
Some Remarks on the Perso-Arabic Nominal Lexica in Bollywood Movie Song Lyrics

TOMASZ GACEK

Abstract: The Indian film industry releases thousands of movies every year, with those in the Hindi language forming a significant part of that number. These movies, and the songs used in them, have been the object of research within the scope of various scientific disciplines, including linguistics. The author of the present paper, however, believes that there are still numerous language phenomena to study within the lyrics of the Bollywood songs. The present text is dedicated to research of the Perso-Arabic lexica in these songs over the period of 80 years (1940–2020). It also analyses the frequency of the Perso-Arabic forms as opposed to native Indo-Aryan ones. Particular attention is paid to the forms of the highest frequency, especially the Persian word *dil*.

Keywords: Bollywood, Hindi, Persian, Arabic, loanwords, frequency

Tomasz GACEK, Jagiellonian University, Cracow, Poland; tomasz.gacek@uj.edu.pl;

 0000-0002-8111-2797



This article is distributed under a Creative Commons Non-Commercial No-Derivatives 4.0 licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The colossal Indian film industry releases thousands of movies every year. An important and the best-known worldwide part is made up of Hindi-language movies produced in the state of Maharashtra by the industry referred to collectively as Bollywood.

As the phenomenon is a very important component of the popular culture of modern India and has growing significance globally, it has been the subject of study from many angles for decades. To give just a few examples, gender norms in Bollywood films have been studied quite often, especially recently (e.g. KHAN and TAYLOR 2018); the problems of globalisation in the context of Hindi movies have been discussed for the last two decades at least (RAI 2009). There have even been attempts to predict the success of particular movies using machine learning techniques (VERMA and VERMA 2019). Other studies focus on the mechanisms behind Bollywood movie production (GANTI 2012).

However, taking into consideration how complicated and specific the language landscape of Bollywood movies is and, especially, of the songs used in them, one could say the number of linguistic studies within the field is surprisingly low. Let us recall just some of the most important aspects of this landscape. The term Bollywood refers mostly, as we have already noted, to Hindi-language movie production. Thus, the dialogues in these films are supposed to be scripted in Hindi.¹ However, here arises the complex relationship between Hindi and Urdu. This is even more so in the case of songs where the subtle, elusive, borderline between Hindi and Urdu is often crossed and Persian words may be used much more often than in an everyday conversation. Moreover, in many cases dialect elements are present, and even use of other languages may be observed. The influence of ethnolects like Punjabi, Braj, Rajasthani, Marathi may also be noticed. Some aspects of this situation, like the problem of code-switching and code-mixing with Hindi and/or Urdu on one side and English on the other (CHANDRA et al. 2016; HABIB et al. 2020) or the use of other languages in these songs (BEHL and CHOUDHURY 2011) are sometimes discussed. Others, however, are still not covered at all. Among these, complex analysis of the problem of the presence and frequency of Persian and Perso-Arabic vocabulary within these texts is important.

At the same time the importance of the question of language interaction and language influences (especially here in the context of Persian and Arabic elements) goes far beyond the linguistic perspective. As language is a vitally important component of every human culture, language contacts are, per se,

¹ In fact, the dialogues are mostly scripted in Romanised Hindi, see e.g. the *Thappad* movie script (SINHA and MRUNMAYEE 2020).

a form of cultural interaction. Language influences are cultural influences.

A linguistic perspective in the culture research may focus on various aspects, however, one of the most interesting and a central one seems to be, as Kramsch puts it, “How are people’s perceptions, beliefs, values encoded in the linguistic sign [...]?” (KRAMSCH 2014: 31). At the same time, the idea that a language influences the way its speakers perceive the world has been around since the seminal work by Whorf (WHORF 1956). The key concept of linguistic relativity and the question of how a language shapes our way of thinking, stress the dependence of our image of the real world on our language (KRAMSCH 2014: 32, 35). Though there are limits to linguistic relativity and the mechanism is not as simple as changing the name of a phenomenon and immediately altering our view of it (KRAMSCH 2014: 36). However, it is still an important factor to be taken into consideration while studying the interaction between language and culture. In other words, language both reflects and influences the culture of its speakers.

The language of Bollywood songs is a very special phenomenon. While based on modern Hindi, it has a number of peculiar features that make it different from actual spoken languages. Some scholars have claimed that, having absorbed influences from many other languages, over the years it was melted into a unified subsystem, possibly to a “universal film song language that appeals across local, regional and international boundaries” (SARRAZIN 2008). Moreover, Behl and Choudhury notice significant deviation of the Bollywood songs’ vocabulary from Zipf’s law² as seen in any natural language, including standard spoken Hindi (BEHL and CHOUDHURY 2011). All this makes research on this particular lect³ a fascinating enterprise.

The author of the present paper believes that tendencies in the sphere of language influences may indicate important factors in the evolution of every culture. Thus, this paper aims at making a step forward in the linguistic research on Bollywood songs, focusing on the phenomenon of the presence of Perso-Arabic nominal lexica within it. We analyse a sample corpus of texts trying to answer the following questions in reference to the nominal lexica within it:

- A. What is the quantity of the words of Persian and Arabic origin within the corpus?
- B. Is it possible to trace changes over time as far as the phenomena mentioned above are concerned?

² As George ZIPF noticed, if we calculate the frequency of particular words in a text and we arrange them in a descending order by frequency, there will be an inverse relationship between each frequency and the rank in the order (KRETZSCHMAR 2010: 190).

³ I am using the term “lect” as defined by Crystal (CRYSTAL 2008: 272).

C. Are there any words that tend to be used in the songs in question particularly frequently? And if so, what is their origin?

D. Is there any significant correlation between the subject of a song and the frequency of the Perso-Arabic lexica in its lyrics?

Methodology

To answer these questions a corpus of over 90 Bollywood songs have been assembled and analysed. To avoid any sort of a bias, the selection of texts has been based on arbitrary criteria. One song from each one of the five most successful movies at five-year intervals starting from 1940 and up to 2020 has been selected. The list of the most successful films was established on the basis of the data provided by the BOX OFFICE INDIA website. The accuracy of these rankings has not been independently checked, although in the author's opinion it does not influence the results of the research, as no language criteria are involved there. In most cases, the first song on the list for a given movie provided by the MYSWAR website has been chosen (this does not have to be necessarily the first song in a movie, but often is). Where the data from either BOX OFFICE INDIA or MYSWAR were unavailable or irrelevant, secondary sources were used, e.g. IMDB database and the HINDI GEET MALA website etc. The lyrics of the songs were collected using a vast number of (mainly online) sources, and transcribed/verified on the basis of available recordings. Where no lyrics were found/retrieved for a particular song/movie, the subsequent ones (according to the same order) were taken into consideration.

The results yielded by the method described above were, however, imperfect. The word count based on non-duplicate verses turned out to be significantly lower for the older songs. The average word count based on the text with rejected duplicate lines was around 78 for pre-1960 songs and above 115 for those from the period of 1960–2020. Thus, an additional criterion was introduced, namely an arbitrary rule that the total word count based on non-duplicate verses for every year should be between 500 and 600 words. To achieve this, a number of supplementary texts were additionally analysed for the older periods, while three texts from more modern times that made an almost 700-word count were rejected.

Once a canonical text for a given song was established, a list of verses without duplicates was created. Lines containing onomatopoeic forms only were rejected, as they can hardly be interpreted as utterances in any language, in the same sense as regular lines. A list of nominal forms was created on the basis of the accepted lines. Every particular form from the same song was accepted only once. To avoid including disputable forms, certain lexemes were excluded from the analysis. This refers especially to infinitives that may be sometimes used as

nouns and various participles. Moreover, in some cases it was not immediately obvious if a given form is a noun or maybe just an ad hoc substantivised adjective. In these cases, the classification in the lexicographical works was decisive. The origin and meaning of every analysed word was studied on the base of the available lexicography (see the bibliography index).

From every song a list of nominal forms was extracted. Repetitions (at the level of a given song) were omitted, though different declension forms were all taken into consideration. Identical forms were taken into consideration whenever they appeared in different songs.

Example: In the song “Naacho Sitaro Naacho Ab Chand Nikalne Wala Hai” from the 1945 movie *Zeenat*, we find the form *rātō* “night” (obl. pl). As it is beyond any doubt a noun, it is included in the database. Exactly the same form (*rātō*) appears again in the same text, but it is not taken into consideration, as it is a repetition. However, another form of the same lemma, *rāt* (rect. sg.) is entered into the database. When either of these forms appears in other songs (e.g. *rātō* in the song “Morni Re Morni Main Jungle Ki Morni” from the movie *Pratiggya* and *rāt* in “Pardesiyon Se Naa Ankhiya Milaana” from *Jab Jab Phool Khile*), they are included again.

While the translations of the texts in question are not relevant for the current study as such, they were essential at some point in the word-selection process and choosing from between homophonous variants.

The list of the analysed texts

The procedures described above have led to the selection of the following 92 songs. To avoid confusion, the titles of songs and movies are given not in the scientific transcription but in the form in which they are known to the public. However, when the same title appears also as a verse of a song and is cited in the analysis either in its entirety or in part, scientific transcription is used instead.

To facilitate the creation and processing of the database, every song was assigned a unique numerical key consisting of three parts: yyyy-fff-sss, where yyyy is the year the movie started showing, fff is the place on the list of the most successful movies (with leading zero digits), and sss is the number of the song. The last two components will be essential if the research is extended in the future. The complete index of the analysed texts has been included in the appendix.

The following table shows the word count based on non-repetitive verses in particular songs:

	Year	1940	1945	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	
Movie - song	001-001	80	93	87	92		207	181	77	88	168	133	93	123	120	106	69	117	
	001-002					49													
	002-001	165	78	69	69	59	66	135	107	152		149	113	112	134		239	81	
	002-002																	69	
	003-001	44	103	76	32	125	81	116	245	181	89	94	78	165		160	112		
	003-002	72			70										75				
	004-001			52	69	146	104	89	88	79	149	110	129	126	81	140			92
	004-002			48															
	005-001	47	65	88	97	102	84	84	84	15	91	91	172		100	140	125	67	
	005-002	79																	110
	006-001	76		65	38	120													
	006-002			108	114														
	007-001		57																
	007-002		115																
	SUM	563	511	593	581	601	542	605	601	515	497	577	585	526	510	546	545	536	

The studied lexical material

The use of the presented methodology allowed us to create a database of 1,908 lexical entries. This consists of 1,092 unique inflected forms and 957 unique lemmata. The author was unable to interpret about five forms, which were ignored in the analysis. Their number is too small to influence the overall results.

Overall proportions of Persian and Arabic lexical items

All the analysed lemmata were classified as either Indo-Aryan (IN), i.e. *tadbhava*, *tatsama* Hindi words or forms borrowed from other Indo-Aryan languages (incl. Marathi, Punjabi, etc.), Arabic (AR), Persian (FA), English (EN), Dravidian (DR) etc. Generally, the basis for the classification is the immediate source of the form, so *patlūn* is classified as a borrowing from English, even though its origins go back to Italian, via French (MCGREGOR 1993: 596; SKEAT 1993: 326), just as *cakleṭ*, which is ultimately derived from Nahuatl *xocoatl*, transmitted via French (SKEAT 1993: 58; PUSTKA 2022: 522) and English. The only exception has been made for the words of Arabic origin. Though, without any doubt, the vast majority of them (if not all) were transmitted into Hindi/Urdu via Persian, they were marked as Arabic (AR) rather than Persian (FA).

Based on the presented criteria, the statistics of the lemmata source languages may be presented in the form of a table.

Language origin	Number
Indo-Aryan (IN)	1,045
Persian (FA)	426
Arabic (AR)	323
English (EN)	35
Indian or Persian – uncertain (IN/FA)	20
Other, unknown	11
Mixed: Arabic and Persian (AR+FA)	11
Mixed: Indian and Persian (IN+FA)	9
Onomatopoeic	7
Indo-Aryan: uncertain (IN?)	6
Onomatopoeic – uncertain	6
Persian: uncertain (FA?)	4
Dravidian	2
Dravidian – uncertain (DR?)	2
Either Indian or mixed Indian and Persian (IN/IN+FA)	1
Total Result	1,908

Let us simplify this table by putting all the Persian, Arabic and Perso-Arabic forms together, as well as grouping the rest of the lexemes (i.e. non-Indo-Aryan and non-Perso-Arabic) under the category “Other”.

Source language	Number
IN	1,045 (54.77%)
AR/FA/AR+FA	760 (39.83%)
Other	103 (5.40%)
Total Result	1,908

Two types of complexities are obvious on the basis of the table above. First of all, in the case of some forms we are dealing with the mixed origin (AR+FA or FA+IN). Another problem is that Persian and Hindi, due to their common origin, in spite of the long time passing since their ancestor languages had parted, still share a number of very similar or even identical forms. This is the case of the words like *bār* (MCGREGOR 1993: 725) (cf. NP *bār*), *rang* (MCGREGOR 1993: 847) etc.

Anyway, we are able to see that at least in the case of 769 either Persian or Arabic influence is unquestionable. In other words, 40.3% of identified nominal lexemes in the lyrics of 92 analysed songs are either entirely of Persian or Arabic origin, while those with unquestionable Indo-Aryan roots constitute 54.7% (1,044 examples). No other vocabulary source even approaches these two categories as far as the numbers are concerned. The next most important source is English, with only 35 nominal lexemes identified (1.8%).

Most frequent lexical items

Let us now see which nominal lexemes are the most frequent in the whole corpus, i.e. what words appear in the maximum number of songs. We will take into consideration only those lemmata that are used at least in 10% of the songs.

lemma	Count - lemma	Comments
<i>dil</i> (FA)	51	“heart, soul, etc.” ⁴ (MCGREGOR 1993: 496) < FA <i>dil</i> “heart, mind, soul, etc.” (STEINGASS 1892: 530)
<i>āmkh</i> (IN)	40	“eye, sight, look, etc.” (MCGREGOR 1993: 754), ultimately derived from Vedic <i>ākṣi</i> (TURNER 1966: 32)

⁴ The meanings provided in this table are based upon the definitions but are not precise citations. Please, refer to the original publications for complete definitions.

lemma	Count - lemma	Comments
<i>pyār</i> (IN)	26	“love, affection” (MCGREGOR 1993: 651) < Indo-Aryan <i>priyakāra</i> - (TURNER 1966: 503)
<i>bāt</i> (IN)	24	“something said, a word, remark, speech, etc.” (MCGREGOR 1993: 722) < Indo-Aryan <i>vārttā</i> (TURNER 1966: 674)
<i>din</i> (IN)	19	“day, daytime, etc.” (MCGREGOR 1993: 495) < Indo-Aryan <i>dina</i> (TURNER 1966: 362)
<i>man</i> (IN)	19	“mind, heart, etc.” (MCGREGOR 1993: 788) < Indo-Aryan <i>manas</i> (TURNER 1966: 564)
<i>nain</i> (IN)	15	“eye” (MCGREGOR 1993: 581) < Indo-Aryan <i>nayana</i> (TURNER 1966: 401)
<i>rāt</i> (IN)	15	“night” (MCGREGOR 1993: 861) < Indo-Aryan <i>rātri</i> (TURNER 1966: 619)
<i>duniyā</i> (AR)	14	“world, people, etc.” (MCGREGOR 1993: 503) < FA <i>dunyā</i> “world, people” (STEINGASS 1892: 539) < AR
<i>jādū</i> (FA)	11	“magic, enchantment, etc.” (MCGREGOR 1993: 367) < FA <i>jādū</i> “conjuration, magic” (STEINGASS 1892: 349)
<i>jān</i> (FA)	11	“life, spirit, etc.” (MCGREGOR 1993: 367) < FA “soul, mind, self, life” (STEINGASS 1892: 352)
<i>jīvan</i> (IN)	11	“life, existence, etc.” (MCGREGOR 1993: 375) < Indo-Aryan <i>jīvana</i> (TURNER 1966: 289)
<i>sapnā</i> (IN)	11	“dream, vision” (MCGREGOR 1993: 981) < Indo-Aryan <i>svāpna</i> (TURNER 1966: 804)
<i>bāmh</i> (IN)	10	“the upper arm; sleeve” (MCGREGOR 1993: 719) < Indo-Aryan <i>bāhū</i> (TURNER 1966: 521)
<i>hāth</i> (IN)	10	“hand, wrist, etc.” (MCGREGOR 1993: 1067) < Indo-Aryan <i>hāsta</i> (TURNER 1966: 811)
<i>khvāb</i> (FA)	10	“dream, vision” (MCGREGOR 1993: 247) < FA <i>xvāb</i> “sleep, dream, etc.” (STEINGASS 1892: 478)

As we can see, while the unquestionable leader is *dil* “heart”, a word of Persian origin, still most of the frequent lexemes have Indo-Aryan roots.

As an example, let us look at contexts in which the word *dil* is used within the analysed corpus. The form in question is used exclusively with the meaning “heart”, which is mostly understood as the organ to which emotions (in particular elevated ones) are ascribed. In perfect accordance with this intuition, love is the most frequent of these emotions, and it happens to be described in various ways. In the text “Yeh Kaun Aaj Aaya Savere Savere” (1940-005-002) the heart (*dil*) experiences a shiver or a shock on seeing unexpected persons’ arrival:

yeh kaun āj āyā savere, savere |
*ke **dil** caumk uṭhā savere, savere ||*

However, a few lines later we can hear that somebody stole the heart of the lyrical persona (***dil** chīn liyā*), which is a universal metaphor for falling in love.

In the song “Naacho naacho sitaaro naacho” (1945-001-001) it is said that a person is entering the heart through the eyes (*ām̄kh se **dil** mem āne vālā*), which conveys the idea of somebody so attractive and charming that their appearance has a lasting impact on others’ hearts. It is also from hearts that desires will emerge (*armān, armān dilom ke nikaleṅge*).

The lovers in “Humko Humise Chura Lo” (2000-002-001) are so afraid of being separated from each other that they encourage each other to hide themselves somewhere inside the other’s heart: ***dil** mem kahīm tum chupā lo*.

The heart is used not only in the image of a romantic love. It is also important in describing the relationship between mother and her child, as in “Tumse Milkar Naa Jaane” Kyun (1985-004-001): *mem tere **dil** kā ṭukrā hūm, e mām* “I am a part of your heart, mum”.

According to “Rani Khol De Apne Dwaar” (1945-005-001), the heart may be confused, perplexed (***dil** merā ghabrā gayā*) and a storm may rise in it (*āj **dil** mem uṭhā hai tūfān*). There are other dangers to which a heart is exposed: it may get under a spell, as is described in “Ae Ji, Dil Par Hua Aisa Jadu” (1955-005-001): ***dil** par hua aisā jādū*, or even worse, a lightning may strike inside it: ***dil** mem camaktī haiṁ bijliyām*.

Love is not always happy, and the lyrical person of “Hum Se Bhool Ho Gaye Humka Maafi Dai Do” (1980-003-001) accuses his beloved one that she has broken his heart: *merā **dil** toṛ jāte ho*. The same idea appears in “O Priya, Priya” (1990-001-001), where the lyrical person goes even further, encouraging his love (a bit masochistically) to come and see his torn apart heart: *āke zarā dekh le **dil** merā cīr ke*. He also wants to burn himself and take death into his heart: *mawt ko apne **dil** se lagā lūm*.

The heart (inner self) may be contrasted with the face (appearance, what is visible), as in “Gore Gore O Banke Chore” (1950-001-001): *cehre pe ḡussā hai, **dil** mem safāī* “There’s rage on the face and clearness (openness, sincerity) in the heart”.

Of course, the heart beats when it experiences strong emotions or simply when we are alive, as we find out e.g. from “Dhadke Mera Dil Mujhko” (1950-002-001) *dharke merā **dil***.

The heart may also dance, presumably when full of joy, as in “Ta Ra Ri Aa Ra Ri Aa Ra Ri...” (1950-003-001): *dil nāce*. It may also speak, thus revealing the true desires of a person: *kahtā hai dil merā calie vahām* “My heart says let’s go there” or send messages as in “Mera Salaam Le Ja Dil Ka Payaam Le Ja” (1955-004-001): *dil kā payām le jā*. The heart may also be sure about something or somebody: *dil ko yaqīn hai*.

If the heart can dance, it can sing as well for sure, and this is attested by the song “Gata Rahe Mera Dil” (1965-005-001): *gātā rahe merā dil*. A singing heart appears in “Tujhe Bulayen Yeh Meri Baahen” (1985-001-001), too.

In “Mera Juta Hai Japāni” (1955-001-001), the heart of the lyrical subject is Indian in spite of every part of his outfit coming from a different part of the world: *phir bhī dil hai Hindustānī*. At the same time, the lyrical persona (and the protagonist of the movie at the same time) and those like him are ruin-hearted princes (*ham bigarē dil śahzāde*).

While in “Mera Juta...” the Indian heart serves as a declaration of patriotism on the national level, the lyrical person of “Ye Lucknow Ki Sar Zameen” (1960-004-001) states that in his city there is love in the hearts: *yahām dilom meṁ pyār hai*. He goes even further declaring that there are wealthy (or noble, influential) people possessing hearts living in Lucknow. The last expression is quite interesting, as it uses the phrase with the Persian *izāfa*: *ahl-e dil*. The same song also associates the notion of honour (*śān*) with the heart. Then, in the song “Qurbāni, Qurbāni, Qurbāni” (1980-001-001) the heart is juxtaposed with another Perso-Arabic word, *īmān*, which may designate “honour” as well, but also “belief, faith, honesty”.

A more original idea is comparison of the lyrical subjects’ heart to the whole world, which happens in the song “Kaisi ye muhobbat ki saza haaye dee hai kisi ne” (1955-003-002): *duniyā hī mere dil kī, dil kī*. But when he feels sad, disappointed, etc., his heart may shrink: *dil ghaṭṭā hai*.

In “Tan Rang Lo Ji” (1960-003-001), the singer encourages his listeners to colour not only their bodies but also their hearts during the Holi festival: *dil rang lo*.

The lyrical person of the “Chand Si Mehbooba” (1965-003-001) is lucky as his beloved one doesn’t dream about palaces as she prefers to reside in his heart: *dekhe nā khvāb mahalom ke mere dil meṁ rahnā cāhe*.

In “Duniya Me Pyaar Ki Sabko Zarurat Hai” (1970-002-001), the phenomenon of love is described as giving and taking the heart at the same time: *dil de, dil le, yahi mohabbat hai*.

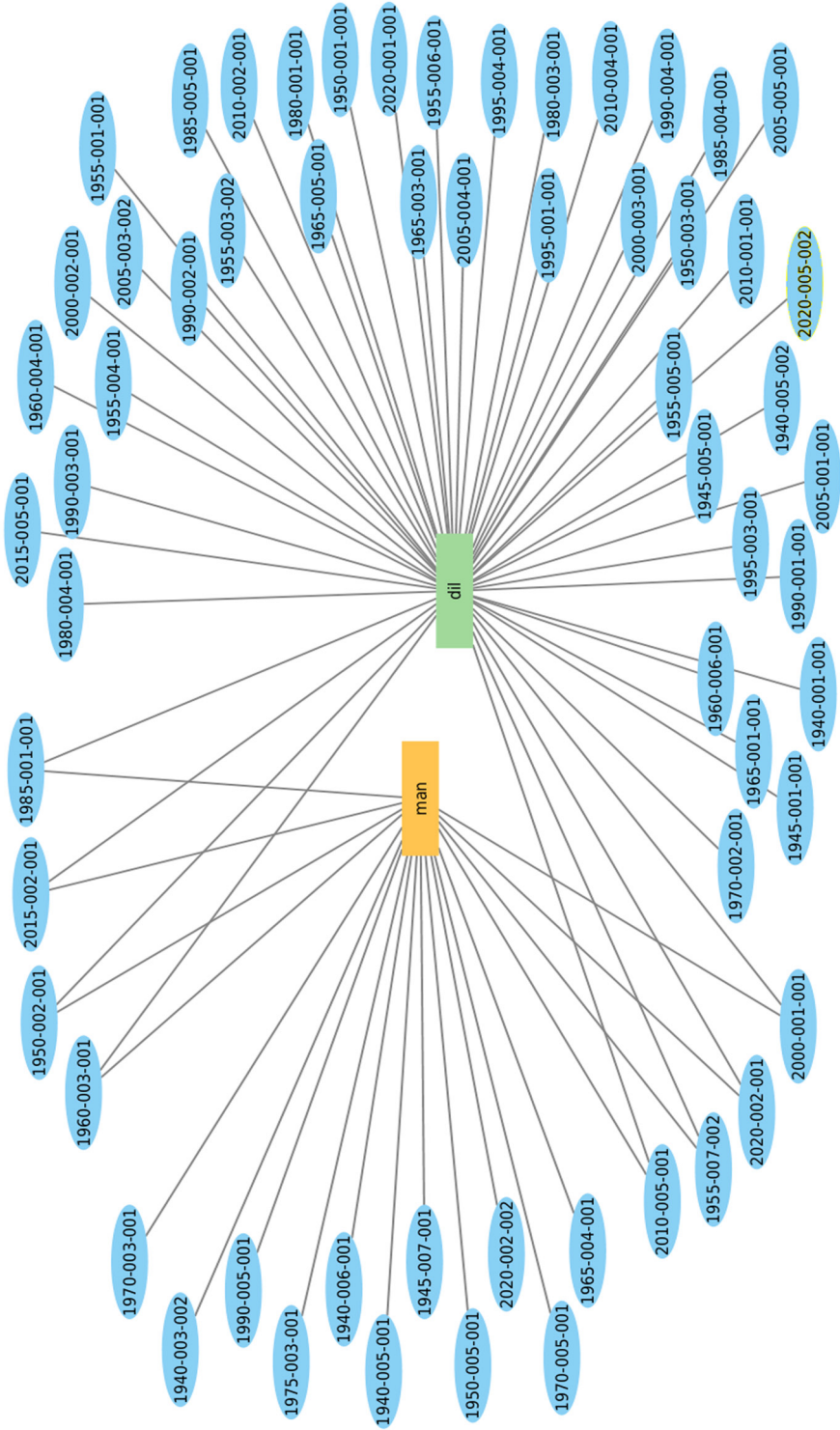


Fig. 1. Distribution of the lexemes *man* and *dill* in the analysed texts

For obvious reasons, the word *dil* is used most often in lyrical songs. However, we may find examples even in those with a religious background, like “Selfie Le Le Re” (2015-002-001), where we can hear that the worshippers keep their hearts as big as that of Hanumanji, who holds the mace: *dil barā rakhte haim jaise hanumānjī kī ho gadā*.

Finally let us note that the word *dil* is also used as a part of a number of compounds used in the texts in question (*dilrubā*, *dildār* etc.).

It is worth noting that we can see a number of synonymic pairs on the list above, namely *dil* (FA) and *man* (IN), *sapnā* (IN) and *k̄hvāb* (FA), *āmkh* (IN) and *nain* (IN), *jān* (FA) and *jīvan* (IN).⁵ Apart from *āmkh* and *nain*, these are pairs containing a word of Indo-Aryan origin and its Perso-Arabic equivalent. Let us have a look at some charts showing the distribution of the elements of such pairs:

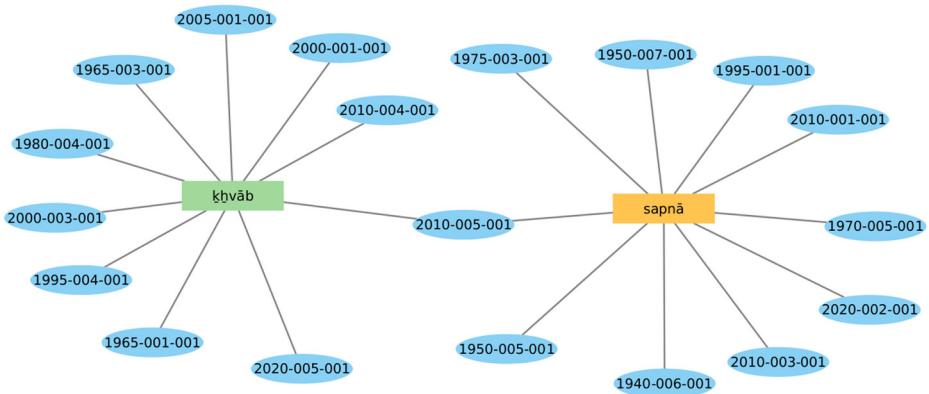


Fig. 2. Distribution of the lexemes *k̄hvāb* and *sapnā* in the analysed texts

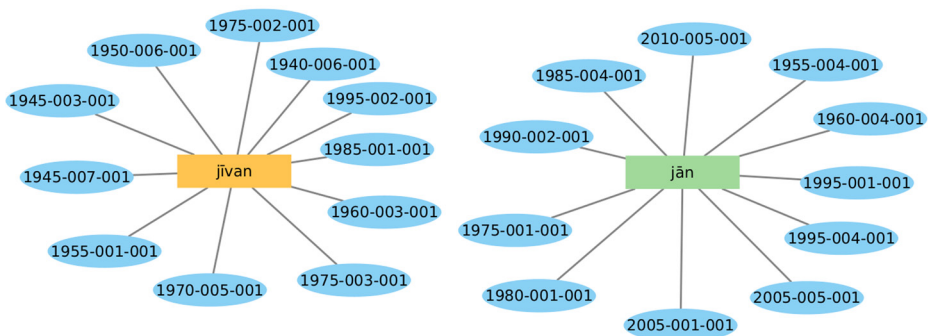


Fig. 3. Distribution of the lexemes *jīvan* and *jān* in the analysed texts

⁵ As the reviewer of the paper rightly pointed out, the last two forms are not perfect synonyms, as *jīvan* can refer to one’s whole life, whereas *jān* cannot.

As we can see, there is no correlation with time. Moreover, the greater the number of occurrences, the bigger the chance of coexistence of both equivalents in the same text.

Lexemes origin as a function of time

Now the most fascinating question is probably of whether there are any observable differences in the proportions between nominal lexemes of Persian and Arabic origins and those with the Indo-Aryan roots at different points of time. The table below includes occurrences of Arabic (AR), Persian (FA), mixed Perso-Arabic (AR+FA) forms on the one hand and of those of the Indo-Aryan and Dravidian origin (IN) on the other.

Year	IN	AR, FA, AR+FA	Other
1940	84 (64.12%)	38 (29.00 %)	9 (6.87%)
1945	58 (68.24%)	21 (24.71%)	6 (7.06%)
1950	90 (69.23%)	31 (23.85%)	9 (6.92%)
1955	78 (51.32%)	66 (43.42%)	8 (5.26%)
1960	75 (47.17%)	75 (47.17%)	9 (5.66%)
1965	54 (55.10%)	44 (44.90%)	0 (0%)
1970	80 (63.49%)	43 (34.13%)	3 (2.38%)
1975	111 (84.09%)	18 (13.64%)	3 (2.27%)
1980	40 (47.06%)	44 (51.76%)	1 (1.18%)
1985	53 (61.63%)	29 (33.72%)	4 (4.65%)
1990	57 (57.58%)	33 (33.33%)	9 (9.09%)
1995	43 (41.75%)	55 (53.40%)	5 (4.85%)
2000	38 (42.70%)	46 (51.69%)	5 (5.62%)
2005	63 (58.88%)	39 (36.45%)	5 (4.67%)
2010	53 (46.49%)	55 (48.25%)	6 (5.26%)
2015	62 (57.41%)	43 (43%)	3 (2.78%)
2020	40 (38.46%)	46 (38.46%)	18 (17.31%)

The amount of the Persian and Arabic lexical items within the analysed vocabulary is variable, ranging from over 13% to over 53%. However, the data presented above show no trace of any clear tendency. While the proportions vary from one tested period to another, there is no one direction in these changes. It will be probably even more visible in the following chart:

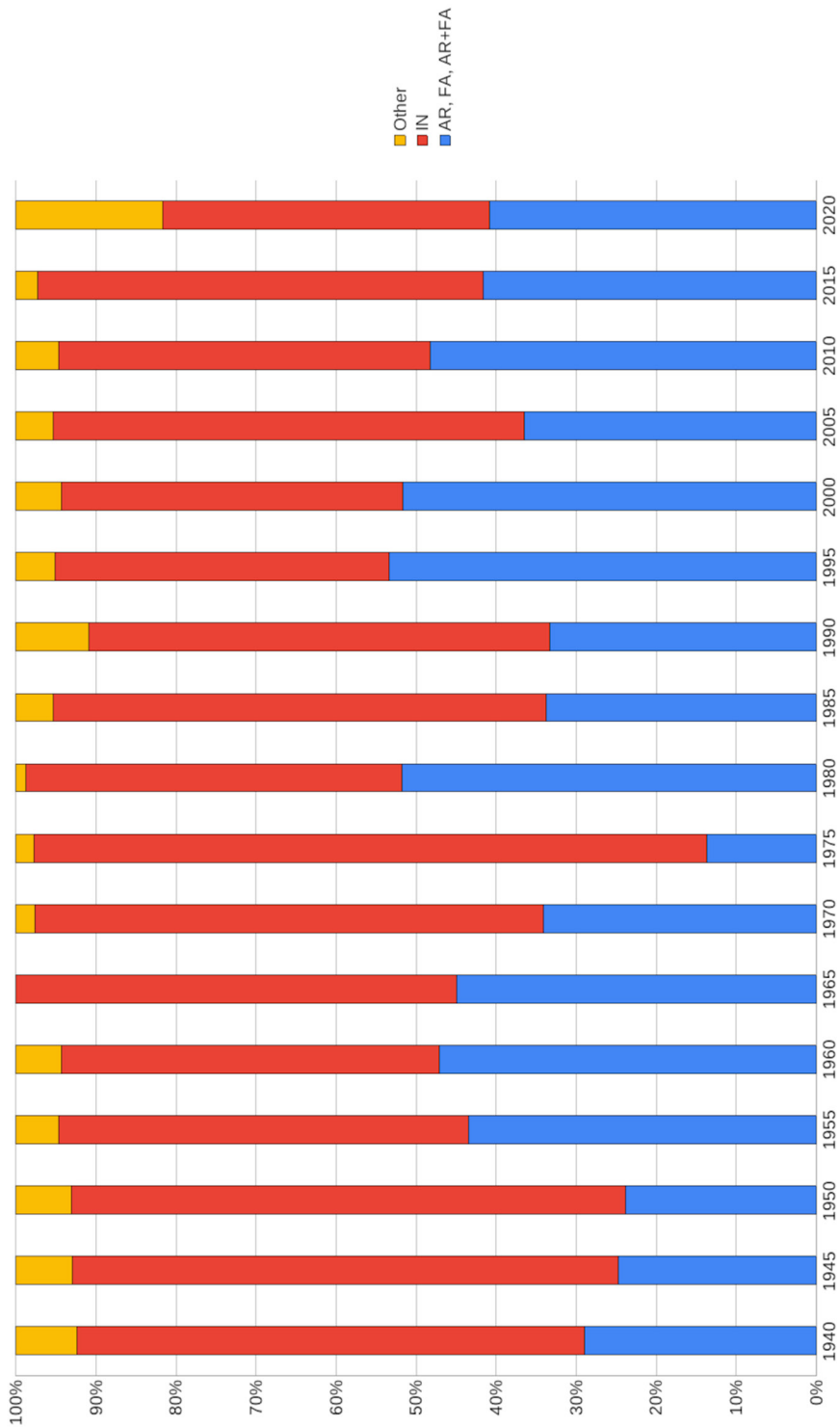


Fig. 4. Proportion of forms of various origin in specific years

As we can see, the differences are noticeable, even significant. They are even more extreme if we have a look at a particular single song. Let us consider a few examples:

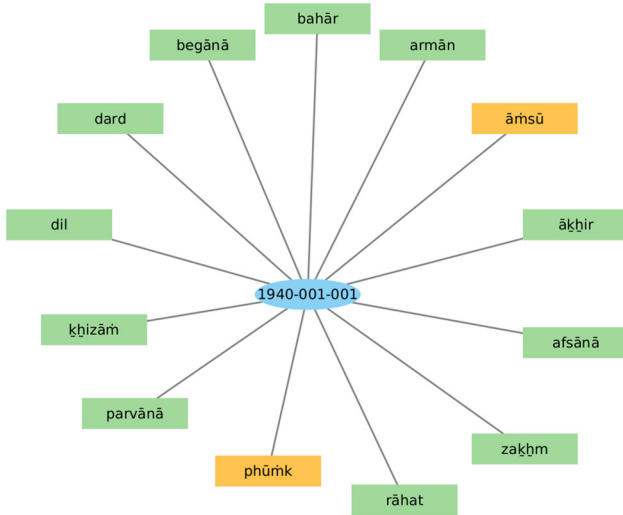


Fig. 5. Nouns of Indo-Aryan and Perso-Arabic origin in the song “Hey Diwana Hoon, Diwana Hoon” (1940-001-001).

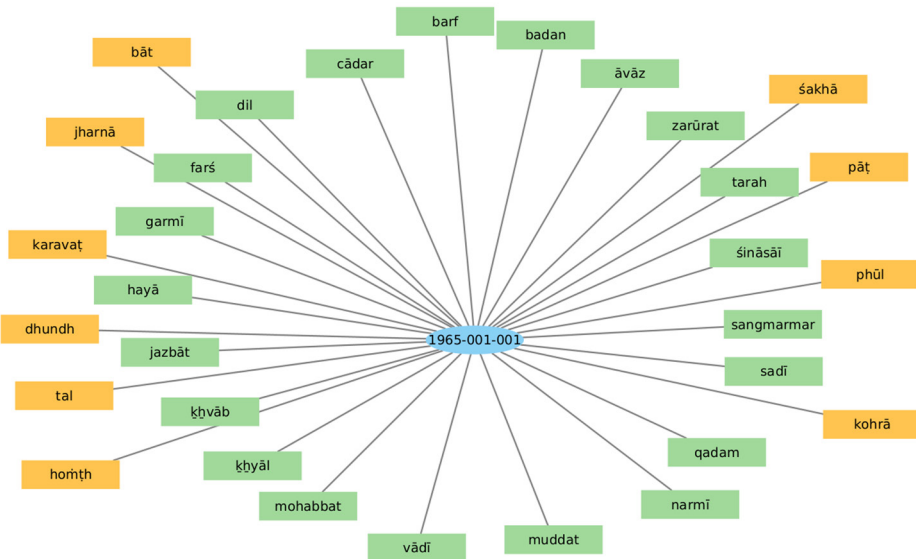


Fig. 6. Nouns of Indo-Aryan and Perso-Arabic origin in the song “Maine Ek Khwab Sa Dekha Hai” (1965-001-001).⁶

⁶ We have to bear in mind that some of the words in question possess variant forms, e.g. *karvaṭ*, *ḍhūmdh*, *sākh*, *kḥayāl(āt)*, *śanāsāi*.

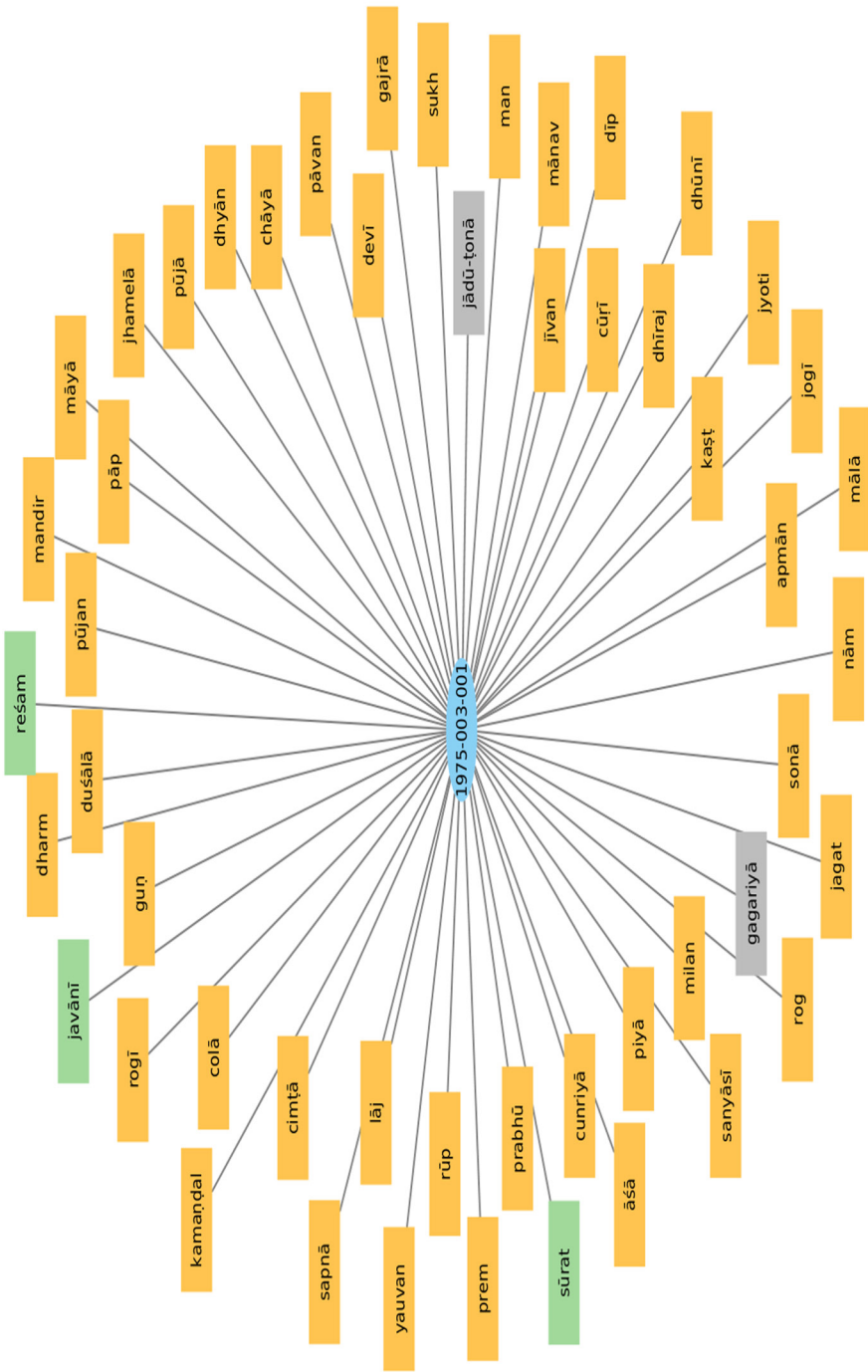


Fig. 7. Nouns of Indo-Aryan and Perso-Arabic origin in the song “Chal Sanyasi Mandir Men” (1975-003-001).

Topic – source language correlation

While we can observe no correlation between the time and proportion of words of Persian and/or Arabic origin in the analysed texts, let us see if there is any link between the topic(s) of a song and the proportion of lexemes of particular origin used in it. The contents of all the analysed songs were classified as belonging to at least one and at most three of the following categories: *love*, *social_problems*, *Hindu_religious*, *joy_and_celebration*, *life_and_fortune*, *friendship*, *nature*, *Muslim_religious*, *other*. The number of songs belonging to each of those categories varies greatly.

Category	Number of songs
<i>love</i>	51
<i>joy_and_celebration</i>	20
<i>Hindu_religious</i>	13
<i>nature</i>	8
<i>other</i>	7
<i>social_problems</i>	6
<i>Muslim_religious</i>	5
<i>friendship</i>	4
<i>life_and_fortune</i>	3

Before we go on with the analysis of the topic – source language correlation – let us make two important remarks. Firstly, the classification is based on the contents of the analysed songs and not of the whole movies. Secondly, it is, of course, to a large extent arbitrary. Other categories could be put forward and in some cases the classification may be disputed. Nevertheless, the author of the present article believes that this is a promising start of research into this subject.

Let us now have a look at the proportions of Indo-Aryan and Persian and/or Arabic vocabulary in the songs classified to selected categories. The forms present in more than 10 songs were chosen plus *Muslim_religious*, which is much less frequent, but interesting for obvious reasons.

	Indo-Aryan	Persian and/or Arabic	Other
<i>love</i>	549 (53.09%)	434 (41.97%)	51 (4.93%)
<i>joy_and_celebration</i>	217 (51.42%)	154 (36.50%)	51 (12.09%)
<i>Hindu_religious</i>	233 (88.26%)	22 (8.33%)	9 (3.40%)
<i>Muslim_religious</i>	37 (38.54%)	54 (56.25%)	5 (5.20%)

The results for the religious texts is as could have easily been predicted: Indo-Aryan terms are clearly dominating in the Hindu religious songs, while Perso-Arabic vocabulary is dominant (though less decisively) in the songs tagged as *Muslim religious*. Let us also note, that the domination of Perso-Arabic elements over Indo-Aryan ones in the songs classified as Muslim religious is less significant than the predominance of Indo-Aryan lexemes in Hindu religious texts, which is, however, hardly surprising, as it is the Indo-Aryan substrate that provides the basis for Hindi/Urdu.

It may be interesting to see what particular Persian and/or Arabic words are used in the songs classified as Hindu religious:⁷

Lemma	Source	Meaning
<i>badlā</i>	AR > FA <i>badal</i> “change, substitution, retaliation” ⁷ (STEINGASS 1892: 164)	“change, exchange, substitute, retaliation, etc.” (MCGREGOR 1993: 703)
<i>bandā</i>	FA <i>banda</i> “bound; slave” (STENGASS 1892: 202) “= I”	“1. servant; slave. 2. your servant (self-deprecatory term used in addressing a superior)” (MCGREGOR 1993: 692)
<i>bīmā</i>	FA <i>bīma</i> “insurance” (STEINGASS 1892: 225)	“1. insurance; insurance premium, or rate. 2. colloq. sthg. insured” (MCGREGOR 1993: 739)
<i>dar</i>	FA <i>dar</i> “door, gate, etc.” (STEINGASS 1892: 506)	“door, gate, etc.” (MCGREGOR 1993: 479)
<i>daurā</i>	AR > FA <i>daura(t)</i> “epicycle, revolution, etc.” (STEINGASS 1892: 543)	“turn, tour, session, etc.” (MCGREGOR 1993: 515)
<i>dil</i>	FA <i>dil</i> “heart, mind, soul, etc.” (STEINGASS 1892: 530)	“heart, soul, etc.” (MCGREGOR 1993: 496)
<i>duniyā</i>	AR > FA <i>dunyā</i> “world, people” (STEINGASS 1892: 539)	“world, people, etc.” (MCGREGOR 1993: 503)
<i>duśman</i>	FA <i>duśman</i> “enemy, etc.” (STEINGASS 1892: 526)	“enemy, etc.” (MCGREGOR 1993: 505)
<i>ġarīb</i>	AR > FA <i>ġarīb</i> “foreigner, a stranger; adj. poor, humble; etc.” (STEINGASS 1892: 886)	“adj. & m. 1. adj. poor. 2. humble, modest” (MCGREGOR 1993: 257)
<i>jādū</i>	FA <i>jādū</i> “conjuration, magic” (STEINGASS 1892: 349)	“magic, enchantment, etc.” (MCGREGOR 1993: 367)

⁷ The meanings provided in this table are based upon the definitions but are not precise citations. Please refer to the original publications for complete definitions.

Lemma	Source	Meaning
<i>javānī</i>	FA <i>javānī</i> “youth; youthful; a young man” (STEINGASS 1892: 376)	“youth, etc.” (MCGREGOR 1993: 364)
<i>kinārā</i>	FA <i>kanār</i> , <i>kinār</i> “side, brink, coast, limit, etc.” (STEINGASS 1892: 1052)	“side, edge, etc.” (MCGREGOR 1993: 196)
<i>lālā</i>	? FA <i>lālā</i> “incomparable; chief servant; etc.” (STEINGASS 1892: 1112)	“a respectful designation for members of some communities” (MCGREGOR 1993: 893)
<i>mahal</i>	AR > FA <i>mahal</i> “abode, mansion, palace, etc.” (STEINGASS 1892: 1189)	“abode, palace, etc.” (MCGREGOR 1993: 799)
<i>maṃzil</i>	AR > FA <i>manzil</i> “inn, house, mansion, etc.” (STEINGASS 1892: 1326)	“halting place, destination, floor, etc.” (MCGREGOR 1993: 778)
<i>mastī</i>	FA <i>mastī</i> “intoxication, lust, love” (STEINGASS 1892: 1236)	“intoxication, passion, delight, etc.” (MCGREGOR 1993: 798)
<i>mohabbat</i>	AR > FA <i>mahabbat</i> “love, etc.” (STEINGASS 1892: 1182)	“love” (CATURVEDI 1970: 622, 827)
<i>nazariyā</i>	associated with AR > FA <i>nazar</i> “look, vision, sight, etc.” (STEINGASS 1892: 1410)	“point of view, viewpoint; angle” (CATURVEDI 1970: 349)
<i>rahmat</i>	AR > FA “divine mercy, compassion, forgiveness, etc.” (STEINGASS 1892: 571)	“divine mercy; gift from Allāh, etc.” (MCGREGOR 1993: 858)
<i>rāstā</i>	FA <i>rāsta</i> “just, true; market place; level road” (STEINGASS 1892: 563)	“road, street, way, etc.” (MCGREGOR 1993: 863)
<i>rešam</i>	FA <i>rešam/rīšam</i> “silk” (STEINGASS 1892: 603)	“silk” (MCGREGOR 1993: 870)
<i>sīnā</i>	FA <i>sīna</i> “bosom, breast, etc.” (STEINGASS 1892: 719)	“chest, breast” (MCGREGOR 1993: 1020)
<i>sitārā</i>	FA <i>sitāra</i> “star, horoscope, etc.” (STEINGASS 1892: 654)	“star, fortune, etc.” (MCGREGOR 1993: 1013)
<i>sūrat</i>	AR > FA <i>sūrat</i> “image, form, face, etc.” (STEINGASS 1892: 795)	“form, appearance, face, etc.” (MCGREGOR 1993: 1036)
<i>umariyā</i>	AR > FA <i>umr</i> “life, lifetime, age” (STEINGASS 1892: 866)	obviously related to <i>umr</i> “life, lifetime, age” (MCGREGOR 1993: 131)
<i>vādā</i>	AR > FA “promise, vow” (STEINGASS 1892: 1473)	“promise, agreement, etc.” (MCGREGOR 1993: 913)
<i>yād</i>	FA “remembrance, memory, etc.” (STEINGASS 1892: 1524)	“memory, recollection” (MCGREGOR 1993: 843)
<i>zor</i>	FA <i>zūr/zor</i> “strength, power, violence, etc.” (STEINGASS 1892: 628)	“strength, effort, etc.” (MCGREGOR 1993: 383)

Interestingly, the data do not strongly support the stereotype of “Urdu as a language of love”, as the numbers for this category (53.09% for Indo-Aryan vocabulary and 41.97% for Perso-Arabic) are close to the ratio within the whole database (54.77% and 39.83% respectively).

Conclusion

The analysed material shows that the lexemes of Persian and Arabic origin constitute a significant part of the nominal vocabulary of Bollywood songs. Their proportion varies significantly from song to song. However, looking at a sample from a given year, this proportion is a bit more balanced, with the average presence of Persian and Arabic nominal lexemes at 38.17%, though in some years within the analysed corpus they might have been well above 50%.

While relatively big differences between the outcome from particular years seem to indicate that more data should be used to obtain precise results, still the studied material seems to indicate that there is no observable tendency to either diminish or expand the use of the Persian and Arabic nominal lexemes over time. The discussed elements remain a vital part of the vocabulary of Bollywood songs with no observable tendency to eradicate them.

On the other hand, it is possible to show some correlation between the subject of a song and the ratio of the Perso-Arabic vocabulary it contains. This has been shown on the basis of religious-themed texts, probably the most suitable for this.

As we have noticed at the very beginning of the present text, the language should be interpreted as an element of culture. What does this situation tell us about modern Indian popular culture? Or at least one of its most important aspects, i.e. Bollywood movies?

To answer this question, let us have a look at the Persian language but from a different angle than before. For decades, the foundations of the discussion of the Western influence on Iranian culture has been the book *Qarbzadegi* by Jalāl Āl-e Ahmad (first published in 1962). The title is self-explanatory. It is composed of two elements *qarb* “west” and *zadegi*, which may be translated as “sickness” or “struckness”, hence the translations of the name as “weststruckedness” or “occidentosis”. Indeed, the book presents the Western influences as a sickness attacking the vernacular Iranian culture. Jalāl Āl-e Ahmad presents the language and linguistic influence among the important elements of this disease (eg. ĀL-E AHMAD 2006–2007: 107, 156).

As Āl-e Ahmad’s ideas became an important component of the ideology of the Islamic Republic of Iran, it is not surprising that we find relevant tendencies in its linguistics policies, too. The latest incarnation of the Farhangestān or the

Academy of Persian Language and Culture (in existence since 1987) posits that only widespread internationalisms of Western origin already well established in the language can be accepted, while all the others should be substituted with Persian equivalents (JAZAYERI 1999). Such puristic tendencies are not only aimed at the new vocabulary items, but also result in numerous revisions of the already existing terminology (JAZAYERI 1999). This is reflected in the legal sphere, as since at least 2006, the use of the forms coined by Farhangestān has been mandatory for all Iranian institutions (MARSZALEK-KOWALEWSKA 2011: 100).

In other words, what we see in the case of Iran is the officially adopted policy of the rejection of Western cultural influences, with language borrowings seen as an important element of these influences.

The attitude towards Persian and Arabic vocabulary in India, and Bollywood songs in particular, is entirely different. While puristic tendencies are not unknown in Hindi (especially the re-Sanskritisation movement), we see no powerful drive to eradicate the lexica of Perso-Arabic origin, and the present work proves this on the basis of Bollywood songs' texts. Moreover, "obscurantism" and "purism" are sometimes presented as negative factors in the development of Hindi vocabulary (MALLIKARJUN 2004). In other words, Bollywood songs remain open to Perso-Arabic culture or at least to one of its aspects, i.e. the language influences, and this Perso-Arabic component remains an important part of Indian culture.

The language of the Bollywood songs certainly deserves more research, broadening both the corpus of the analysed texts, as well as including more problems, e.g. code-switching, syntactical influences, presence of vocabulary borrowed from other languages (especially other Indo-Aryan languages and English).

Appendix – list of the analysed songs

No.	Id (incl. year)	Movie	Song
1	1940-001-001	<i>Zindagi</i>	Hey Diwana Hoon, Diwana Hoon
2	1940-002-001	<i>Bandhan</i>	Chane Jor Garam Babu Main Laya Mazedar
3	1940-003-001	<i>Achhut</i>	Deen Dukhi Ko Daan Diya To Kya Mila Hari Ram Mila

4	1940-003-002	<i>Achhut</i>	Naahi Bolun Naahi Bolun Lakh Manaye
5	1940-005-001	<i>Nartaki</i>	Madbhari Rut Jawan Hai
6	1940-005-002	<i>Nartaki</i>	Yeh Kaun Aaj Aaya Savere Savere
7	1940-006-001	<i>Punar Milan</i>	Naacho Naacho Pyare Man Ke Mor
8	1945-001-001	<i>Zeenat</i>	Naacho Sitaro Naacho Ab Chand Nikalne Wala Hai
9	1945-002-001	<i>Gaon Ki Gori</i>	O Pardesi Raja
10	1945-003-001	<i>Badi Maa</i>	Mata Tere Charnon Mein Guzar Jaye Umariya
11	1945-005-001	<i>Tadbir</i>	Raani Khol De Apane Duvar Milane Kaa Din Aa Gayaa
12	1945-007-001	<i>Humayan</i>	Naina bhar aaye neer
13	1945-007-002	<i>Humayan</i>	Jo Desh Kal Nahin Tha
14	1950-001-001	<i>Samadhi</i>	Gore Gore O Baanke Chhore
15	1950-002-001	<i>Babul</i>	Dhadke Mera Dil Mujhko Jawani...
16	1950-003-001	<i>Dastaan</i>	Ta Ra Ri Aa Ra Ri Aa Ra Ri...
17	1950-004-001	<i>Jogan</i>	Ghunghat Ke Pat Khol
18	1950-004-002	<i>Jogan</i>	Main To Girdhar Ke Ghar Jaaon
19	1950-005-001	<i>Har Har Mahdev</i>	Man Naa Maane Man Ki Batiyaan
20	1950-006-001	<i>Sangram</i>	Kas Ke Kamar Ho Ja Taiyyar
21	1950-007-001	<i>Beqasoor</i>	Chale Jaiho Bedarda Main
22	1955-001-001	<i>Shree 420</i>	Mera Joota Hai Japani
23	1955-002-001	<i>Azaad</i>	Naa Bole Naa Bole Naa Bole Re

24	1955-003-001	<i>Jhanak Jhanak Payal Baje</i>	Jhanak Jhanak Payal Baaje
25	1955-003-002	<i>Jhanak Jhanak Payal Baje</i>	Kaisi Yeh Mohabbat Ki Saza
26	1955-004-001	<i>Uran Khatola</i>	Mera Salaam Le Ja Dil Ka Payaam Le Ja
27	1955-005-001	<i>Mr & Mrs 55</i>	Ae Ji Dil Par Hua Aisa Jadoo
28	1955-006-001	<i>Insaaniyat</i>	Tumhare Mohabbat Ke Hum Hain Ghulam
29	1955-006-002	<i>Insaaniyat</i>	Main Bandar Hoon Shahar Ka
30	1960-001-002	<i>Mughal E Azam</i>	Mohe Panghat Pe Nandlal
31	1960-002-001	<i>Barsaat Ki Raat</i>	Garjat Barsat Saawan Aayo Re
32	1960-003-001	<i>Kohinoor</i>	Tan Rang Lo Ji Aaj Man Rang Lo
33	1960-004-001	<i>Chaudvin Ka Chand</i>	Yeh Lucknow Ki Sar Zamin
34	1960-005-001	<i>Jis Desh Mein Ganga Behti Hai</i>	Mera Naam Raju Gharana Anaam
35	1960-006-001	<i>Dil Apna Aur Preet Paraya</i>	Sheesha-E-Dil Itna Na Uchhaalo
36	1965-001-001	<i>Waqt</i>	Maine Ek Khwab Sa Dekha Hai
37	1965-002-001	<i>Jab Jab Phool Khile</i>	Pardesiyon Se Naa Ankhiya Milaana
38	1965-003-001	<i>Himalay Ki God Mein</i>	Chand Si Mehbooba Ho Meri
39	1965-004-001	<i>Arzoo</i>	Bedardi Baalma Tujhko Mera Man Yaad Karta Hai
40	1965-005-001	<i>Guide</i>	Gaata Rahe Mera Dil
41	1970-001-001	<i>Johny Mera Naam</i>	O Mere Raja
42	1970-002-001	<i>Sachaa Jhutha</i>	Kar Le Pyar Kar Le Aankhen Char Kar Le
43	1970-003-001	<i>Purab Aur Pachhim</i>	Om Jai Jagdish Hare

44	1970-004-001	<i>Kati Patang</i>	Ye Shaam Mastani
45	1970-005-001	<i>Jeevan Mrityu</i>	Jhilmil Sitaron Ka Aangan Hoga - Duet
46	1975-001-001	<i>Sholay</i>	Yeh Dosti Hum Nahin Todenge
47	1975-002-001	<i>Jai Santoshi Maa</i>	Main Toh Aarti Utaaroon Re Santoshi Mata Ki
48	1975-003-001	<i>Sanyasi</i>	Chal Sanyasi Mandir Men
49	1975-004-001	<i>Deewaar</i>	Idhar Ka Maal Udhar
50	1975-005-001	<i>Pratigya</i>	Morni Re Morni Main Jungle Ki Morni
51	1980-001-001	<i>Qurbani</i>	Qurbani Qurbani Qurbani Allah Ko Pyari Hai Qurbani
52	1980-002-001	<i>Aasha</i>	Tune Mujhe Bulaya Shera Waliye (Jai Mata Di)
53	1980-003-001	<i>Ram Balram</i>	Hamka Maafi Dai Do
54	1980-004-001	<i>Dostana</i>	Salamat Rahe Dostana Hamara – Part 1
55	1980-005-001	<i>Shaan</i>	Doston Se Pyar Kiya
56	1985-001-001	<i>Ram Teri Ganga Maili</i>	Tujhe Bulayen Yeh Meri Baahen
57	1985-003-001	<i>Geraftaar</i>	Dhoop Mein Nikla Na Karo
58	1985-004-001	<i>Pyar Jhukta Nahin</i>	Tumse Milkar Naa Jaane Kyun (Female)
59	1985-005-001	<i>Ghulami</i>	Zihal-E-Maskeen
60	1990-001-001	<i>Dil</i>	O Priya Priya
61	1990-002-001	<i>Ghayal</i>	Maahiya Teri Kasam
62	1990-003-001	<i>Aaj Ka Arjun</i>	Gori Hai Kalaiyan Tu La De Mujhe Hari Hari Choodiyan
63	1990-004-001	<i>Agneepath</i>	I Am Krishnan Iyer M.A.
64	1990-005-001	<i>Kishen Kanhaiya</i>	Krishna Krishna Haye Krishna

65	1995-001-001	<i>Dilwale Dulhania Le Jayenge</i>	Tujhe Dekha To Ye Jaana Sanam
66	1995-002-001	<i>Karan Arjun</i>	Jaati Hoon Main Jaldi Hai Kya
67	1995-003-001	<i>Coolie No 1</i>	Aa Jaana Aa Jaana
68	1995-004-001	<i>Raja</i>	Nazrein Mili Dil Dhadka
69	1995-005-001	<i>Rangeela</i>	Rangeela Re
70	2000-001-001	<i>Kaho Naa Pyaar Hai</i>	Kaho Naa Pyar Hai
71	2000-002-001	<i>Mohabbatein</i>	Humko Humise Chura Lo
72	2000-003-001	<i>Mission Kashmir</i>	Bumbro
73	2000-004-001	<i>Josh</i>	Sailaru Sailare
74	2005-001-001	<i>No Entry</i>	Ishq Di Gali Wich No Entry
75	2005-002-001	<i>Bunty Aur Babli</i>	Dhadak Dhadak
76	2005-003-002	<i>The Rising: Ballad Of Mangal Pandey</i>	Main Vari Vari
77	2005-004-001	<i>Maine Pyaar Kyun Kiya</i>	Just Chill
78	2005-005-001	<i>Garam Masala</i>	Ada Aay Haay Ada
79	2010-001-001	<i>Dabangg</i>	Tere Mast Mast Do Nain (Male)
80	2010-003-001	<i>Raajneeti</i>	Bheegi Si Bhaagi Si
81	2010-004-001	<i>Housefull</i>	Oh Girl You're Mine
82	2010-005-001	<i>My Name is Khan</i>	Sajdaa
83	2015-001-001	<i>Talvar</i>	Insaaf Insaaf Insaaf Hoga
84	2015-002-001	<i>Bajrangi Bhaijaan</i>	Selfie Le Le Re
85	2015-003-001	<i>Baby (I)</i>	Beparwah
86	2016-005-001	<i>Bajirao Mastani</i>	Deewani Mastani
87	2020-001-001	<i>Baaghi 3</i>	Dus Bahane 2.0

88	2020-002-001	<i>Is Love Enough, Sir?</i>	Jeene Mein Kya Jaata Hai
89	2020-002-002	<i>Is Love Enough, Sir?</i>	Zindagi Ka Feel
90	2020-004-001	<i>Sayonee</i>	Sayonee
91	2020-005-001	<i>Shakeela</i>	Woh Lamha
92	2020-005-002	<i>Shakeela</i>	Taaza

Abbreviations

AR	Arabic
DR	Dravidian
EN	English
FA	Persian (Fārsi)
IN	Indian
obl. pl.	Oblique case plural
rect. sg.	Direct case singular
SH	Solar Hijri

References

- ĀL-E AHMAD, Jalāl 2006–2007 (1385 SH). *Gharbzadegi*. Qom: Nashr-e Xorram.
- BEHL, Aseem and Monojit CHOUDHURY 2011. “A Corpus Linguistic Study of Bollywood Song Lyrics in the Framework of Complex Network Theory”. [In:] *Proceedings of ICON-2011: 9th International Conference on Natural Language Processing*. Macmillan Publishers, India. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=994ac9ad1054b8691f7fcb939f18723d0ba99a71> (accessed 23 February 2024).
- BOX OFFICE INDIA. <https://boxofficeindia.com> (accessed between 12 March 2022 and 20 February 2023) including its older content preserved at <http://web.archive.org> (accessed between 12 March 2022 and 12 February 2023).
- CATURVEDI, Mahendra 1970. *A practical Hindi-English dictionary*. Delhi: National Publishing House.
- CHANDRA, Subhash, Bhupendra KUMAR, Vivek KUMAR and SAKSHI 2016. “Acute Sporadic English in Bollywood Film Songs’ Lyrics: A Textual Evidence based Analysis of Code-Mixing in Hindi”. *Language in India* 16(11): 25–34. <http://www.languageinindia.com/nov2016/bollywoodhinglishsongs1.pdf> (accessed 10 February 2023).

- CRYSTAL, David 2008. *A Dictionary of Linguistics and Phonetics*. Sixth edition. Malden, Oxford, Victoria: Blackwell Publishing.
- GANTI, Tejaswini 2012. *Producing Bollywood: Inside the Contemporary Hindi Film Industry*. Durham: Duke University Press. <https://doi.org/10.1515/9780822395225>
- HABIB, Muhammad A., Muhammad SAEED, Arshad A. KHAN and Mohsin A. HANJRA 2020. "An Analysis of English-Urdu Code-Switching in Bollywood Songs". *Jahan-e-Tahqeeq* 3(2): 55–66. <https://jahan-e-tahqeeq.com/index.php/jahan-e-tahqeeq/article/download/287/210> (accessed 10 February 2023).
- HINDI GEET MALA. <https://www.hindigeetmala.net> (accessed between 12 March 2022 and 20 February 2023).
- IMDb. <https://www.imdb.com> (accessed 3 February 2022).
- JAZAYERI, Mohammed Ali 1999. "Farhangestān". [In:] Yarshater, Ehsan et al., eds, *Encyclopædia Iranica*. <https://iranicaonline.org/articles/india-xiv-persian-literature-in-india> (accessed 30 March 2021).
- KHAN, Subuhi and Laramie D. TAYLOR 2018. "Gender Policing in Mainstream Hindi Cinema: A Decade of Central Female Characters in Top-Grossing Bollywood Movies". *International Journal of Communication* 12(2018): 3641–3662. <https://ijoc.org/index.php/ijoc/article/view/8701/2448> (accessed 27 December 2023).
- KRAMSCH, Claire 2014. "Language and Culture". *AILA Review* 27(1): 30–55. <https://doi.org/10.1075/aila.27.02kra>
- KRETZSCHMAR, William A., Jr. 2010. *The Linguistics of Speech*. Cambridge: Cambridge University Press.
- MALLIKARJUN, B. 2004. "Fifty Years of Language Planning for Modern Hindi The Official Language of India". *Language in India* 4(11). <http://www.languageinindia.com/nov2004/mallikarjunmalaysiapaper1.html> (accessed 31 March 2021).
- MARSZALEK-KOWALEWSKA, Katarzyna 2011. "Iranian Language Policy: A Case of Linguistic Purism". *Investigationes Linguisticae* 22: 89–103. <https://doi.org/10.14746/il.2010.22.5>
- MCGREGOR, Ronald S. 1993. *The Oxford Hindi-English Dictionary*. Oxford (England), New York: Oxford University Press.
- MYSWAR. myswar.co (accessed between 12 March 2022 and 20 August 2023).
- PUSTKA, Elissa 2022. "Loanword phonology in Romance". [In:] Gabriel, Christoph, Randall Gess and Trudel Meisenburg, eds, *Manual of Romance Phonetics and Phonology*. Berlin, Boston: De Gruyter, pp. 503–530. <https://doi.org/10.1515/9783110550283-017>

- RAI, Amit S. 2009. *Untimely Bollywood. Globalization and India's New Media Assemblage*. Durham: Duke University Press. <https://doi.org/10.2307/j.ctv125jmcz>
- SARRAZIN, Natalie 2008. "Celluloid love songs: musical *modus operandi* and the dramatic aesthetics of romantic Hindi film". *Popular Music* 27(3): 393–411. <https://doi.org/10.1017/S0261143008102197>
- SINHA, Anubhav and Laagoo MRUNMAYEE 2020. *Thappad*. https://images.assettype.com/filmcompanion/2022-10/ac804418-2120-433b-8d94-bcc6165f716d/Thappad_Shootinhg_Draft_2.pdf (accessed 30 April 2023).
- SKEAT, Walter W. 1993. *The Concise Dictionary of English Etymology*. Ware (Hertfordshire): Wordsworth Editions.
- STEINGASS, Francis J. 1892. *A Comprehensive Persian-English dictionary, including the Arabic words and phrases to be met with in Persian literature*. London: Routledge & Kegan Paul.
- TURNER, Ralph L. 1966. *A Comparative Dictionary of Indo-Aryan Languages*. London: Oxford University Press.
- VERMA, Garima and Hemraj VERMA 2019. "Predicting Bollywood Movies Success Using Machine Learning Technique". [In:] Khatri, Sunil Kumar, Ajay Rana and P. K. Kapur, eds, *2019 Amity International Conference on Artificial Intelligence (AICAI)*. Dubai: IEEE, pp. 102–105. <https://doi.org/10.1109/AICAI.2019.8701239>
- WHORF, Benjamin L. 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Ed. by Carroll, John B. Massachusetts: MIT Press.