

Acute Sporadic English in Bollywood Film Songs' Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi

Subhash Chandra, M.Phil., Ph.D.
Bhupendra Kumar, Ph.D. Candidate
Vivek Kumar, Ph.D. Candidate
Sakshi, Ph.D. Candidate



Abstract

Bollywood (BW) movies are very rich source of songs with various features of corpora. This study focuses on the English mixing in the songs of BW movies. The song corpora have been collected from various websites that archive Bollywood song's lyrics in Devanagari and roman scripts. BW produces more than thousand Hindi movies in every year. A single movie contains at least 2-6 songs. The major goal of the study is to identify the language mixing pattern in BW movies songs and then identify the linguistics phenomena for automatic language

detection and processing. The songs from the 1008 number of movies released during 2000-2015 were selected, collected and analyze for this study.

Keywords: Code Mixing, Linguistic Pattern in Bollywood Song, Language Mixing, Bollywood songs

1. Introduction

Mixing of two or more languages is very common phenomenon in day to day informal conversation and writing especially in social networks like Facebook, google+ and Tweeter, blogs, social media and chats etc. Dominance of English language in the world everybody specially youngsters are mixing words, phrases and sentences in their languages for chat, informal writings and bloggings through sharing their ideas to making fashionable and interesting. Bollywood (BW) is Mumbai, India, Bollywood based film industry generally produces more than thousand Hindi movies in every year. Each movie may contain up to 6 songs. BW films are musicals and contain attractive music in the form of song and dance numbers woven into the script.

Film's victory often depends on the quality of such musical numbers. Indeed, a film's music is often released before the movie and helps increase the audience [1] The language of the BW The films depend on the subject matter, location and the characters involved, the code switches between sociolects, standard languages and distinct Persian or Sanskrit features, jargons with regional variants right through to other Indian national languages such as Panjabi, Marathi, Gujarati [2, 3]. English mixing of the BW films is very common.

Bollywood is the sobriquet for the Hindi language film industry, based in Mumbai, India. The term is often incorrectly used as synecdoche to refer to the whole of Indian cinema; however, it is only a part of the large Indian film industry, which includes other production centers producing films in many languages. Bollywood is one of the largest film producers in India and one of the largest centers of film production in the world [4-5].

Language in India www.languageinindia.com ISSN 1930-2940 16:11 November 2016

Subhash Chandra, M.Phil., Ph.D., Bhupendra Kumar, Ph.D. Candidate,

Vivek Kumar, Ph.D. Candidate and Sakshi, Ph.D. Candidate

Acute Sporadic English in Bollywood Film Songs' Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi

Dialogues are usually written in an unadorned Hindi that would be understood by the largest possible audience. Some movies, however, have used regional dialects to evoke a village setting, or old-fashioned, courtly, Persian-influenced Urdu in Mughal era historical films. Contemporary mainstream movies also make great use of English [6].

BW movies are very rich source of songs with various features of corpora. This study focuses on the English mixing in the songs of BW movies. The song corpora have been collected from various websites that archive Bollywood song's lyrics in Devanagari and roman scripts. BW produces more than thousand Hindi movies in every year [7]. A single movie contains at least 2-6 songs.

The major goal of the study is to identify the language mixing pattern in BW movies songs and then identify the linguistics phenomena for automatic language detection and processing. The songs from the 1008 number of movies released during 2000-2015 were selected, collected and analyze for this study.

2. Background

It is clear that in post-globalization India, English is an essential component of upward mobility [8]. It is also seen that the growing popularity of Indian culture around the world, including BW movies, means that Hinglish (English word, phrase mixed with Hindi language) will soon become more widely spoken outside the continent [9].

Bollywood Songs (BS) started their journey as soon as 1931 with the movie 'Alam Ara', the first Indian sound film, and Hindi music never turned back and grew into a giant industry. Within a few decades, Bollywood became the most energetic film industry in Asia. Therefore lyricists are mixing words from other language to compose songs and the songs have become very famous. Hindi is the official language of India but still very few Indians love to speak and read pure Hindi. Language mixing is taking everyone in its grip very rapidly [9-10]. Mixing of

Language in India www.languageinindia.com ISSN 1930-2940 16:11 November 2016

Subhash Chandra, M.Phil., Ph.D., Bhupendra Kumar, Ph.D. Candidate,

Vivek Kumar, Ph.D. Candidate and Sakshi, Ph.D. Candidate

Acute Sporadic English in Bollywood Film Songs' Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi

English makes the songs catchy and very entertaining and the audience loves them [11]. However, the trend of language mixing is not new, it has been going on for a while now. The mixing started in the late 50s, seen in the songs like ‘Mera naam chin chin chu’ from the movie (Howrah Bridge, 1958).

3. Language Resource and Corpora

The www.lyricsmasti.com website was used for corpora which archive songs’ lyrics in Roman Script only. Devanagari Script was not considered in this study. A *python* based program was developed to collect raw corpora from the website through *urllib2*. This program has been run on www.lyricsmasti.com for the collection of texts of lyrics.

Table 1 Corpora Details

Sr.	Year	No. of Movies	No. of Songs	Unique Words
1.	2000	68	216	6748
2.	2001	68	283	8415
3.	2002	104	342	10054
4.	2003	69	209	8458
5.	2004	53	225	9075
6.	2005	71	248	9075
7.	2006	107	406	12818
8.	2007	116	467	17021
9.	2008	103	441	16723
10.	2009	75	309	11287
11.	2010	53	210	8492
12.	2011	55	185	8887
13.	2012	60	232	10428
14.	2013	60	232	9200
15.	2014	50	190	8120
16.	2015	10	48	1890
Total		1062	4011	146263

Language in India www.languageinindia.com ISSN 1930-2940 16:11 November 2016

Subhash Chandra, M.Phil., Ph.D., Bhupendra Kumar, Ph.D. Candidate,

Vivek Kumar, Ph.D. Candidate and Sakshi, Ph.D. Candidate

Acute Sporadic English in Bollywood Film Songs’ Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi

4. Methodology

A dictionary based checking method has been applied to extract English words in the collected song lyrics. 4011 songs written in Roman script were collected from the 1062 number of movies released during 2000-2015 through a *python* based program. A list of total unique words containing 1,46,263 with frequency was created. Then it was checked in English lexicon (contains 50,428 English words) and 2383 unique English words were extracted with frequency with the help of *python* based program. Sample of the English words is shown in table 2. It was also checked in songs and 217 songs were extracted which were written in pure Hindi. 3794 songs were found containing English words in lyrics. Then a manual effort was taken to verify the result.

Table 2 Sample of extracted English words with frequency

English word	Frequency
You	2584
Love	1547
My	1092
No	988
Be	615
Baby	558
Your	555
Am	540
All	449

5. Results and Discussion

The mixing of English was highly observed in very popular and hit songs. The English word 'you', 'love', 'my' and 'no' were found with 2584, 1547, 1092, and 988 frequency respectively. 734, 350 and 185 words were used once, twice and thrice frequency.

Language in India www.languageinindia.com ISSN 1930-2940 16:11 November 2016

Subhash Chandra, M.Phil., Ph.D., Bhupendra Kumar, Ph.D. Candidate,

Vivek Kumar, Ph.D. Candidate and Sakshi, Ph.D. Candidate

Acute Sporadic English in Bollywood Film Songs' Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi

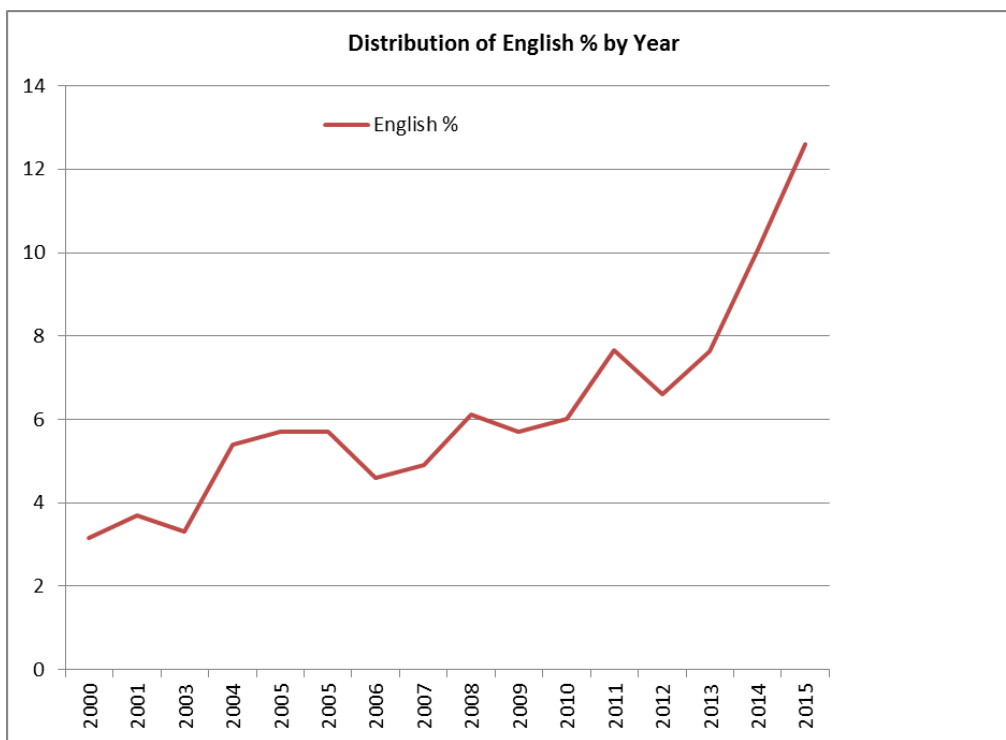


Fig. 1 Distribution of English in Songs

The mixing of English in songs is increasing day by day as shown in figure 1. It was 3.15% in songs composed in 2000 and it increased to 3.61%, 5.7%, 6%, 7.65%, 7.63%, 10.86% and 12.59% in 2001, 2005, 2010, 2011, 2012, 2013, 2014 and 2015 respectively. After observation it was found that language mixing is seen more in hit and popular songs. The manual review revealed same linguistics features discussed by Kundu et al. [9], Chandra et al. [10], Sinha et al. [12] and Chandra [13].

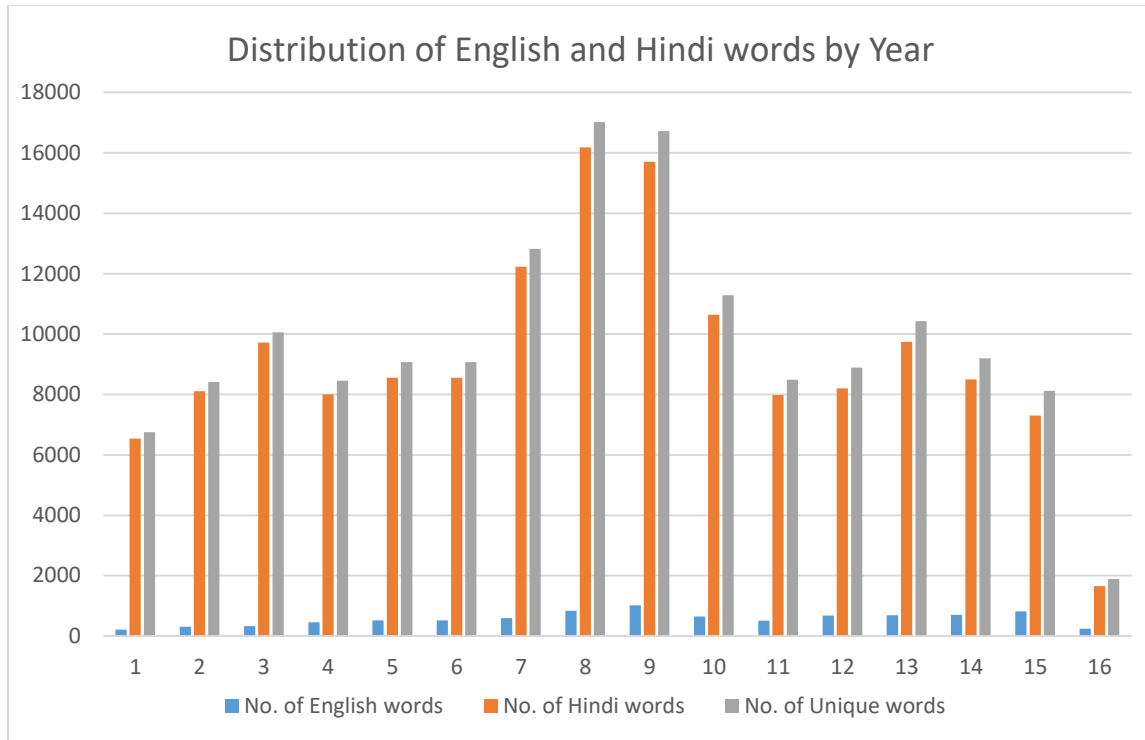


Fig. 2 Distribution of English and Hindi by Year

Distribution of the English words mixing year by year in the BW songs is shown in Table 3. A manual verification process was also applied on extracted 2384 English words to check whether these words are really used as English or transliteration of any Hindi words. It was found a few words are used as Hindi in a few songs and as English in a few songs. The frequency of English words listed in Table 2 may be decreased by those types of words. For example “so” appeared as Hindi word in maximum songs which means ‘sleep’ but in a few songs it was used as an English word. It was found that no standard transliteration was followed to spell a word of Hindi in Roman script. A widespread spelling variation was observed [14-15]. It was also found a few English words were used as fashionable form. For example, “good”, “wanna”, etc., were not extracted due to spelling differences. But these were highly seen in song lyrics [13].

Table 3 English Mixing in Songs by Year

Year	No. of English words	No. of Hindi words	No. of Unique words
2000	213	6535	6748
2001	311	8104	8415
2002	333	9721	10054
2003	456	8002	8458
2004	518	8557	9075
2005	518	8557	9075
2006	589	12229	12818
2007	836	16185	17021
2008	1021	15702	16723
2009	643	10644	11287
2010	510	7982	8492
2011	680	8207	8887
2012	688	9740	10428
2013	702	8498	9200
2014	819	7301	8120
2015	238	1652	1890

6. Conclusion

This study has demonstrated that the manner and frequency of the English word mixing in the song lyrics of Bollywood movies increased every year. The data shows a strong turnover in the language of choice among the youngsters. It is also evident that either the volume of English use has increased or a new type of language is emerging.

References

1. Jha SK. *The Essential Guide to Bollywood*. Roli Books; 2005.

Language in India www.languageinindia.com ISSN 1930-2940 16:11 November 2016

Subhash Chandra, M.Phil., Ph.D., Bhupendra Kumar, Ph.D. Candidate,
Vivek Kumar, Ph.D. Candidate and Sakshi, Ph.D. Candidate

Acute Sporadic English in Bollywood Film Songs' Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi

2. Gulzar N. Chatterji GS. *Encyclopaedia of Hindi Cinema*. India: Encyclopaedia Britannica Pvt Ltd: 2003. pp.10–18.
3. Aung S. A diachronic investigation of Hindi–English code-switching, using Bollywood film scripts. *International Journal of Bilingualism* 2010; 15(4) 388-407.
4. Gupta T. Triumph of Hinglish: How shuddh Hindi lost its groove. Firstpost; 2011, <http://www.firstpost.com/ideas/triumph-of-hinglish-how-shuddh-hindi-lost-its-groove-48098.html>. Retrieved on 20.05.2016.
5. Kundu B. Chandra S. Automatic Detection of English Words in Benglish Text: A Statistical Approach, In: *4th International Conference on Intelligent Human Computer Interaction 2012 (IHCI 2012)* at the Indian Institute of Technology Kharagpur, India; 2012.
6. Chandra S. Kundu B. Hunting Elusive English in Hinglish and Benglish Text: Unfolding Challenges and Remedies, In: *10th International Conference on Natural Language Processing (ICON-2013)*. India: Macmillan Publishers; 2013.
7. Chaudhuri R. Maine Karoo to Character Dheela Hai! Business is Marketing, Rajita Chaudhuri Blogger, Thursday, June 16, 2011. <http://rajitachaudhuri.blogspot.in/2011/06/maine-karoo-to-character-dheela-hai.html>. Retrieved on 05.06.2016.
8. Sinha RMK. Thakur A. Machine Translation of Bi-lingual Hindi-English (Hinglish) Texts. In: *10th conference on Machine Translation*. Phuket: MT Archive; 2005, pp.149-156.
9. Chandra S. Main ho gaya single I wanna mingle...: An Evidence of English Code-Mixing in Bollywood Songs Lyrics Corpora. In: *First Workshop on Language Technology for Indian Social Media Text with The Eleventh International Conference on Natural Language Processing (ICON-2014)*, 18-21 December, 2014, Goa University, Goa, India; 2014.
10. Gupta P. Bali K. Banchs RE. Choudhury M. Rosso P. Query Expansion for Multi-script Information Retrieval. In: *37th Annual ACM SIGIR Conference, SIGIR-2014*, Gold Coast, Australia, June 6-11; 2014.



11. Roy RS. Choudhury M. Majumdar P. Agarwal K. Overview and datasets of fire 2013 track on transliterated search. In: *FIRE 2013 Shared Task on Transliterated Search*; 2013.

Dr. Subhash Chandra, M. Phil., Ph.D.

Assistant Professor, Computational Linguistics



Department of Sanskrit, Faculty of Arts, University of Delhi
Delhi-110007, India

subhash.jnu@gmail.com

schandra@sanskrit.du.ac.in

Mr. Bhupendra Kumar

Doctoral Research Scholar, Computational Linguistics

Department of Sanskrit, University of Delhi



Delhi-110007, India

bhupendrakmr87@gmail.com

Mr. Vivek Kumar

Doctoral Research Scholar, Computational Linguistics

Department of Sanskrit, University of Delhi

Delhi-110007, India

vishnorapatyam@gmail.com



Ms. Sakshi

Doctoral Research Scholar, Computational Linguistics

Department of Sanskrit, University of Delhi

Delhi-110007, India

sakshi911003@gmail.com

Language in India www.languageinindia.com ISSN 1930-2940 16:11 November 2016

Subhash Chandra, M.Phil., Ph.D., Bhupendra Kumar, Ph.D. Candidate,

Vivek Kumar, Ph.D. Candidate and Sakshi, Ph.D. Candidate

Acute Sporadic English in Bollywood Film Songs' Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi