



Grau de Estudis Anglesos

Treball de Fi de Grau

Academic Year: 2019-2020, G4

A Comparative Analysis of Colour Lexicalization Patterns
in Urdu and Hindi

STUDENT'S NAME: Sehar Ambreen Rashid

TUTOR'S NAME: Emilia Castaño

Barcelona, 11th June of 2020



Declaració d'autoria

Amb aquest escrit declaro que sóc l'autor/autora original d'aquest treball i que no he emprat per a la seva elaboració cap altra font, incloses fonts d'Internet i altres mitjans electrònics, a part de les indicades. En el treball he assenyalat com a tals totes les citacions, literals o de contingut, que procedeixen d'altres obres. Tinc coneixement que d'altra manera, i segons el que s'indica a l'article 18, del capítol 5 de les Normes reguladores de l'avaluació i de la qualificació dels aprenentatges de la UB, l'avaluació comporta la qualificació de "Suspens".

Barcelona, a 11 de juny de 2020

Signatura: 

I would like to express my humble gratitude to Dr. Emilia Castaño whose countless feedback, guidance, advice and motivation has helped me through the whole process of writing my End of Degree Paper (EDP)..

Abstract

The aim of this Corpus-based study is to compare and contrast colour lexicalization patterns in two Indo-Iranian languages, Hindi and Urdu. Both languages derive from the same ancestor language, Sanskrit and are so similar that they are practically indistinguishable. However, despite their similarities, previous researches have shown that significant differences can be observed in technical terms and high-level vocabulary. This study analyses colour lexicalization in an attempt to explore whether Hindi and Urdu might also exhibit significant lexical differences in basic vocabulary. In order to conduct this study, I have analysed in a corpus twelve colour terms whose counterparts have already been analysed by Drocco and Risato's work on colour lexicalization in modern Hindi (2019). Results show that most of the differences between Hindi and Urdu colour lexicalization patterns are related to what Drocco and Risato (2019) call "approximation", colour words that are formed either by suffixation or reduplication and that are used to indicate inexactness or uniqueness. This could be due to the fact that many "approximation type" colour terms combine colour words from different origins, mainly Hindi and Persian words.

Keywords: Lexicalization, colour words, high-level vocabulary, basic vocabulary, corpus.

Resumen

El propósito de mi trabajo es observar la diferencia entre dos lenguas Indo-Europeas, Hindi y Urdu que comparten la misma lengua materna; Sanskrit. La similitud de las dos lenguas es tanta que a veces son prácticamente indistinguibles, aunque la diferencia se puede percibir en el registro más culto de las lenguas. Para poder confirmar esta hipótesis, doce colores de Hindi han sido elegidos de un trabajo escrito por Drocco y Risato sobre la lexicalización de términos de colores en Hindi moderno (2019). Una vez escogidos los colores, el siguiente paso ha sido examinar los en un corpus basado en Urdu. Los resultados encontrados demuestran que la mayoría de las diferencias se encuentran en "approximation type" lexicalización (Drocco and Risato, 2019). Los resultados demuestran que la mayoría de las diferencias se perciben en "approximation type lexicalization".

Palabras claves: Lexicalización, términos de colores, registro culto, lenguaje coloquial, corpus

Table of contents

1. Introduction	1
2. History of Urdu and Hindi	2
2.1 Urdu	2
2.1.1 Language family	
2.1.2 Origin and influence	
2.1.3 Where is it spoken	
2.2 Hindi	4
2.2.1 Language family	
2.2.2 Origin and influence	
2.2.3 Where is it spoken	
2.3 Main differences between Urdu and Hindi	6
3. Lexicalization pattern of colour	7
3.1 Culture, thought and language	7
3.1.1 Colour and culture	
3.2 Drocco and Risato's colour lexicalization pattern in modern Hindi	8
4. Colour lexicalization pattern drawn from an Urdu corpus	10
4.1 Metonymy type lexicalization	10
4.1.1 Corpus findings on Urdu colours	10
4.1.1.1 Khāki	
4.1.1.2 Gulābī	
4.1.1.3 Badāmi	
4.1.1.4 Gājri	
4.1.1.5 Zāfrānī	
4.1.1.6 Gandumī	

4.1.2	A brief comparison of data found “on metonymy type” lexicalization between Urdu and Hindi	13
4.2	“Approximation type” lexicalization	14
4.2.1	Corpus based findings on Urdu colours	14
4.2.1.1	Lāl	
4.2.1.2	Nīla	
4.2.1.3	Harā	
4.2.1.4	Kāla	
4.2.1.5	Pīla	
4.2.1.6	Harā	
4.2.2	A brief comparison of data found on “approximation type” lexicalization between Urdu and Hindi	21
5.	Conclusion	22
6.	References	23

1. Introduction

As a speaker of Urdu and Punjabi, I have always admired Indo-European languages, specifically south Asian languages. My admiration for Indo-European languages comes from a course I took in high school on Latin. Our teacher gave us a tree-diagram of Indo-European languages, something I was not aware of, and familiarise us to the origin of Latin. My curiosity took me to investigate a bit more on my mother tongue Urdu and when it was time to select a topic to work on for our last project in Batxillert (Trell de Recerca), I took this opportunity and compared Urdu and Hindi based on the usage of grammar. Years later, selecting Urdu and Hindi as a topic for my end of degree paper was not challenging to me because I wanted to continue my research on how different Urdu and Hindi are.

My end of degree paper focuses on a comparison of colour lexicalization patterns in Hindi and Urdu. On the one hand, both languages belong to the Indo-Iranian branch of the Indo-European family and descend from Sanskrit. On the other hand, despite their similarities, both languages show many differences, particularly, in cultivated language and administrative vocabulary. The aim of this paper is to detect if differences between Urdu and Hindi on low-level vocabulary exist by creating a corpus-based analysis in Urdu on lexicalization of twelve colour terms (Khāki Gulābī, Badāmi, Gājri, Zāfrānī, Gandumī, Lāl, Nīla, Harā, kāla, pīla and harā).

This paper is divided into two parts: first of all, the *Literature review* concentrates on explaining the history of both languages and how they have developed into two different languages. Moreover, it will also tackle the causes behind their separation and the linguistic repercussion it had on the population of India and Pakistan. Secondly, the *Discussion of the paper* is a corpus-based analysis, in which I have compared the lexicalization of twelve Urdu colour terms whose counterparts in Hindi have already been analysed by Andrea Drocco and Orsola Risato (2019). The purpose of this comparison is to detect whether disagreements between Urdu and Hindi exist on the low-level vocabulary. Finally, the *Results* expose that Urdu and Hindi show some differences to Hindi colour lexicalization (Drocco and Risato, 2019) in some aspects.

2. History of Urdu and Hindi

2.1. Urdu

2.1.1. Language family

Urdu is one of the daughter languages of Proto-Indo-European. It belongs to the Indo-Iranian branch, which is in turn divided into two sub-branches called Iranian and Indic. The former includes the Semitic languages, that is, Persian and Arabic, to mention a few. The Indic sub-branch embraces the languages spoken in the subcontinent of India; nowadays India, Pakistan, Bangladesh and Sri-Lanka. Urdu belongs to the Indic languages.

2.1.2. Origin and influences

To start with, it is very important to take into account that Urdu is treated as dialect of Hindi and not as an individual language. This fact might be motivated by its obvious similarities with Hindi, which could be justified by their common origin: Sanskrit. Sanskrit, being the ancestor language of both Hindi and Urdu has lent words to both languages. However, in the case of Urdu, as opposed to Hindi, which is richer with Sanskrit words, only 7-10% of its vocabulary has its origins in Sanskrit. If Urdu has borrowed only 7-10% of its vocabulary from Sanskrit, what is the source of the other 90%? In order to answer this question, we have to take into consideration the present state of Urdu vocabulary and the historical events that have contributed to its development through time. Apart from Sanskrit, Urdu has also received the influence of other languages such as Arabic and Persian. Urdu not only borrowed vocabulary from these languages but also adopted the Arabic-Persian script as well.

Schmidt (2003, p. 288) claims that “Urdu has developed primarily from the sub-regional language of the Delhi area, which has been variously called Khari Boli”. The emergence of this language can be traced back to the Mughal empire after the conquest of Delhi. Even though the roots of the Urdu language are found in Dilli (Delhi), it was the city of Agra, which is located 200 km south from Delhi, that gave hostage to the Mughal empire and where it blossomed. During the Mughal conquest of Agra, the elites spoke and wrote in Persian while the military personnel spoke several dialects of Braj Bhasa, such as Kauravi and Hariyanvi (2003, p. 289).

I would like to refer to Nespital’s five periods of classical Urdu literature found in Schmidt (2003, p. 289) to provide a chronological description of the development of

Urdu. The first period, from 1648 to 1700, is defined by the “Rekhta Poetry of Sufis”. This type of poetry involved the use of the Devanagari and Perso-Arabic alphabet by the poets of the time, called “Sufis”. The second period, spanned from 1700 to 1720, is characterized by the Dakhni poetry by poets such as Vali Dakhni, who came to Delhi bringing with him the Dakhni language, which is considered as an archaic form of Urdu. This is the moment when the Dakhni language merged with Persian, giving way to a new language which later would be called Urdu. The next period (1770-1740) led to the “standardization of grammar of the new literary language” (Schmidt, 2003: 289). During these years, Urdu adopts the Perso-Arabic script, Ugarit. Nevertheless, it adapts the Akkadian version of the Perso-Arabic script and abandons the Ugarit version later (Khan, 2006, p. 350) in order to reflect the phonetic characteristics of Urdu.

The fourth period (1740-1780) was the “golden age of Urdu school of poetry and disintegration of the Mughal empire” in Delhi (Schmidt, 2003, p. 290). The last period started in 1780 and ended in 1810. During this span of time, the standardization of the language, that started in 1720, ended and the emergence of the Urdu language as a poetic language, started. In the late 18th century, the arrival of the East India company in India and the decline of the Mughal empire paved the path to the conquest of the Indian subcontinent by the British empire. To this point, Urdu and Hindi were not being developed under the influence of the Mughal empire but under the influence of the British empire. Under the British tutelage, the differences between Urdu and Hindi became more obvious, to the extent that they split up into two different languages. In fact, in 1947 the decision of the British empire to divide India into two countries -India and Pakistan- provoked a split between Hindi and Urdu, which were selected as the official languages of these two new countries.

As Khan mentions in his book, *The Politics of Language. Urdu/Hindi: An Artificial divide*, “Language-Culture is integrative, dynamic, and holistic, while religions are generally divisive and divisible into sects” (2006, p. 197). The division of India into India and Pakistan was a matter of religion while the distinction between Hindi and Urdu was merely based on politics. In order to understand this better, we need to rewind to the origin of British Raj tutelage in 1858 (Schmidt, 2003, p. 255). Even before this moment of complete control over the Indian population, the British exerted their power on them through their jurisdiction on language policy, which allowed the adoption of “Urdu as a successor of Persian (1835), as the administrative language and as a medium of education because of the national and global status of Persian script” (Khan, 2006, p. 6). This choice

had religious and political repercussions on India. On the one hand, languages were understood as religious symbols: Urdu represented the Muslim population of the country while Hindi represented the Hindu population. This partition between languages after the arrival of the British empire asked for the official recognition of both languages taking into account their religious symbolism, which ended up being a source of violence and conflict. On the other hand, “conflicts over language were also part of the long and often riotous politics of identity in the region” (King, 1994, as cited in Aneesh, 2010, p. 87). Once the languages had been labelled by religions and caused conflicts between Hindus and Muslims, it affected the politics. The idea of “two nation states” (Aneesh, 2010, p. 87) arose keeping in mind the “the production of modern Hindi in connection with the nationalist project of India” (Aneesh, 2010, p. 87). This idea of “nationalist Project” is valid for both countries. Both India and Pakistan were found of this idea and wanted to preserve religion, language and different political point of views. Thus, it is reasonable to say that the partition was a “political decision in their self-interest (Khan, 2003, p. 197).

2.1.3. Where is it spoken nowadays?

Urdu was proclaimed as the official language of Pakistan in 1947 and this was just because of the country’s Muslim religion. However, according to a study carried out in 1993 by the Summer Institute of Linguistics (SIL), *Ethnologue*, mentioned by Schmidt (2003, p. 302) in his chapter on the *Indo-Aryan languages*, only “7.5% of the population” considered Urdu as a mother tongue and only “7.6%” of the Muslims in Pakistan consider Urdu as their mother tongue (Jain, 2003, p. 57). This could be due to the fact that many regional languages are spoken in the country. In India, the percentage is smaller but still considerable. “5.13%” of the Indian population speaks Urdu (Schmidt, 2003, p. 302) and, according to a study conducted in 1991, “43%” of all Muslims in India consider Urdu as their mother language (Jain, 2003, p. 57).

2.2.Hindi

2.2.1. Language Family

Hindi belongs to the Proto-Indo-European family. It is also part of the Indo-Iranian branch and the Indo-Aryan sub-branch, like Urdu. Both languages are identical when referring to their origins, which makes very difficult to characterize them individually.

2.2.2. Origin and influences

In this section of my paper, I will explain the origins of Hindi and the influence that other languages have exerted on it. As mentioned above, Hindi is very similar to Urdu except high-level vocabulary such as administrative or poetic vocabulary.

Having explained the origin of Urdu, it is much easier to explain the roots of Hindi. Nevertheless, Hindi goes way back in history to find its roots. Hindi's history starts with the arrival of the Aryans in India around "1000 BC" (Khan, 2006, p. 16). At that time, on the one hand, Sanskrit was at its peak of development. It was normally aimed at the elite community of India, the Brahmins. On the other hand, another slightly less prestigious language was also being developed. This second language was considered to be the language of the farmers, in other words, the language of the lower class. Even though, it was the language of the farmers and lower classes, it managed to overtake Sanskrit. In 280 BC King Aśoka used Pali, one of the Prakrits, as his official language and made "Pali — not SKT — the oldest written language of India" (Khan, 2006, p. 16). This made Sanskrit a little less prestigious language but it was still kept as the language of the Brahmins, therefore, as a religious language.

During the first period of Urdu (1648-1700), the Mughal empire expanded to include Delhi and found a language spoken by the vast majority of the population: Pali and Khari Boli. After the arrival of the Mughal empire in Delhi, these languages shared history and linguistic features but only one of them managed to take over. Urdu ended up being the successor to the throne while Pali and Khari Boli lost ground. What's more, Pali even adopted the Persian-Arabic script during this time. This situation changed during the Fourth period in the history of the Urdu language because the Mughal empire started to lose its power (Schmidt, 2003, p. 209). The fall of the Mughal empire led the population to question their religious identity. Furthermore, differences between Hindus and Muslims caused conflicts and we can say that this could be the beginning of the religious consciousness created by the "two nation state" (Aneesh, 2010, p. 87).

2.2.3. Where is it spoken nowadays?

After the end of the British Raj in the Indian Sub-Continent, Hindi was declared as one of the official languages of India. A very fair amount of people speak Hindi in Pakistan as well. The Hindus in Pakistan have learnt this language in order to acquire more knowledge about their religion, as the holy books are mainly written in Hindi,

2.3 Main differences between Urdu and Hindi

Urdu and Hind can be called two different languages that emerged from one same dialect, Khari Boli. As I have mentioned above, the origin of both languages is found in the same mother language, Sanskrit, and the purposes of their division are merely religious and political. The speakers of Urdu and Hindi are able to understand the basic vocabulary of both languages but differences are made obvious when high level and administrative vocabulary is being used. We could barely differentiate these languages if it was not for administrative level vocabulary and in order to prove this hypothesis, I will focus on the domain of colour lexicalization as colours are part of the basic vocabulary of languages and, therefore, should not exhibit significant differences.

3. Lexicalization patterns of colours

3.1 Culture, thought and language

In this part of my paper I will emphasize on the influence of culture on language and thought. Even though our perceptual experience of colors is similar because our visual system is similar, the color continuum can be carved up differently in the world's different languages, which may be motivated by the characteristics of the physical environment and other cultural factors, as pointed out by Malta and Majid (2013). These linguistic differences may influence how their speakers think about colors. Languages also vary on the lexicalization patterns that they use to encode colors.

The cultural background is very important because according to what culture a language belongs, it will be more or less difficult to find a literal translation of a word into another language. All source languages show culturally-based lexical idiosyncrasies which cannot or are not easy to translate into the target language. For instance, the Arabic word "MashAllah", which can be roughly translated as "beautiful", does not hold the full essence of the word when translated into English. The English translation lacks the cultural component that the meaning of this word contains in Arabic. Nevertheless, languages that share more or less the same cultural background, like Urdu, will find it much easier either to borrow this word directly or to adapt it and integrate it into their vocabularies. The ethnic background creates environments which affect thoughts and stress the necessity to provide speakers with the necessary lexical resources to encode those thoughts. Thus, "Culture affects thoughts not language affects thoughts" (Malt and Majid, 2013, p. 584).

3.1.1. Colour and culture

As I have mentioned above, some lexicalization patterns are more dependent on culture than others and that is the case of colour lexicalization patterns in different languages.

The first studies on colour categorization and lexicalization (Berlin & Kay, 1969) pointed that there are "similarities in the ways that different languages with the same number of basic colour terms carve up colour space" (Hardin, 2013, p. 1). Such similarities, according to Berlin and Kay, were motivated by the fact that there is a hierarchy to the way languages organize colours with basic colours at the centre. This means that "colour

categories are not random and the centres of basic colour terms are very similar in all languages” (1969, p. 8).

A very important part of Berlin and Kay’s study was that they claimed that “colour words translated rather too easily among various pairs of unrelated languages” (1969, p. 1). This statement was questioned by Lyons (1997) who argued that “word-for-word translation of colour terms across languages is frequently impossible” (as cited in Hardin, 2013, p. 585). In this respect, my main claim is that “word by word” translation becomes even more difficult when there are substantial cultural differences between the source and the target languages. The fact that two languages are culturally different makes translation even more difficult than it actually is and colour terms are not an exception. They cannot be literally translated because background information and culture play an important role in the process. According to Malta and Majid (2013) the diversity of colours that languages exhibit could be due to these facts: the perception of physical environment, individual differences within population and different cultural practices. Thus, colour lexicalization cannot escape cultural variation.

To sum up, the very first study on colours was conducted by Berlin and Kay, however, their study has been criticised and revised. Many critics rejected their claim that “colour words translation” is quite straightforward because “colour terms are very similar in all languages” (1969, p. 8). The revised version of their theory acknowledges that “Basic colour terms” are not entirely similar in all languages because speakers’ cultural background plays a very important role in the way they name a certain colour. Therefore, some cultures will name a colour shade in a certain way because they don’t or do have that colour in their language and the same colour will be called by some other cultures differently because lexicalization of colour words takes into consideration social and cultural conditions.

3.2 Drocco and Risato’s colour lexicalization patterns in modern Hindi

Drocco and Risato’s study on modern Hindi colour patterns takes as a starting point basic colour terms and describes the lexicalization strategies that Hindi uses to describe different shades of these colours. It is a synchronic study based on corpus and dictionaries. The authors have focused on the colour terms used nowadays and have divided the lexicalization patterns that have identified into two categories: the metonymy type and the approximation type.

The colour terms grouped under the category *metonymy* are derivative, that is, “they are result of derivation by suffix -i” (Drocco and Risato, 2019, p. 192). This strategy has been defined as “object colour for colour as the suffix leads to the creation of a colour term” (2019, p. 192). The meaning of the suffix -i in Hindi is “having to do with”, for instance, the word *gulābī* [gulāb “rose” + i] which is translated into English as “pink” (Drocco and Risato, 2019, p. 194). Their data show that the suffix -i is typically used with fruit terms but it is not restricted to only those words. Thus, for instance, we can find words such as [Asmān “sky” + I = Asmāni] which means “sky blue”.

The second and the most extended part of Drocco and Risato’s study focuses on the colour terms lexicalised by means of approximation. The process of lexicalization by “approximation” is based on two different methods. On the one hand, the suffix -sā is used to make reference to a colour that is not exactly the colour being described. For instance; [lāl + sā = lāl-sā] “red-ish” or “red-like” makes reference to a colour which looks like red but it is not exactly the colour red. Indeed, this idea of not exactness of colours is related to perception and how different culture would perceive colours differently (Malt and Majid, 2010). On the other hand, the “reduplication” of colour terms have also been identified as a way of indicating colour approximation. It is distinguished from the use of the suffix -sā by the fact that reduplication is speaker and context dependent. According to the authors “the reduplication consists of the repetition of phonological material that marks individual and meaningful morphemes” (2019, p. 199). For example, colour words such as “pilē-pilē” which is translated into English, depending on the context, as “yellow-ish” or “pale” is a clear example of a reduplicated colour word in Hindi. Reduplication conveys context-related semantic connotations, which Abbi (1985), has classified into four groups: exclusiveness, augmentation of referents or variation of emphasis, temporal edition and distributiveness (as cited in Drocco and Risato, 2019, p. 200). However, I will only consider “augmentation of referents or variation of emphasis” a key property to determine colour lexicalization as the others refer to reduplication in general. The augmentation of referents is applied when a non-duplicated word fails to convey the meaning a reduplicated word would effectively do. A reduplicated colour form that increases the referent is making reference to a special colour word and not a regular one. For example, “Nilā” means “blue” but the reduplicated form “Nilē-Nilē” shows more, it is blue but not the regular blue; rather a specific one, an intense one.

The purpose of this study is to make observations on the lexicalization patterns used to describe some Hindi colour words.

4. Colour lexicalization pattern drawn from an Urdu corpus

This part of my study has taken Drocco and Risato's work as reference. In order to prove my thesis statement, I have based this part of my paper on Drocco and Risato's findings on colour lexicalization patterns in modern Hindi. As mentioned in section 3.2, the authors found two types of lexicalization patterns in Hindi colour terms. The first one is "metonymy type" and the second one is "approximation type". My aim is to analyse whether similar patterns are found in Urdu. I have selected twelve colour terms from Drocco and Risato's work: six of them belong to the "metonymy type" and the other six are examples of the "approximation type". In order to explore how these colours are lexicalized in Urdu, I have collected data from a corpus of Urdu texts.

First of all, Hindi terms were translated into using several dictionaries: (a) John Plats' dictionary, which contains Hindi words, their Urdu translation and their etymological origin; and (b) Lal Narain's Urdu dictionary, which is a bilingual Urdu-English dictionary that concentrates on Persian-origin words. Once the colour words had been translated in to Urdu, their lexicalization patterns were examined in a corpus of Urdu texts. The corpus used for Urdu is available on "Sketch Engine". It contains 53,269,273 words drawn from websites that have parts especially dedicated to Urdu such as "bbc.co.uk" and other websites written in Urdu. The study has been based on the analysis of concordances.

4.1 Metonymy type lexicalization

The colour terms that I have chosen are: *khāki*, *gulābi*, *bādāmī*, *gājri*, *zāfrānī* and *gandumī*.

4.1.1 Corpus based findings on Urdu colours

4.1.1.1 Khāki

According to Plats, the word *Khāk* means "dust" (2004, p. 1013) but it can also be used as colour term, in this case the suffix -i must be attached. Lal has translated *khaki* as "dusty, dust-coloured" (1897, p. 284). Based on my corpus analysis, only in 5 out of the

375 instances that appear in the corpus, *khāki*, is used as a colour term One of the example of *khāki* as a colour word is:

1. کوئی پرانا راز خاکى رنگ کی پریشانی میں

In + distress + of + colour + khāki + secret + old + some

Some old secret in a **dusty coloured** distress

[19661]

In this example the colour *khāki* is being post-modified by *rang* which means “colour. The Data shows that when this word is used as a colour term, it is always post-modified by the word “rang”. One of the reasons why colour *khāki* is not found with other post-modifiers is that it is not usually used as a colour term. As shown by the data in the corpus, *khāki* quite often is used as a proper noun such as, in the case of عدیل خاکی “Adil Khāki”, where Khāki is a surname.

4.1.1.2 Gulābī

This colour term has been translated as “rose coloured” (Plats, 2004:1862) and its origins are in the Persian language. Steingas had translated the word *gulāb* as “rose water” (1998, p. 844). In Urdu the suffix-i has been added in order to make a colour term *Gulābī* which means “pink” in Urdu

There are 173 instances of this term in the corpus and they all refer to the colour *gūlabī*. That is the case of *sūrkh gūlabī hōnth* which means “redish-pink lips”. It is quite frequent to find clusters that contain two different colours in Urdu. In this example, *Sūrkh* is working as a pre-modifier of *gūlabī* which is quite common in the data. In Urdu, different colours work as pre-modifiers in order to add more emphasis to the main colour, which in this case is *gūlabī*. Furthermore, this colour is also related to the weather as shown in *rut gūlabī*, which can be roughly translated into English as “pink weather” or “pleasant weather”. This leads us to the conclusion that other colours besides “blue” are also related to the sky and weather.

4.1.1.3 Bādāmī

In Urdu, *bādām* means “almond” and as translated in Lal’s dictionary, *bādāmī* means “almond coloured” (1955, p. 90). According to the corpus, *bādāmī* as a colour term is not frequently used. What I have found about this colour term is that it is highly used as a noun which is being modified by another noun; and specially to refer to the name of a mosque such as *Bādāmī masjid*. In this example *Bādāmī* is a proper noun, which is working as a noun modifier of *masjid* “mosque” which is a noun. I found 35 instances of this colour in the corpus and 5 of them take the noun *rang* “colour” as a post-modifier, which indicates that only in 5 examples it is used as a colour.

As I have mentioned above, I found that Urdu tends to combine colour terms to refer to a different colour or to emphasize one of them, normally the second one. In the case of *bādāmī*, data show combinations such as *bādāmī kāli* “almond coloured black”, which is associated to a different kind of black, perhaps brownish black; but it still makes reference to the black skin tone², a concept that holds negative connotation in south Asian countries.

4.1.1.4 Gājri

Gājri is one of the few colour terms that is not recorded in the Urdu dictionary that has been used as a source for this analysis. Nonetheless, its meaning can be deduced from the word *gājar* (Platts, 2004, p. 1822), which means “carrot” in Urdu. According to Drocco and Risato’s work, the “metonymy type lexicalization” adds the suffix -i to a word to form a colour term. The lexicalization of word *gājar* demands the suffix -i in order to form the colour term *gājri*. The fact that this term is not found in the Urdu or Hindi dictionaries that have been consulted could be the reason why no instances of this word are found in the Urdu and Hindi samples that have been explored. Even though the colour word is not found in the dictionaries, the word *gājar* appears in 132 instances in the corpus.

4.1.1.5 Zāfrānī

This colour term finds its origin in the Persian language colour word *z’afārāni*, which means “saffron-coloured” (Steingas, 1998, p. 1617). Based on corpus observation, 32 instances have been found of this word and only 6 of them refer to the colour *zāfrānī*. In addition, this colour takes the word *rang* “colour” as post-modifier when it is used as a

¹ *Bādāmī* is considered a proper noun in *Bādāmī masjid* because a particular mosque is named by the Word *Bādāmī* which makes it a proper noun.

² In many south Asian countries, “white skin” is considered a privilege and it is given preference over the “black skin”, opening a debate on racism in the south Asian culture.

colour term. In the rest of the examples, *zāfrānī* appears in the expression *zāfrānī korma*³. The use of this word as a colour term is restricted to a few examples and even though it is not recorded in the Urdu and Hindi dictionaries that have been consulted, naturally occurring data extracted from real-world contexts show that it is used as a colour.

4.1.1.6 Gandumī

Lal translates this colour as “brown” (1955, p. 510). *Gāndum* originates in the Persian word *gāndum* which means “a kind of wheat” (Steingas, 1998, p. 556). In the corpus, there are 50 instances of this word and in all of them *gandumī* is used as a colour term. In many examples it is preceding the word *rang* “colour” and the Urdu word *chehra* which means “face”. We can find many examples in the corpus that use *gandumī* as an adjective of *chehra*. This leads us to the conclusion that the colour term *gandumī* is often used to refer to a skin tone, specifically, a brown skin tone.

4.1.2 A brief comparison of data found “on metonymy type” lexicalization between Urdu and Hindi

The main point of this comparison is to see whether Drocco and Risato’s findings on colour lexicalization patterns in modern Hindi can also be applied to the lexicalization pattern in Urdu. As far as I have observed, there are not significant differences between Urdu and Hindi lexicalizations of “metonymy type” word colour however, we can find some differences which will be explained below.

The analysis has been done taking into consideration the origin of these colour terms. Some of these colours are of Hindi origin while others come from Persian and Arabic. Drocco and Risato state in their work that the suffix *-i* is a Persian borrowing that has been taken by Hindi, so it is found in the colour terms related to food and vegetables but with exceptions (2019, p. 192-193).

In the list of colours analysed above, two of them are of Hindi origin: *bādāmī* and *gājri*. These terms either show very few instances as colour terms or cannot be found in the corpus, which is the case of *gājri*. In my opinion, this might indicate that the origin of the word influences how frequently it is used. Hence, Persianized terms are more prominent

³ A Mughal inherited dish in which “saffron” soaked in milk is one of the main ingredients.

in Urdu vocabulary. Furthermore, other colours such as *khāki*, *zāfrānī* and *bādāmī* show more instances as referring to food and names rather than colour terms.

4.2 Approximation type lexicalization

Reduplication of colour words and the use of suffix -sā is what Drocco and Risato have called “approximation”. In order to observe where these patterns of lexicalization repeat in Urdu, I have chosen six “approximation type” colour terms from Drocco and Risato’s work: *lāl*, *nīla*, *harā*, *kāla*, *pīla* and *gōra*.

4.2.1 Corpus based findings on Urdu colours

4.2.1.1 Lāl

In Hindi and Urdu, *lāl* means “red” but Urdu has another term with similar meaning. Lal translates the Hindi colour word *lāl* into Urdu as *lāl* and *sūrkh*, which comes from Persian (1955, p. 529). Urdu uses both Sanskratized and Persianized colour terms. However, one of them is used more frequently than the other. In Urdu, the use of the word *sūrkh* to refer to the “red colour” is more frequent than the use of the word *lāl*. For that reason, *sūrkh* has thousands of instances in the corpus and it is always used as a pre-modifier of nouns such as *gulāb* “flower”, *libās* “clothes” and *rang* “colour”. Having said that, his colour is often related to the skin tone as well. In south Asian cultures *sūrkh rangat* “red coloured skin” is related to health and it is considered a privilege to have this type of skin tone. Taking into consideration Drocco and Risato’s classification of this colour term, it can be stated that, in Urdu, *sūrkh* also appears in reduplication structures similar to those used in Hindi in the case of *lāl*. For instance, this sentence shows a clear example of reduplication in Urdu,

2. کھجانے سے سرخ سرخ نشانات نمودار ہوجاتے ہیں۔

appear + marks + red + red + like + scratching

Red red marks appear from scratching.

[747]

What this example shows is that the perception of the colour is “red” but it is not a regular type of “red”, it can be a little lighter or brighter than regular “red” depending on the

context. As Drocco and Risato mention in their study, “it could also be defined as reddish” (2019, p. 207).

On the other hand, Urdu also uses the suffix -sā with this colour term, something which is not observed in Drocco and Risato’s analysis of this colour term. The sentence below provides the necessary context to explain the use of the suffix -sā with the colour term *sūrkh*:

3. آج مٹی کا رنگ بھی بڑا سرخ سرخ سا نظر آتا ہے

Looks + reddish + a lot + colour + as well + mud + today

Mud’s colour looks a lot reddish today as well.

[10955]

In example 3, -sā brings a hint of uncertainty about the colour of the “mud”. If we look closely, both reduplication and the suffix-sā are used when there is doubt about the exact colour being observed. Both indicate “approximation” but depending on the context, they may be used in other construction such as in example 2.

The Hindi word *lāl* is also widely used in the Urdu corpus, however it is usually used as a noun such as in *Lāl masjid*⁴ “Red mosque” and Joahar *Lāl* Nehru⁵ in which *Lāl* is again presented as a proper noun; a surname. As a matter of fact, these two words are homonyms of the Hindi colour word *lāl*.

Another observation that is worth mentioning is that *lāl* is exceptionally used as colour term in the expression *lāl mirch* which means “red chilli”. As we can see, *lāl* is used as a colour here mainly because Urdu has inherited this word from Hindi and it has not changed throughout the years. In fact, many spices preserve their original name from the times of British India.

4.2.1.2 Nīla

⁴ *Lāl* is the name of a particular mosque. That mosque is known by the name of *Lāl*, which in this example is a proper noun.

⁵ Jahar *Lāl* Nehru is one of the most important Indian political figure during the time of partition and later, the first Prime Minister of India.

Sanskrit has given origin to this Hindi colour name which is also used in Urdu. According to Lal’s dictionary, *nīla* is translated as *nīla* “blue” in Urdu (1955, p. 639). Taking into account the origin of this word, the lexicalization patterns of this colour term in Urdu should be similar to those in Hindi. Drocco and Risato identified reduplication as a basic pattern of lexicalization for this colour in Hindi, however, I would like to add the use of suffix -sā in Urdu. Very few instances have been found in the corpus but they are crucial to the study.

Drocco and Risato have classified *nīla* as a “situational” colour term (2019, p. 206) which means that the image we get when we hear this colour term is speaker-specific, that is, it is linked to the personal mental image that the speaker has (2019, p. 206).

4. اور یہ نیلا آسمان کس (کی عظمت) پر حیران ہے

Is + amazed + on + (glory) + whose + sky + blue + on + and.

And whose/who (glory) has amazed the blue sky.

[1385]

In the example above, the highlighted word is *nīla asmān* which means “blue sky”. This is in line with Drocco and Risato’s analysis of *nīla* as a “situational colour” term. In Hindi and Urdu, “sky” is often used as an example for “blue” colour. In fact, sometimes *asmāni rang* which is translated as “the colour of the sky” (not “sky blue”) substitutes *nīla* because *asmāni* comprehends all “blues” and not only the lighter versions of “blue”; *asmāni nīla* “sky blue”. Therefore, *nīla* activates associations related to natural objects, such as *asmān* “sky” or *sāmāndar* “the sea”.

4.2.1.3 Harā

Drocco and Risato’s have described the colour term *harā* “green” as a “situational colour term”. In Urdu dictionaries we can find two different colour terms that refer “green”; *harā* and *sābz*.

Firstly, *harā* is not widely used as a colour term in the Urdu language, even though examples such as that below show that it can be used in some collocations.

5. محمد علی جناح کا پیرا بھرا پاکستان تھا آج وہ قبرستان بن چکا ہے

Had + made + burial ground + that + today + was + Pakistan + full of Greenery + of +
Muhammed Ali Jinnah.

It was Muhammed Ali Jinnah’s green Pakistan, which today has converted into a burial
ground.

[8025]

In example 5, we can see *harā bharā Pakistan*, which is roughly translated into Eng¹⁶
as “green Pakistan” or “lively Pakistan, a country full with life”. It is definitely not a
prototypical collocation for the colour term, even though it supports Drocco and Risato’s
hypothesis that green conveys “positive emotions” (2019, p. 206).

Secondly, *sābz* shows many more instances in Urdu than *harā*. As in the case of *harā*,
sābz is also duplicated in this case.

6. محلہ میں بھی سبز سبز پرچم لہرائے

Waving + flag + green + green + as well + in + neighbourhood.

The green green flag is waving in the neighbourhood as well.

[9041]

In example 6, *sābz* is used as a reduplicated form. Observing the context in which it is
used and the noun (*parcham* “flag”) it is modifying we can say that this colour term has
a “situational” meaning because in Pakistan the colour “green” is often related to the flag
of the country.

7. اس وقت سبز گھاس کی رنگت زرد ہو چکی ت

Had + yellow + colour + of + grass + green + time + this.

This time the green coloured grass had gone yellow.

[1963]

The example above shows a comparison between *zārd* “yellow colour” and *sābz* “green
colour”. This comparison highlights the contrast between the vividness of *sābz ghās*
“green grass” and the dullness of *zārd ghās* “yellow grass”. In conclusion, the Urdu colour

word *sābz* represents “life” and “youth”. This is in line with Drocco and Risato’s analysis of *sābz* as a colour term that conveys a “positive meaning”.

Moreover, data suggest that the colour *sābz* does not take the suffix -sā but the suffix -tar. However, Urdu dictionaries do not record -tar as a suffix. Thus, -tar might be part of a word. In order to know from which word it descends, we need to look at the synonyms of *sābz* (or noun phrases that represent the same idea as *sābz* or convey the same positive connotations as *sābz*) so we can observe whether or not any of them use this suffix. Shakespear describes the Persian colour term *sābz* and uses the noun phrase “tar-o-taza” which means “fresh” (1843, p. 2310). It can be understood that *sābz* takes the first part of the Persian word “tar-o-taza” as a suffix for the purpose of increasing the intensity of the colour as it can be observed in example 8. In this example the colour designated by *sābz* is intensified by the following noun (parcham “flag”). The colour is emphasized because the author is talking about the national flag and the love towards it, which should be everlasting and positive.

8. آباد نظر آتی ہے جس پر سبز تر پرچم ستارہ ہلال

Pure + star + flag + Green + on + is + look +inhabitable

Inhabitable look (places) that have pure flag (with a star on it).

[993]

4.2.1.4 Kāla

Drocco and Risato have categorised this colour term as an “emphasized colour” in modern Hindi. Emphasized colours intensify the essence of the colour being described by means of reduplication and the suffix -sā. The colour words *kāla* and *gōra* are perfect examples of “emphasized colours”.

Lal describes *kāla* in his dictionary, as “black” (1955, p. 68) and uses the Persian word “siyāh”, which also means “black”, as a synonym (1955, p. 68). As expected, in the corpus *siyāh* is more frequent than *kāla*.

Kāla is not normally used as a colour term, even though, I have found some instances of it being used as a colour term in Urdu. Moreover, it is used as a reduplicated form and takes the suffix -sā in some instances.

Siyāh shows a lexicalization pattern very different from *kāla*. It is not found in structures of reduplication or in combination with the suffix -sā. However, it is used as an intensifier of the colour word *kāla*.

9. جب ان کو لڑکی کی خبر سنائی جاتی ہے تو غم کے مارے ان کا چہرہ کالا سیاہ
بوجاتا

When + they + girl + of + news + heard + is + then + worry + of + they + of + face +
black + becomes.

When they are given girls' news, their face become black (darker than black) out of
worry.

[2268]

Example 9 illustrates the use of *siyāh* as an intensifier of *kāla*. In Urdu, the combination of *kāla* and *siyāh* in the same sentence exaggerates the perception of the colour “black”. *Kāla siyāh* is not a normal “black” anymore, it is darker than “black”. Furthermore, this example also makes reference to a *kāla chehra* which means “black face” but it does not refer to the skin tone. Therefore, we can say that the colour black is connected to “worry” because, as it can be observed in the example, faces become “black” out of worry. Thus, both *kāla* and *siyāh* transmit negative feelings.

4.2.1.5 Pīla

Drocco and Risato have characterized this colour term as a “softened colour” and a “changing colour”. Firstly, according to Drocco and Risato “speakers can use the reduplicated colour terms to describe something with a paler or more subdued colour than the prototypical one” (Drocco and Risato, 2019, p. 205). “Softened colours” are not used to emphasize a colour but to represent pleasant or some unpleasant versions of a colour term.

As in the case of many other colour terms, Urdu includes a Persian version of *pīla* as well. Lal translates *pīla* and *zārd* as “yellow” (1955, p. 183, 708).

According to Drocco and Risato, *pīla* conveys an unpleasant shade of yellow when it is concerned about a particular skin tone. *Pīla* equates “paleness” when it refers to the skin tone. As for skin tones, it is crucial to mention that in many South Asian countries, skin tones are linked to positive and negative connotations; some are better than the others

(see page 12). For example, *Pīla* represents an unpleasant shade in this type of context. In Urdu, the corpus does not show any instance of the word *pīla* representing the skin tone or “paleness” and reduplication is not applied in Urdu as opposed to Hindi.

Another word used to describe the “yellow colour” is *zārd*. Even though Urdu does not resort to the reduplicated forms of *pīla*, it uses the reduplicated forms of *zārd* as it is illustrated in example 10.

10. لیکن کامریڈ نور محمد کا وہ زرد زرد سا چہرہ سالگ رام کی نگاہوں سے
اوجھل نہیں

But + comrade Noor Muhammed (N) + of + that + yellow yellow + like + of + face +
Salag Ram (N) + of + eyes + from + away + no.

But comrade Noor Muhammed’s yellow yellow (pale) face could not go away from
Salag Ram’s eyes.

[26342]

11. وہ مطمئن لیکن زرد سا چہرہ

That + content + but + yellow-like + face.

That content but yellow (pale) like face.

[1166]

Furthermore, Urdu also uses the suffix *-sā* with this colour word as it is illustrated in example 11; *Zard sā chahera*, which means “a pale (like) face”. The suffix *-sā* is used as “approximation” which means that “face” is not yellow but “yellow-like” or “pale”, to be more precise. As it can be observed in example 10 and 11, the Urdu colour term *zārd* is used in both in reduplication structures and in combination with the suffix *-sā*. Moreover, both examples are compatible with Drocco and Risato’s claim that the “yellow colour” conveys an “unpleasant shade” in specific contexts, especially when it is related to the skin tone. Therefore, it can be said that this “softened colour” projects a negative feeling.

4.2.1.6 Gōra

Plats translates this colour word as “a fair complexion” (2004, p. 1885). In Urdu or even in Hindi, *gōra* is not used as a general colour but to describe skin complexion specifically. Normally, when we talk about someone who is “gōra”, we are talking about someone who is fair of complexion. It is worth mentioning that in south Indian culture, a fair skin

complexion is considered desirable and respected in comparison to brown or black skin tones.

Urdu uses *gōra* in the same context as it is used in Hindi and no synonyms are found for this word in Urdu. In this case both reduplication and -sā are common in the examples extracted from the corpus.

12. کیا دیکھتی ہوں ایک گورا سا شخص بال لہراتا میری ہی طرف چلا آرہا ہے
What + looking + was + one + White -like/looking + person + hair +waving + me +
towards + walking
What am I looking at! One white-looking person, waving his hair, was coming towards
me.

[12301]

In the example above, we can observe how *gōra* is being used with the suffix -sā with positive connotations. Indeed, this colour term is always used to describe someone's beauty since beauty is related to "white complexion" in south Asian countries.

4.2.2 A brief comparison of Hindi and Urdu "approximation type" lexicalization

According to the corpus research conducted for this paper, firstly, it can be seen that most of the Hindi origin colour words that are used in Urdu follow the same lexicalization patterns in Urdu as in Hindi (i.e., "reduplication" and the suffixation of -sā). That is the case of *nīla*, *kāla* and *gōra*. As observed in the data, not only Hindi origin words in Urdu show the same lexicalization patterns as in Hindi, some persianized colour word such as *sūrkh* show the influence of Hindi lexicalization of colour words. Secondly, some Hindi colour words such as *lāl* are not found in the corpus, instead we find it's homonyms that are working as proper nouns.

The colour terms that show the most significant differences are those that derive from Persian or Arabic. They totally dismantle the lexicalization patterns described by Drocco and Risato. Colour terms such as *sābz* take their own suffixes, that are not found in any other colour terms. Moreover, constructions that combine a Persianized and a Sankratized word, such as *kāla siyāh* are found only because Persian words are allowed to work as intensifier of Hindi colour term.

5 Conclusion

The analysis of the data suggests that there exist discrepancies between Hindi and Urdu as regards the pattern of lexicalization that Drocco and Risato (2019) call “approximation type”. This pattern of lexicalization, in Urdu, is applied not only to colours inherited from Hindi but also from Persian and Arabic. Hindi colour words such as *lāl*, *harā*, *kāla* and *Pīla* are a perfect example of Hindi colour terms that have a Persian or Arabic equivalent in Urdu. In fact, the data drawn from the corpus show that Urdu speakers prefer Persian or Arabic colour word instead of Hindi colour words, even though there are countless similarities between these two south Asian languages.

What I wanted to explore in this paper was that Urdu and Hindi are not only distinct when it comes to High register vocabulary but also in low-level vocabulary such as colour terms. As it can be observed above, numerous differences have been found in the use of Urdu colour words compared to Hindi colour words. On the one hand, some colour terms that derive from Hindi do not follow the same lexicalization patterns as in Hindi (Drocco and Risato, 2019), for example, *lāl* and *harā*. On the other hand, Persian and Arabic inherited colour terms in Urdu show some similarities and many disagreements to Drocco and Risato’s colour lexicalization of Hindi colour terms. There is not a single pattern of lexicalization that can be applied to describe how Hindi inherited colour terms and Persian or Arabic acquired colour terms are lexicalized. Finally, it is important to highlight that the results reported here reflect how people actually use colour terms on the internet. Thus, these results might deviate from the prescriptive rules recorded in grammars and dictionaries.

To conclude with, it has been an honour to work on Urdu and Hindi. This research, sometimes, has resulted a tremendously difficult task. However, I have come to find peace in the process of researching something so precious to me. Personally, Urdu has a very special place in my heart not only because it is my mother tongue but because I consider it as one of the most euphonious and interesting Indo-European languages.

6. References

- Aneesh,A. (2010). Bloody Language: Clashes and Constructions of Linguistic Nationalism in India. *Sociological Forum*, Vol. 25. No. 1. pp. 86-109. <https://www.jstor.org/stable/40542542>
- Berlin, Kay, P. (1969). *Basic Colour Terms: their Universality and Evolution*. Berkeley: University of California Press,
- Drocco, A. & Risato, O. (2019). Lexicalization patterns in colour naming: the case of Modern Hindī. In Ida Raffaelli, Daniela Katunar and Barbara Kerovec (Eds), *Lexicalization Patterns in Colour Naming: A cross Linguistic Perspective* (p. 191-212). Amsterdam: John Benjamins publishing Company.
- Hardin, C. L. (2014). Berlin and Kay Theory. *Luo, Ronnier; Encyclopedia of Color Science and Technology*, 1(4).
- Jain, D. (2003). Chapter two: Sociolinguistics of the Indo-Aryan languages. In George Cardona and Dhanesh Jain (Eds), *Indo- Aryan Languages* (p. 47-66). London and New York: Routledge.
- Khan, A. J. (2006). *The Politics of Language. Urdu/Hindi: An Artificial Divide*. New York: Algora Publishing.
- Malt, Barbara C. & Majid, A. (2013). How thoughts are mapped into words. *WIREs Cognitive Science* 4(6). 583–597.
- Schmidt, L. R. (2003). Chapter eight: Urdu. In George Cardona and Dhanesh Jain, *Indo- Aryan Languages* (p. 286- 350). London and New York: Routledge.
- Trenholm, S. (2017). *Thinking through communication: An introduction to the study of human communication*. New York, NY: Routledge.

Dictionaries

- Lal, R. N. (1955). *A student's practical dictionary: containing English words, with English and Urdu meaning in Persian characters and Hindustani word with English meaning together with a list of Latin and Greek words and phrases with their equivalent English and Urdu equivalent in Persian characters* (18th., ed). Allahabad, India: Ram Narayan Lal Beni Madho publishers and book sellers.
- Plats, T. J. (2004). *A dictionary of Urdu, classical Hindi and English* (3rd., ed). New Delhi, India: Munshiram Manoharlal Publishers
- Steingas, F.J. (1998). *A comprehensive Persian English dictionary : including the Arabic words and phrases to be met with in Persian literature ; being Johnson and*

Richardson's Persian, Arabic, and English dictionary (1998., ed). Beirut, Lebanon: Librairie du Liban.

- Shakespear, J. (1834). *A dictionary, Hindustani and English, with a copious index, fitting the work to serve, also, as a dictionary of English and Hindustani* (3^{th.}, ed). London: Parbury, Allen and Co.