Maria Isabel Maldonado Garcia*

A Corpus Based Quantitative Survey of the Etymological Elements in the Basic Vocabulary of Urdu Language

Abstract: This article is the last of a series on the topic of the etymological components of the basic vocabulary of Urdu language as well as their density within the language. Each of the articles was aimed to calculating the density of a particular etymological element in the basic vocabulary of Urdu. The research is significant since it is the first corpus-based survey of the density of the etymological elements in the basic vocabulary of Urdu. The calculation is conducted utilizing a corpus of one thousand words of the basic vocabulary and extracting the origin of each word through an etymological analysis. The words are then classified by language of origin and the precise percentage of each language element obtained. While the previous research (Maldonado García 2014, 2015, 2015a, 2015b), dealt with the density of individual etymological elements, this one treats this topic as a whole, dealing with all of the language elements at one time and comparing the etymological analysis with historical facts. The study debunks the widespread theory that Urdu was a language of the armies, artificially created at some point in history when the Arab, Persian and Turkish armies met at an unidentified geographical location producing the birth of Urdu as a mixture of their languages. After the identification phase of this project, it was revealed that from an etymological and historical perspective, Urdu derived from Sanskrit into the Prakrit languages through a process of evolution and has received influence from Persian, Arabic and English during the period of the invasions. No elements of Turkish language were found during the analysis.

1. Introduction

English and Arabic coexist alongside Urdu and other regional languages in Pakistan. English is the official language while Urdu has traditionally been the national language. Arabic is the language of the religion of Islam and hence, Muslims around the world are bound to learn it in some measure.

Language contact can be defined as the use of two or more languages in one particular location in a specific period of time (Thomason & Kaufman, 2001). Persian lived alongside Urdu for centuries, as well as Arabic, and English principally in a non-bilingual environment. Religions have been responsible for the spread of languages

*Associate Professor, Department of English Language and Literature, University of Management and Technology, Pakistan.

almost since the beginning of times. This is the case of Christianity and Latin as well as Islam and Arabic. This particular type of language contact is significant since Arabic is in influenced by many languages of the world, in fact a large number of Muslims do not speak it nor understand it. In this sense, this is a type of contact without full bilingualism.

Another of these cases is that of English. The Pakistani population is exposed to English language mainly through television, internet, printed media, science and technology with some private schools and universities use it as the medium of instruction. While a reduced percentage of the population is able to speak it and has a good proficiency of the language, in many cases, the knowledge of the population receives is passive. This means that people may understand it when they see it written or listen to it but are unable to reproduce it independently in a correct manner.

Through language contact changes in languages occur. Languages receive influences from other languages. This can be reflected in lexical borrowing, transfer of grammar structures, etc. This all happens due to social history. Languages come into contact for many reasons; through trading, through wars, through invasions, colonization, through meetings of people from the same profession, education, etc. The Subcontinent has suffered from multiple events which propitiated the onset of language contact and have influenced Urdu. The purpose of this study is to analyze some of these events as the cause of innumerable loanwords and calculate the density of those loanwords in the basic vocabulary of Urdu.

2. Literature Review

Urdu is a language, unlike English, Spanish and others, which has not been studied in depth from a linguistic point of view. It is for this reason that some myths still exist in the minds of those who have engaged in the literature studies of Urdu and who believe that Urdu is a mixture of Persian, Arabic and Turkish: "Urdu is one of the popular Indian script in the Indian subcontinent and national language of Pakistan evolved in the subcontinent from the mixture of Arabic, Turkish, Farsi and Hindi Languages." (Ahmad et al., 2007). In addition, this mistaken information is even included in the national curriculum of Pakistan and the textbooks that children study in school.

Initially, the fact that Urdu and Hindi are the same language has been stated by King,(1994) and King,(2001). Furthermore, the linguists who propose that Urdu derived from Sanskrit are multiple (Dittmer, 1972; Kelkar, 1968; Chatterji, 1969; Kachru, 1987; Cardona, 1987; King, 1994; King, 2001; Rahman,2008). Maldonado Garcia (2015) proved it by extracting the etymological elements of this language as well as Prakrit proving that these elements do, in fact, exist in Urdu. With the available information it is now clear that the myth which exists and which states that Urdu is a mixture of Persian, Arabic and Turkish has already been disproven, since this theory cannot explain the presence of Sanskrit and Prakrit elements in the language.

However, the search for the truth began with a doctoral dissertation about the etymological elements present in Urdu language, where this step was a part of the process

of an etymological analysis of a corpus of one thousand basic words of Urdu, Spanish and English in order to find their origin as well as the similarities between the word classes in these languages (Maldonado Garcia, 2013). This study revealed the etymological components of Urdu, apart from a vast array of linguistic information related to the other languages not relevant to this study.

Furthermore, it inspired additional research such as one on the reforms the Urdu language has undergone (Maldonado Garcia, 2014) in which the origin and evolution of Urdu was discussed as well as another about Urdu as the official language of Pakistan (Maldonado Garcia, 2016).

The previous study was followed by a study on the similar lexical elements in Turkish and Urdu or those which where borrowed by Urdu from Turkish. This study returned the discovery that the words that are similar in Turkish and Urdu, are not of Turkish origin. Rather, the words are loanwords that Turkish and Urdu borrowed from Persian and Arabic (Maldonado Garcia, 2014a). This means that the Turkish element of the before mentioned theory was proven wrong and there was no Turkish lexical elements in the composition of Urdu language.

The subsequent study treated the lexical elements of Sanskrit and Prakrit languages that still remain in Urdu (Maldonado Garcia, 2015) which were discovered out of a corpus of one thousand Urdu words. The analysis resulted in a large portion of Sanskrit and Prakrit lexical elements which is another point in debunking the traditional theory of the composition of Urdu language. In fact, if that theory was true these lexical elements of Sanskrit and Prakrit languages would not have been present in Urdu. These two discoveries are enough to contradict the traditional theory which states that Urdu is a language created when the Persian, Arabic and Turkish armies were together in one particular location. The theory does not state the time frame of that occurrence either. It is not really possible to create a language from elements of other languages and make it work effectively in a societal context. In addition, this association tends to ignore approximately four hundred years of the history of Urdu (Rahman, 2013). Language suffers a process of evolution and that process is what produces language change. Urdu, is indeed not a reduced form of language, rather, this language was not born spontaneously from the mixing of the armies' languages in an army camp. It is obvious that entire populations do not stop speaking their own language just to adopt a new one. Consequently, the population had a language which was the ancestor of Urdu and that was present in India before the Muslims arrived (Rahman, 2013).

A pivotal part of the analysis was the identification of the lexical elements of Persian and Arabic. The analysis returned that the largest component of Urdu language belongs to lexical elements that were borrowed from Persian and Arabic (Maldonado Garcia, 2015a).

The last study was a survey of the etymological elements of English language in Urdu (Maldonado Garcia, 2015b). The study returned a small number of loanwords that Urdu has taken from the English language.

The present research is devoted to reassess the etymological components of Urdu from the standpoint of comparative dialectology. It is assumed that the historical data demonstrates the feasibility and possibility of language transfer of the elements revealed in the etymological study through language contact. Loanword (or lexical borrowing), as defined by Haspelmath, is "as a word that at some point in the history of a language entered its lexicon as a result of borrowing (or *transfer*, or *copying*)" (Haspelmath, 2009). This definition will be followed for the purpose of the results of this article.

3. Methodology

This investigation is a comparative analysis of the density of the etymological components of Urdu language as a whole. For this purpose, important historical data related to the progression of language contact and the evolution of Urdu will be collected so as to compare the findings of the etymological analysis with the possible causes of the presence of a particular element. In addition, the etymological data was collected as a whole from a corpus of one thousand words as in Fang & Cao (2010). The percentages of each element then will be compared so as to determine the densities of each language in the composition of Urdu. The following diagram illustrates the methodology that will be followed.

Fig 1. Methodology
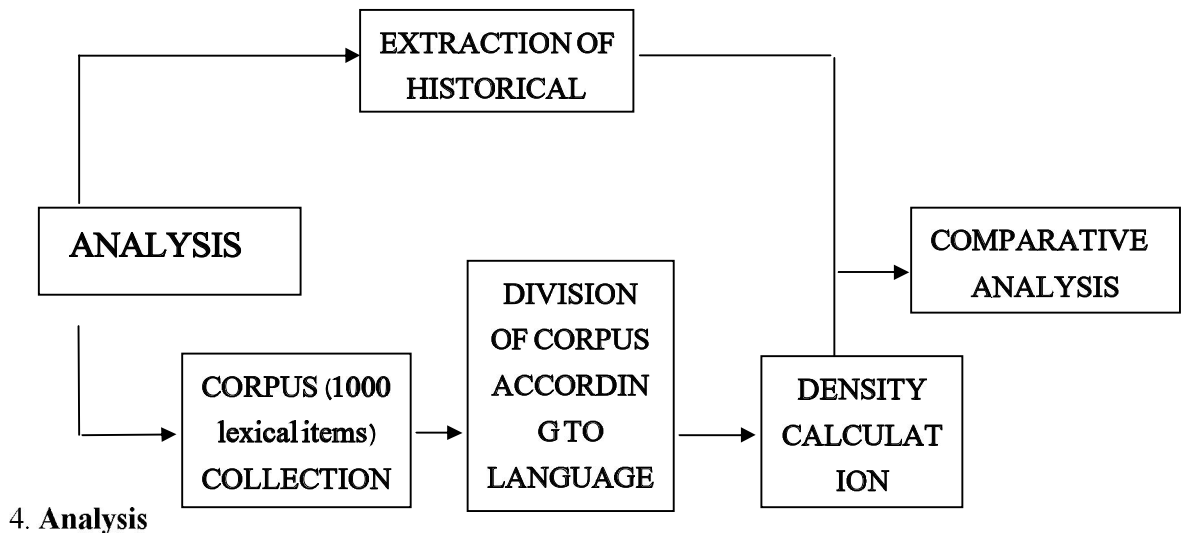
4. Analysis

Part 1. Extraction of Relevant Historical Data

As a first step of the analysis the historical data was extracted:

Key Historical Data Related Language Borrowings. The Evolution of Urdu to its Current Form

Presence of Sanskrit and Prakrit in the Subcontinent

Sanskrit is an old Indo-Aryan language from which Prakrit, a middle Indo-Aryan language derived. Prakrit underwent several stages of development:

a. An early era epitomized in the Bhuddist Theravada.
b. A subsequent stage characterized by the Standard Literary Prakrit in which the language was established in written form.
c. A Middle Indo-Aryan stage where Bengali, Hindi, Marathi, Gujarati and other languages commenced to be documented from the first millennium A.D. This is also the stage of documentation of Hindi and Urdu as well as other languages.
d. Appearance of Modern Indo-Aryan languages where the languages evolved to their present form (Burrow, 2001).

Persian Presence in the Subcontinent

Persian is an Indo-European language from the Iranian family of languages. It derives from Proto-Iranian and has endured three stages of development and the following facts are significant:

Persian developed in three different eras. (a) Old Persian era (circa 525-300 B.C.). During this period Persian utilized a cuneiform script. (b) Middle Persian era (circa 300 B.C.-800 A.D.). This area of language development is characterized for the use of three scripts Pahlavi, Manichaean and the Avestan script. (c) Modern Persian era (800 A.D.-onwards). During this time the Persian language utilized a Perso-Arabic script (Skjærvø, 2002).

Rahman (2013) points out that "the ancestor of Urdu-Hindi was pervasive in the eleventh century when the Persian-speaking Turkish and Persian invaders, mendicants, merchants, and holy men arrived in India". The influence of Persian on the ancestors of Urdu starts as the Persian dynasties were maintaining their rules on some parts of India (Pourjafar & Taghvaee, 2006). At this point, Indian medical books were translated from Sanskrit to Pahlavi as well as other science works from Greek to Persian (Rahman, 2008).

Arabic Presence in the Subcontinent

Arabic belongs to the Afro-Asiatic family of languages and within it, to the Semitic branch. It is a language which presents no genetic relation with Urdu. The arrival of the Arabic language in India and its coexistence alongside Persian is marked by the following historical events:

Arabs traders arrived to the Subcontinent in 712 A.D. In the 8th century Mohammad Bin Qasim conquered Sindh and Multan and starting the development of an Islamic rule in the Subcontinent. By the 10th and 13th century the northern part of India was almost in its entirety dominated by the Muslim rule (R.D. King, 2001). In this period Urdu literature developed from elements of Sanskrit and Persian.

The invasions of the 8th century led to subsequent invasions during the period from the 10th to the 15th century which gave birth to the Muslim empire and the Mughal Empire.

In the 11th century Prakrit is replaced by some of the dialects derived from it (Oriya, Bengali, Marathi, Gujarati, and Khari Boli, a branch of Western Hindi and Hindvi from which Urdu and Hindi had developed (Schimmel, 1975).

By the 1400's Nastaliq script had been developed by Mir Ali Tabrizi which later is adopted by Urdu as a language of the Muslims, hence developing its Muslim identity (Encyclopaedia Britannica, 2014). The rationale of this usage was to have a script similar to the script of Arabic, so that the population at large would be able to read Arabic easily which in turn was a facilitating factor for the study of the religious texts.

The Muslim rulers during the 14th century promoted Urdu language development with several important literary authors of the time.

The beginning of the Mughal rule starts in the 16th century with the Mughal Emperor Babur in the year 1526. In 1530 Persian was chosen as the language of the court in India by the Mughals. During this time major dictionaries of Persian language started to be compiled from the fifteen century to the eighteen century. In this sense Urdu was developing with a substantial impact and influence from Persian (Skjærvø, 2002).

English Presence in the Subcontinent

English is an Indo-European language from the Germanic branch. It is the widest spoken language in the world, in terms of geographical locations (only to be surpassed by Chinese in terms of number of speakers) and it does have a genetic relationship with Urdu, as the latter belongs to the Indic branch of Indo-European languages. The British invasion brought with it the English language which started spreading in the following manner:

A group of merchants from the British East India Company arrived in India in the year 1600. Through a series of battles the Company began its ruling over large area s of India by the mid 1700's. Fort William College is founded by Lord Wellesley in Calcutta in the year 1800. In this College students learnt English as well as different subjects including local languages such as Bengali, Hindustani, Marathi, Tamil, etc. as well as Sanskrit, Persian and Arabic (C. King, 1994). In 1835 the Macaulay Minute (or the Minute of February 2nd) was approved by Lord Bentinck, the Governor-General of India, establishing the English language as the official language of the administration (Macaulay, 1835). In 1837 Persian was removed as the language of the court in the areas under British rule.[i]

The last Mughal ruler, Bahadur Shah, lost power to the British by 1857. By this time, the Muslim had remained in power in the Subcontinent for nearly 800 years. The Muslims continued speaking and producing literary works in Urdu.

From 1947 onwards the Arabic language continues its influence as an Islamic language in Pakistan and Bangladesh. At the same time that Urdu alongside English became national and official languages of the newly founded Pakistan.

The emergence of Bangladesh occurred in 1971 which had partially been spurred by the Bangladeshi's unhappiness with relation to their language not becoming an official language of Pakistan.

In 2014 English and Urdu continued to be the official and national languages of Pakistan, spoken widely, being taught in schools. English has become the language of the elite, of private schools, higher education, international business, technology and science, as well as a class marker. In the current times it has even been designated as a variety of the English language called 'Pakistani English' (Khan, 2012). Urdu, on the other hand is reserved for the education of the masses, a vehicle of transmission of Islam, the first language for millions of people, the language of conservative politics, local television, radio, a certain segment of printed media, etc., as well as a class marker.

On September 8th, 2015 a Supreme Court order tried to implement the Constitution's wish in article 251, to have Urdu as the official language of Pakistan. However, even though the order has put some level of pressure on the government officials to engage in the writing of Urdu in the official communications and websites of the offices. However, due to various factors, this has been impossible so far.

These are in a synthesized manner the main historical events which have affected the evolution of Urdu as a language of the Subcontinent. This situation of multiple language contact has resulted in a vast array of language borrowings or loanwords from Persian, Arabic and English mainly. The density of these loanwords or lexical borrowings will be measured next to be able to assess the impact of these languages on Urdu.

Part 2. Corpus Collection and analysis

The second part of the methodology of this investigation is based on the identification of 1,000 basic words of Urdu language and measuring the lexical density of each etymological element through the extraction of the etymology of each word. This was done by locating each word in Urdu Dictionary and extracting the etymology.

Terms with the same etymology were grouped together in other researches. The etymological elements found in the composition of Urdu were Sanskrit and Prakrit (Maldonado García, 2015) Persian and Arabic (Maldonado García, 2015a) and English (Maldonado García, 2015b) where the results were extracted for each etymological element that composes Urdu language. This research compiled all of these results as to find out the total density of each etymological component in relation with the other etymological components. The sum of all of them is the total lexical density.

Measuring the lexical density of each etymological element in the basic vocabulary of Urdu.

Etymological Analysis and Results

In this sense, this study measures the complete density of each etymological component of Urdu. In other words, one thousand words of the basic vocabulary of English [ii] were translated into Urdu[iii] and analyzed from an etymological point of view as well as classified according to their origin (Maldonado Garcia, 2013).

The results were as follows:

Fig. 1. Percentages of terms from Sanskrit and Prakrit present in Urdu language

	NUMBER OF TERMS	PERCENTAGE
Terms with etymology in Sanskrit	269	26.9%
Terms with etymology in Prakrit	64	6.4%
Terms from Sanskrit which arrived in Urdu through Prakrit	9	0.9%

Source: Maldonado Garcia (2015)

As it can be seen in the above table out of 1000 words of Urdu, 269 derived from Sanskrit and 64 from the Prakrit group of languages. 9 words of Prakrit derived from Sanskrit and later on through a process of language evolution are now a part of Urdu. If we add these to the Sanskrit element the total percentage of Sanskrit becomes 27.8%

Fig. 2. Percentages of terms from Persian and Arabic present in Urdu language

	NUMBER OF TERMS	PERCENTAGE
Terms with etymology in Persian	217	21.7%
Terms with etymology in Arabic	299	29.9%

Source: Maldonado Garcia (2015a)

The previous table depicts the number of terms derived from Persian and Arabic and their percentage. There are 217 terms which derived from Persian and 299 terms derived from Arabic. Until now Arabic language terms since to have a larger density than the other languages.

Fig. 3. Percentage of terms from English present in Urdu language

	NUMBER OF TERMS	PERCENTAGE
Terms with etymology in English	78	7.8%

Source: Maldonado Garcia (2015b)

The density of English language terms found in the 1000 words of Urdu set for analysis was only of 78 terms which make 7.8%.

Other results derived from this study:

Fig. 4. Percentage of terms from Urdu-Hindi present in Urdu language

	NUMBER OF TERMS	PERCENTAGE
Terms with etymology in Urdu-Hindi	72	7.2%

Fig. 5. Percentage of terms from unknown origin

	NUMBER OF TERMS	PERCENTAGE
Terms with unknown etymology	1	0.1%

Once the density of each etymological component has been extracted a complete overview of all the components which compose Urdu. As in the table below:

Fig. 6: Complete overview of the density of the etymological elements of Urdu language
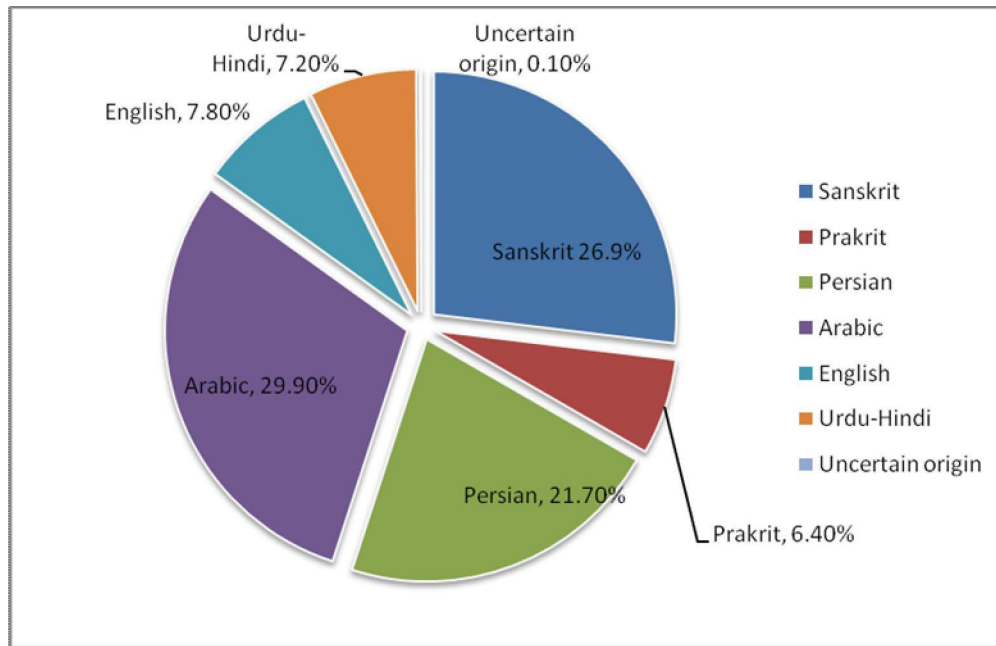
The following table shows the results in order of influence:

Fig. 7: Density of the etymological elements of Urdu language in order of highest influence.

ETYMOLOGICAL COMPONENTS OF URDU	ETYMOLOGICAL DENSITY
ARABIC	29.9%
SANSKRIT	26.9%
PERSIAN	21.7%
ENGLISH	7.8%
URDU-HINDI	7.2%
PRAKRIT	6.4%
UNKNOWN ORIGIN	0.1%

Comparison of the historical facts with the etymological analysis

The result of the preceding survey yielded a very mixed verdict on the etymological composition of Urdu's basic vocabulary, although highly interesting from a sociolinguistics perspective.

While usually the highest etymological component of a language is that of the language it derives from, it is important to note here that the strongest density in the basic vocabulary of Urdu, is that of the Arabic language, element surpassing even Sanskrit, the language from which Urdu developed. From a sociolinguistics and historical linguistics point of view the spread of a religion has often led to the spread of a language. In this case, the fact that the first Arabs arriving in the Subcontinent had been merchants and traders come to signify that the original motives for their arrival were not to create an empire. However, the historical data as well as the results indicate the spread of their culture, way of life, their scientific developments, their religion and of course, this all led to a high level of linguistic influence. In this case the situation of language contact was extensive with cities in the Subcontinent in a situation of bilingualism such as Mansura and Multan where it seems Arabic and Sindhi were being spoken (Schimmel, 1980). The influence of Arabic language on Urdu spans a period of 1,300 years.

In the second place in terms of etymological density, Sanskrit shows the highest influence. Sanskrit was another language which arrived to the Subcontinent with the purpose of trade and cultural encouragement, rather than military campaigns (Ostler, 2005) and spread to become the main language, especially in the areas of religion and administration. However, Vedic Sanskrit had become antiquated by the fifth century B.C. and updated forms of Sanskrit were being developed. It seems that with the acceptance of Buddhism and Jainism by the Crown of India the Prakrits had become popular while Sanskrit had been relegated to religious ritual commitments by 300 BC. This evolution occurred within a millennium and it was mainly due to language contact between the

merger of the speakers of Aryan languages and Non-Aryan languages (mainly Dravidian languages) (Bhadriraju, 2003). At present Sanskrit is an almost extinct language, alive only through the religious scriptures of Hinduism and Buddhism (*14,100 in India (2001 census). Population total all countries: 15,770. L2 users: 194,000 in India* (Ethnologue, 2013)). Still Urdu basic vocabulary presents 27.8% etymological density from this language. This shows that other etymological elements have had a higher level of influence on the language. This logic seems acceptable considering the fact that Sanskrit had ceased to be a widespread spoken language approximately 2,300 years ago and currently it is only spoken on a limited basis in India and Nepal due mainly to Buddhism and Hinduism (Ethnologue, 2013). In fact, it is remarkable that such share of influence still remains in Urdu, considering the aggressive impact of Arabic and Persian and the language reforms Urdu has undergone and indicates a high degree of resistance to change of those particular Sanskrit terms.

The next important etymological component in the basic vocabulary of Urdu is Persian with a density level of 21.7%. The influence of Persian in Urdu commenced in South Asia during the pre-Islamic era (Pourjafar & Taghvaee, 2006) and it continued until 1837 when Persian language was discontinued as the language of the administration. This means that the language has been present in the Subcontinent for more than 1500 years exercising influence. The Persian component, together with Sanskrit and Arabic, make up for 78.5% of the total etymology of the basic vocabulary of Urdu.

The English component shows a density of 7.8%. English arrived in the year 1600 and even though neither Pakistan nor India are in a colonial setting anymore and the British are long gone, English is alive and thriving in Pakistan with the help of universities, private schools, print and electronic media, technology and science and of course the internet. Urdu and English coexist with each other and in many occasions can be found within the same sentence or conversation in a code-switching manner (Nadeem, 2012). It is for this reason that the exact lexical influence of English over Urdu from a percentage perspective is difficult to measure.

During the search for the etymological components of Urdu we found that 7.2% of the basic vocabulary of Urdu derived from Hindi and Urdu itself. This means that Urdu has certain words originally from Hindi language as well as there are others which are indigenous from Urdu.

The next component in order of influence is Prakrit with 6.4% influence. As stated before Sanskrit developed into the Prakrits which in turn developed into Hindi-Urdu. Still elements of this language are present in the basic vocabulary of Urdu. In the same manner as the terms from Sanskrit seem to display a high level of resistance to change, so do those derived from Prakrit. In this sense, the influence of the Prakrits lasted from 300 B.C. (Bhadriraju , 2003) to the 11th century (Schimmel, 1975). Furthermore, some of the Sanskrit terms were transmitted into Urdu through the Prakrits.

These constitute 0.9% of the terms. To conclude 0.1% of the terms show an unknown etymology.

Interestingly enough, in this particular corpus of the basic vocabulary of Urdu no elements or loanwords from Turkish language were found. Keeping in mind the fact that the *Urdu* word itself comes from *Ordu* of Turkish origin, the matter of loanwords purely from Turkish language into Urdu was investigated.

Basic vocabularies of a language especially if they are high frequency are usually the most resistant to change. After the results showed not a single word of Turkish origin in the basic vocabulary of Urdu it can be safely assumed that Turkish language is without any doubt not a fundamental component of Urdu. However, a low number of words of Turkish origin have been noted in Urdu which does not constitute an extensive corpus (Anwer, 2011).

In another research, while looking for shared vocabulary between Urdu, Arabic and Persian, three words were found in Urdu with etymology in Turkish language (Maldonado Garcia, 2013). However, these also do not belong to the basic vocabulary of Urdu and in turn are not a part of the fundamental vocabulary of this language. In any case, if a language borrows three words from another, it does not make up in any way a considerable density.

Conclusion

As the table indicates the highest, density is that of Arabic language borrowings with 29.9%, this is due to the loanwords acquired during the Arab invasion of the Subcontinent. The Sanskrit element follows with 27.8% density. If the armies' theory was correct this element should not have been found during the etymological analysis. However, this element constitutes a large percentage of etymological density, only possible if, in fact, the Urdu language derived from Sanskrit. The Persian etymological element is the next one in terms of density with a percentage of 21.7%. Comparing historical events the presence of this element makes sense since the Persian language was a strong influencer in the Subcontinent for a long period of time. The English invasion resulted in a percentage of 7.8%. 7.2% belongs to terms that are original of Urdu/Hindi, 6.4% belong to the Prakrit element. This is another clue of the evolution of Urdu, since elements of Sanskrit could have been borrowed at a later date due to the religious influence. However, the presence of the element of the Prakrit group of languages ensures the feasibility and accuracy of this study proving the process of evolution. At last 0.1% has an unknown origin. Furthermore, Maldonado Garcia (2014a) proved that Turkish language is in no way a considerable element of Urdu language. This study has proven the same. There are no Turkish elements in the etymological composition of the basic vocabulary of Urdu language.

As this stance, it has been proven that the theory of the birth of Urdu through the mixing of the armies is flawed and the fact that Urdu evolved from Sanskrit is correct.

End Notes

[1] Some authors like Skjærvø, Prods Oktor, say it was abolished by those managing the East India Company (*An Introduction to old Persian.* 418). Others like Christopher R. King states that the language was removed by the Indian Government (in One Language, Two Scripts, 9). Tariq Rahman states (in Urdu and the Muslim Identity: Standardization of Urdu in the Eighteenth and Nineteenth (2008) 83.) that Persian was discontinued in India as an administrative language in 1834. However, as a matter of fact, it continued until 1849, the year in which the British annexed Punjab.

[1] See annex 1, Basic Vocabulary of English Language http://simple.wikipedia.org/wiki/Wikipedia: List_of_1000_basic_words, Accessed 28th August, 2014.

[1] See Annex 2. Basic Vocabulary of Urdu Language. Maldonado García, María Isabel (2013) Comparación del Léxico Básico del Español, el Inglés y el Urdu. Unpublished Doctoral Dissertation, Madrid: UNED.

References

Ahmad, Z., Orakzai, J. K., Shamsher, I., & Adnan, A. (2007, December). Urdu nastaleeq optical character recognition. In *Proceedings of world academy of science, engineering and technology.* 26, 249-252.

Anwer, S. M. (2011). 'Language: Urdu and the borrowed words.' Dawn Newspaper. http://www.dawn.com/news/672945/

Bhadriraju, Krishnamurti (2003) *The Dravidian Languages.* Cambridge: Cambridge University Press. 37.

Burrow, Thomas (2001) The Sanskrit Language. Motilal: Banarsidass Publications, 2.

Cardona George, (1987) Indo-Aryan languages: In the World's Major Languages. New York: Oxford University Press, 440-447.

Chatterji, S. K. (169) *Indo-Aryan and Hindi.* Calcutta: KL Mukhopadhyay. 149.

Dittmer, Kerrin. (1972) Die Indischen Muslims und die Hindi-Urdu Kontroverse in den United Provinces. Wiesbaden: Otto Harrassowitz.

Encyclopaedia Britannica (online version) (2018) http://www.brita nnica.com/EBchecked/topic/384749/ Mir-Ali-of-Tabriz,

Haspelmath, Martin. (2009) 'Lexical Borrowing: Concepts and Issues,' in Martin Haspelmath & Uri Tadmor, *Loanwords in the World's Languages: A Comparative Handbook.* Berlin: De Gruyter Mouton, 36.

Kachru, Yamuna. (1987) 'Hindi-Urdu,' in The World's Major Languages (New York: Oxford University Press, 470-489.

Kelkar, Ashok (1968) Studies in Hindi-Urdu. Poona: Deccan College Postgraduate and Research Institute.

Khan, Humaira Irfan (2012).'The Evolution of Pakistani English (PakE) as a Legitimate Variety of English,' *International Journal of Applied Linguistics & English Literature* 1, 5. 90-99.

King. Christopher R. (1994) *One Language, Two Scripts: The Hindi Movement in Nineteenth Century North India.* Bombay: Oxford University Press. 22.

King, R. D. (2001) 'The Poisonous Potency of Script: Hindi and Urdu,' *International Journal of the Sociology of Language.* 150, 43-60.

Lewis, M. P., G. F. Simons, & C. D. Fennig, (Eds.). (2013). *Ethnologue: Languages of the World.* 17th ed. Dallas, TX: SIL International. https://www.ethnologue.com/language/san

Macaulay T. B. (1835). *Minute on Indian Education*, Frances Pritchett's Website at Columbia University. http://www.columbia.edu/itc/mealac/pritchett/00fwp/#fwpt/00generallinks/maca_ulay/txt_minute_education_1835.html+&cd=1&hl=en&ct=clnk&gl=pk&client=firefox-a

Maldonado Garcia, M.I. (2016) Debate on Urdu as the Official Language of Pakistan: Urdu vs English. *Almas* 18; 700-712.

Maldonado Garcia, M. I. (2015b) A Corpus Based Quantitative Survey of the English Elements in the Basic Vocabulary of Urdu language. International Journal of Humanities and Social Science. 5(4.1); 29-36.

Maldonado García, M.I. (2015a) A Corpus Based Quantitative Survey of the Persian and Arabic Elements in the Basic Vocabulary of Urdu Language. *Pakistan Vision.* Vol.16. N. 1. 2015.

Maldonado García, M.I. (2015) A Corpus Based Quantitative Survey of the Sanskrit and Prakrit Elements in the Basic Vocabulary of Urdu Language. *Annual Pakistan Journal.* 51. 1-27.

Maldonado García, M. I. & Yapici, M. (2014a) Common Vocabulary in Urdu and Turkish Languages. A Case of Historical Onomasiology. *Journal of Pakistan Vision* 15(1) 193-221.

Maldonado García, M.I. (2014) The Urdu Language Reforms. *Almas* 15. 14-24.

Maldonado García, M. I. (2013) *Comparación del Léxico Básico del Español el Inglés y el Urdu.* PhD Thesis. Madrid: UNED.

Nadeem, Mubashar. (2012) 'Urlish: A Code Switching/Code Mixing Pedagogical Approach in Teacher Education.' *JRRE* 6, 2. 154.

Ostler, Nicholas.(2005). Empires of the Word: A Language History of the World. New York: Harper Collins.84.

Pourjafar M. Reza and Taghvaee,(2006). 'Indo-Iranian Socio-Cultural Relations at Past, Present and Future,' *Web Journal on Cultural Patrimony.* 1. 77-78.

Skjærvø, Prods Oktor, (2002) *An Introduction to Old Persian.* Cambridge: Harvard University Press, 13.

Schimmel Anna Marie. (1975). *Classical Urdu Literature from the Beginning to Iqbāl* Wiesbaden: Otto Harrassowitz Verlag, 127.

Schimmel, Anna Marie (1980) *Islam in the Indian Subcontinent.* Brill: Leiden. 5.

Rahman, Tariq (2013) *From Hindi to Urdu. A Social and Political History.* Lahore: Oxford University Press, 50.

Rahman, Tariq (2008) 'Urdu and the Muslim Identity: Standardization of Urdu in the Eighteenth and Nineteenth Centuries,' *Annual of Urdu Studies* 25. 90.

Thomason, S. G., & Kaufman, T. (2001). *Language contact* (pp. 8325-8329). Edinburgh: Edinburgh University Press.

Annex 2

Basic Vocabulary of Urdu

ایک , بارے میں , اوپر , پار , ختم , عمل , سرگرم , سرگرمی , صحیح , اپنانا , شامل , مناسب , ڈرکے بعد , پھر , عمر , پہلے , ہوا ,
تمام , اکیلے , ساتھ , پہلے ہی , ہمیشہ , ہوں , شمار , ایک , غصہ , دوسرے , جواب , چیونٹی , کوئی , کسی , کچھبی , میجہیوقت , ظاہر ,
سیب , منظور , ہیں , علاقہ , بازو , فوج , گرد , پہنچ , آرٹ , جیسا , پوچھنا , میں , حملہ , خالہ , خزاں , دور , بے ٹی , بنیاد , کمر , برا ,
بستہ , گیند , بینک , ٹوکری , غسل , ہونا , دانہ , برداشت , خوبصورت , شراب , بستر , سونے کا کمرا , برتاؤ , پہلے , شروع , کھٹی دریچے ,
علاوہ , بہترین , بہتر , درمیان , بڑا , پرندہ , پیدائش , سالگرہ , تھوڑا , کاٹ , سیاہ , خونبہنا , بلاک , خون , پھونکنا , نیلا , بورڈ , کشتی ,
جسم , ابال , ہڈی , کتاب , سرحد , پیدا , قرض , دونوں , بوتل , نچلا , کٹورا , ڈبہ , لڑکا , شاخ , بہادر , روٹی , لوٹنا , ناشتہ , سانس , پل ,
روشن , لانا , بھائی , برش , تعمیر , جلا , کاروبار , بس , مصروف , لیکن , خریدنا , نے کیک , بلانا , کر سکتا , موبتی , ٹوپی , کار ,
دیکھ بھال , احتیاط , لاپروا , گاجر , اٹھانا , کیس , بلی , مرکزی , صدی , بعض , پیچھا , تبدیل , موقع , کرسی , سستا , پنیر , چکن , بچہ ,
چاکلیٹ , انتخاب , انتخاب کرنا , دائرہ , شہر , جماعت , چالاک , صاف , واضح , چڑھنا , گھڑی , کپڑے , کپڑوں , بادل , ابر آلود , بند , کافی , کوٹ ,
سکہ , سردی , جمع , رنگ , سنگھی , آرامدہ , عام , موازنہ , آنا , مکمل , کمپیوٹر , شرط , جاری , کنٹرول , پکانا , ٹھنڈا , تانبہ , کئی , کونا ,
درست , لاگت , شامل , شمار , ملک , راستہ , ڈھانکنا , تباہی , پار , رونا , کپ , الماری , کاٹنا , رقص , خطرناک , اندھیرا , بیٹی , دن , مردہ ,
فیصلہ , کم , گہرا , ہرن , انحصار , ڈیک , تباہ , ترقی , مرنا , مختلف , مشکل , ڈنر , سمت , میلہ , دریافت , ڈش , کرنا , کتا , دروازہ , نقطہ , دگنا ,
نیچے , کھینچنا , خواب , کپڑے , پینا , چلانا , خشک , قطرہ , لچ , مٹی , فرض , ہر , کان , ابتدائی , کمانا , زمین , مشرق , آسان , کھانا , تعلیم , اثر ,
انڈہ , آٹھ , کوئی بھی , بجلی , ہاتھی , اور , خالی , آخر , دشمن , لطفاندوز , بہت , داخلہونا , برابر , داخلہ , قرار , بھی , شام واقعہ , ہمیشہ , ہر ,
سب , قطعی , سب , امتحان , مثال , سوائے , غلطی , پرجوش , ورزش , توقع , مہنگی , وضاحت , انتہائی , آنکھ , چہرہ , حقیقت , ناکام , جعلی , کرنا ,
خاندان , مشہور , دور , فارم , چربی , والد , تیز , غلطی , خوف , کھلانا , محسوس , مؤنث , بخار , چند , لڑنے , بھرنے , فلم , تلاش , ٹھیک , انگلی , ختم ,
آگ , پہلے , موزوں , پانچ , ٹھیک کرنا , پرچم , ہموار , تیرنا , فرش , آٹا , پھول , پرواز , ٹھہرنا , کھانا , بیوقوف , پاؤں , فٹبال , کے لئے , مجبور ,
غیر ملکی , جنگل , بھول , معاف , کانٹا , فارم , لومڑی , چار , مفت , آزادی , محمد , تازہ , دوست , دوستانہ , سے , سامنے , پھل , بھر اہوا ,
مزہ , فرنیچر , فرنیچر , مزید , مستقبل , کھیل , باغ , دروازہ , عام , شریف آدمی , حاصل , تحفہ , دینا , خوش , شیشہ , جانا , بکری , خدا , اچھا ,
الوداع , دادا , دادی , گھاس , قبر , عظیم , سبز , سرمئی , زمین , گروپ , بڑھنا , بندوق , بال , نصف , ہال , ہتھوڑا , ہاتھ , ہونا , خوش ,
مشکل , ٹوپی , نفرت , رکھنا , اس , سر , صحتمند , سن , دل , بھاری , اونچائی , آداب , مدد , مدد گار , مرغی , اس , یہاں , اسکا , چھپانا ,
اونچا , پہاڑ , اس , انکی , مارا , شوق , پکڑنا , چھید , چھٹی , گھر , امید , گھوڑا , ہسپتال , گرم , ہوٹل , مکان , کیسے , سو , بھوکا , گھنٹہ , جلدی ,
شوہر , تکلیف , میں , برف , خیال , اگر , اہم , میں , قتل , قسم , بادشاہ , باورچی خانہ , چاقو , دستک , معلوم , سیڑھی , خاتون , چراغ , زمین ,
شامل , رس , چھلانگ , صرف , رکھنا , اہم , قلم , قسم , بادشاہ , باورچی خانہ , چاقو , دستک , معلوم , سیڑھی , خاتون , چراغ , زمین ,
بڑا , آخری , دیر , جامیں , بعد , ست , قیادت , پتی , سیکھنا , چھوڑ دینا , ٹانگ , بائیں , قرض , لمبائی , کم , سبق , اجازت , خط , لائبریری ,
جھوٹ , زندگی , روشنی , جیسے , شیر , ہونٹ , فہرست , سننا , تھوڑا , رہنا , تالا , تنہا , طویل , نظر , کھو , بہت , محبت , کم , نیچے کرنا , قسمت ,
مشین , اہم , بنانا , نر , آدمی , بہت , نقشہ , نشان , مارکیٹ , شادی , بات , شاید , میں , کھانا , مطلب , پیمائش , گوشت , دوا , ملاقات ,
رکن , ذکر , طریقہ , درمیان , دودھ , ذہن , ملین , ملانا , ماڈل , غلطی , کھونا , جدید , لمحہ , پیسے , بندر , ماہ , چاند , مزید , صبح , زیادہ ,
ماں , پہاڑ , منہ , حرکت , بہت , موسیقی , ضروری , میرے , نام , تنگ , قوم , نوعیت , قریب , تقریباً , گردن , ضرورت , سکچی پڑوسی ,
نہیں , جال , بچھیں , نئے , خبر , اخبار , اگلا , اچھا , رات , نو , نہیں , عظیم , شور , کو بھیں , نہی , شمال , ناک , نہیں ,
اطلاع , اب , تعداد , اطاعت , اعتراض , بجر , کا , پیشکش , دفتر , اکثر , تیل , پرانے , پر , ایک , صرف , کھول , مخالف , یا , نارنجی , حکم ,
دیگر , ہمارے , باہر , اوپر , اپنے , صفحہ , درد , پینٹ , جوڑی , پین , کاغذ , والدین , پارک , حصہ , ساتھی , پارٹی , گزرنا , گزشتہ , تنھی ,
راستہ , ادا , امن , قلم , پنسل , لوگ , مرچ , فی , کامل , مدت , شخص , پٹرول , فوٹوگراف , پیانو , چننا , تصویر , ٹکڑا , سور , پین , گلابی ,

جگہ , طیارہ , پودا , پلاسٹک , پلیٹ , کھیلنا , پلیز , خوش , بہت , جیب , زہر , نقطہ , پولیس , شائستہ , پول , غریب , مقبول , مقام , ممکن , آلو , ڈالنا , طاقت , پیش , پریس , خوبصورت , روکتھام , قیمت , شہزادہ , جیل , ذاتی , انعام , شاید , مسئلہ , پیدا , وعدہ , مناسب , حفاظت , فراہم , عوامی , کھینچنا , سزا , طالبعلم , دھکا , ڈال , ملکہ , سوال , فوری , خاموش , ریڈیو , قطعی , بارش , برسات , بڑھانے , پہنچنا , پڑھنا , تیار , اصلی , واقعی , وصول , ریکارڈ , لال , یاد , ہٹانا , کرایہ , مرمت , دہرانا , جواب , رپورٹ , آرام , ریسٹوران , نتیجہ , واپس , چاول , امیر , سواری , رائٹ , بجنا , سڑک , لوٹ , پتھر , کمرہ , گول , ربر , بدتمیز , حکمرانی , حکمران , بھاگنا , جلدی , اداس , محفوظ , نمک , تھکین , اسی , ریت , بچانا , کہنا , اسکول , سائنس , تلاش , سیٹ , دوسری , دیکھنا , لگتا , فروخت , بھیجنا , سزا , خدمت , سات , کئی , جنس , رنگ , سایہ , ہلا , شکل , حصہ , تیز , وہ , صرف , بھیڑ , چادر , شیلف , چمک , قمیض , جہاز , جوتا , شوٹ , دکان , مختصر , کندھا , چننا , ظاہر , بیمار , کنارہ , اشارہ , خاموشی , حماقت , چاندی , ملتے جلتے , سادہ , ایک , بعدسے , گانا , ڈوبنا , بہن , بیٹھا , چھ , سائز , مہارت بہت جلد , سکرٹ , کیمبی , آسمان , سونا , پھسلنا , ست , چھوٹا , بو , مسکرانا , دھواں , تمباکونوشی , برف , تو , صابن , موزا , نرم , پچھ , کسی , کچھ , معافکرنا , آواز , سوپ , جنوب , خلا , بات , خصوصی , رفتار , جادو , خرچ , پنچ , کھیل , پھیل , موسمہ بہر , مربع , مہر , کھڑا , ستارہ , شروع , ریاست , سٹیشن , رہنا , چوری , بھاپ , قدم , اب , بھی , پیٹ , پتھر , روکنے , ذخیرہ , طوفان , کہانی , عجیب , گلی , مضبوط , ساخت , طالبعلم , مطالعہ , پاگل , مضمون , مادہ , کامیاب , ایسے , غیرمتوقع , شکر , مناسب , موسمگرما , سورج , دھوپ , حمایت , یقین , حیرت , میٹھی , تیر , تلوار , میز , لینا , اونچا , ذائقہ , ٹیکسی , چائے , پڑھانا , ٹیم , آنسو , ٹیلیفون , ٹیلیویژن , بتانا , دس , ٹینس , خوفناک , امتحان , سے , کہ , ان , تو , اُدھر , لہذا , ان , گاڑھا , تیلی , چیز , سوچتے , تیسری , اس , اگرچہ , خطرہ , تین , صاف , باندھنا , عنوان , پر , آج , پنجہ , ملکر , کل , آجرات , بھی , آلہ , دانت , اوپر , ٹوئل , چھو , ٹاؤن , ٹرین , ٹرام , سفر , درخت , مصیبت , ٹرک , سچ , اعتبار , دوبار , دو , کوشش , باری , قسم , نیچے , چچا , سمجھ , اکائی , بھتک , اپ , استعمال , مفید , جسممعمول , اکثر , سبزی , بہت , گاؤں , آواز , دورہ , انتظار , جاگنا , چلنا , چاہنا , گرم , دھونا , ضائع , دیکھنا , پانی , راستہ , ہم , پہننا , موسم , شادی , ہفتہ , وزن , عجیب , خوشآمدید , تھیکتھاک , مغرب , کیلی , کیا , پہیا , جب , جہاں , جس , جبکہ , سفید , جو , کیوں , وسیع , بیوی , جنگلی , وصیت , جیت , ہوا , کھڑکی , شراب , موسمسرما , تار , دانشمندانہ , خواہش , ساتھ , بغیر , عورت , تعجب , لفظ , کام , دنیا , فکر , بہترا , لکھنا , غلط , سال , جی , کل , ابھیتک , آپ , جوان , آپکا , صفر , زیبرا.

☆☆☆☆☆