



## ORTHOGRAPHICAL DIFFERENCE OF PUNJABI LANGUAGE AND SARAIKI VARIETY IN PUNJAB PROVINCE

**Sohaib Khalid**, M.Phil in Applied Linguistics, University of Management and Technology, Lahore, Pakistan  
[sohaibkhalidbajwa@gmail.com](mailto:sohaibkhalidbajwa@gmail.com)

**Amara Batool**, M.Phil in Applied Linguistics, Govt. College University Faisalabad, [amarabatool377@gmail.com](mailto:amarabatool377@gmail.com)

**Muhammad Imran**, Lecturer in English, University of Sahiwal, Pakistan, [imranjoyia76@gmail.com](mailto:imranjoyia76@gmail.com)

**Muhammad Younas**, School of Education, Soochow University, Suzhou, China, [younaskherani@stu.suda.edu.cn](mailto:younaskherani@stu.suda.edu.cn)

**ABSTRACT-** The objective of this study is to identify the lexical differences between Punjabi and Saraiki. It also helps those students who want to learn Punjabi language and the second language of Saraiki. The numeral amount of studies have carried out which estimate the degree of comparability between two separate languages and two dialects of the same language, but the level of similarity between Saraiki and Punjabi has not yet been investigated. One thousand words of high-frequency use of the Levenshtein's algorithm (1966) to examine the discrepancy between these words (the new corpus fulfilled by the investigator) in Punjabi and Saraiki. The theory is that, owing to the hereditary relationship between the two language systems, the lexical gap between the two systems is less than 75%.

**Key words:** Punjabi language, Saraiki variety, lexical differences, cross-language similarities, lexical distance, language acquisition.

### I. INTRODUCTION

According to the Ethnologue, 73 languages are spoken in Pakistan from which Urdu is the national language. According to The World Fact book six main languages are spoken in Pakistan with fifty-nine smaller languages. Central Intelligence Agency declared that the major languages in Pakistan are: Punjabi with 48% of the total speakers, Sindhi is 12%, Saraiki is the third biggest language of Pakistan with 10%, Pashto and Urdu speakers are 8% in the Pakistan and Balochi is 3% of the total population of the country. Additionally, Hindko has 2% speakers, and Brahui has only 1% speakers of the total population of the country. Moreover, Burushaski and other languages collectively have 8% speakers of total population of the country. The two most important languages of the country are Punjabi and Urdu. Mostly, the languages, which are present in Pakistan, belong to Indo-Iranian families of language, which has its roots in the Indo-European language families. The Indo-Iranian families of language divided into Urdu, Punjabi, Sindhi, Hindko, Saraiki and other languages (Ethnologue, 2017). The distributions of cognates and cross-language frequency diverge according to language contact and evolutionary change. While there is variation of the studies that estimate the comparison level among two changed languages and the two dialects of a same language, While now the lexical resemblance or difference level among Saraiki and Punjabi language is not explored?

As stated before Saraiki and Punjabi languages belongs to Indo-Aryan families of language due to this reason they might have the extraordinary connection of mutual intelligibility between them. Sometimes, the people who speak different languages can communicate with one another i.e., speakers of Spanish and Portuguese can understand each other because the similarity level between both languages is 89% (Maldonado García, 2014) and both languages have roots in the same family. The higher level of similarity between two languages or dialects, the linguistically closer to each other the speakers of that languages or dialects are. The higher level of distance between two languages or dialects, the linguistically farther to each other the speakers of that languages or dialects are. Saraiki and Punjabi are from the Indo-Aryan family of languages which mean that both languages share genetic relations. The speakers of Saraiki language claim that it is a separate language but Maldonado García (2016) as well as others proposed that it is a dialect of Punjabi language. The main

objective of the present study is to estimate the orthographic differences between Punjabi language and Saraiki variety.

### **Language Patterns in Pakistan**

Census of the population (1984) in Pakistan comprises a detailed questionnaire column, which keeps the eight languages options for selecting a desired language from the respondents. Eight major dialects have been qualified as Pakistan's national idioms and these languages are subjected to represent the Pakistani residents. This specified approach is best suited for Punjab with contains more than the three-quarters of targeted population prefers Punjabi as their mother tongue. This important relation may be reported for the Sindhi and Pashto languages (Rahman, 1995). In Baluchistan, Baluchi is the main tongue. The value of language stems from fact that Hindko, Barahui and Saraiki spoken in north side of the province Punjab.

### **Punjabi Language**

As Punjabi is themother tongue of most of the Punjabi people, it isspoken across the cities of Punjab province i.e., Lahore, throughout Gujranwala, and on the border areas with India and some of the smaller cities of the north Punjab. The Punjabi language has millions of speakers throughout Pakistan as Liudmalia (2014)states that the Punjabi language is spoken almost 48% of the total population of the country as mother tongue language and 57.56% in cities of Punjab Province. According to Ethnologue (2017),basically, the Western Punjabi contains two scripts, Arabic and a Landa script, which is no longer used. Shahmukhi is derived from any of these. Shahmukhi is in which the mouth of the Guru for which Gurmukhi refers to used to write Punjabi language and it often used to write Urdu. In the history of Punjabi, Sufi poets in Punjabi and after the independence of Pakistan used Shahmukhi alphabet for the first time, it became the common writing form in the language (Goyal and Lehal, 2008). According to Ethnologue, Punjabi has different varieties such as Bhatyiana, Bathi, Majhi, Doab, Powadhi, Malwa, Shahpuri and Lahori (2017). Punjabi and Saraiki have about 70 percent-85 percent parallels with each other (Ethnoligues, 2017). Although according to Punjabi standard dialect is Majhi in Pakistan. However, in Haryana, Delhi, Himachal Pradesh, and district of Jammu and Kashmir, it is spoken as mother tongue.

### **SaraikiVariety**

Saraiki variety is also recognized as Multani, Landhi Seraiki and Western Punjabi (Ethnoligues, 2017). It is also acknowledged by Maldonado García(2016) that Saraiki is a dialect of Punjabi. It is spoken by 10% speakers of the total population of Pakistan (Liudmalia, 2014). Ethnologue states that the Saraiki variety is written in Arabic script mostly, while "Devanagari" script and "Gurmukhi" script were used to be written for the Saraiki language but now, no longer in use (2017). The Arabic script of Saraiki is commonly used on computers for the writing.Kazim (2015), a well-knownSaraiki scholar, states that Saraiki has 45 alphabets for writing and Hindus and Sikhs used Gurmukhi and Devanagari scripts which were written left to right. Although, these scripts are no longer in use in Pakistan still in India the Gurmukhi and the Devanagari scripts are in use. In the past, "Landhi" script was used more prominently by the bookkeepers, traders, writers and other people, but in these days,a significant reduction came in the use of this script (Liudmalia, 2014). On the contrary, Multani; which belongs to the Landa script was also used to write down Saraiki. The dialects that are commonly associated with Saraiki language are Bhawalpuri, Jafri, Jatki, Derwali, Jangli, Multani, Khatki, Reasati, Hindki, Southern Panjabi and Thali (Ethnologue, 2017).

### **Lexical Similarity**

Lexical resemblance is very measure to the extent thatgroups of words of two different particular dialects are similar(Renfrew and Tarim, 2003). If two words have 100% similarity, it means thatthe words are the same in both languages, if words have 0% similarity; it shows that there is no resemblanceat all. Ethnologue, (2017) explains thatratio of similarity between two languages can checked by associating the meaning and form of two differenttype'slists of words. If two dialects or languages have 85% similarities or more than that, it shows that the one which called dialect of another language. Younas, M. & Noor, U. (2020) obtain various

results by associating letters and the phonetic symbols of the different inquiry level studies. For example, French language and English have 27% resemblance while the English and the German language have almost 60% resemblance level. When the languages have similarity level more than 85%, it shows that there is particular relation, syntactically, either phonetically and morphologically (Maldonado García, 2014).

In this present study, the scholars has selected 1,000 words of Punjabi language from Majhi dialect, which called the standard dialect form of Punjabi and then interpreted these words into Saraiki variety. Until now the orthographical differences between Punjabi and Saraiki have not been recognized. Maldonado García (2014) states, the lexical connection level between the Portuguese and the Spanish is about 89% and both languages are Romance languages. The method adopted in this study based on Levenshtein's (1966) algorithm model to discover the orthographical distance among Punjabi and Saraiki language.

### **Objective of the Study**

One of the objectives of the research presented in this study is to explore the lexical gap between Punjabi and Saraiki by collecting the data. A researcher has made the distinction of 1000 high-frequency words in Punjabi. The data gathered in Majhi language and then converted to Saraiki. The aim of this research is to compare and contrast the two languages.

### **Significance of the Study**

As we know that Pakistan is a multi-lingual country, it accommodates a number of major and minor languages, some of them consider as a major mother tongue, and Punjabi and Saraiki are also included in them. This study is intended to help the students of Saraiki and Punjabi language in learning second language setting. Younas et al. (2020) examples that learners who need to learn Punjabi or Saraiki as their subordinate language may get the list of words of similar words in both (Punjabi and Saraiki) and can use them in the acquisition of the vocabulary.

## **II. LITERATURE REVIEW**

Paul, Christopher and Stephen (2002) state the language family involves an inter-related linguistic groups invented from a specific ancestor (of common history), also labelled as protolanguage (Greek word proto meaning 'early'). There is no direct consideration regarding ancestor language, though; there is a possibility to ascertain its diverse features through proportional methods' application that can express status. As per Irene (2015), Romance, Germanic, and Slavic are among numerous branches of the family of Indo-European language.

The development of Turkish language as an administrative language was mostly in Ottoman Empire (of 600 plus years) by borrowing words from Persian and Arabic (Renfrew and McMahon, Trask, 2000). It was later on named Ottoman Turkish, mainly an official version of routine day life spoken Turkish at that time.

Comparative linguistics is implemented for establishing languages' membership for the specific language family. Younas, M (2020) explains that linguistics theme mostly focuses on the nature of various modern languages. Linguists who work in this field of study typically use family identity to classify the languages (Hardcastle and Beck, 2005). Currently formed families of the languages are the more certain results of common past of their language origins. The authors referred this data as "cultural and linguistic groups". Besides, there are a huge variety of GPGs around the world (Gahl and Johnson, 2012). However, this method is not useful for languages whose status is indeterminate. Korean and Japanese speakers are not similar to each other and to speakers of other East Asian languages.

## **Orthographical Difference Among Different Languages**

Orthography is a term referring to the collection of conventions used for writing a language (Gahl, Yao and Johnson, 2012). It contains phrase splits, hyphenation, spelling guidelines, capitalization, emphasized words and so on. Recently several languages are being written in specifically built systems and writing standards. Findings by Shany et al., in November 2011 showed that those who are able to interpret words correctly with diacritics are better able to read words without diacritics. So far, it has not been discussed or shown just how distinct Saraiki dialect is from Punjabi dialect because of its orthographical distinction. For example, Saraiki share a significant portion of vocabulary and morphology of Punjabi language. Younas, M, & Noor, U (2020) argued that language has yet to record by the linguists.

## **Ethno-Lingual Distinctiveness of Saraiki Identity and Rest of Punjab**

The majority of linguists are not accepting the fact that Saraiki and Punjabi are two separate languages that could be mutually interrelated and understood i.e. one language can replace other easily in the context of understanding (Rehman, 1995). It is clearly indicated by Asif, that out of all languages of Punjab, Saraiki is the most important language but the Saraiki narrative is claimed as the opposite (2005). There are two belts that keep differing norms and cultures that are Saraiki and Punjabi. Saraiki keeps the distinctive association with other parts of other provinces like Sindh and central Punjab and also with north India. Asif has viewed that Saraiki and Punjabi are almost same in many ways and share context on the basis of morphology (2005). Saraiki has become the important language and symbol of ethnicity among the residents of South Punjab and their living associations in the province. The areas of South Punjab and Sindh keep mutual similarities in context to the geographical locations, political backgrounds, and cultural inference.

## **Mutual Intelligibility**

Rowe and Levine (2015) concluded that two or more languages in which speakers of different languages (but genetically related) will very readily understand each other without prior understanding are called reciprocal intelligibility. Sometimes it may be used as a criterion to distinguish different languages. The extent of mutual intelligibility of languages varies greatly. How to improve the accent takes a lot of effort. The speakers of such languages of comprehension because of their shared genes. Being able to comprehend and learn a foreign language is not an automatic thing (Brown, 2015).

## **III. METHODOLOGY**

To obtain the orthographic difference level between Punjabi language and Saraiki variety data was compiled. Since we want to find the orthographical difference, it is substantial to utilize those terms, which the speaker mostly finds in that language. For this drive, one thousand regularly used words of Punjabi (Majhi dialect) have been recognized. The motive behind this is that the most frequent words are used by the most number of speakers. Maldonado Gracia, (2014) states that, the words which are frequently used by the speakers have a resistance to change. To obtain the high-frequency words, the researcher got the "BHULEKHA" newspaper as a corpus, which is in Punjabi language (dated August 28 and 29, 2017). Then, by the help of most frequent word finder software (online word counter), the researcher has obtained the list of one thousand most frequent words in Punjabi.

## **Measuring Orthographic Distance**

In 1995, Kendall used the Levenshtein distance to measure language distance between varieties. He used the algorithm to analyse different dialects in Irish. Levenshtein distance is a single-character edit distance metric. According to the verbal gap between varieties of English, there is a disparity between spellings. The present thesis is focused on Ethnologue and Automated Similarity Judgment Program (ASJP). In order to form the similarities index between Spanish and English, the Levenshtein distance algorithm was also used.

Levenshtein algorithm was used to equate the English term "interest" with the Swedish word "intresse." These four operations were applied to compare two Swedish words. The actual distance is 33% of the necessary radius. Interest may be calculated in several forms but Levenshtein distance still offers the expense of the cheapest mapping. In each character, we notice a base and a diacritic. For eg, an acute accent sits on the e syllable. Two characters may have different foundation and diacritics. We multiply variations with a unitary number: 1. If two characters have the same basis, but separate diacritics, we consider it to be 0.3 and calculate each other. For example: e or é, è or é. We think that the total system weights should be skewed to the left, mostly because the pronunciation is dependent on diacritics. If two bases are the same, the product is always 1, regardless of whether there are separate diacritics at the beginning or the end. This is counted as one. As all the languages of the region have similar scripts it is determined phonetically and not orthographically.

To uncover the differences between Punjabi and Saraiki, it is divided into various groups.

#### IV. DATA ANALYSIS

To get the corpus of recent date, we have got the newspaper in Punjabi language: "Bulekha". The corpus of fifteen thousand words was collected and from which by using the software, researcher has got the one thousand most frequent words. Once the words in Punjabi (Majhi dialect) were found, their equivalents in Saraiki language were recognized. When the words list in both; Punjabi and Saraiki was prepared than researcher also translate them in the English language. So, in first column we have serial no, second is reserved for words in English language, third for words in the Punjabi language, the fourth column is for words in Saraiki language and in the last column, we have Levenshtein distance. Terms are categorized according to the methodology described in previous chapter.

(A) Terms which have only one character or distance 1

The words which have distance of one character are 18 out of 1000. They are given below.

List of words which have one character distance

| Sr.no | Words in English | Words in Punjabi | Words in Saraiki | Levenshtein Distance |
|-------|------------------|------------------|------------------|----------------------|
| 1     | On               | اتے              | تے               | 1                    |
|       |                  | اتے              | تے               |                      |
| 2     | To say           | آکھنا            | آکھیا            | 1                    |
|       |                  | آکھنا            | آکھیا            |                      |
| 3     | Time             | بار              | وار              | 1                    |
|       |                  | بار              | وار              |                      |
| 4     | Make             | بنان             | بنڑن             | 1                    |
|       |                  | بنان             | بنڑن             |                      |
| 5     | First            | پہلے             | پہلا             | 1                    |

|    |                          | پھل ا  | پھل ے  |   |
|----|--------------------------|--------|--------|---|
| 6  | Where                    | کتھے   | جتھے   | 1 |
|    |                          | کتھ ے  | جتھ ے  |   |
| 7  | Sparrow                  | چڑی    | چڑیا   | 1 |
|    |                          | چڑی    | چڑیا   |   |
| 8  | Give                     | ڈتا    | دتا    | 1 |
|    |                          | ڈتا ا  | دتا ا  |   |
| 9  | To tell                  | ٹسیا   | دسیا   | 1 |
|    |                          | ٹسیا ا | دسیا ا |   |
| 10 | Put                      | رکھ    | رکھن   | 1 |
|    |                          | رکھ    | رکھن   |   |
| 11 | To do                    | کرڈے   | کردے   | 1 |
|    |                          | کرڈے   | کردے   |   |
| 12 | To meet                  | ملنا   | ملن    | 1 |
|    |                          | ملنا ا | ملن ا  |   |
| 13 | Punch                    | مکھ    | مکا    | 1 |
|    |                          | مکھ    | مکا    |   |
| 14 | Introduction             | تعارف  | متعارف | 1 |
|    |                          | تعارف  | متعارف |   |
| 15 | Near                     | نیڑے   | نیرے   | 1 |
|    |                          | نیڑے   | نیرے   |   |
| 16 | New                      | نوی    | نئی    | 1 |
|    |                          | نوی    | نئی    |   |
| 17 | There (pronoun of place) | الے    | والے   | 1 |
|    |                          | الے    | والے   |   |
| 18 | Look at                  | ڈیکھنا | ویکھنا | 1 |

|  |  |         |         |  |
|--|--|---------|---------|--|
|  |  | وی کھنا | ڈی کھنا |  |
|--|--|---------|---------|--|

(B) Terms which have difference of 02 characters

The words which have distance of two characters are 12 out of 1000. They are illustrated by the following table.

List of words which have difference of 02 characters

| Sr.no | Words in English | Words in Punjabi | Words in Saraiki | Levenshtein distance |
|-------|------------------|------------------|------------------|----------------------|
| 1     | Here             | اے               | ایہ              | 2                    |
|       |                  | اے               | ای ہ             |                      |
| 2     | To do            | کردیاں           | کر ڈیتائی        | 2                    |
|       |                  | کر دیاں          | کر ڈی تائی       |                      |
| 3     | Happened         | ہوی              | تھی              | 2                    |
|       |                  | ہوی              | تھی              |                      |
| 4     | When             | جوں              | جڈاں             | 2                    |
|       |                  | ج دوں            | ج ڈاں            |                      |
| 5     | Here             | ایتھ             | ایتھاں           | 2                    |
|       |                  | ایتھے            | ایتھاں           |                      |
| 6     | Hit              | ماریا            | مارے             | 2                    |
|       |                  | ماریا            | مارے             |                      |
| 7     | Force            | فورس             | فوج              | 2                    |
|       |                  | فورس             | فوج              |                      |
| 8     | New              | نیا              | نوا              | 2                    |
|       |                  | نیا              | نوا              |                      |
| 9     | Hundreds         | سیکڑوں           | سیکڑے            | 2                    |
|       |                  | سیکڑوں           | سیکڑے            |                      |
| 10    | Top              | اوپر             | اوتے             | 2                    |
|       |                  | اوپر             | اوتے             |                      |

|    |          |       |        |   |
|----|----------|-------|--------|---|
| 11 | Where    | کیتھے | کیتھاں | 2 |
|    |          | کیتھے | کیتھاں |   |
| 12 | Increase | ودھان | ودھن   | 2 |
|    |          | ودھان | ودھن   |   |

(C) Terms which have difference of 03 characters:

The words which have distance of three characters are 26 out of 1000. They are illustrated by the following table.

List of words which have difference of 03 characters

| Sr.no | Words in English | Words in Punjabi | Words in Saraiki | Levenshtein distance |
|-------|------------------|------------------|------------------|----------------------|
| 1     | And              | اینڈ             | بیاء             | 3                    |
|       |                  | اینڈ             | بیاء             |                      |
| 2     | Such as          | ایسی             | ایجھی            | 3                    |
|       |                  | ایسی             | ایجھی            |                      |
| 3     | Twelve           | بارہ             | بارھاں           | 3                    |
|       |                  | بارہ             | بارھاں           |                      |
| 4     | Peshawar         | پشور             | پشاور            | 3                    |
|       |                  | پشور             | پشاور            |                      |
| 5     | Power            | پاور             | طاقت             | 3                    |
|       |                  | پاور             | طاقت             |                      |
| 6     | Three            | تین              | تریہ             | 3                    |
|       |                  | تین              | تریہ             |                      |
| 7     | Justice          | جسٹس             | جج               | 3                    |
|       |                  | جسٹس             | جج               |                      |
| 8     | Therefosre       | جکھ              | جڈاں             | 3                    |
|       |                  | جکھ              | جڈاں             |                      |



|    |              |        |        |   |
|----|--------------|--------|--------|---|
| 9  | Go           | جاوے   | ونجے   | 3 |
|    |              | جاوے   | ونجے   |   |
| 10 | Place        | جگہ    | جہاء   | 3 |
|    |              | جگہ    | جہاء   |   |
| 11 | They         | جناں   | جنہاں  | 3 |
|    |              | جناں   | جنہاں  |   |
| 12 | Both of them | دونوں  | ڈونہاں | 3 |
|    |              | دونوں  | ڈونہاں |   |
| 13 | System       | سسٹم   | نظام   | 3 |
|    |              | سسٹم   | نظام   |   |
| 14 | Terms        | شرائط  | شرطاں  | 3 |
|    |              | شرائط  | شرطاں  |   |
| 15 | Citizens     | شہری   | شہرآلے | 3 |
|    |              | شہری   | شہرآلے |   |
| 16 | In case      | صورت   | حالت   | 3 |
|    |              | صورت   | حالت   |   |
| 17 | People       | عوام   | لوک    | 3 |
|    |              | عوام   | لوک    |   |
| 18 | Duty         | فریضہ  | فرض    | 3 |
|    |              | فریضہ  | فرض    |   |
| 19 | Butcher      | قصاباں | قصابی  | 3 |
|    |              | قصاباں | قصابی  |   |
| 20 | Out          | کٹیا   | کٹھیا  | 3 |
|    |              | کٹیا   | کٹھیا  |   |
| 21 | Any          | کسی    | کوئی   | 3 |
|    |              | کسی    | کوئی   |   |

|    |              |        |         |   |
|----|--------------|--------|---------|---|
| 22 | To write     | لکھنا  | لکھڑنا  | 3 |
|    |              | لکھنا  | لکھڑنا  |   |
| 23 | To take      | لینا   | گھننا   | 3 |
|    |              | لینا   | گھننا   |   |
| 24 | Mission      | مہم    | مقصد    | 3 |
|    |              | مہم    | مقصد    |   |
| 25 | Done         | ہو گیا | تھی گیا | 3 |
|    |              | ہو گیا | تھی گیا |   |
| 26 | Has happened | ہو گئے | تھی گئے | 3 |
|    |              | ہو گئے | تھی گئے |   |

The words which have distance of four characters are 17 out of 1000. They are illustrated by the following table.

List of words which have difference of 04 characters

| Sr.no | Words in English | Words in Punjabi | Words in Saraiki | Levenshtein distance |
|-------|------------------|------------------|------------------|----------------------|
| 1     | Landlord         | امراء            | سردار            | 4                    |
|       |                  | امراء            | سردار            |                      |
| 2     | Announcement     | بیان             | اعلان            | 4                    |
|       |                  | بیان             | اعلان            |                      |
| 3     | Full             | بھرنا            | بھریچنا          | 4                    |
|       |                  | بھرنا            | بھریچنا          |                      |

|    |           |       |        |   |
|----|-----------|-------|--------|---|
| 4  | Exterior  | بیرون | باہر   | 4 |
|    |           | بیرون | باہر   |   |
| 5  | Out       | باہر  | بارہوں | 4 |
|    |           | باہر  | بارہوں |   |
| 6  | Base      | بیس   | بنیاد  | 4 |
|    |           | بیس   | بنیاد  |   |
| 7  | Side      | پاسے  | پرہاں  | 4 |
|    |           | پاسے  | پرہاں  |   |
| 8  | Trader    | تاجر  | وپاری  | 4 |
|    |           | تاجر  | وپاری  |   |
| 9  | Animal    | جانور | مال    | 4 |
|    |           | جانور | مال    |   |
| 10 | Foreign   | فارن  | بارہوں | 4 |
|    |           | فارن  | بارہوں |   |
| 11 | Work      | کاری  | کم کاج | 4 |
|    |           | کاری  | کم کاج |   |
| 12 | Streaming | گامزن | جاری   | 4 |

|    |         |        |       |   |
|----|---------|--------|-------|---|
|    |         | گامزن  | جاری  |   |
| 13 | Planted | لگائے  | لائے  | 4 |
|    |         | لگائے  | لائے  |   |
| 14 | Animals | مویشی  | مال   | 4 |
|    |         | موشی   | مال   |   |
| 15 | Month   | ماہ    | مہینہ | 4 |
|    |         | ماہ    | مہینہ |   |
| 16 | Meeting | ملاقات | ملنا  | 4 |
|    |         | ملاقات | ملنا  |   |
| 17 | Done    | ہوئے   | تھیوے | 4 |
|    |         | ہوئے   | تھیوے |   |

The words which have distance of five characters are 02 out of 1000. They are illustrated by the following table.

List of words which have difference of 05 characters

| Sr.no | Words in English | Words in Punjabi | Words in Siraki | Levenshtein distance |
|-------|------------------|------------------|-----------------|----------------------|
| 1     | Revenue          | ریونیو           | پیسہ            | 5                    |
|       |                  | ریونیو           | پیسہ            |                      |
| 2     | Parents          | والدین           | ماں پُئو        | 5                    |
|       |                  | والدین           | ماں پُئو        |                      |

## V. DISCUSSION

One thousand most frequent words in Punjabi language were identified from the large corpus which was obtained in Punjabi language. There was a purpose to select the most frequent words in Punjabi language because they are used mostly by the native speakers in the daily routine of life and have more resistance to change. That is why they evolve slower than the others. Once the words list was established, their equivalents in Saraiki variety were recognized and their distance was calculated. Now, the data were categorized according to the different orthographical distance measures between the two words or strings. If the one pair of string has a difference of 1 character put separate and the other string has the difference of 3 characters put separately. The basic cause to do this is that one string may have the similar difference and the other string might have the more difference than the first one. This is just because of the length of every word. If the words have a number of characters in their formation, the difference between the two words will be less. The analysis of the strings according to the above mention categories revealed the following pairs.

Total number of pairs with orthographical difference

| Sr.no | Levenshtein distance | Number of pairs |
|-------|----------------------|-----------------|
| 1     | 0                    | 753             |
| 2     | 1                    | 18              |
| 3     | 2                    | 12              |
| 4     | 3                    | 26              |
| 5     | 4                    | 17              |
| 6     | 5                    | 02              |
| 7     | Completely different | 172             |
| 8     | Total                | 1000            |

The investigation of the words showed that 753 pairs have the same orthography or the same words in both languages; Punjabi and Saraiki out of 1000. Further, 18 pairs have one character difference between both languages. The twelve strings have the difference of two pairs and twenty-six pairs have the difference of three characters. 17 strings have four character difference and just two pairs have five character differences out of 1000 pairs of words. From one thousand pairs of words, 172 strings are completely orthographically different.

Total orthographical Similarity between Punjabi and Saraiki

| Language pairs                  | Punjabi and Saraiki |
|---------------------------------|---------------------|
| Total no of pairs               | 809                 |
| Total orthographical similarity | 80,9%               |

The analysis shows that Punjabi and Saraiki are 80.9% orthographically similar with one another.

Table4.10 Total orthographical difference between Punjabi and Saraiki

|                                 |                     |
|---------------------------------|---------------------|
| Language pairs                  | Punjabi and Saraiki |
| Total no of pairs               | 191                 |
| Total orthographical difference | 19.1%               |

The analysis shows that Punjabi and Saraiki are 19.1% orthographically different from one another.

## VI. CONCLUSION

To figure out the orthographical difference between Punjabi language and Saraiki variety which are genetically related languages a statistically defined algorithm was used which calculates the distance on making the strings as a basis. To obtain the orthographical distance between Punjabi and Saraiki, measurements and comparisons of orthographic segment connection were used. The strings is placed according to the degree of difference, since the level of difference rises; the level of diversity between the pairs often increases. The formula reveals the connection between the strings. The orthographical study which was performed on one thousand strings shows that Punjabi language and Saraiki variety have 80.9 percent orthographical similarity and 19.1 percent orthographical variation (hypothesis revealed true). With orthographical similarity of 80.9 percent , we can infer that the Punjabi and Saraiki are strongly linked to each other. From the point of view of second language acquisition, the significance of this study relies on the assumption that the resulting similarity list of words or shared vocabulary can be a valuable reference for students who are studying Punjabi or Saraiki as an L2. It is an effective help for the students whose L1 is either Saraiki or Punjabi and who want to study the other language as an L2

## VII. REFERENCES

1. Asif, S. I. (2005). Saraiki Language and Ethnic Identity. *Journal of Research*, 7, p. 9-17.
2. Brown, K. (2015). The influence of explicit instruction on failure to acquire a phonological rule due to orthographic input: the case of native English speakers learning German (published master's thesis). University of Utah.
3. Cutler, Anne, and Bowers, M. (2005). Phonetic precision in listening. A figure of speech: A festschrift for John Laver, ed. by W. J. Hardcastle and J. M. Beck, chap. 4, 63-91. Erlbaum
4. Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66, p. 789-806.
5. Goyal, V., & Lehal G. (2008) Comparative Study of Hindi and Punjabi Language Scripts. Research Gate.
6. Irene, T. (2015) Language family. Retrieved November 2, 2017, from <http://aboutworldlanguages.com/language-families>.
7. Kazim, A. U. (2015). *Jamaee Saraiki Qwaied*. Multan: Usman Publications.
8. Kessler, B. (1995). Computational dialectology in Irish Gaelic. *Proceedings of the European*, p. 60-66.
9. Levenshtein, V. I. (1996). Binary codes capable of correcting, deletion, insertions and reversals. *Cybernetics and Control Theory*, 10, p. 707-710.
10. Liudumalia V. khokhlova.(2014). Majority Language death. *Language Endangerment Preservation in South Asia*, ed. by Hugo C. Cardoso, pp. 19-45
11. Maldonado Gracia, M., & Borges De Souza, A. (2014). Lexical similarity level between English and Portuguese. *Graassroots*, 49(1), p. 203-18.
12. Maldonado Gracia. M. (2016). Saraiki: language or dialect? *Eurasian Journal of Humanities*, 1(2), p. 40-53

13. Oxford English Dictionary. (n.d.). Retrieved November 3, 2017, from <https://en.oxforddictionaries.com/definition/cognatef>
14. Pakistan Languages Ethnologue. (2017). Pakistani Languages. Retrieved September 21, 2017, from <https://www.ethnologue.com/country/PK/languages>
15. Paul, S., Christopher, S., Stephen, L. (2002) Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance* 28: p. 661–681
16. Rehman, T. (1995). *Language Planning and Politics in Pakistan*. The Sustainable Development Policy Institute.
17. Renfrew C., Tarim B., (2003) Indo-European origins: a view from the West. [Review of the book *The Bronze Age and Early Iron Age peoples of Eastern Central Asia*. By Karen S., 1998].
18. Renfrew, C., McMahon, A., & L. Trask, Eds. (2000). *Time Depth in Historical Linguistics*. Cambridge, England: The McDonald Institute for Archaeological Research
19. Rowe, B. M., Levine, D. P. (2015). [A Concise Introduction to Linguistics](https://books.google.com.pk/books?hl=en&lr=&id=ePQ5CgAAQBAJ&oi=fnd&pg=PP1&dq=+A+Concise+Introduction+to+Linguistics.&ots=DUKedgekQx&sig=q0pidfmYb-tyN7YLq2IU61dxxQs#v=onepage&q=A%20Concise%20Introduction%20to%20Linguistics.&f=false). Routledgepp. Retrieved 02, November 2017 from <https://books.google.com.pk/books?hl=en&lr=&id=ePQ5CgAAQBAJ&oi=fnd&pg=PP1&dq=+A+Concise+Introduction+to+Linguistics.&ots=DUKedgekQx&sig=q0pidfmYb-tyN7YLq2IU61dxxQs#v=onepage&q=A%20Concise%20Introduction%20to%20Linguistics.&f=false>
20. Shany, M., & Breznitz, Z. (2011). Rate and accuracy disabled subtype profiles among adults with dyslexia in the Hebrew orthography. *Developmental Neuropsychology*, 36 (7) p. 889–913.
21. The World Fact book - Central Intelligence Agency. (n.d.). Retrieved November 5, 2017, from <https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html> [ons/the-world-factbook/fields/2098.html](https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html)
22. Younas, M. (2020). English Language Teacher Education: A Sociocultural Perspective on Preservice Teachers' Learning in the Professional Experiences. *Journal of Teaching and Learning*, 14(2), 67-69. [HTTPS://DOI.ORG/10.22329/JTL.V14I2.6322](https://doi.org/10.22329/JTL.V14I2.6322)
23. Younas, M., & Noor, U. (2020). Inquiry-based learning-undergraduate research: the German multidisciplinary experience, *Educational Gerontology*, 46:12, 830-832 <https://doi.org/10.1080/03601277.2020.1828761>
24. Younas, M., & Noor, U. (2020). Teaching for Excellence and Equity. *Journal of International Students*, 10(4), 1114-1116. <https://doi.org/10.32674/jis.v10i4.2758>
25. Younas, M., Afzaal, M., Noor, U., Khalid, S., & Naqvi, S. (2020). Code Switching in ESL Teaching at University Level in Pakistan. *English Language Teaching*, 13(8), 63-73. <https://doi.org/10.5539/elt.v13n8p63>