them with sufficient pre- and post- contexts. The occurrences that demand details for further analysis are derived with full contexts. All instances, then, are classified as per our working definitions. Code Switching instances include full sentences, main clauses, and subordinate clauses. Code-mixing category, on the other hand, consists of the register-bound expressions and Content words/items. These code-mixed words are further classified broadly as per the field or the register they belong to. For this classification, at primary level, we referred to the works of Kachru, Yule and Burnell. A brief mention of the classification method used by them should not be out of context here.

Yule and Burnell’s (1886) *Hobson-Jobson* is ‘a Glossary of Colloquial Anglo-Indian Words and Phrases and of Kindred Terms, Etymological, Historical, Geographical and Discursive’. The Glossary, as the editors claim, deals with : (a) Words that seem to have been admitted to full franchise (curry, toddy) ; b) Words familiar enough to the English ear, though hardly yet received into citizenship (compound, pucka); c) Words, long since fully assimilated, which really originated in the adoption of an Indian word, or the modification of and Indian proper names (shawl, monsoon) ; d) Phrases of different character (‘that is the cheese’, ‘I don’t care a dam’); e) Administrative terms which are in such familiar and quotidian use as to form part of the common Anglo–Indian stock ; f) Terms adopted in scientific nomenclature, botanical and zoological; g) Words of Asiatic origin which have disappeared from the colloquial use, but which occur in old writers on the East. h) Geographical names which are or have
been familiar use in books on the Indies (Bombay, Madras, Malbar, Zanzibar);
i) Terms which had, in some shape or other, found their way at an early date into
use on the shores of the Mediterranean ( bazar, brinjal).

Kachru (1983) has listed the semantic areas in which the hybrid items
occur. This classification, as he notes, is based on the restricted data. The areas
he has considered are: 1) Administration: (city kotwali), 2) Agriculture: (akkulu
paddy), 3) Animals/reptiles: (basavan-bull), 4) Arms: (curved kukri), 5) Articles
of use: (angrezi furniture), 6) Art/music: (bazaar musician), 7) Buildings: (dak
bungalow), 8) Clothing/dress: (choli-piece), 9) Concepts: (counterfeit kismet),
10) Edibles/drinks: (angrezi sweets), 11) Education: (dakshina fund), 12)
Evaluation (attitude): (babu-mentality), 13) Furniture: (angrezi furniture), 14)
Habits: (beedi-smoking), 15) Medicine: (ayurveda system), 16) Modes of
address/reference: (babu manager), 17) Money/banking: (anna-coin), 18)
Occupations: (beediseller), 19) Place names: ( jungle path ), 20) Politics:
(ahimsa soldier), 21) Religion and rituals: (akasshti holidays), 22) Social
(general): (aam session), 23) Speech/langu age: (angrezi speech), 24)
Trees/flowers: (aruni field ), 25) Villages (general): (panchayat board), 26)
Vehicles/carriages: (coolie-car).

We have located in our data as many as 24 socio-cultural registers (The
details are given in the next chapter), along with their sub-categories. Then, the
code-mixed expressions, belonging to each register, are taken for analysis. The
analysis of each Register includes:

1) Name of the Register
2) Name of the Sub- Group (if any)
3) Words that are Grouped in that Register
4) Their Frequency, provided in the Brackets
5) Broad Semantic Components of the Register
6) Examples from the Data
7) Sociolinguistic and Pragmatic Impressions
8) Grammar Peculiarity (if any)
While determining the semantic components, Leech’s (1974:96) definition of componential analysis as a process of breaking down the sense of a word into its minimal distinctive features, i.e. into components which contrast with other components, has been closely considered. He cites the words *man, women, boy, girl* and other related words which belong to the semantic field ‘the human race’. He presents the relations of these words by using feature symbols like ± HUMAN, ± ADULT, ± MALE, ±FEMALE.

\[
\begin{align*}
\text{Man:} & \quad + H \ +A +M \\
\text{Women:} & \quad + H \ +A -M \\
\text{Boy:} & \quad + H \ -A +M \\
\text{Girl:} & \quad + H \ -A -M
\end{align*}
\]

These formulae are called the componential definitions of the items concerned; they can be regarded, as Leech claims formalized dictionary definitions.

While analyzing the code-mixed expressions, the semantic components that are peculiar to a certain register are provided. For example, the code-mixed words like *didi, akka, jija* belong to the register of ‘kinship’. The common feature marked for these words is the ‘Indian kinship term’.

Regarding examples from the data, the selection is made as per the frequency of their occurrences. That is to say, if there are, twenty examples in our data in which the mixed word *pooja* occurs, we have selected one of them as representing these occurrences. Every example quoted from the data is supported by its location in the text. It is possible, because the computer-based Corpus provides this facility by using Code Symbols. For example:

```
Lekin apko interest dena hai to muzhe kuch <ICE-IND:S1A-094#365:3:A>
kamana bhi hai to with that one I'm requesting
you that you please, release this one karke.
```

Here, the code ‘S1A’ represents section of Spoken English in the Corpus. It is followed by ‘094’ which is the number allotted to the text from which the extract has been taken. This is followed by ‘365:3’ which is a line and passage number of the text, and the letter ‘A’ refers to the particular speaker.
3) **Methodology for Analysis and Assessment of Code Switching:** The procedure used for the analysis and the assessment of Code Switching occurrences is a bit different from the one used for Code Mixing. This analysis is done, unlike that of Code Mixing, as per the situation in which Code Switching occurs. The main reason, as discussed in Chapter II, is that the Code Mixing phenomenon is more register-bound, whereas Code Switching is more situation-bound. Of course, it again depends on the scope of CM and CS one considers. As Fishman (1972: 439) observes, “CS in multilingual settings is regulated by the topic of discourse and stylistic (formal, informal) considerations….The implication of topical regulation of language choice is that certain topics are somehow handled ‘better’ or more appropriately in one language than in other in particular in the multilingual context.”

The procedure for the analysis of Code Switching:

As mentioned above, the CS occurrences are presented as per situation. The analysis and the assessment of CS in each situation include:

1) Text Number in the Corpus Serial  
2) Switched Count (i.e no. of examples in that corpus text)  
3) Examples from Data  
4) Social Setting of the Text  
5) Language Switched  
6) Boundary Level of the Occurrence  
7) Theme of Conversation  
8) Striking Pragmatic Consideration.

4) **Sociolinguistic/Pragmatic Impressions:** The sociolinguistic and/or the pragmatic impressions, referred to above, further deal with the assessment of the data. In other words, in CM they elucidate the meaning in Indian socio-cultural context, as well as the pragmatic factor that provides the context category of the speech event in which these expressions are used.

The code switching occurrences are assessed by highlighting the theme of the text, the socio-cultural factors involved and the prominent pragmatic considerations.
For the analysis of the data, the authoritative dictionaries of Hindi-English and Marathi-English, Sanskrit glossaries, Books, News Paper Weekly Supplements on Indian culture, recipes have been referred to. Sometimes, internet encyclopedia like Wikipedia and such other websites have also been browsed. Further, the analysis is supported also with Tables and Charts.

3.6 Hypotheses of the Study

The hypotheses underlying this work have been recorded on the basis of the primary observation of the texts in the data as well as the common experience of CM/CS being used in day-to-day life of the Indian speakers of English.

1. CM and CS are basically the products of language contact situation which appear mostly in bilingual and multilingual countries. India is a multilingual, multicultural country and that’s why Indians, while speaking in English, mix the Indian language codes or switch to Indian language.

2. Speech and writings are varieties according to the medium. They have different typical linguistic characteristics. However, as Leech (1982) notes, there is some overlap. Speech is generally unprepared and spontaneous. But sometimes, it is ‘prepared’ for lecture, debate or news broadcasting. In such cases, it displays the characteristics of the writing mode. In this context, we may hypothesize that the prepared IE speech is marked by less Code Switching and Discourse Markers than the spontaneous speech.

3. The Spoken form of language is used more than the written form. The writing is generally explicit, prepared and does not have monitoring features. This makes us hypothesize the frequency of CM/CS appears more in the Spoken form of IE than in the Written one.

4. The Code-mixed lexical items such as paan, salwar, acres, Pooja in Indian English speech acts are specific to certain registers. The speakers mix them because alternatives for these items are not available in English. In the context of IE, the mixing of words from Mother Tongue, which is
register-oriented, becomes inevitable. This makes us hypothesize that the Code Mixing in IE speech is more register-bound.

5. One more striking feature of the Indian English speech is the frequent use of Discourse Markers. The mixed items such as Accha, Haan, yaar are the main indicators of conversation, which make the spoken discourse different from the written one. They essentially belong to the speech activity. The characteristic of these responses, apart from the fact that they express emotions, is that they possess the regional identity. Their *prima-facie* behaviour makes us hypothesize that in Indian English, the code-switched Discourse Markers play a crucial role in the speech activity.

6. The switching from English to the Mother Tongue or to the Indian languages is detected at the interclausal and the intersentential levels. For example, the expressions such as “aap baithiye, Please sit down”; or, “the shirt you gave me you know, mere pant ko matching nahi hota”, are used several times in the discourse. This makes us hypothesize that switching from English to Mother tongue is not only registered-bound but also pragmatically and situationally determined.

There have been individual studies on the aspects of CM and CS in Indian English. Many of them use the written data (Kachru 1978b, Shastri 1988, Verma 1976), as elaborated in Chapter II. Only a few studies deal with the spoken data. Malik’s (1994) study deals partly with the Spoken and partly with the Written data. The present work is perhaps, the only full scale and comprehensive study of the sociolinguistic and the pragmatic features and the frequency of CM and CS in IE speech.