

# **Lexical similarities between punjabi and siraiki**



**Submitted By**

Sohaib Khalid

M.Phil Applied Linguistics

ID # S2016084004

Submitted in partial fulfillment of the requirements for the degree of Master of Science in

Applied Linguistics

Supervisor: Dr. María Isabel Maldonado García

School of Social Sciences and Humanities

Department of English Language and Literature

University of Management and Technology, Lahore

2018

## **ABSTRACT**

The present study aims at finding the lexical similarities between Punjabi language and Siraiki. The study also determines at supporting those students who want to learn Punjabi language and Siraiki as a second language. It has been noted that there is a variety of studies which estimate the similarity level between two different languages and two dialects of the same language, but the similarity level between Siraiki and Punjabi has not been explored yet. The researcher collected one thousand high-frequency words (the latest corpus compiled by the researcher) in Punjabi and Siraiki and analyzed by using the Levenshtein's (1965) algorithm to explore the difference between them. The study is based on the hypothesis: lexical difference between both selected languages will be inferior to 75% due to the genetical relationship between both linguistic systems.

**Key words:** Punjabi language, Siraiki, lexical differences, cross-language similarities, lexical distance, language acquisition

## **CERTIFICATE OF APPROVAL**

Accepted by the Faculty of the Department of English Language and Literature, School of Social Sciences & Humanities, University of Management and Technology, Lahore in partial fulfillment of the requirements for the degree of M.Phil in Applied Linguistics

---

Supervisor

Dr. María Isabel Maldonado García

Prof. Dr.

---

External Examiner

---

Chairman

Dr Muhammad Shaban

Prof. Dr.

---

Dean, SSS&H

Date: \_\_\_\_\_

## **ACKNOWLEDGEMENTS**

I owe gratitude to Dr. María Isabel Maldonado Garcíafor being such a supportive supervisor. I am obliged for her support that assisted me throughout this research. She has always been very kind and generous with her ideas and time. She always encouraged me with her constructive remarks, valuable advice and kind assistance. It would not be an exaggeration to say that she is the one who taught me how to conduct research.

I would like to thank my friends Abdul Qayyum, Muhammad Younis and Zainab Faruki without their support I would have been unable to complete this study.

## **DEDICATION**

To my parents,

For their tireless efforts in assisting me while achieving my life goals.

# TABLE OF CONTENTS

ABSTRACT.....	ii
CHAPTER ONE	
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
INTRODUCTION .....	1
1.1	Introduction
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.1.1	Punjabi Language
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.1.2	Siraiki
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.1.3	Cognates or Loan Words
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.1.4	Lexical Similarity
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.2	Purpose of the Study
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.3	Goals and Contributions of the Study
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	
1.4	Significance of the Study
.....	<b>Er</b>
<b>ror! Bookmark not defined.</b>	

1.5	Limitations	of	the	Study	
.....					<b>Er</b>
<b>ror! Bookmark not defined.</b>					
1.6	Research Questions	.....			7
CHAPTER TWO .....					8
LITERATURE REVIEW .....					8
2.1	Relation between Linguistic Systems that Belong to the Same Family of Languages	.....			8
2.1.1	Language Family	.....			8
2.2	Lexical Difference among Different Languages	.....			13
2.2.1	Phonological and Orthographic Neighborhood Densities	.....			16
2.2.2	Phonetic Similarity	.....			17
2.2.3	Orthographic Similarity	.....			19
2.3	Punjabi, a Background	.....			20
2.4	Siraiki, a Background	.....			21
2.5	Intra Lingual Difference in Pakistan	.....			22
2.5.1	Language Patterns in Pakistan	.....			22
2.5.2	Punjabi and Ethnic Identities	.....			23
2.5.3	Siraiki and Ethnic Identity	.....			23
2.5.4	Ethno-Lingual Distinctiveness of Siraiki Identity and Rest of Punjab	.....			24
2.6	Mutual Intelligibility	.....			24
2.7	Levenshtein Distance	.....			25
CHAPTER					THREE
.....					<b>Er</b>
<b>ror! Bookmark not defined.</b>					
METHODOLOGY .....					27
3.1	Compiling the Database	.....			27
3.2	Finding the Cognates	.....			27
3.3	Measuring Lexical Distance	.....			28
Category					1:
.....					<b>Er</b>
<b>ror! Bookmark not defined.</b>					

Category 2:.....	30
Category 3:.....	30
CHAPTER FOUR.....	32
DATA ANALYSIS.....	32
4.1 Category 1.....	32
(a) Terms Which are Exact Cognates (distance 0 or 100% similarity).....	32
(b) Terms Which have Only One Character or Distance 1.....	76
(c)Terms Which have Difference of 02 Characters.....	78
4.2 Category 2.....	79
(d)Terms Which have Difference of 03 Characters:.....	79
Distance 03:.....	79
Distance 4:.....	81
Distance 5:.....	82
Distance 6:.....	83
Distance 7:.....	83
Distance 8:.....	83
Distance 9:.....	83
Distance10:.....	83
4.3 Category 3.....	83
(e) Terms Which are Completely Different.....	83
4.4 Discussion.....	94
CHAPTER FIVE.....	96
CONCLUSION.....	96
REFERENCES.....	97

## LIST OF TABLES

Table 2.1 Levenshtein distance between the strings ‘Lend’ and ‘Lens’.....	26
Table 4.1 Terms which are exact cognates.....	32
Table 4.2 List of words which have one-character distance.....	77

Table 4.3 List of words which have difference of 02 characters.....	78
Table 4.4 List of words which have difference of 03 characters .....	79
Table 4.5 List of words which have difference of 04 characters.....	81
Table 4.6 List of words which have difference of 05 characters.....	83
Table 4.7 Terms which are completely different.....	83
Table 4.8 Total number of pairs with lexical difference.....	94
Table 4.9 Total lexical Similarity between Punjabi and Siraiki.....	95
Table 4.10 Total lexical difference between Punjabi and Siraiki.....	95

### **LIST OF FIGURES**

Figure 3.1 Leveshtein Algorithm.....	28
Figure 3.2 Research Design.....	31

## **INTRODUCTION**

According to Ethnologue (2017), 73 languages are spoken in Pakistan from which Urdu is the national language. According to The CIA World Fact book, there are six main languages which are spoken in Pakistan with fifty-nine smaller languages (2017). Central Intelligence Agency declares that the major languages in Pakistan are: Punjabi with 48% of the total speakers, Sindhi with 12% of the total speakers, Siraiiki is the third biggest language of Pakistan with 10% of speakers, Pashto and Urdu (native) speakers are 8% in Pakistan and Balochi speakers are 3% of the total population of the country (2017).