

# MORPHOLOGICAL ANALYSIS OF NOUN PATTERNS IN SARAIKI: A CORPUS-BASED STUDY

Zubair Majeed<sup>\*1</sup>, Luqman Manzoor<sup>2</sup>, Hafiz Muhammad Qasim<sup>3</sup>

<sup>\*1-2</sup>M.Phil. Scholar, Department of Applied Linguistics, Government College University Faisalabad, Pakistan.

<sup>3</sup>Assistant Professor, Department of Applied Linguistics, Government College University Faisalabad, Pakistan.

<sup>1</sup>majeedzubair1056@gmail.com, <sup>2</sup>luqmanmanzoor875@gmail.com, <sup>3</sup>muhammadqasim@gcuf.edu.pk

Corresponding Author: \*

Luqman Manzoor

DOI: <https://doi.org/10.5281/zenodo.20961015>

Received	Accepted	Published
25 April 2026	04 June 2026	21 June 2026

## ABSTRACT

This descriptive study investigated and categorized the morphological patterns of noun formation in Saraiki, an under-researched language in the Indo-Aryan family spoken in Pakistan. To fulfil these goals, a vast corpus of two million words was created. The data that was not machine-readable was turned into editable documents using OCR. The text was evaluated using the AntConc 3.5.9.0 procedure, which entails extracting words that end in specified morphological patterns and analyzing them for their syntactic role as nouns. The productivity and frequency of the patterns were used to classify the results. The results demonstrated that patterns such as **س**, **ق**, and **غ** were much more productive in creating nouns than other patterns. These patterns were important in Saraiki's grammatical structure, and their production was higher in formal and academic settings. In contrast, patterns like **ئی** exhibited reduced productivity, primarily functioning as adjectives or adverbs. This **ے** pattern advances our knowledge of Saraiki grammar and offers insightful information on the morphology of a language that has received little attention in linguistic research. According to this research, in addition to basic nouns, the Saraiki language also has derived nouns, a small number of compound nouns with unique regional vocabulary, and words borrowed from Urdu and English. The results also have applications in computational linguistics and the creation of Natural Language Processing (NLP) tools for Saraiki. Future studies should broaden the corpus and delve more into the grammatical functions of these noun-forming patterns.

**Keywords:** Saraiki, computational linguistics, morphological patterns, noun formation, and corpus analysis.

## Introduction

### 1.1 Background of the study

Language is not just used for communication purposes, but it is also used for making connections and creating identity. It is language that makes it possible for us to communicate, share ideas, and develop social bonds. In this universe, thousands of languages are spoken, with a few languages that are dominant and act as international languages. Pakistan is a multilingual nation with over 74 languages, just like other nations. (Rahman, 2006). Other languages are widely spoken in Pakistan, but Urdu is the national language. Although Urdu

is Pakistan's national language, English is nevertheless an official language and used as a medium of instruction in educational institutions across the country. (Rahman, 2006; Mahboob, 2009). Although Punjabi, Pashto, Sindhi, Saraiki, and Balochi are all extensively spoken in Pakistan, Urdu is the national language and English is the official language. (Renschler & Kleiner, 2013).

This study uses a comprehensive corpus-based analysis to look at the morphological patterns of nouns in the Saraiki language. The study examines the formation of nouns by methodically examining compound structures,

derivative forms, and simple noun formulations using a two-million-word Saraiki corpus. Morphology, the study of word development and its internal structure, is a crucial component of linguistics. (Aronoff & Fudeman, 2022). Morphology is concerned with the internal structure of words, their relationships with other words in the same language, and how meaning is expressed in word structure. (Katamba & Stonham, 2006). This branch of linguistics is important for examining the grammatical structures of languages, particularly those like Saraiki that have strong cultural roots but have received little investigation. (Shackle, 1976; Rahman, 1996). A thorough examination of Saraiki's morphological features, particularly the noun patterns that it possesses, may yield useful information on its structure and function in Linguistics and other related disciplines. Thus, this study will use a corpus-based approach to evaluate the morphological trends of Saraiki nouns.

### 1.2 Statement of the problem

Although Saraiki is not a dying language, printed Saraiki remains underrepresented in digital technology. In the technology sector, 'Morphological analyzer,' 'VerbNet,' 'Sense relations,' and 'Semantic tagger' are unavailable for the Saraiki language.

Despite having many speakers and significant cultural significance, Saraiki is still less understood, especially in terms of noun morphology. Scholars have focused on its phonological, syntactic, and sociolinguistic aspects, but little emphasis has been made to how nouns are structurally formed from a morphological perspective. Previous studies on the Saraiki language have not given importance to grammar. Therefore, Saraiki grammatical information is not available on Google platforms.

For doing analysis in the English language, special software like the semantic tagger, VerbNet and Complete lexical tutor are available, but for analysis of the Saraiki language, these tools do not work. Moreover, as Saraiki is not a documented language, it was very difficult to collect the Saraiki corpus. From our study, researchers can get help in making Saraiki dictionaries, Saraiki grammar books and other Saraiki analysing softwares.

### 1.3 Objective of Research

This research aims to:

1. Investigate and analyze the morphological patterns of nouns in the Saraiki language.
2. Examine the prevalence of specific word patterns in the Saraiki corpus.
3. Recognize and analyze the structure of simple, derived, and compound nouns in Saraiki.

### 1.4 Research Questions

1. What morphological patterns regulate the Saraiki language's noun formation?
2. How frequently do particular noun patterns appear in the Saraiki corpus?
3. What is the structure of Saraiki's simple, derived, and compound nouns?

### 1.5 Significance of the study

This work fills a gap by providing the frequency of Saraiki nouns along with their structures of simple, derived and compound nouns. The findings of this work establish the framework for developing crucial language resources, including dictionaries, grammars, and computational tools that are needed for scholarly research, education, and technology, all of which are essential to development in both the humanities and innovation.

## Literature Review

### 2.1 Overview of Saraiki Language

Saraiki is an Indo-Aryan language spoken by around 20 million people, primarily in southern Punjab, Pakistan. The Indian subcontinent is home to several significant language families, such as Indo-Aryan, Indo-Iranian, Dravidian, Munda, and Tibeto-Burman (Jain and Cardona, 2007; Beames, 1875). Saraiki belongs to the Indo-Iranian language family, which also includes Urdu, Hindi, and Punjabi, all of which are widely spoken throughout South Asia. An estimated 78.7% of people in the subcontinent speak languages belonging to the Indo-Aryan branch. (Berton, 1999). According to the Pakistan Bureau of Statistics (2022), around 25.4 million people spoke Saraiki in various parts of Pakistan. It includes the Punjab province's southwest as well as the nearby districts of Sindh, Balochistan, and the KPK. (Shackle, 2003).

Saraiki is crucial to cultural identity and communication, especially in southern Punjab.

Supporters of the Saraiki language movement contend that Saraiki is a different language with distinctive linguistic characteristics, despite the fact that it shares lexicon, grammar, and sound with Punjabi. (Hussain & Khan, 2016). Its phonological, morphological, and syntactic structures all exhibit these distinguishing characteristics. Saraiki is a major linguistic community in many southern Punjabi cities, especially in Multan, Bahawalpur, and Dera Ghazi Khan.

According to Atta (2019), based on the research conducted by Shackle (1976), there are six dialects of the Saraiki language that include Southern Saraiki, Northern Saraiki, Sindhi Saraiki, Jhangi Saraiki, Shahpuri Saraiki, and Central Saraiki.

Saraiki is among the top 100 languages according to the percentage of native speakers worldwide, making it the 60th most spoken language by native speakers. (Gillani, 2013). Despite its enormous number of speakers, Saraiki is substantially underrepresented in linguistic studies and digital content. Saraiki has not gotten the same level of scholarly and technological attention as Urdu and Punjabi, which has led to little progress in several linguistic studies and digital technology domains.

## 2.2 Relevant past studies

In order to create a supporting lexicon database, Azher (2023) carried out a socio-corpus-based study that looked at the morphological and syntactic characteristics of Saraiki verbs in relation to Urdu and in various dialects. Azher (2023) found significant dialectal variations in verb usage influenced by social and regional factors after assembling a sizable corpus from a variety of Saraiki literature. Despite the fact that all Saraiki verbs are morphologically complex, the study found that urban and rural dialects differ structurally due to factors like occupation and education.

Gilani (2013) also examined the sociolinguistic function of Saraiki in southern Punjab, focusing on the language's use as an ethnic identity marker. With an emphasis on speakers' pride in the language and how it strengthens their feeling of community, the study sought to comprehend the social relevance of the language. Gilani (2013) collected the research data by

interviewing Saraiki speakers. Researchers also highlighted that Saraiki people feel pride when they speak the Saraiki language, showing a strong connection with their native language. But this study has just given importance to sociolinguistic elements and ignored morphological elements of language.

Similar to this, Arslan et al. (2023) 9 million-word Punjabi corpus to check the morphological trends of nouns in Shahmukhi Punjabi. Their work focused on the construction of nouns specially they analysed compounding and affixation rules. The findings demonstrated that Shahmukhi Punjabi nouns follow inflectional and derivational rules. But this study has just focused on Punjabi and completely ignored Saraiki.

Building on previous studies of the language, Syed and Kula (2020) examined the impact of nasal spreading on morphophonemic patterns in Saraiki. In their phonological analysis of Saraiki nouns and verbs, they looked at how nasal features were distributed among syllables. The results showed a considerable relationship between morphological traits and nasalization, especially in particular syntactic contexts. However, more general morphophonemic phenomena were left out of the discussion due to the study's focus on nasalization.

Malik (2023) examined how grammatical metaphors operate in Saraiki across multiple registers, drawing comparisons with English along the way, in contrast to earlier studies that mostly concentrated on English. The study employed a corpus-based methodology to classify metaphoric structures and pinpoint differences between the two languages. The findings demonstrated Saraiki's distinctive metaphor usage habits, particularly when it comes to abstract nouns. The authors concluded that these metaphors had a considerable impact on Saraiki's linguistic and cultural identity, despite the constraints of the study's concentration on specific text genres.

In a comparable study, Zamir et al. (2021) examine the lexico-semantic correlations of Saraiki verbs. They collected data of 2 million Saraiki words, mostly from newspapers. By examining patterns of synonymy, antonymy, and semantic shifts, their research demonstrated that the Saraiki language has its own verbs with lexical borrowing from other languages.

Although the study provides valuable information, it has given importance to verb morphology and ignored noun morphology. Similarly, Nazeer et al. (2024) used a corpus to quantitatively examine the lexico-semantic connections of Saraiki nouns. The study used a two-million-word corpus from the Saraiki daily Jhoke to look at patterns like synonymy and antonymy. The findings highlighted Saraiki's limited linguistic resources and lexical borrowing challenges. Although these works demonstrate computational rigor, they do not provide a thorough examination of morphological structures.

### 2.3 Research Gap

Previous studies on the Saraiki language have not given importance to grammar and morphology. Therefore, Saraiki grammatical information is not available on Google platforms. This study helps to fill the gap by providing information about Saraiki nouns and their structures. This study will help in making Saraiki dictionaries and Saraiki grammar books.

### 2.4 Theoretical Framework

Theoretical models provide a framework for understanding what you need to know and where to begin while studying linguistic aspects. This study is descriptive and uses an integrative framework that combines a frequency-based model and the concepts of inflectional and derivational morphology by Aronoff & Fudeman (2022) as its theoretical foundation.

### Methodology

A corpus-based, quantitative linguistic approach was used in the present research, which is based on a descriptive-analytical perspective.

#### 3.1 Overview of the Data Collection Procedures

The data collection is a sequential procedure. A two-million-word Saraiki corpus is created, and generic searches are used to examine the morphological patterns of nouns. The data collection process involved multiple processes. First, corpus compilation was carried out, and information was collected from websites, blogs, newspaper content, and books. The data was digitized and scanned. Next, pre-processing took place, where the text was filtered, and

corrections were made to the spelling and diacritics of the OCR outputs.

This study made use of digital tools such as OCR and text editors. OCR software was used to convert scanned documents into editable versions. Text editors like Notepad and Word were used to clean and format data.

#### 3.2 Overview of the Data Analysis Procedures

The first step in the investigation was pattern-based concordance searches in Antconc. Antconc 3.5.9.0 is used for keyword-in-context (KWIC) searches, counting frequencies, and building concordances. First of all, the corpus was loaded on Antconc, then suffixes with \* were searched. To ascertain the grammatical function of the target word, each line was manually reviewed. The collected sentences after searching were then exported to Excel. The sentences were then carefully categorized. The words that are acting as nouns in sentences were labeled "yes "; other words that are not nouns functioning in sentences are labelled as 'no'. This tagging technique was essential for differentiating between noun-functioning and noun-non-functioning words. After that, each suffix's noun productivity was calculated with the help of tables. Two tables were designed. One table was about pattern productivity, showing the total number of instances for every pattern and noun-acting instances for every suffix. The other was the lexical inventory table, which included a list of all words that had been verified as nouns within each pattern. With these tables, productivity ratios were then calculated, and cross-references were made among various suffixes.

### Results and Discussion

The result section consists of three tables: Table 4.1 shows patterns with their total occurrences and instances acting as nouns, Table 4.2 displays patterns with examples functioning as nouns, and Table 4.3 contains a list of common Saraiki nouns with Urdu and English translation.

#### 4.1 Morphological Patterns and Noun Frequency

Table 4.1 presents a basic outline of the quantitative analysis performed on the morphological patterns of nouns in the Saraiki language. Additionally, this table provides a

systematic overview of the information obtained from a large corpus (2 million words), calculated using AntConc 3.5.9.0. It is structured by five central columns that start with a serial number, the name of the pattern (that is mainly formed by the last letters or suffixes), the overall number of all the instances of the specific pattern that can be found in the corpus, the number of sentences which are used as nouns functioning, and the last column describing the percentage of nouns for every pattern.

The approach to constructing this table involves several stages of analysis. Firstly, after searching through AntConc, the results are exported into an Excel spreadsheet. To check whether words are acting as Noun -functioning or Noun non -

functioning, each sentence was then carefully studied. After these words, acting as nouns are labelled as 'yes' and words acting as non-functioning nouns are labelled as 'No'. Table 1 is a linguistic map that illustrates how morphological patterns function in proportion and how language is naturally employed in Saraiki. It aims to address significant linguistic issues, such as the most common Saraiki word ending patterns that result in nouns and the effectiveness and consistency of these patterns in other word classes.

The table will therefore be used as an empirical instrument to investigate both frequency and the practical integrity of morphological structures in noun formation.

*Pattern with Frequency of all Words and Words Acting as Nouns*

Sr. No.	Pattern	Total No. of instances	Instances acting as Noun	Percentage
1.	ا	45033	32750	72.7
2.	ب	364	304	83.5
3.	پ	36	32	88.9
4.	ت	1113	1004	90.2
5.	ٹ	167	154	92.2
6.	ج	83	45	54.2
7.	چ	49	14	28.6
8.	خ	37	30	81.1
9.	د	148	128	86.5
10.	ڈ	75	62	82.7
11.	ر	2669	2295	86.0
12.	ڑ	3069	2843	92.6
13.	س	14514	13805	95.1
14.	ش	3252	2734	84.0
15.	ع	4052	3774	93.1
16.	غ	155	153	98.7
17.	ف	3439	2947	85.7
18.	ق	5565	5563	99.9
19.	ک	112	101	97.6
20.	گ	500	446	89.2
21.	ل	4128	3858	93.4
22.	م	1571	1429	91.0
23.	ن	3305	2981	90.2
24.	ں	1038	597	57.5
25.	ھ	2704	2075	76.7
26.	ی	1747	1430	81.8
27.	ے	1584	381	24.0
28.	ٹ	701	34	4.8
29.	ٹا	124	6	4.8
30.	ٹی	2028	33	1.6
31.	یو	1272	1170	92.0

#### 4.2 Morphological Patterns and Corresponding Noun List

Table 4.2 is created, which provides a detailed breakdown of the morphological suffixes that

contribute to noun formation in Saraiki, along with their corresponding lexical items.

Pattern	Words acting as nouns
ا	بابا، اجوگا، خدا، سنیہا، جوار، بہیا، پٹکا، دیگڑا، گوگڑا، کوہاڑا، کھپسا، مونڈھا، ویہڑا، کھاڈا، ادا، کوکا، کنڈا، جھنڈا، لسوڑا، ڈوہڑا، کھوتا، بہتریجا، جوڑا، ڈبوا، ہماچا، خلا، ابا، کاندھا، ڈونگا، پھلکا، نلکا
ب	جیب، وسیب، خواب، خطاب، ویب، جنوب، مکتوب، قلب، ادیب، سیب، اعصاب، غریب، کرتب، تیزاب، رابب، خاب، شاداب، نقاب، مغرب، عجائب، کتاب، خطاب، خاکروب
پ	وڈپ، ڈھپ، کوپ، سوہٹپ، جیپ، کھپ، سٹاپ، پانپ، مالدیپ، گروپ، کیمپ، یورپ، ورکشاپ
ت	مسیت، تریمت، رت، اگست، عدالت، شخصیات، مذاکرات، ملاقات، رات، بھارت، مصنوعات، زراعت، نشست، موت، سیاست
ٹ	رپورٹ، فرسٹ، بجٹ، مارکیٹ، ووٹ، پلانٹ، ڈسٹرکٹ، سرٹیفیکیٹ، ٹویٹ، ٹیلنٹ، پائلٹ، اکاؤنٹ، سینیٹ، پورٹ، کرکٹ، شرٹ، کینٹ، فورٹ
ج	احتجاج، کالج، مفلوج، فوج، جیلنج، گاج، پنج، پیکج، حج، افواج، سٹیج، پیج، اندراج، تاج، اخراج، مہاج، لاج، کوریج، ننانج، اعلاج، انچار، عروج، ارج، امتزاج، گونج، بنج، محتاج، دھج، اج، بیج، علاج
چ	چمچ، سوچ، میچ، بلوچ، مارچ، کوچ، برانچ، ریسرچ، کیچ
خ	شیخ، تاریخ، منورخ، ملخ، شاخ، سرخ، منسوخ، تلخ، نرخ، تریخ، میونخ، مشانخ، تنسیخ، فرخ، سوراخ
د	وفا، اندھاندہ، مسجد، ورد، واحد، وجود، ہند، گیند، مقصد، گود، والد، محمد، اعتماد، آمد، تعداد، کھاد
ڈ	گانڈ، کھنڈ، رائونڈ، اسٹینڈ، ریکارڈ، فلڈ، ورلڈ، بورڈ، لینڈ، فنڈ، ڈیمانڈ، ریپڈ، کارڈ، فیلڈ، لوڈ، روڈ، کمانڈ، سٹینڈ، گولڈ، پریڈ، ایوارڈ، گریڈ، ڈیڈ، تھرڈ، آکسفورڈ، کمپیوٹرائزڈ، گھنڈ، گرینڈ، بوڈ، ریٹائرڈ، ڈیکریڈ، انرلینڈ، گراونڈ، فوڈ، کوڈ
ر	کردار، ہزار، گھر، ڈار، لاپور، دفتر، سپیکر، اظہار، وزیر، صدر، ڈنگر، ڈاکٹر، کشمیر، غیر، سفیر، امور، ٹپر، قطر، شہر، اکتوبر، فریئر، تنور، عمر، تاجر، انور، امیوار، کھیر
ڑ	گالھ مہاڑ، تارڑ، کاکڑ، کروڑ، ڈیہاڑ، چھیکڑ، کوڑ، باجوڑ، ایکڑ، کوبیڑ، اکھنڑ، کاوڑ، جھکڑ، تروڑ، پہاڑ، آڑ، دوڑ، لوڑ، جڑ، کھلوڑ، جھڑ، دراڑ، واچھڑ، اجاڑ، وگاڑ، دھوڑ، ساڑ، سانگھڑ، چھڑ، گڑ، پکڑ، موڑ، بھاڑ، لونڑ، سڑ، وگڑ، آکھنڑ، ڈیہاڑ، بھڑ
س	اجلاس، ایپس، کانفرنس، پولیس، افسوس، افس، جسٹس نوٹس، گیس، فنانس، باکس، سائنس، بس، فورس، فرانس، گیٹس، فیس، وائرس، کانگریس، کورس، فزکس، نس، جاسوس
ش	سفارش، سازش، داعش، لاش، سویڈش، سکواش، درویش، فرمائش، دوش، بینش، برش، فرش
ع	موقع، دفاع، جامع، ضلع، وسیع، موضوع، مواقع، اجتماع، شجاع، شمع، مجمع، سماج، طلوع
غ	فروغ، بلغ، باغ، ابلاغ، بالغ، دماغ، چراغ، داغ، تبلیغ، نابالغ
ف	شریف، آصف، چیف، تعریف، انصاف، ریلیف، یوسف، حلف، شہباز شریف، سٹاف، عارف، سیف، معروف، اشرف، برف، مصنف، عرف، انصاف، واصل، طواف، تصانیف
ق	عاشق، سبق، احمق، حق، مشرق، رزق، فریق، صدیق، اشفاق، خناق
ک	لوک، ٹرک، ناک، ملک، بینک، اکیڈمک، چاک، پلاسٹک، نمک، ناک، تحریک، بلیک، الیکٹرانک، ٹریک، اکاٹامک
گ	ڈنگ، ونگ، انگ، جگ، بیگ، ملنگ، جنگ، ونگ، بریفنگ، لیگ، سمگلنگ، بینکنگ، فائرنگ
ل	گوڈل، بٹھل، سفیل، سنہل، ہڑتال، زال، بدل، درسال، پرنسپل، نیپال، بال، جنرل، کونسل، سال، اپریل، اسرائیل، اقبال
م	مٹام، ڈرم، وزیراعظم، اقوام، اسلام، اعظم، عوام، ابراہیم، فورم، حکام، مریم، علم، مرحوم، شام، مستحکم، مسلم، الزام، پروگرام

ن	مکان، مسلمان، پاکستان، بیان، ایران، خان، ترجمان، چین، قانون، امن، خواتین، اعلان، وطن، سامان، اپوزیشن، انفارمیشن، راشن، فریقین، چیئرمین، رن، کن،
ں	لوکان، دھوان، ملاں، دریڑاں، ایانڑاں، سومان، ڈاتریاں، گھنگڑیاں، اوپریاں، مسیتاں، امان، چنگیراں،
ھ	سمجھ، اکٹھ، سلھ، نچھ، انسٹھ، وانجھ، ڈینھ، مینھ، اٹھ، مٹھ، سانجھ، کٹھ، سنگھ، وٹھ، سٹھ، بیٹھ، سوہھ، کندھ، ڈھڈھ، مونجھ، باسٹھ
ی	ماسی، آپاسی، پھاسی، فارسی، پالیسی، بہتریجی، بہاجی، منجی، بجاجی، نتھلی، ڈندیلی، پلاالی، سوالی، کندھلی، اکھلی، کو یلی، چونکی، وڈکی، گھڑونجی، پواندی، سراندی، کچری، بالڑی، ڈاتری، مندری، وردی، چھری، تھگڑی، دیگرڑی، پوتری، کوھاڑی، ڈپھاڑی، امڑی، اروڑی، پاتری، میستری، چاچی، اندھاری، تاڑی، سنگتی، کنی، ذنی، گوانڈھی، مالچھی، پڑوپی، سنگتی، تھاپی، بوھا ری
ے	پلوٹے، کوٹھے، پرنالے، انگارے، ساھورے، کیارے، کالے، جھنڈے، مظاہرے، دھماکے، ویلھے، اوندے، قبضے، عہدے، کھالے، مصلے
ن	ساوٹ، بھین، لوٹ، کھمٹ، چھٹ، چھٹ، اگواٹ، سیہوٹ، سنجانٹ، سہٹ، وطن، سنٹ، انجٹ، سچٹ، مکھٹ، افسرانٹ، اگواٹ
ٹا	اپٹا، سوھٹا، راٹا
ٹی	پاٹی، چھاوٹی، چھاٹی، ملکائی، کہاٹی، مندھاٹی، نواٹی، راٹی
یو	ریڈیو، ریسکیو، انٹرویو، ویو، ویڈیو، پولیو، ریونیو، پیو

Table 4.2 addresses the lexical view of the identified noun patterns. The main point of Table 4.2 is to provide form and content to the patterns that are studied in Table 4.1 as abstract. It shows how each morphological ending is used in a real-life context. Additionally, the examples in the table highlight how different suffixes contribute to the formation of nouns, which can represent objects, actions, qualities, places, or abstract concepts. As a result, the table provides valuable insight into the linguistic structure of the language, and the following explanation covers each pattern comprehensively. From all patterns searched on Antconc, a few patterns with screenshots from Antconc are

given below. It is easy to explain the findings with some examples of nouns created with the preponderant pattern. For example, "امریکا" "America" (America) is a proper noun and denotes a country. The word دعا (prayer) is an abstract noun; it is a concept or an activity. The pattern of flexibility is demonstrated in semantic domains since, in this case, the (دنیا) (world) is a concrete thing. These illustrations reflect the flexibility of the pattern of the ۱ in making diverse nouns in terms of proper names and common nouns alike. The keyword search for this word on AntConc is given below in the form of an image.

Hit	KWIC	File
165	سفير مسعود خان گزریل ڈینہ شکاگو دے پاکستان اچ امریکا . مسئلہ فلسطین دا حل نوشتہ دیوار اے۔ اے گالھ	SAI
166	سفير مسعود خان گزریل ڈینہ شکاگو دے پاکستان اچ امریکا . مسئلہ فلسطین دا حل نوشتہ دیوار اے۔ اے گالھ	SAI
167	سفير مسعود خان گزریل ڈینہ شکاگو دے پاکستان اچ امریکا . مسئلہ فلسطین دا حل نوشتہ دیوار اے۔ اے گالھ	SAI
168	سفير مسعود خان گزریل ڈینہ شکاگو دے پاکستان اچ امریکا . مسئلہ فلسطین دا حل نوشتہ دیوار اے۔ اے گالھ	SAI
169	بدنام کریندا بنے ، امریکی کانگریس ا کوں پاکستان اچ دنیا آکھئے بی ٹی آئی دا سوشل میڈیا سیل پوری	SAI
170	نہ آخرت اچ ، اسلام امانت اچ تھیں کامیاب اچ دنیا زکوٰۃ ادا کرو۔ انہاں آکھئے والدین دا نافرمان نہ	SAI
171	بھارت دے یوم آزادی کوں کالا اچ کشمیری اچ دنیا پی پی (پی) کنٹرول لائن دے ڈوبیں یاسوں تے پوری	SAI
172	استعمال کوں روکئ ، حیاتیاتی ، کیمی دے ہتھیاراں اچ خلا ، دن ، میگوں یقین اے اسان عالمی سلامتی اے استحکام	SAI
173	معمولی واقعہ کانٹی، اساکوں پتہ واقعہ دا اگست 26 ، پیا محسن نقوی آکھئے بلوچستان اچ کوئی آپریشن نی تھیندا	SAI
174	اچ 5 دہشت گرد ماریے گئں جڈاں پریشن اے این۔ کیتا اچ انفارمیشن بیسڈ آپریشن (آئی بی او)	SAI
175	” وزارت اطلاعات دے فیکٹ چیکر وی اچ توں این ۔ کریندا جیڑھا پاکستان دے عوام دے جذبات کوں ظاہر نی	SAI
176	ایجھا قانون تے اصلاحات کیتیاں ونجن اچ نظام این لہذا اچ بلدیاتی نظام جمہوریت دا حک بنیادی ستون اے	SAI
177	گوای اے۔ شیخ ڈاکٹر ماہرین حمد دی توحید ایہا کیتا عبادت کرو جیں تساکو تے تسان توں ہلے پیدا	SAI
178	دار کوں حق ادا کرئ دا حق اسلام ، اے آلا چاہدے ، اللہ تعالیٰ عظیم اے تے بہوں وڈی حکمت	SAI
179	چیز دا مالک اے تے انہاں ہر اللہ ، اے ہویا اے ، اللہ فرمائے میڈی رحمت پر چیز کوں لکایا	SAI
180	مقابلہ عزم نال کرئ آلیاں قومان دا اوکھ اے ودھتا یوٹ شامل اے ۔ وزیراعظم آکھئے اساکوں محنت نال اگوں	SAI
181	مقابلہ عزم نال کرئ آلیاں قومان دا اوکھ اے ودھتا یوٹ شامل اے ۔ وزیراعظم آکھئے اساکوں محنت نال اگوں	SAI
182	اچ تھیوٹ آلی برقسم دی دہشت ملک ایوان اے دا ۔ قرارداد اچ موقف اختیار کیتا گئے جو پنجاب اسمبلی	SAI
183	تے پورے پاکستان دے عوام اچ واقعہ این اے آلا دا جیڑھا واقعہ پیش اے او بہوں دل ڈکھاوٹ	SAI
184	تے پورے پاکستان دے عوام اچ واقعہ این اے آلا دا جیڑھا واقعہ پیش اے او بہوں دل ڈکھاوٹ	SAI

Nouns built on the pattern with the "ب" pattern, such as "کتاب" (book) and "خطاب" (address), reveal that this pattern is very important when it comes to building nouns that denote a certain entity or a position. Although the occurrence of these

nouns is not as much compared to those ending with the pattern of use ۱, these words form an important part of Saraiki words, and they speak a lot about the significance of the use of the ب pattern in the language.

Hit	KWIC
6	اے 23.3°N گئے تے عرض البلد ودھ باسے آلے مغرب عرب اچ گزریل 6 گھنٹے اچ مزید مغرب/جنوب
7	دے بہوں ضلعیں اچ زور دا پنجاب گھنٹے 24 آندے موجب اے حافظ آباد اچ زور دا مینہ وٹھے ۔ ترجمان
8	اچ ملوٹ لوک تے کمبیاں دے فراڈ سیلرٹیکس آکھئے نگزیب ے پی پی) وفاقی وزیر خزانہ ومحصولات سینیٹر محمد اور
9	عمل اے۔ خطبے دے بعد مسجد شیطانی جوا اے شراب ، قتل نی کرئ، اللہ فرمائے فیصلے عدل نال کرو
10	وارم اب میچاں دے بعد پاکستان کوں ستمبر 30 اے 28 بالترتیب انڈیز نال اے۔ سکاٹ لینڈ اے بنگلادیش دے خلاف
11	اچ بناہ گھنٹے تے غور کریندیاں لینڈ فن اے عرب وزیر اعظم شیخ حسینہ واجد متحدہ عرب امارات، سعودی
12	تعلقات نویں دورے اچ داخل تھی دے پاکستان اے عرب ڈٹی جڈاں جو وزیراعظم شہباز شریف آکھئے جو سعودی
13	معاشی تعلقات ہک نویں دور اچ دے پاکستان اے عرب تے محنت دی تعریف کیتی ۔ انہاں آکھئے جو سعودی
14	تعلقات نویں دورے اچ داخل تھی دے پاکستان اے عرب ڈٹی جڈاں جو وزیراعظم شہباز شریف آکھئے جو سعودی
15	معاشی تعلقات ہک نویں دور اچ دے پاکستان اے عرب تے محنت دی تعریف کیتی ۔ انہاں آکھئے جو سعودی
16	کلومیٹر جنوب مشرق اچ موجود اے۔ 300 توں گوادر اے مغرب رب اچ ، کراچی اور ماڑہ توں 230 کلومیٹر جنوب ، جنوب
17	جو غزہ اچ جاری جنگ دے کیتا خیردار اردوان طیب کریندن۔ انقرہ 29 جولائی (اے پی پی) ترک صدر رجب
18	جمع کراڈٹی ۔ قرارداد اچ موقف اچ قرارداد اچ اسمبلی پنجاب اچ دہشت گردی مکاوٹ تے بھری دے ودھارے سانگے
19	تے عوام کوں 46 ارب روپے دا پلان بجلی اسمبلی پنجاب ریلیف عوام سانگے ٹھڈی ہوا دا جھوکا اے۔ سبیکر
20	تے عوام کوں 46 ارب روپے دا پلان بجلی اسمبلی پنجاب ریلیف عوام سانگے ٹھڈی ہوا دا جھوکا اے۔ سبیکر
21	ایوان ملک اچ تھیوٹ آلی برقسم اے دا اسمبلی پنجاب جمع کراڈٹی ۔ قرارداد اچ موقف اختیار کیتا گئے جو
22	احمد خان ملاقات کیتی۔ ملاقات اچ محمد ملک اسمبلی پنجاب پی پی) وزیر اعلیٰ پنجاب مریم نواز شریف نال سبیکر
23	احمد خان قائد محمد نواز شریف محمد ملک اسمبلی پنجاب سیکرٹری زاہد اختر زمان تے بٹے موجود ہن ۔ سبیکر
24	احمد خان ملاقات کیتی۔ ملاقات اچ محمد ملک اسمبلی پنجاب پی پی) وزیر اعلیٰ پنجاب مریم نواز شریف نال سبیکر
25	احمد خان قائد محمد نواز شریف محمد ملک اسمبلی پنجاب سیکرٹری زاہد اختر زمان تے بٹے موجود ہن ۔ سبیکر

The figure shows the reduced frequency of the pattern, "ب", which emphasizes the intricacy of noun formation in Saraiki. While these patterns are present in the corpus, they are not as commonly utilized as others, implying that their use may be more limited. The Antconc results are given below for this pattern.

Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 701

Hit KWIC

128 کون سلام کرن الیاں نال الاون برجم فومی ای منن سیاسی ای دینی فریضہ اے، پاکستان دے اتین کون ،

129 جیوں سوکھا کرن حکومت دی ترجیح دا لوکان ای کھتاوون تارڑ آکھے جو بجلی دے بلاں اج ریلیف، مہانگ

130 جیوں سوکھا کرن حکومت دی ترجیح دا لوکان ای کھتاوون تارڑ آکھے جو بجلی دے بلاں اج ریلیف، مہانگ

131 عہدے توں ہٹاؤن اج مصروف ہن۔ کون لوکان ای کھنن ڈھی۔ انہاں آکھیا جو وزیراعلیٰ خیر بختونخوا کمیشن

132 دا حکومتی عزم دوہرائے، کمالیہ ودھاوون محصولات ای کھتاوون نیاداں تے مستحکم کرن دی جامع کوششاں سانگے اخراجات

133 لوڑھ اے جنیں اج پٹی دیر دی ونڈن ای بھیجنن چاہیدی اے ای غزہ دے لوکان کون فوری امداد

134 لوڑھ اے جنیں اج پٹی دیر دی ونڈن ای بھیجنن چاہیدی اے ای غزہ دے لوکان کون فوری امداد

135 سلام کرن الیاں نال ہی گالھ کون برجم ای منن سوال ہی نہیں ہندا، صرف پاکستان دے اتین کون

136 | تھیسن۔ سال پہلے جھل اج سوچھل شریک گائیک ای اکواون ن نشایر سرائیکی ادیب، دانشور، شاعر، سیاسی، سماجی

137 رہائی بارے مذاکرات دا چھیکڑی موق دی برعالمیاں ای کراوون کیتے بہتر راہ تے ٹورن، غزہ اج جنگ بندی

138 رہائی بارے مذاکرات دا چھیکڑی موق دی برعالمیاں ای کراوون کیتے بہتر راہ تے ٹورن، غزہ اج جنگ بندی

139 ،قومی، سیاسی ای دینی فریضہ اے دا ساریاں اسان رلاوون پیدا نئیں تھیندا، انہاں دے ناپاک عزائم مٹی اج

140 ڈسائے جو اسان این کانفرنس ا مشوری شاہنواز انجینئر اکواون پی نال گالھ مہاڑ اج سوچھل دھرتی واس دے

141 گالھ نی ہضم تھیندی پٹی جو اے کون انہاں ہن لکھ تے ملک کون ڈیفالٹ کراوون چاہندے ہن۔

142 ڈیکھدیاں ہن۔ اخبار دی بندو دے آہشن ڈوجھے او ہن تے امریکا دے دروازے فی الحال بند تھیوون باروں

143 جھل نال بننن سانگے جنگے اقدامات آلی اوٹن اج ساوون زخمیاں کون چنگیاں طبی سہولتاں ڈیوون دی ہدایت کیتی ۔

144 ادا کرسکدے ۔ پروفیسر محمد یونس ٹیلی کردار اہم ای ہٹاوون تعاون جنوبی ایشیا دے لوکان دی زندگیاں کون چنگیاں

145 نہیسی۔ ڈویس ملکان دے تعلقات دی کردار اہم اج ودھاوون وفد دا موجودہ دورہ روس تے پاکستان وچال تعاون

146 ساوتھ فرنٹ لائن تے ہے، فرٹینئر سی ایف اج ہٹاوون یف کریندے آکھے فتنہ الخوارج دے مذموم عزائم ناکام

147 تقریباً ہک کروڑ 20 لکھ روپے دا حکومت اج بننن کون سنہیندی اے ۔ خواجہ محمد آصف آکھے ہک ڈاکٹر

Search Term  Words  Case  Regex Search Window Size 50

Advanced

Start Stop Sort Show Every Nth Row 1

Kwic Sort

Level 1 1R  Level 2 2R  Level 3 3R

#### 4.3 Saraiki Nouns with Urdu, English Translation

A list of words that have occurred repeatedly in corpora and have been used in our real life many

times, for this purpose, is given in Table 4.3, comprising common Saraiki nouns, with Urdu and English translation. So that any person who wants to learn Saraiki words can learn easily.

Saraiki	Urdu	English
ادا	بھائی	Brother
مونڈھا	شانہ	Shoulder
کھپسا	جیب	Pocket
ویہڑا	صحن	Courtyard
پٹکا	پگڑی	Turban
ڈیوا	چراغ	Lamp
کوپ	کپ	Cup
مسیت	مسجد	Mosque/Masjid
تریمت	عورت	Woman
ڈند	دانت	Teeth
کھنڈ	چینی	Sugar
ڈنگر	جانور	Animal
ٹبر	خاندان	Family
تنور	تندور	Oven
کوبیڑ	دھند	Fog
جھڑ	بادل	Cloud
گالھ مہاڑ	بات چیت	Discussion
لونڑ	نمک	Salt
بوچھڑ	دوپٹہ	Long Scarf
نک	ناک	Nose
ذال	بیگم	Wife
پیو	ابو	Father

مینه	بارش	Rain
دینہ	دن	Day
کنده	دیوار	Wall
دھڈھ	پیٹ	Abdomen
سلہ	اینٹ	Brick
ماسی	خالہ	Aunty
گوانڈھی	پڑوسی	Neighbour
لون	نمک	Salt
بھین	بہن	Sister
پانی	پانی	Water
مندھائی	مدھانی	Churner
امری	ماں	Mother
کوہاڑی	کلہاڑی	Axe
دھی	بیٹی	Daughter
سنگنی	دوست	Friend
میستری	معمار	Mason

#### 4.4 Simple, Derived and Compound Nouns

The study concentrates on the following three main goals. The first goal is to analyze the morphological patterns governing noun formation in the Saraiki language. The second goal is to explore the frequency and distribution of these patterns within the Saraiki corpus, identifying which patterns are more commonly used to form nouns. Tables 1 and 2 answer the first and second questions. The third goal is to investigate the structure of simple, derived, and compound nouns in Saraiki.

Simple nouns are independent words that do not undergo any affixation or combination. They directly refer to tangible objects, places, people, or abstract concepts. For example, in the examples from Table 2, the word ادا (brother) is a simple noun that does not undergo any affixation. Words like رب (Lord), عرب (Arab), یورپ (Europe), گروپ (group), زمین (earth), آسمان (sky), عدالت (court), کوپ (cup), are acting as simple nouns. Derived nouns are formed by adding affixes, usually suffixes, to root words. The process of derivation involves transforming the base word into a more complex or specialized noun. For example, from the word دیگ (cauldron), adding ڈا yields دیگڑا (small cauldron). Another example is where the suffix ا has changed the word from "handwoven blanket" (کھیس) to "pocket" (کھیسا). Similarly, شخصیات (personalities), derived from شخص (person) with the suffix یات to indicate the plural form, حکومت (government), derived from حکم (command) with the suffix تہ, and وزارت (ministry), derived from وزیر (minister) with the

suffix تہ. plural objects, words can be easily made with the help of the suffix. Moreover, both of these patterns predominantly use the derivation process. By adding these meanings, words are changed. For example, by adding in all the given words, grammatical categories changedی, such as سرکاری حکومتی, مذہبی, سماجی, پاکستانی, اسلامی, and a few words by adding this, meanings are changed, برادری (brother) changed into برادری (community). Similarly, by adding ے to a few words, their meanings are changed. For example, فیصل (Faisal) is changed into فیصلے (decisions), the same with word خط (letter), which is converted into خطے (region), and word فرق (difference) is also changed into فرقے (sects). Compound nouns are formed by combining two or more independent words to create a new noun with a distinct meaning. These compound nouns represent more complex ideas and often consist of two words joined together. For instance, وزیر اعظم (Prime Minister) is a compound noun formed by combining وزیر (minister) with اعظم (great), indicating a high-ranking government official. Similarly, پنجاب (Punjab), formed from پنج (five) and آب (rivers), refers to the land of five rivers. Other compound nouns include سعودی عرب (Saudi Arabia), a compound noun formed from سعودی (Saudi) and عرب (Arab). پاکستان (Pakistan) is a compound noun formed by combining پاک (pure) and ستان (land), signifying "Land of the Pure." Workshop (ورکشاپ), a combination of ورک (work) and شاپ (shop), is an example of a compound noun that is predominantly present in Saraiki corpora. Other examples of compound nouns in Saraiki

corpora include پیلز پارٹی، لیڈرشپ خودکش، ٹیلیگرام، بنگلادیش، وزیراعلیٰ، اندھادھند، and ٹیلیگرام، بنگلادیش، وزیراعلیٰ. Regarding our third study question, the results show that Saraiki includes derived nouns and a very limited number of compound nouns in addition to simple nouns. In addition to borrowing words from both Urdu and English, Saraiki also has its unique regional lexicon.

This study has significant results in comprehending the structure of the Saraiki nouns and their functions in constructing sentences. The ۱ pattern and its frequency load of noun examples highlight its position as the central part of the noun morphology in Saraiki. This points to the idea that ۱ is a vital morphological marker in noun formation in the language, probably connected to inflection or derivation.

The relatively low rate of nouns in the ب and پ patterns shows that although it is less compared with ۱, it nevertheless plays a major role in enhancing the system of nouns in Saraiki. The existence of this type of pattern indicates that Saraiki uses a large set of affixation tactics to make nouns, relying on the syntactic and semantic setting.

The results also point out the obscure typology of nouns. Examples of the usage of various affixes to make a verb, an adjective, and other word classes of nouns depict the effect that the Saraiki language uses the morphological process to attribute blending meaning and purpose to words. The morphological variety is a marked characteristic of the Saraiki grammar, providing information about the way the language, in its affixation system, forms meaning.

In addition, the findings imply that the patterns of these nouns are very closely conjoined with gender and case marking, where some of the affixes have been used to depict the grammatical functions of the nouns in the sentences. As an example, the ۱ pattern could be linked with certain gender differences or other functions in a sentence like the nominative case, which is a core feature of noun morphology in the Indo-Aryan languages.

### Conclusion

Based on the results, it is concluded that the system of noun formation in Saraiki is highly regulated by a minimal number of highly fruitful

affixal schemas, the "۱" scheme being the most effective. This suffix is a primary indicator of a considerable number of nouns, suggesting that Saraiki adheres to the morphological economy phenomenon, where a small number of affixes are used to construct a vast number of nouns.

The results also indicate that the system exhibits a pattern and predictability in noun formation in Saraiki rather than a random collection of affixes. Such a conclusion corresponds to the general theory of morphology in linguistics, according to which the presence of affixes is essential to indicate special grammatical categories, including number, gender, and case, and the affixes are employed systematically throughout the language. Furthermore, the results of the study demonstrate that the existence of other patterns, such as ب and پ, can be limited to the specific types of nouns, such as a particular object, role, or action, which also explains their slower frequency of use.

An additional and exciting discovery was the identification of less productive morphological patterns, such as ٹی and ے, which, although occasionally used in noun formation, were more frequently observed in the creation of adjectival or adverbial forms.

### 5.1 Limitations and Scope for Improvement

Though this corpus study on the Saraiki language has tried to provide valuable information about Saraiki nouns, certain limitations are also present. This study used just a corpus of 2 million Saraiki words, so future research can be extended by expanding the corpus size.

Moreover, as most of the corpus is from literary and newspaper texts, it doesn't carry the language of ordinary people, so this study cannot provide information about the variety of the Saraiki language. Additionally, this study is unable to give objective statements about the Saraiki language.

### 5.2 Recommendation

This work tries to pave a way for future research by providing the frequency of Saraiki nouns along with their structures of simple, derived and compound nouns. Future researchers can carry out this research by doing a comparative study of Saraiki, Urdu, English and any other regional language. Moreover, researchers can

also conduct parts of speech analysis on the Saraiki language.

### References

- Anthony, L. (2004). *AntConc: A learner and classroom-friendly, multi-platform corpus analysis toolkit*.
- Anthony, L. (2023). *AntConc (Version 3.5.9.0) [Computer software]*. Tokyo: Waseda University. Available from <https://www.laurenceanthony.net/software/antconc/>
- Aronoff, M., & Fudeman, K. (2022). *What is morphology?* John Wiley & Sons.
- Arslan, M. F., Mahmood, M. A., Shoaib, M., Idrees, S., & Tariq, Z. (2023). Morphological description of nouns in Shahmukhi Punjabi: A corpus-based study. *Journal of Positive School Psychology*, 7(2), 1259–1269.
- Atta, F., & Rasheed, S. (2019). Morphophonemic variations in the Saraiki language. *International Journal of Linguistics, Literature and Translation*, 2(3), 290–299.
- Azher, M. (2023). Developing lexical resources of Saraiki verbs: A corpus-based study. *Linguistic Forum*, 5(3), 137–148. <https://www.linguisticforum.com/113940486>
- Beames, J. (1875). *A comparative grammar of the modern Aryan languages of India (Vol. 2: The Noun and the Pronoun)*. London: Trübner & Co
- Berton, P. (1999). *My times: Living with history, 1947–1995*. Library Services Branch, Province of British Columbia.
- Gilani, M. H. (2013). Historical background of Saraiki language. *Pakistan Journal of Social Sciences (PJSS)*, 33(1).
- Gilani, M. H. (2013). Historical background of Saraiki language. *Pakistan Journal of Social Sciences (PJSS)*, 33(1), 1-15.
- Hussain, S. S., & Khan, B. M. (2016). Integrated linguistic links of Pushto and Saraiki languages. *Pakistan Journal of Social Sciences (PJSS)*, 36(2).
- Hussain, S. S., & Khan, B. M. (2016). Integrated linguistic links of Pushto and Saraiki languages. *Pakistan Journal of Social Sciences (PJSS)*, 36(2).
- Jain, D., & Cardona, G. (2007). *The Indo-Aryan languages*. Routledge.
- Katamba, F., & Stonham, J. (2006). *Morphology* (2nd ed.). Palgrave Macmillan.
- Kula, N., & Syed, N. (2020). Non-myopic nasal spreading in Saraiki. *Linguistic Inquiry*, 51(3), 421–452. [https://doi.org/10.1162/ling\\_a\\_00346](https://doi.org/10.1162/ling_a_00346)
- Mahboob, A. (2009). English as an Islamic language: A case study of Pakistani English. *World Englishes*, 28(2), 175–189. <https://doi.org/10.1111/j.1467-971X.2009.01583.x>
- Malik, S. (2023). Identification of ideational grammatical metaphor in Saraiki language and its deployment in different registers of Saraiki and English: A contrastive study. *NUML Journal of Critical Inquiry*, 21(3), 1-20. <https://doi.org/10.52015/numljci.v21iii.261>
- Nazeer, M., Azher, M., Pervaiz, A., & Yasmeen, I. (2024). Developing lexico-semantic relations of Saraiki nouns: A corpus-based study. *The University of Chitral Journal of Linguistics and Literature (JLL)*, 8(1), 162–182.
- Rahman, T. (1996). *Language and politics in Pakistan*. Karachi: Oxford University Press.
- Rahman, T. (2006). Language policy, multilingualism and language vitality in Pakistan. *Trends in Linguistics Studies and Monographs*, 175, 73–88.
- Renschler, I., & Kleiner, B. (2013). Considering dialect in survey research. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 118 (1), 51–59.
- Shackle, C. (1976). *The Saraiki language of Central Pakistan: A reference grammar*. London: School of Oriental and African Studies, University of London.
- Shackle, C. (2003). Punjabi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 581–621). London: Routledge.