

## RESEARCH ARTICLE



# Morphology-Assisted Sindhi Text Analysis for Natural Language Processing Applications

 OPEN ACCESS

Received: 10-07-2023

Accepted: 02-08-2023

Published: 25-09-2023

**Citation:** Sodhar IN, Sulaiman S, Buller AH (2023) Morphology-Assisted Sindhi Text Analysis for Natural Language Processing Applications. Indian Journal of Science and Technology 16(35): 2894-2901. <https://doi.org/10.17485/IJST/v16i35.1719>

\* **Corresponding author.**

[iram10akber@gmail.com](mailto:iram10akber@gmail.com)

**Funding:** None

**Competing Interests:** None

**Copyright:** © 2023 Sodhar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.iseeindia.org/))

**ISSN**

Print: 0974-6846

Electronic: 0974-5645

**Irum Naz Sodhar**<sup>1\*</sup>, **Suriani Sulaiman**<sup>2</sup>, **Abdul Hafeez Buller**<sup>3</sup>

**1** Post-Doctoral Fellow, Department of Computer Science, Kulliyah (Faculty) of Information and Communication Technology, International Islamic University, Malaysia

**2** Assistant Professor, Department of Computer Science, Kulliyah (Faculty) of Information and Communication Technology, International Islamic University, Malaysia

**3** Post-Doctoral Fellow, Department of Civil Engineering, Kulliyah (Faculty) of Engineering, International Islamic University, Malaysia

## Abstract

**Objectives:** Understanding word construction and internal structure, especially in the Sindhi language, requires knowledge of the linguistic field known as morphology. In this study, Sindhi morphology is examined with particular attention paid to its structure, function, nature, word categories, and writing system. Natural Language Processing (NLP) relies on morphological analysis to identify words and their grammatical features, enabling applications like spell checkers and machine translation. A comparative analysis is done to comprehend how Sindhi Morphology developed. Because research on morphology analysis lack proper classification and cover both modern and conventional methodologies, Sindhi morphology variances present difficulties for computerization. **Methods:** Morphological analysis is crucial in Natural Language Processing (NLP) domains like spell checkers and gadget translation, studying word production and phrase shape using morphemes, the smallest grammatical elements in a language. Morphemes are the building blocks of words and are divided into free and fixed morphemes. **Findings:** Sindhi's rich morphology and complexity enable borrowing and lending of words, but ambiguity is high due to similar patterns and vowel deletions. Morphological analysis influences semantic and syntactic analysis. Computerization is challenging due to prefixes, suffixes, and stem positions. Primary and secondary words can be subdivided into compound and complicated terms. The language uses initial, middle, and end writing styles. **Novelty:** This research aims to develop an automatic Sindhi morphological analyzer for future NLP applications, ensuring compatibility with existing Information Technology world applications. It will help understand Sindhi word structure and benefit software developers in developing Sindhi natural language and speech processing applications.

**Keywords:** Sindhi Morphology; Morphological Analysis; NLP; Communication and information sharing; machine Learning

## 1 Introduction

Morphology in Semitic languages hinders Sindhi natural language processing (SNLP). Sindhi words consist of roots derived from stem patterns and affixes, forming various stems. Studying and categorizing methods used in Sindhi morphological analysis is crucial for better understanding and effective morphological methodology. Sindhi is an Indo-Aryan language with similarities to Hindi, Urdu<sup>(1)</sup>, and other languages spoken in the northwest Indian subcontinent. Perso-Arabic<sup>(2)</sup> script is used in South Asian nations like Pakistan and India, with Devanagri script also used in some regions<sup>(2,3)</sup>.

Pakistan’s National Identity Card (NIC) is issued in Sindhi and Urdu, with Urdu being the third most widely spoken language. It is the top language in South Asia because of its phonology, which is similar to many other Indo-Aryan languages. The primary difficulty in computerizing Sindhi is the language’s extensive morphological variety, which is increased by the concatenation of complicated words, especially in morphological analysis<sup>(3)</sup>.

This paper explores Sindhi morphological analysis methodologies to improve NLP applications and ensure interoperability with diverse IT applications<sup>(4)</sup>. The long-term objective is to create an automatic Sindhi morphological analyzer, as few studies accurately classify contemporary and historical methodologies. Morphological analysis is crucial for processing Sindhi natural language<sup>(5,6)</sup>.

## 2 Proposed Approach

Morphological analysis and generation are crucial components of Natural Language Processing (NLP)<sup>(7)</sup> domains, such as spell checkers and gadget translation. Morphological analysis studies the shape of phrases and identifies how a word is produced using morphemes, which are the smallest elements of a phrase with grammatical features. Morphological analysis is based on the procedure of morphology, text (Sindhi), morphemes, analysis of the alphabet, and construction of words. Morphology is crucial for NLP tools like spell-checkers, POS tagging, parsing, data retrieval, and device translation.

Sindhi, an Indo-Aryan language with an inflectional morphology, is a top language in South Asia due to its phonology being similar to many other Indo-Aryan languages<sup>(3)</sup>. Morphological analysis is critical for automatic morphological systems to separate stems from prefixes and phase sentences into token morphemes as shown in Figure 1. Applications of NLP that use large terminology sets, such as speech recognition, dictionary automation, machine translation, and text retrieval, can benefit from morphological analysis. Morphemes are the smallest grammatical issues<sup>(8)</sup> in a language, and they are the building blocks of words. Morphology is divided into free, fixed morphemes and Null morpheme.

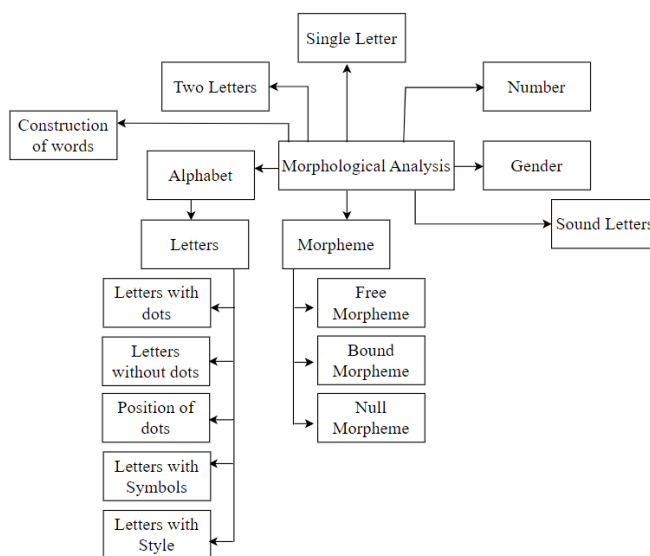


Fig 1. Sindhi Text Morphology Analysis

### 3 Sindhi Morphology

Sindhi is a complex language with a rich morphology that allows for word borrowing and lending<sup>(9,10)</sup>. It has a high rate of ambiguity due to similar patterns and vowel deletions. Morphological analysis has an impact on higher levels of semantic<sup>(11)</sup> and syntactic analysis. Although they are recognized as part of great languages, the variety of affixes in Sindhi makes it more complex. The significant variety in Sindhi’s morphology caused by different prefix, suffix, and stem placements in words makes it difficult to computerize<sup>(10)</sup>.

The extensive use of inflections and derivations utilizing prefixes and suffixes reveals its rich morphology. The two kinds of words in Sindhi are primary and secondary, which can be further broken down into compound and complex terms<sup>(12)</sup>. Complex words combine the prefix, suffix, and stem, whereas compound words are combinations of two words or free morphemes.

#### 3.1 Free morpheme

Examples of separate morphemes are the words ”profitable”, ”overcrowd” and ”uplift”. The words ”in a position”, ”over”, and ”up” are distinct morphemes that can stand alone as phrases and also have strong meanings in their free forms<sup>(2)</sup>, in Table 1, a few examples of free morphemes are discussed.

Table 1. Example of free morpheme

Word	Roman Sindhi	English
گلاب	Gulab	Rose
ڏينهن رات	Deenh raat	Day Night

#### 3.2 Bound morpheme

Sindhi has a rich morphology because, like many other Indo-Aryan languages<sup>(13)</sup>, it uses suffixes for derivational and inflectional morphemes. The branch of morphology known as morpheme studies morphemes, the smallest meaningful grammatical unit. Bound morphemes are unable to stand on their own and can only have significance when they are a part of a word<sup>(11)</sup>. Combining different morphemes, such as particular and uplifting morphemes, can give a phrase meaning. A morpheme is a brief piece of meaning-filled grammar.

The English terms ”worse” and word replacement are distinguished in the phonological representation of morphemes. Morphemes, the lowest meaningful grammatical unit, are impermeable. In Table 2, a few examples of Bound morphemes are discussed.

Table 2. Examples of Bound morpheme

Word	Root word	Suffix/prefix	English
پڙهڻ	پڙهه	ڻ	study

#### 3.3 Null morpheme

A null morpheme is an invisible affix that can be converted into verbs in languages with distinctions in phonological content. It is common in languages with singular numbers, present tense, and third persons, as well as in the least-marked cases like the nominative in nominative-accusative languages and the absolutive in ergative-absolutive languages. Null morphology, which is disputed in morphological theory, describes morphological units without any phonological content. It is not simple to define what constitutes a null morpheme since it requires careful distinction from other circumstances in which there is no obvious morphological indication. The definition of a null value may also vary according to the particular theories under which it is made.

Null morphology, in the strict sense, refers to a complete morpho-syntactic representation connected to null phonological content. Other definitions of null morphology, such as the complete lack of morphological expression and certain theory-internal interpretations of what qualifies as null, differ from this. Null morphology is the subject of a complicated and nuanced argument, with proponents and opponents of it making claims about how it affects language and linguistics.

In Sindhi, homophonous words, such as sheep, ’fish’, and ’deer’, remain identical in both plural and singular paperwork. These words are also referred to as null morphemes, as their phonological illustrations are null in each style. In Sindhi, no such styles

are determined, but some homographic words do alternate their shape for changing into past or plural, but do so by changing their diacritics with the same set of letters<sup>(2)</sup>. In Table 3, a few examples of Null morphemes are discussed.

**Table 3.** Examples of Null morpheme

Singular	Plural	English
ڪَ + مَ	ڪَ مَ	Work + Works
ڏَ + ڏَ	ڏَ ڏَ	Door + Doors
ڪِتابَ	ڪِتابَ	Book+Books

### 3.4 Gender and Number

Sindhi nouns have masculine and feminine genders. Similar to Urdu, this gender category includes both animate and inanimate nouns. Gender is artificially determined for non-animate nouns that don't have genders that are clearly indicated. The Sindhi language has its own set of numerals, including those shown in the Table 4.

**Table 4.** Numerical Numbers

Sindhi Number	Sindhi Pronunciation	English Number
۱	هڪَ	1
۲	ٻه	2
۳	ٽي	3

### 3.5 Alphabet and Letters

In Sindhi, the alphabet of the Sindhi language contains a total of 52 Letters<sup>(14)</sup> with different shapes, placements of dots, symbols, and numbers<sup>(14)</sup>. This language has its own writing and is also morphologically rich. This language is a right-handed language, just like Urdu<sup>(2)</sup> and Arabic<sup>(2)</sup>. Letters in Sindhi are of two types, such as Letters with dots and Letters without dots<sup>(15)</sup>.

#### 3.5.1 Letters with Dots

Thirty three of the letters have dots these letters have different styles with different placement of dots. Started nine letters have same shape (ب ب پ ت ت ٺ ٺ پ) it means writing style-I, after that Style-II six letters have same shape (ج ج ڄ ڄ خ), Style -III have five letters (ڏ ڏ ڍ ڍ ڻ) Style-IV have two letters (ڙ ڙ) Style-V have one letter, (ش) Style-VI have one letter, (ض) Style-VII have one letter (ظ) Style-VIII have one letter(غ), Style-IX have two letters (ف ڦ), Style-X have one letter Style-XI,(ق) have two letters (گ ڳ), Style-XII have one letter, (ن), Style-XIII have one letter ,(ي) style-XIV have one letter (جھ) are shown in Table 5.

#### 3.5.2 Letters without Dots

Seventeen of the letters have without dots these all letters have different styles such as: style-I have one letter ,(ل) style-II have one letter ,(ر) style-III have one letter ,(ح) style-IV have one letter ,(د) style-V have one letter ,(ء) style-VI have one letter ,(ء) style-VII have one letter ,(و) style-VIII have one letter ,(م) style-IX have one letter ,(ڄ) style-X have one letter ,(ڪ) style-XI have one letter ,(گ) style-XII have one letter ,(ڪ) style-XIII have one letter ,(ع) style-XIV have one letter ,(ط) style-XV have one letter ,(ص) style-XVI have one letter ,(س) style-XVII have one letter (ڳھ) with no any dots<sup>(13)</sup>, in Table 5 of Letters with and without dots are discussed.

#### 3.5.3 Positions of Dots

In this language have major three positions of dots in letter, upper, middle and down with different number of count dots<sup>(13)</sup>. These letters contain one, two three, four dots letters with X-axis, Y- axis directions and detail description shown in Table 5.

#### 3.5.4 Letters with Symbol

In this language also have used symbols in letters. Basically the letter symbols are, (ء, ٺ) and other are zabar, zair, peeshu (ـَ, ـِ, ـُ) used in letters to pronounce properly otherwise did not differentiate the proper meaning of words<sup>(13,14)</sup>.

**Table 5.** Letters With and Without Dots and Positions of Dots

Letters With and Without Dots			
Letters with dots		Letters without dots	
دءوہم ل ک گ ک ع ط ص س گھ رح ا		ب پ پ ت ٹ ٹ ٹ پ ج چ چ چ خ ذ ڈ ڍ ڏ ڙ ش ض ظ غ ف ق گ گ ن ي جي	
Positions of Dots	Dots position with letters	Directions of Dots	
No: Dots	Upper	خ ض ظ غ ف ن ز ذ	X-axis
One dot Letters	Middle	ج	X-axis
	Down	ب ڍ	X-axis
Two Dots Letter	Upper	ت گ ق ڏ	X-axis
	Middle	چ	Y-axis
	Down	پ ي گ ڍ	Y-axis
Three Dots Letter	Upper	ٹ ش ڙ	X-axis + Y-axis
	Middle	چ	X-axis + Y-axis
Four Dots Letter	Upper	ڙ ڙ	X-axis + Y-axis

### 3.6 Letter Styles

Basically the styles of letters are three forms initial, middle and end in this language as shown in below Table 6.

**Table 6.** Example of Letter Styles

Whole Letter	Initial Style	Middle Style	End Style
ا	ا	No any	ا
ب	ب	ب	ب
ت	ت	ت	ت
ج	ج	ج	ج
د	د	د	د

#### 3.6.1 Single letters in Alphabet

In alphabet contains (fifty-50) single letters with no any combination of another letter such as: (خ چ چ چ چ چ خ) Those letters have own shape and with and without dots<sup>(13,14)</sup>.

#### 3.6.2 Combination of two letters in Alphabet

There are two letters were found combination of two different letters. One letter is jaha (جھ) used in Sindhi alphabet basically this letter is combination of jeem and ha (ج+ه). Second letter is gaha (گھ) is also combination of ga and ha (گ+ه). In both letters have one letter is common that is ha (ه).

## 4 Sound of Sindhi letters with English letters

Sindhi letters have a similar initial sound but a variant of an English alphabetic order. According to sound, 19 letters of English are shown in Table 7. Missing letters of English are (e,i,o,q,u,v,x) in the initial sound but are also used in the middle of words when Sindhi roman is used in daily life for sending messages from person to person, uploading photos on social media, and so on<sup>(11)</sup>.

In the above-mentioned Table 7 , there are 19 rows that contain all letters of the Sindhi alphabet, with the initial sound of English letters and the complete sound of Sindhi letters. No. of rows described as: R1 contains two letters, alif and aeen, both of which have a similar initial sound, "a." R2 contain three letters Ba, be, bha so initial sound "b". R3 contains five letters: te, ta, tuwe, th, and tha, so the initial sound is "t." R4 contains four letters: saeh, seen, sheen, and suwad, with the initial sound "s."

**Table 7.** Sound of Sindhi letters with English letters

No. of Rows	Sindhi Sound			Initial Sound Letter	Complete Sound of Sindhi Letters
R1	ا		ع	a	alif, aeen
R2	ب	ب	پ	b	Ba, be, bha
R3	ت	ت	ط	t	te, ta, tuwe, th, tha
R4	ث	س	ش	s	saeh, seen, sheen, suwad
R5	پ			p	pa
R6	ج	چ	ڄ	j	jeem, jah , jaha, ja
R7	چ			c	ch, ch
R8	ح		ه	h	ha, he, hamza
R9	خ	ق	ڪ	k	kha, kaaf, kaf, khe
R10	د	ڌ	ڏ	d	Dal, dha, dah, da, daa
R11	ذ	ز		z	zal, ze, zuwad, zuwe
R12	ر	ڙ		r	Re, ree
R13	غ	گ	ڳ	g	gheen, ghaf, ghe, gha, gaen
R14	ف		ڦ	f	fe, fa
R15	ل			l	laam
R16	م			m	meem
R17	ن		ڻ	n	noon, naen
R18	و			w	wao
R19	ي			y	yeh

R5 contains one letter, pa, so the initial sound is "p". R6 contains four letters: jeem, jah, jaha, and ja, so the initial sound is "j". R7 contains two letters, ch, so the initial sound is "c". R8 contains three letters: ha, he, and hamza, so the initial sound is "h". R9 contains four letters: kha, kaaf, kaf, and khe, so the initial sound is "k". R10 contains five letters: dal, dha, dah, da, and daa, so the initial sound is "d". R11 contains four letters: zal, ze, zuwad, and zuwe, so the initial sound is "z". R12 contain two letters re, ree so initial sound "r". R13 contains five letters: gheen, ghaf, ghe, gha, and gaen, so the initial sound is "g". R14 contain two letters fe, fa so initial sound "f". R15 contains one letter, laam, so the initial sound is "l". R16 contains one letter, meem, so the initial sound is "m". R17 contains two letters, noon and naen, so the initial sound is "n". R18 contains one letter, wao, so the initial sound is "w". R19 contains one letter, yeh, so the initial sound is "y". R5, R15, R16, R18, and R19 contain one letter. R1, R7, R12, and R14 contain two letters. R2 and R8 contain three letters. R4, R6, R9, and R11 contain four letters. R3, R10, and R13 contain six letters.

In Sindhi, words were constructed through three basic shapes or styles of writing (Initial, Middle, and end). Initial style is used when words start, middle style is used in the middle of words, and end style is used at the end of the appropriate words. The Construction of words through combinations of one two, three, four, five, and six letters shown in Table 8.

Every language needs improvements to survive, and the present era has introduced additional necessities including electronic developments, contemporary communication standards, virtualization, and computer compatibility. These languages need to be fed and saved into the computer's virtual world, where software programs and linguistic resources will help them survive. The non-concatenative structure of Sindhi, like that of the majority of Arabic script-based languages, makes computational morphological processing difficult. Due to vowel deletion, similarity of affixes, and the use of root pattern schemes to create words, Sindhi contains a significant level of ambiguity. Higher levels, such as syntactical and semantic analyses, are impacted by this uncertainty, as well as morphological analysis.

**Table 8.** Examples of words through combination of one two, three, four, five, six letters

<b>Example 1.</b> combination one letters				
English Word	Sindhi Word	Initial Letter	Middle Letter	End Letter
Duck	بدڪ	ب	دَ	ڪَ
Monkey	پولڙو	پ	ولَ	ڙو
Sparrow	جهرڪي	جھ	ر	ڪي
Star	تارو	تَ	ا	رو
String	يڪتارو	يَ	ڪ ت ا	رو
<b>Example 2.</b> combination two letters				
English word	Sindhi word	L1 +L2		
Flower	گل	گ + ل		
Judge	جج	ج + ج		
Sun	سج	س + ج		
<b>Example 3.</b> combination three letters				
English word	Sindhi word	L1 +L2+L3		
Gratitude	شڪر	ر ڪ + ش		
Reason	سبب	ب + ب + س		
City	شهر	ر + ه + ش		
<b>Example 4.</b> combination four letters				
English word	Sindhi word	L1 +L2+L3+L4		
sugarcane	ڪمند	د + ن + م + ڪ		
rise	اڀرڻ	ا + پ + ر + ا		
bad	خراب	ب + ا + ر + خ		
<b>Example 5.</b> combination five letters				
English word	Sindhi word	L1 +L2+L3+L4+L5		
Karachi	ڪراچي	ڪ + ر + ا + چ + ي		
Ladyfinger	پينڊي	پ + ي + ن + ي + پ		
tomato	ٽماٽو	ٽ + ا + م + ا + ٽ		
<b>Example 6.</b> combination six letters				
English word	Sindhi word	L1+L2+L3+L4+L5+L6		
Door	دروازو	د + و + ز + ا + و + ر + د		
Disliked	ناپسند	د + ن + س + پ + ا + ن		
Best	بهترين	ب + ه + ت + ر + ي + ن		

The future survival of a language with a strong literary, historical, and cultural foundation, as well as its own alphabet, grammar and other linguistic features has not received enough attention from native speakers. This work introduces the Sindhi Morphological Analyzer, which intends to improve the morphology and analysis of Sindhi. Typically, a morphological analyzer is the major instrument used for introductory computational operations.

## 5 Conclusion

The grammar and structure of Sindhi, a complex and distinctive language made up of basic, secondary, and compound words, are distinctive. Due to its distinctive morphology, internal structure, and interactions between prefixes, suffixes, and word kinds, it presents difficulties for artificial natural language processing (NLP) applications. Both derivational and inflectional morphological constructions are possible in Sindhi. By ensuring Sindhi's compatibility with all NLP application types not yet created in the world of information technology, the study intends to enhance the language's morphology and analysis. The smallest grammatical components in a phrase, known as morphemes, are examined in morphological analysis to determine how words are formed.

The beginning, middle, and end of the Sindhi alphabet each have their own distinctive shape, thus Sindhi has a distinctive grammar and writing style. Initial, middle, and end are the three primary word-building patterns in Sindhi. Native speakers haven't given the future survival of a language with a solid literary, historical, and cultural foundation enough thought. The Sindhi Morphological Analyzer, a tool commonly used for initial computational processes, attempts to enhance the morphology and analysis of Sindhi.

## Acknowledgement

This research was supported by the International Islamic University Malaysia to provide a platform for the Post-Doctoral research. We are thankful to our colleague, Dr. Suriani Sulaiman, who provided expertise that greatly assisted the research. This research is not funded by any institution, agency, or company; it is self-funded by the corresponding author, and there is no conflict of interest.

## References

- 1) Khattak A, Asghar MZ, Saeed A, Hameed IA, Hassan SA, Ahmad S. A survey on sentiment analysis in Urdu: A resource-poor language. *Egyptian Informatics Journal*. 2021;22(1):53–74. Available from: <https://doi.org/10.1016/j.eij.2020.04.003>.
- 2) Allothman A, Alsaman A. Arabic Morphological Analysis Techniques. *International Journal of Advanced Computer Science and Applications*. 2020;11(2). Available from: <https://doi.org/10.14569/IJACSA.2020.0110229>.
- 3) Nathani B, Joshi N. Part of Speech Tagging for a Resource Poor Language: Sindhi in Devanagari Script using HMM and CRF. *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*. 2021;p. 611–618. Available from: <https://aclanthology.org/2021.icon-main.75/>.
- 4) Faizullah S, Ayub MS, Hussain S, Khan MA. A Survey of OCR in Arabic Language: Applications, Techniques, and Challenges. *Applied Sciences*. 2023;13(7):4584. Available from: <https://doi.org/10.3390/app13074584>.
- 5) Ali Z, Khan SU, Khanda G. A Morphosemantic Analysis of An Agent Theta Role in English and Sindhi. *Southern Journal of Arts & Humanities*. 2023;1(01):14–25. Available from: <https://sjah.isp.edu.pk/index.php/sjah123/article/view/2>.
- 6) Amin M, Ali Z. Phonological and Morphological Variations between Lasi and Standard Sindhi. *Journal of Humanities and Social Sciences Research*. 2021;3(2):181–194. Available from: <https://doi.org/10.37534/bp.jhssr.2021.v3.n2.id1109.p181>.
- 7) Rajan A, Salgaonkar A. Survey of NLP Resources in Low-Resource Languages Nepali, Sindhi and Konkani. In: *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Springer Singapore. 2022;p. 121–132. Available from: [https://doi.org/10.1007/978-981-16-0739-4\\_12](https://doi.org/10.1007/978-981-16-0739-4_12).
- 8) Sodhar IN, Hussain A, Ibrahim M, Nawaz D. Identification of Issues and Challenges in Romanized Sindhi Text. *International Journal of Advanced Computer Science and Applications*. 2019;10(9):10. Available from: <https://doi.org/10.14569/IJACSA.2019.0100929>.
- 9) Sodhar IN, Jalbani AH, Channa MI, Hakro DN. Romanized Sindhi Rules for Text Communication. *Mehran University Research Journal of Engineering & Technology*. 2021;40(2):298–304. Available from: <https://doi.org/10.22581/muet1982.2102.04>.
- 10) Sodhar IN, Jalbani AH, Channa MI, Hakro DN. Parts of speech tagging of Romanized Sindhi text by applying rule based model. *International Journal of Computer Science and Network Security*. 2019;19(11):91. Available from: [http://paper.ijcsns.org/07\\_book/201911/20191113.pdf](http://paper.ijcsns.org/07_book/201911/20191113.pdf).
- 11) Amur ZH, Hooi Y, Sodhar IN, Bhanbhro H, Dahri K. State-of-the Art: Short Text Semantic Similarity (STSS) Techniques in Question Answering Systems (QAS). In: *International Conference on Artificial Intelligence for Smart Community*. Springer Nature Singapore. 2022;p. 1033–1044. Available from: [https://doi.org/10.1007/978-981-16-2183-3\\_98](https://doi.org/10.1007/978-981-16-2183-3_98).
- 12) Sodhar IN, Sulaiman S, Buller AH, Sodhar AN. Hybrid Approach Used to Analyze the Sentiments of Romanized Text (Sindhi). *International Journal of Advanced Computer Science and Applications*. 2023;14(3). Available from: <https://doi.org/10.14569/IJACSA.2023.0140362>.
- 13) Sodhar IN, Jalbani AH, Buller AH, Channa MI, Hakro DN. Sentiment analysis of Romanized Sindhi text. *Journal of Intelligent & Fuzzy Systems*. 2020;38(5):5877–5883. Available from: <https://doi.org/10.3233/JIFS-179675>.
- 14) Sodhar IN, Sulaiman S, Buller AH. Exploration of Sindhi Corpus Through Statistical Analysis on the Basis of Reality. *Indian Journal Of Science And Technology*. 2023;16(12):924–931. Available from: <https://doi.org/10.17485/IJST/v16i12.236>.
- 15) Talpur N, Talpur MJ, Samar T. Researching on Analysis and creating Corpus from Primary level Sindhi language Book for Sindhi. *Repertus: Journal of Linguistics, Language Planning and Policy*. 2023;p. 37–48. Available from: <https://rjllp.muet.edu.pk/index.php/repertus/article/view/24>.