# Etymological and String Analysis of Portuguese-Urdu Shared Vocabulary

**María Isabel Maldonado García**
**Ana Borges**

## Abstract

The current study is framed within the discipline of applied linguistics and falls within the scope of contrastive analysis between Portuguese and Urdu. Portuguese and Urdu are Indo-European languages, hence, share a genetic relationship. In addition, they integrate elements from other languages such as Arabic and Persian. The originality of this study is based on the fact that until now, comparative studies of both languages are inexistent. In addition, both have integrated different elements belonging to other languages, such as Arabic and Persian. Lexical borrowing plays a fundamental role in this study considering the Portuguese presence in India for centuries. A comparative analysis of phonetically similar terms in Portuguese and Urdu is performed in order to confirm their common origin, and level of similarity in the form through an etymological and string analysis. This investigation is a very interesting and important didactic instrument for the Pakistani students of Portuguese language and the Portuguese speaking students of Urdu, who will learn how to identify shared vocabulary between these languages, and utilize this vocabulary for learning a second language. It is equally important for those interested in the studies of Portuguese and Urdu.

***Keywords***: shared vocabulary, linguistic similarities, bilingualism, historical linguistics

## Introduction

The purpose of this research is to analyze 10 sets of terms which present phonetic similarities in Portuguese and Urdu languages. The rationale behind the comparison is to assist language students in the identification of shared vocabulary. In this case, it will be revealed whether the pairs have a common origin or not and possess other elements of similarity, such as semantics. The words from Portuguese language are *balde, braço, chá, chave, dez, girafa, hospital, sono, toalha, tu.*

As we move forward within the study, we should mention briefly the history of comparative linguistics. This is a branch of historical linguistics that deals precisely with the comparison between languages and thus establishes among them historical relations, for example their genetic

relationships. The studies of the common origin of languages date back to the eighteenth century, with William Jones, (1746-1794), British orientalist and jurist who worked with Indo-European languages and launched the hypothesis that among them there was a common origin. He was not only a linguist; he was a self-taught polyglot. In fact, he conducted research highlighting a number of similarities between Sanskrit, and Greek including a common origin (Olchewski, 2002). Many others followed such as Johang Christoff Adelung, of whom it is said that he contributed to linguistics creating the term "Indo-European," as well as Franz Bopp who was another German linguist, known for his extensive work within the frame of Indo-European comparativism. It was him, though; the one to propose through a detailed comparison to show that within the Indo-European languages there was a common origin.

Languages which are related, belong to the same family and one original language denominated in linguistics i.e., proto-language. According to Robert Rankin:

> While the principal goal of most linguists who are also historians has been to learn as much as possible about earlier languages and about past cultures through their languages, other branches of linguistics have benefited a great deal from the by-products of comparative work. Many who are philosophically synchronic linguists have looked to comparativists to inform them about the possible types and trajectories of language change. The study of attested and posited/reconstructed sound changes has played an important role in the formulation of notions of naturalness in phonological theory, and modern theories of markedness and optimality often rely, implicitly if not explicitly, on historical and comparative work. (2003)

According to Countinho (1976), "The Comparative-Historical method is based on relating the facts of a language similar to another in the same family so they discover the source or origin." In this context it is possible to reconstruct a proto-language through the similar characteristics of derived languages.

Portuguese language has been influenced by several other languages such as Persian and Arabic. The number of Arabisms found in Portuguese language is considerably less than in Spanish language. In addition, contact between Portugal and Persia began in the late fourteen hundreds and due to this reason Portuguese language has been influenced by Persian language.

The Portuguese at that time in history began their maritime expansion, spreading to many regions of Africa, Asia and America. From the sixteenth to the eighteen century, Portuguese language became the Lingua Franca of Asia and Africa being it used for the administration of the colonies or for trade and communication among the local officials.

In the early twentieth century, the political presence of Portugal in Asia was limited to the territories of Goa, Daman and Diu, in India; a part of the island of Timor, Indonesia; and the area of Macau, on the shores of China. But the Portuguese had controlled much more extensive regions formerly, especially in Ceylon and Malacca. Today, the Portuguese sovereignty disappeared in the East: Macau definitely went to China in 1999, the "Portuguese India" was recovered by the Indian Union in 1961; Timor was annexed by Indonesia in 1974. Still Portuguese is still present in some of these areas. In the state of Goa, in India, the Portuguese language is currently taught in official and private schools. The Goa University has a Master's Degree in Portuguese Studies since 1988. It is also an official language of the "Special Chinese Administrative Region" Macau (alongside Chinese). It is not strange though, that Urdu, originally an Indian language has been influenced by Portuguese.

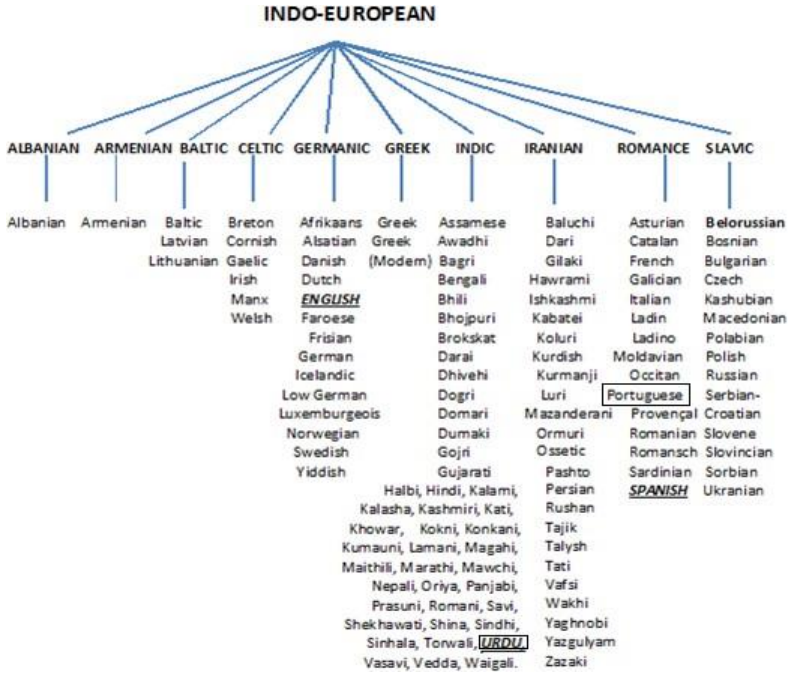


Figure: Genetic Relationship between Portuguese and Urdu Languages[1]

The figure illustrates genetic relationship between Portuguese, a Romance language and Urdu, an Indic language.

It is assumed, given the previous information, that during this research some percentage of borrowing as well as phonetic adaptation to the new language will be encountered. Campbell states “there are many different kinds of language-contact situation, and the outcome of borrowing can vary according to the length and intensity of the contact, the kind of interaction, and the degree of bilingualism in the populations” (1998). Haspelmath & Tadmor on the other hand state that when assessing genealogical relatedness it is fundamental to separate or identify inherited material from that material that constitutes a borrowing. While the loan words are an indication of historical contact, they are not of genealogical relatedness (2009, p. 1).

The motivation of our research is to embark in comparative linguistic research involving Urdu and Portuguese. Urdu is a language which has not been investigated thoroughly from a linguistics point of view. The Urdu departments of the Pakistani Universities focus on the study of Urdu literature rather than Urdu linguistics. Rahman, a well-known Pakistani linguist stated “Pakistan is perhaps the most backward country of South Asia in the field of linguistics” (1998). Fifteen years later, the situation has considerably improved, yet, the above statement still holds true.

When carrying this research we should not fail to mention the previous research of Maria I. Maldonado titled Estudio Etimológico de Cuatro Pares de Cognados en Español y Urdu, published in 2013 in the journal *Revista Iberoamericana de Lingüística,* Vol 8. In her work, Maldonado’s hypothesized that since Arabic and Persian are languages which have influenced Spanish and Urdu, there must have been a common origin in some of the sets. The end result of the etymological study was a common origin in at least two sets of the analyzed vocabulary which had their origin in Arabic and Persian and in the other two sets there was some level of uncertainty, although the evidence pointed to a common origin in both, one Latin and the other Germanic.

According to Maldonado, “En el ámbito de este trabajo, los cognados son en realidad, sinónimos en lenguas diferentes; distintos significantes que representan el mismo significado” (2013). This means that in line with the author, cognates are synonyms in different languages, different signifiers with the same meaning. This definition was obtained from *Curso de Gramática Española* written by the renowned Spanish linguist Francisco Marcos Marín (1980). The previous research produced positive results, for that reason in the present research the same

procedure will be followed in order to identify the origins, semantic and phonetic similarities present in the ten sets of cognates, so that students can utilize the procedures for cognate identification.

## Methodology

The sample object of our investigation consists of ten term sets of Portuguese and Urdu. The words were selected due to their synonymy as well as their apparent phonetic similarity. This is only a sample set as multiple other cognates have already been identified. The degree of similarity with reference to different aspects of linguistics will be assessed. These aspects are:

1. **Etymological Aspects**
   The etymology of each word in Portuguese language will be extracted and compared with its counterpart in Urdu language in order to contrast the origins of both terms.

2. **Interlingual Synonymy Related Aspects**
   **Semantic Analysis:** Definitions will be compared in order to find out if the shared vocabulary is synonymic or not.
   **Phonetic Analysis:** The phonetics of the word pairs will be compared according to the following parameters:

   a. There is no difference in phonetics.

   b. The difference is of one or two sounds, usually at the end.

   c. The difference is found in two or more different sounds, sometimes at the initial position.

   d. The difference is more than half of sounds.

   e. The difference is based on the fact that most of the sounds are different and have an uneven layout.

   f. The Levenshtein distance will also be used as a factor to determine the level of phonetic similarity.

Once the information has been extracted, a contrastive analysis of these results will be conducted.

## Data Presentation

The data is presented in the form of tables. First, the sets of Urdu-Portuguese words are presented. Then their etymology, semantic and phonetic similarity is traced and presented in the form of tables.

Table I: Urdu-Portuguese Sets

| Sets | Portuguese | English Meaning | Urdu & Latin Script |
|---|---|---|---|
| Set 1 | balde | bucket | balti (بالٹی) |
| Set 2 | braço | arm | bazoo (بازو) |
| Set 3 | chá | tea | chai (چائے) |
| Set 4 | chave | key | chabi (چابی) |
| Set 5 | dez | ten | daz (دس) |
| Set 6 | girafa | giraffe | zarafa (زرافہ) |
| Set 7 | hospital | hospital | aspatal (اسپتال) |
| Set 8 | toalha | towel | tolia (تولیہ) |
| Set 9 | sono | sleep | sona (سونا) |
| Set 10 | tu | you | tu (تو) |

### Etymological Aspects

Table 2: Etymology

| Portuguese | Urdu |
|---|---|
| 1.Signifier: **balde**<br>Word of uncertain origin. | Signifier: **بالٹی** Phonetics: **['bal.te.e]**<br>Originally derived from the Portuguese word *balde*. |
| 2.Signifier: **braço**<br>From the Latin *bracchium, -ii* and this one from the Greek *brakhion*. | Signifier: **بازو** Phonetics: **['ba.zoo]**<br>Originally a Persian word and this one from Old Avestan *baazu, baazaau*. |

| | |
|---|---|
| 3.Signifier: **chá**<br>From Mandarin *ch'a.* | Signifier: **چائے** Phonetics: **[ʧ a.'e.e]**<br>From Chinese Language. |
| 4.Signifier: **chave**<br>From Latin *clavis, -e.* | Signifier: **چابی** Phonetics: **[ʧ a.'β̞e.e]**<br>Portuguese word derived from the original. |
| 5.Signifier: **dez**<br>From the Latin *decem.* | Signifier: **دس** Phonetics: **['ɖaz]**<br>In Urdu, the word arrived through Prakrit. |
| 6.Signifier: **girafa**<br>From Arabic *zarafa*, "*girafa"* through Italian *giraffe*. | Signifier: **زرافہ** Phonetics: **[z˖a.'ra.fa]**<br>Word borrowed from the English *giraffe*. Ultimately from Arabic *zarafa*. |
| 7.Signifier: **hospital**<br>From Latin *hospitalis [domus]*, guest house. | Signifier: **اسپتال** Phonetics: **[ass.pa.'tal]**<br>Word taken from the English *hospital*. It came to Middle English via Old French from Medieval Latin *hospitales* neuter of Latin h*ospitalis*. |
| 8.Signifier**: toalha**<br>From Provencal *toalha*, and this one from Frankish (a Germanic language) *thwahlja*. | Signifier: **تَولیا** Phonetics: **[to.'le.a]**<br>The author is unsure. Probably from English. |
| 9.Signifier: **tu**<br>From Latin *tu*, "*tu.*" | Signifier: **تو** Phonetics: **['to.o]**<br>The word comes from the Sanskrit *tao.* |
| 10.Signifier: **sono**<br>From Latin *somnus, -i.* | Signifier: **سونا** Phonetics: **['so.na]**<br>Word coming from Prakrit *shu*. In Urdu it is used as a noun with *so.* |

## Inter-Lingual Synonymy Related Aspects

Table 3: Semantics and Form

| Sign 1: Portuguese | Common Meaning | Sign 2: Urdu |
|---|---|---|
| 1.Signifier: **balde**<br>Phonetics: **['bal.ɖ̪]** | An oval or cylindrical open container, made of metal or plastic used to hold and carry liquids. | Signifier: **بالٹی**<br>Phonetics: **['bal.te.e]** |

| | | |
|---|---|---|
| 2.Signifier: **braço**<br>Phonetics: **['bra.zo]** | Arm. Any of the extremities or limbs of the human upper torso. | Signifier:**با زو**<br>Phonetics: **['ba.zoo]** |
| 3.Signifier: **chá**<br>Phonetics: **[ʃa]** | Tea. Hot drink made by boiling an infusion with the dried leaves of the tea plant. | Signifier: **چائے**<br>Phonetics: **[tʲ̪ a.'e.e]** |
| 4.Signifier: **chave**<br>Phonetics: **[ʃa'b]** | Key. A small piece of shaped metal with individual shapes used to open or close a lock. | Signifier: **چابی**<br>Phonetics: **[tʲ̪ a.'βe.e]** |
| 5.Signifier: **dez**<br>Phonetics: **['ɖ̪eʃ˖]** | Ten. Number. Equivalent to the product of five and two; one more than nine; 10 | Signifier: **دس**<br>Phonetics: **['ɖ̪az]** |
| 6.Signifier: **girafa**<br>Phonetics: **[ɖ̪ʒ̂i. ɾa'.fa]** | Giraffe. A tall African mammal with a long neck. This animal's skin exhibits brown patches separated by lighter lines. | Signifier: **زرافہ**<br>Phonetics: **[z˖a.'ra.fa]** |
| 7.Signifier: **hospital**<br>Phonetics: **[oʃ.pi.'tal]** | Hospital. Medical and surgical institution providing care for the ill or injured. | Signifier: **اسپتال**<br>Phonetics: **[ass.pa.'tal]** |
| 8.Signifier: **toalha**<br>Phonetics: **[tu.'a.lea]** | Towel. Absorbent cloth or paper used for drying oneself or things, usually of rectangular shape. | Signifier: **تَولیا**<br>Phonetics: **[to.'le.a]** |
| 9.Signifier: **tu**<br>Phonetics: **[to.o]** | You. Used to refer to the second person of the singular. | Signifier: **تو**<br>Phonetics: **['to.o]** |
| 10.Signifier: **sono**<br>Phonetics: **['so'.no]** | Sleep. A condition of body and mind which usually recurs for a few hours every night. The nervous system presents inactivity, suspension of consciousness and body rests during this state. | Signifier: **سونا**<br>Phonetics: **['so.na]** |

## Analysis of the Results

### Identification of the Sets

The identification of the ten sets of shared vocabulary between Portuguese and Urdu was performed during our interaction with Pakistani individuals as well as Pakistani and Portuguese speaking students at University of the Punjab. Ten pairs of terms were selected which presented similarity in relation to semantics and phonetics. Basic terms have been included in the list, such as low cardinal numbers, body parts, etc., since there is a higher probability of these terms being inherited from a proto-language, hence being identified as a cognate.

### Etymological Aspects

**SET 1: Balde – بالٹی ['bal.te.e]**

***Balde*** in Portuguese has an uncertain origin. In Urdu, the word **بالٹی** was taken from Portuguese. The first record of it in the written form dates to 1900. The particular book was Mirza Rusva's *Sharif Zada.* The novel is about a man who turns entrepreneur through a happy and content life with new social values. According to the findings **بالٹی** is a borrowed term from Portuguese language.

**SET 2: Braço – بازو ['ba.zoo]**

**Braço** in Portuguese has come from the Latin *bracchium, -ii* and this one from the Greek *brakhion*, upper arm (which is from the Indo-European *bhaaghu*). On the other hand, **بازو** is originally a Persian word, from Old Avestan *baazu, baazaau*. The Sanskrit cognate is *baahu*. It was used for the first time in 1503 in the book *Nausar Har* written by Shah Ashraf Bayani. The book is a long epic describing the martyrdom of the prophet of Islam's grandson (2006, p. 169).

**SET 3: Chá – چائے [ʧ a.'e.e]**

The set shares a common Chinese origin. In Portuguese language it came from the Mandarin *ch'a* (Amoy dialect). The first report of the word we have is from the 1550s when the term arrived through Macao. In Urdu the term **چائے** was first recorded in the book *Lecturaun Ka Majmua* by the author Muhammad Nazeer Ahmad Khan in 1892. The pair shares the same origin since they both borrowed the term from Chinese language. Other cognates which came from Mandarin are Arabic: *shay,* Greek: *tsai*, Persian: *cha*, Russian: *chai*, Turkish: *çay.*

**SET 4**: **Chave – چابی [ʧ a.'βe.e]**

In Portuguese the term *chave* comes from the Latin *clavis-e* 'door-key, bar.' In Greek, the oldest form we can reconstruct is *klau-*; assuming this to be the oldest form in Italic it is explainable why Latin had a stem *(clavis)* as

well as an ostem *(clavus).* Other cognates from Indo European are: Myc. *ka-ra-wi-po-ro* /KǍąFl-фóроç/, PGr. *klāuī-* with base on the noun *klāu(o).*

In Urdu the word **چابی** has been borrowed from the original Portuguese term and underwent an adaptation. In this language its presence was reported for the first time in 1869 in the book *Khutut-i-Ghalib* by Mirza Asadullah Khan Ghalib. The pair shares the same origin.

**SET 5: Dez – دس [ˈd̪az]**

The set shares apparently a different origin. In Portuguese, the term is derived from the Latin *decem or decern.* In Urdu, the word arrived from Prakrit which is the vernacular form of Sanskrit. The Indo-European root is shared by Sanskrit *daśa*, Greek *deka*, and Latin *decem.* It is for this reason that the set shares a common origin. Other cognates are from Proto Indo-European (Germanic): *dekm*, Albanian: *djetu,* Armenian: *tasn,* Avestan: *dasa*, Breton: *dek,* Greek: *deka*, Lithuanian: *desimt*, Old Church Slavonic: *deseti,* Old Irish: *deich*, Sanskrit: *dasa*, Welsh: *deg.*

Arabic language also used it in the same way as Prakrit during the time of the prophet of Islam. The first record of the word in Urdu is dated to 1635 in the book called *Sabras* which is the first book written in Urdu language recorded in history. It was written by Asadullah Wajhi. It was actually a translation from a Persian book *Masnavi Dastur-e-Ushshaq and Husn-o-dil* written by the Persian author Mohammad Yahya Ibn-e-Saibak. The printing press had not yet arrived in India so the book was handwritten. Interestingly enough, the first printing press was brought to India by the Portuguese and the first Urdu book printed was published in 1801. Its title was *Bagh-o-Bahar* and the author was Mir Amman.

**SET 6: Girafa – زرافہ [zˠa.ˈra.fa]**

The set is an interesting one. In Portuguese *girafa* came from Arabic *zarafa*, "*girafa,*" through the Italian *giraffe*. The term is derived from an Arabic word although the letter *g* which comes from a derivation of the letter *z* is not a natural phenomenon. We can assume that the word is a common combination or linguistic metathesis. The scientific name of giraffe is *Giraffa camelopardalis*, family *Giraffidae.*

In Urdu the word was borrowed from the English *giraffe* and used for the first time in 1895 in the book *Ilm al-lisan*. Although the Urdu dictionary does not reflect the complete etymology, the word entered English language in the late 16th century from the French *giraffe*, the Italian *giraffa* or perhaps the Spanish or Portuguese *girafa* with origin in the Arabic *zarāfa*. In Middle English it was called camelopard. The pair shares the same origin.

**SET 7**: **Hospital – اسپتال [ass.pa.'tal]**

The set shares a common origin. In Portuguese language it is derived from the Latin *hospitalis [domus]*, guest house. Italic cognates include Pael. *hospus* [nom.sg.] 'stranger' (*-pot-(i)s;* Indo European Cognates: OCS *gospodb;* Russian *gospod'* the Lord, god' *ghost(i)-pot-* (Slav, *-d-* from the voc.sg. *-pofi).*

In Urdu the word is taken from the English *hospital*. It was considered slang until it was used in writing for the very first time in 1869 in the book *Khutut-i-Ghalib* by Mirza Asadullah Khan Ghalib. In English language it came into Middle English via Old French from Medieval Latin *hospitales* neuter of Latin *hospitalis*.

**SET 8**: **Toalha – تَولیا [to.'le.a]**

The set has apparently different origins. In Portuguese it came from Provencal *toalha*, which is from Frankish, a Germanic language, *thwahlja*. The author of the Urdu dictionary is uncertain about the origin and mentions English as a possibility. English is a Germanic language; hence this presumption could be correct and the set could share a common origin. In fact, if it came from the English *towel,* it would have arrived in Middle English from (*towel, towail, towaille*) Old French *toaille,* from the Frankish *thwahlja,* from Proto-Germanic *thwakhlijon,* and from Proto Indo European *tʷak-.* Some cognates are the Old High German *dwahila*, the Modern German *Zwehle*, the Dutch *dwaal* "cloth used for the altar," Middle Dutch *dwale*, in Old English *þwean* "to wash."

**SET 9**: **Tu – تو ['to.o]**

The set shares a different origin. In Portuguese it stems from the Latin Tu, "*tu*" which in turn comes from the Proto Indo-European *tu* "you" (nom. sg.). Some Proto Indo-European cognates are *ti* (H) [nom.], *tue* [acc.], *toi* [gen.dat.], *teue* [gen.], tued [abl.] 'you,' *tu-o-* 'your.' Indo European cognates: Hit. *zik* [nom.], *tu-* [obi.], CLuw. *ti, tu- tiH, tu~;* Sanskrit. *t(u)* Vam [nom.], *t(U)vam* [acc.], *tubhya(m)* [dat.], *t(u)vdt* [abl.], *tdva* [gen.], *tva, tuva* [acc.encl.], *te* [gen.abl.dat.encl.]

In Urdu the term is derived from Sanskrit *tao.* Because of influence of Persian and Sanskrit, it is possible that the word *tao* became *tu*. It was used for the first time in 1739 in the book *Qulyat-e-Siraj* by the Urdu poet Siraj Aurangabadi.

**SET 10**: **Sono – سونا [sona]**

In Portuguese the term *sono* is derived from the Latin *somnus, -i.* Some Proto Indo-European cognates are *swep-no*, from the root *swep-* "sleep."

in Sanskrit *svapnah,* in Avestan *kvafna-, i*n Greek *hypnos,* in Lithuanian *sapnas,* Latin *sopor* "deep sleep," and Old English *swefn.*

Indo-European cognates are Olr. *suan,* W. *hun* 'sleep' *suopno;* Hit. *supp-('f(ri)* 'to sleep' *sup-(t)o, suppariie/a*'to sleep' *sup-r-ie/o supparuant-* /sleepy. In Sanskrit *svapna-* [m.] 'sleep, dream,' *svapnya-* [n.] 'dream, vision,' *dusvapnyam* nightmare,' Av. *xvafna-* [m.] 'sleep, dream.' The Latin *somnium* may derive from Proto Indo-European derivative (already suggested by Schindler) or perhaps be a formation of inner-Latin. There is a chance that the Proto Indo-European *suepno-* is probably a derivation of the *r/n*-stem *suep-r/n-.*

In Urdu the word سونا came from the Prakrit *shu*. It is used as a noun with*so*. The term was used for the first time in 1503 in the book *Nausar Har* from the author Shah Ashraf Bayani.

## Semantic Analysis

The pairs present shared meanings in all sets, although this is not true for all definitions, rather, in at least one of the definitions, while other meanings are not shared.

## Analysis of the form

The form, in which Portuguese and Urdu is written, without any doubt, is completely different. While Spanish utilizes Latin script, Urdu utilizes Arabic-Persian script, Nastaliq style. For this reason the pairs do not share any orthographic characteristics. The form will be analyzed through the phonetics of both languages.

**SET 1: Balde ['bal.d̪] – بالٹی ['bal.te.e]**

Level of phonetic similarity: 3/6 -**bal-**

**'b-a-l-** d̪ ≠ **'b-a-l-** ʈ-e-e

The difference in the form is of more than two different sounds at the end. The common sounds **[b][a] [l]** constitute the similarity level which is of 50**%.**

**Levenshtein distance: 3**

**SET 2: Braço ['bra.zo] – بازو ['ba.zoo]**

Level of phonetic similarity: 4/6 **b-azo-**

**b**-r-**a.-z-o ≠ b-a-z-o**-o

There is a phonetic difference of two sounds [r] and the last [o] of the second term. The common sounds are **[b][a][z][o]**. The level of similarity is of 66.66%

**Levenshtein distance: 2**

**SET 3: Chave [ʃa'b] – چابی [t͡ʃ a.'β̪e]**

Level of phonetic similarity: 1/4 **-a-**

ʃ**-a-**b ≠ t͡ʃ̂ **- a-**β̪-e

The difference is present in more than two sounds. Nevertheless, as in the previous case similarity is found in sounds that are different. In the pair [**ʃ**] versus [**t͡ʃ̂**] like in the previous case and in [**b**] versus [**β̪**], similarities can be found, since the sounds are somehow identifiable with each other and there is correspondence between them.

The difference is based then on the fact that one sound **[a]** is identical while [**ʃ**] and [**t͡ʃ**] are correspondent and in the same location within the word and [**b**] and [**β̪**] also share the same relationship and are located in the same place within the word. The last vowel in the Urdu word is not present in the Portuguese word. Hence, the common sound **[a]** constitutes the level of similarity which is of 66.6%.

**Levenshtein distance: 2**

**SET 4**: **Chá [ʃa'] – چائے** [t͡ʃ **a.'e]**

Level of phonetic similarity: 1/3 -**a-**

ʃ**-a** ≠ t͡ʃ̂ **- a-**'e

Initially, the difference is based on the sound [**ʃ**] versus [**t͡ʃ**] as well as an ending [**e**] sound. The common sound is **[a]**. Similarity can also be found on the fact that between [**ʃ**] and [**t͡ʃ̂**] there is correspondence.

The common sound **[a]** constitutes the level of similarity which is of 66.6%.

**Levenshtein distance: 1**

**SET 5: Dez ['d̪eʃˠ] – دس ['d̪az]**

Level of phonetic similarity: 1/3

**'d̪-**e-ʃˠ **≠ d̪**-a-z

The difference in the form is obvious in the second and third sounds with e-ʃˠ **≠** -a-z. The variation is found in the second and third sounds. The common sound is **[d]** although it is followed by a vowel and there is correspondence in the similar sounds ʃ/z which will not be counted in the percentage. The similarity level is of 33.3 %.

**Levenshtein distance: 2**

**SET 6: Girafa [d̪ʒ̂i. ra'.fa] – زرافہ [zˠa.'ra.fa]**

Level of phonetic similarity: 4/6

d͡ʒ-i-**ɾ-a-f-a ≠** z-a-**ɾ-a-f-a**

The similarity is based on the last four sounds which are identical **[r][a][f][a].**

The level of similarity is of 66.66%

**Levenshtein distance: 2**

**SET 7: Hospital [oʃ.pi.'tal] – اسپتال [as.pa.'tal]**

Level of phonetic similarity: 4/7

o-ʃ**-p**-i-**t̪ a-l ≠** a-s**-p**-a-**t̪ a-l**

The similarity is based on the third sound **[p]** as well as the last three sounds **[t][a][l].**

The level of similarity is of 57.14%.

**Levenshtein distance: 3**

**SET 8: Toalha [tu.'a.lea] – تَولیا [to.'le.a]**

Level of phonetic similarity: 4/7

**t̪**- u-a**-l-e-a ≠ t̪**- o-**l-e-a**

The difference is based on more than two different sounds in the center of the word.

The sounds **[t] [l][e] [a]** constitute the level of similarity which is of 57.14**%.**

**Levenshtein distance: 2**

**SET 9: Tu [to.o] – تو [‘to.o]**

Level of phonetic similarity: 3/3

**t̪- o-o = t̪- o-o**

The sounds of this word are identical. The level of similarity is 100%.

**Levenshtein distance: 0**

**SET 10: Sono ['so’.no] – سونا ['so.na]**

Level of phonetic similarity: 3/4

**'s-o’.-n-o ≠'s-o-.n-a**

The difference is based on one different sound at the end of the word. The sounds **[s] [o] [n]** constitute the level of similarity which is of 75%.

**Levenshtein distance: 1**

## Discussion

In two of the sets, the Urdu term has been borrowed from Portuguese language and in one of them the origin of the Portuguese term is inconclusive. Four of the sets share the same origin and there is a high level of probability that one more shares the same origin making it obvious that these sets have been borrowed. Three of the sets have a different origin. In one of the sets the Urdu dictionary only cites the vehicular language, rather than the origin, although it can be traced to a Sanskrit root. A more exhaustive investigation of the etymon of the unclear terms is necessary, possibly revealing a common origin in some of the sets. In Portuguese language, the origins of the terms have been contrasted in different dictionaries, resulting in the same origins in all of the different sources. In Urdu, the absence of diverse sources makes it impossible to conduct a deeper contrastive analysis. For this reason, we believe that in addition to the creation of additional sources in Urdu, a review of the terms in the Urdu dictionary would assist etymological study of the Urdu lexicon.

The study is of interest to the Pakistani students of Portuguese as well as the Portuguese speaking students of Urdu as it can help in the phonetic and semantic recognition of shared vocabulary and in understanding language borrowing as well as how these languages are related and interact with each other. It is also an important study from the point of view that it opens doors for facilitating language acquisition through the recognition of cognates and shared vocabulary.

The study is also of interest to linguists working in comparative linguistics and genetic studies of Indo-European languages as until now studies comparing both languages, Portuguese and Urdu have never been performed.

## References


Campbell, L. (1998). *Historical linguistics*. Boston: MIT Press.

Coutinho, I. (1976). *Histórica*. Rio de Janeiro: Ao Livro Técnico.

Cunha, A. G. (2001). Dicionário etimológico (2nd ed.). Brazil: *Editora Nova Fronteira*.

DeVaan, M. L. (2008). *Etymological dictionary of Latin and the other Italic languages.* Leiden, Boston: Brill.

Dicionário Priberam língua Portuguesa**.** (2003-2013). Porto, Portugal: Porto Editora. Retrieved from http://www.priberam.pt/

Haspelmath, M., & Tadmor, U**.** (2009*). Loanwords in the world's languages: A comparative handbook.* Berlin, Germany*:* De Gruyter Mouton.

Khan, A. J. (2006). *Urdu/Hindi: An artificial divide, African heritage, Mesopotamian roots, Indian culture & British colonialism.* New York: Algora Pub.

Maldonado García, M. I. (2013). Estudio etimológico de cuatro cares de cognados en Español y Urdu. *Revista Iberoamericana de Lingüística*, *8*, 61-76.

Maldonado García, M. I. (2013). *Comparación del léxico básico del Español, el Inglés y el Urdu.* (Unpublished doctoral dissertation). Madrid: UNED.

Marcos Marín, F. (1980). *Curso de gramática Española.* Madrid: Cincel-Kapelusz.

Nourai, A. (2013). *An etymological dictionary of Persian, English and other Indo-European languages.* Bloomington, In.: Xlibris Corp.

Olschewski, T. (2002). Indo-European Linguistics*: A study on the basic differences between Germanic and Slavic language, exemplary presented on English and Serbo-Croatian*. Munich: Grin Verlag.

Rankin, R. L. (2003). *The Handbook of Historical Linguistics.* UK: Blackwell Publishing.

Rahman, T. (1998). Linguistics in Pakistan [Canada]: A Country Report. In R. Singh (Ed.), *The yearbook of South Asian languages & llnguistics* (pp. 90-123). Delhi: Sage Publications.

Simpson, J., & Weiner, E. (1989). *The Oxford English dictionary*. Oxford: Oxford University Press.

Urdu *Encyclopedia*. (2011). Ministry of Science and Technology. Islamabad, Pakistan. Retrieved from http://www.urduencyclopedia.org/ urdudictionary