ISSN Online: 3006-4708
ISSN Print: 3006-4694
SOCIAL SCIENCE REVIEW ARCHIVES
https://policyjournalofms.com

Core and Periphery in Saraiki Verbal Morphology: A Corpus-Based Study

Luqman Manzoor[1], Zubair Majeed[2], Dr. Hafiz Muhammad Qasim*[3]

[1] MPhil, Dept. of Applied Linguistics, GC University, Faisalabad, Pakistan.
Email: luqmanmanzoor875@gmail.com

[2] MPhil, Dept. of Applied Linguistics, GC University, Faisalabad, Pakistan.
Email: majeedzubair1056@gmail.com

[3] Assistant Professor, Dept. of Applied Linguistics, GC University, Faisalabad
*Corresponding Author Email: muhammadqasim@gcuf.edu.pk

DOI: https://doi.org/10.70670/sra.v3i4.1884

Abstract

The article discusses a morphological analysis of verb patterns in Saraiki, an under-resourced yet significant Northwestern Indo-Aryan language. This study is important because understanding Saraiki's verbal morphology not only helps fill a crucial gap in linguistic study but also provides the basis for developing language technology for millions of its speakers. The study aims to investigate the formal and functional characteristics of Saraiki verbs to support language classification and morphological theory. It relies on a dataset of 552 verb tokens manually annotated and coded from a specially compiled, multi-genre corpus. The quantitative analysis distinguishes 40 morphological patterns and shows that more than 81% of the verbs are inflectional. The data reveal a Zipfian distribution, with a few highly productive core suffixes at the head of a long tail of low-frequency forms, and both inflectional and derivational complexities are evident. The suffix -وݨ /-w-n̲/ is highlighted among the others for its polyfunctionality, serving as both the derivational causative and the inflectional purposive participle, thereby questioning the traditional binary classification of inflection and derivation in morphology. The study also reports a diachronic change, the neutralization of gender agreement in plurals, which signals the process of morphological simplification as one of the characteristics of the language. The implications of these findings are significant, as they provide the linguistic community with extensive empirical data on Saraiki verbal morphology and are thus valuable for the development of natural language processing applications.

Keywords: Saraiki verb morphology, corpus linguistics, inflection derivation, Zipfian distribution, typology of languages, under-resourced languages.

1.1 Introduction

Saraiki is a prominent Northwestern Indo-Aryan language with a speaker population of over 25 million people, mainly in Pakistan; nevertheless, the language remains comparatively under-documented in descriptive and computational linguistics (Masica, 1991; Shackle, 1976). While baseline grammars have identified that Saraiki has an agglutinative, suffix-heavy morphology, scholarship has largely reported qualitative descriptions and has not empirically validated those systems with quantitative data systematically verified with corpus-based methods. (Shackle, 2003). The lack of empirical detail is especially pronounced when examining the verbal morphology in Saraiki, where the functional balance between inflectional and derivational processes and the productivity of specific patterns have not undergone systematic examination.

The lack of fundamental data has impeded the development of Natural Language Processing (NLP) resources for this low-resource language. This paper addresses that gap by providing one of the first comprehensive corpus-based quantitative studies of Saraiki verbal morphology. Beyond computational utility, these findings have already been applied to educational contexts; for instance, Manzoor et al. (2025) utilized this morphological data to propose a pedagogical framework for teaching English verb systems to Saraiki speakers, demonstrating the immediate practical value of empirical linguistic documentation.

This study seeks to elucidate how the frequency distribution of verb patterns complies with universal constraints of economy,-like Zipf's Law (Zipf, 1949), and to explore complex phenomena, like morpheme polyfunctionality, that test a strict dichotomy of inflection or derivation. (Butt, 1995; Haspelmath, 1994). This study provides the informational data of core structures and peripheral structures of Saraiki verbal morphology.

1.2 Research Objectives

1. To identify and describe the inflectional morphological patterns of Saraiki verbs.
2. To analyze the frequency distribution of Saraiki verb patterns in relation to Zipf's Law.

To direct this research, the study responds to these key research questions:

1.3 Research Questions

1. What inflectional morphological patterns are used in Saraiki verbs?
2. How does the frequency distribution of Saraiki verb patterns conform to Zipf's Law?

2. Literature Review

Saraiki has a well-defined but under-researched verbal morphology (Masica, 1991; Shackle, 1976). Most of the earlier research has focused on modalities of classification, dialectology, and historical evolution (Rahman, 1995; Bailey, 1924), leaving an absence of empirical studies based on a corpus. The literature recognizes that the Saraiki language is morphologically synthetic and suffixing (Masica, 1991; Shackle, 1976), but at present, this assertion is not supported by quantitative validation from corpus-based methods.

2.1 Saraiki Morphology in Typological Perspective

Typologically recognized as an agglutinative, suffixing language, Saraiki morphologically marks tense, aspect, and mood (TAM) contrast, agreement, and valency changes via bound morphemes (Masica,1991; Shackle, 1976).

Morphologically, Saraiki is structurally aligned with Punjabi and Sindhi but differs along dimensions of vowel harmony, retroflex series, and retention of some Old Indo-Aryan markers. Comrie (1989) highlights that agglutinative languages typically exhibit high degrees of morphological synthesis — a trend that is also found in Saraiki, with complex systems of inflectional and derivational layering exhibited productively.

2.2 Inflection vs. Derivation in South Asian Verbal Morphology

In South Asian languages, inflectional morphology marks TAM and agreement, and derivation modifies valency or meaning (Bhatia, 1993; Butt & King, 2004). Saraiki has suffixes such as -پیسن / ī s n/ that mark third person plural present agreement, and causatives like -وݨ / w n̲/, which add agents in creating ... (Shackle, 1976). However, as Butt (1995) and Haspelmath (1994) note for Indo Aryan languages, inflectional and derivational morphemes can be polyfunctional, which is also confirmed by Saraiki data.

2.3 Literature on Saraiki Verb Morphology

Descriptive grammars (Shackle, 1976, 2003) presented paradigms, numbers only as frequency counts for elicited data, and Pavri et al. (2016) investigated Saraiki Punjabi verbal parallels, but lack corpus analysis for

numbers. There is no pre-existing study documenting the Zipfian distributions of suffixes in Saraiki, so that the interrelatedness of core, high-frequency patterns with the low-frequency, or “peripheral” morphological forms has not been dealt with until now.

2.4 Computational and Quantitative Strategies

Analysis of morphology based on frequency of occurrence is the groundwork for many NLP resources, including morphological parsers and machine translation engines to low resources language (Hankamer, 2018; Zeldes, 2017). Zipf’s Law (1949), which refers to instances of skewed frequency distribution with a short head and long tail with low frequency, applies across languages (Piantadosi, 2014 including morphologically rich languages like Turkish and Finnish.

2.5 Theoretical basis

This study embraces an integrative theoretical basis for a multi-layered analysis of Saraiki verb morphology, drawing from descriptive, theoretical, and quantitative linguistics.

1. **Construction Morphology:** Construction Morphology (Booij, 2010) uses morphological patterns that take the form of a form-meaning pairing or "constructions". This theory fits particularly well for analyzing some of the derivational processes in Saraiki, such as building causatives, where specific suffixes consistently build onto a verb stem to elaborate the verb's meaning.
2. **Distributed Morphology (Halle & Marantz, 1993):** This theory states that morphological processes take place late in the syntactic derivation, meaning that after defining the whole structure, the abstract features will be phonologically realized. This framework can provide conceptual clarity to the complex affordances of the TAM and agreement marking systems in Saraiki, particularly through morphophonemic changes, where the phonological context determines the realized shape of a morpheme.
3. **Zipf's Law and Corpus Linguistics:** This is a continuing study in the footsteps of Zipf (1949) and Sinclair (1991), both of whom were concerned with usage-based linguistics. Zipf's Law states that the frequency of a given linguistic item is inversely proportional to its rank. By applying Zipf's Law on verb patterns, we could clearly see the core patterns that are productive as well as the vast "long tail" of infrequent, specialized forms. So, corpus-based quantitative methods provide empirical grounding that extends into the domain of pure, grammatical idealizations.

2.6 The Study's Contribution

This study addresses these gaps by providing one of the first large-scale corpus-driven quantitative descriptions of Saraiki verbal morphology, differentiating inflectional versus derivational processes, quantifying polyfunctional suffixes, and providing statistical support for a Zipfian distribution. This study contributes to Indo-Aryan linguistic typology and computational linguistics for a major, low-resource language.

3. Methodology

3.1 Corpus Development

The present investigation is based on a multi-genre corpus of Saraiki created for this project, amounting to a total word count of two million words. There are relatively few digital resources available in Saraiki, which led to the compilation of this corpus through a labor-intensive combination of manual and semi-automated methods. The data set was obtained from the following types of sources:

Literary Texts: Novels, dramas, and poetry were utilized in order to capture the narrative and stylistic diversity in and around Saraiki.

Journalistic Texts: Articles printed in the Associated Press of Pakistan, including local, national, international, and business articles, either in Saraiki or translated from English into Saraiki, were incorporated to demonstrate formal or written usage of Saraiki.

Transcribed Texts: Stories and testimonies orally recorded and transcribed from the various methods of written convention, which included features and elements of spoken language.
This multi-registry method ensures a range of variety and representation of different registers and styles. The process of data collection presented numerous challenges because Saraiki is written in the Perso-Arabic script. We utilized Optical Character Recognition (OCR) tools, such as Google Lens, to convert printed texts into digital formats. These attempts led to a high level of errors requiring multiple rounds of manual corrections and proofreading. Again, to minimize the level of error with OCR, we used a method of taking three separate images of each page and then comparing them to find areas of error. In some cases, we were able to copy digitized texts directly. In the end, everything needed to be captured manually into one large corpus, as there are no methods for auto-digitizing or integrating various files digitally in Saraiki.

Derivational Morphology: Derivational Morphology is that class of affixation in which either a new verb is formed or the valency (argument structure) of the verb is altered, using, for example, causative or intensive marking.

3.2 Data Extraction and Annotation

Due to the complexity of morphological annotation, a controlled sample of 552 verb tokens was extracted from the two-million-word corpus for manual annotation. The verb forms were located and extracted using the corpus analysis toolkit AntConc and regular expressions, which were designed to identify a range of verbal forms. Each of the 552 tokens was manually annotated for a range of morphosyntactic features. A two-level classification system was the foundation of the annotation process:

- **Inflectional Morphology:** This is a category of affixes that mark a verb for grammatical purposes without changing the core lexical meaning. Features annotated included tense, aspect, mood, person, number, and gender.
- **Derivational Morphology:** Derivational Morphology is that class of affixation in which either a new verb is formed or the valency (argument structure) of the verb is altered, using, for example, causative or intensive marking.

A native Saraiki speaker with training and experience in linguistics conducted the annotation process. The ambiguous forms, with a particular focus on multifunctional suffixes, were double-checked in the original context to ensure complete understanding of classification. The manual annotation produced a highly detailed, linguistically rich dataset that comprises the empirical basis of this study.

4. Analysis of Data and Findings

The analysis of the 552 annotated verb tokens provides a very precise quantitative and qualitative picture of the Saraiki verbal morphology; evidence of a system whose major features include the importance of inflection, a highly skewed frequency distribution, and complex relations between grammatical categories.

4.1 Importance of Inflectional Morphology

Concerning a quantitative finding, the most remarkable finding relates to the overwhelming importance of inflectional processes in the Saraiki verbal morphology.

- Inflectional morphology comprised 447 tokens (81.0% of the dataset).
- Derivational morphology comprised 104 tokens (18.8% of the dataset).

- There was 1 token (less than 0.2%) that was ambiguous and could not be classified.

This distribution suggests that Saraiki can be considered to be a highly inflectional language since grammatical relationships are mainly denoted by suffixing the verb root. Although derivational processes are less frequent, they still fulfill an essential role in lexical expansion as well as in changing argument structure. Below is a breakdown of the morphological categories:

Table 4.1: Morphological Categories

Morphological Category	Token Count	Percentage of Corpus
Inflectional Morphology	447	81.0%
Derivational Morphology	104	18.8%
Unclassified	1	0.2%
Total	**552**	**100%**

4.2 Zipfian Distribution and Inventory of Verb Patterns

The analysis revealed 40 distinct morphological patterns present in this data set. The frequency of these different patterns is not evenly distributed. The patterns conform to a Zipfian distribution, where a small number of central patterns are highly frequent. This lopsided distribution conveys a system based on a stable, productive core. Manzoor et al. (2025) argue that this 'core' is essential for language instruction, as prioritizing these high-frequency patterns allows for a maximum communicative payoff for learners before moving into the more complex, lower-frequency morphological periphery. We know this to be true because the five most frequent patterns alone accounted for over 53% of all tokens. Conversely, 23 of the 40 patterns, which make up 57.5% of the patterns, occurred fewer than five times each, resulting in a considerable "long tail" of morphological diversity. This lopsided distribution conveys a system based on a stable, productive core, supplemented by a rich periphery of specialized and infrequent forms.

Table 4.2:

The five most frequently occurring suffix patterns are presented in the table :

Rank	Suffix Pattern	Token Count	Percentage of Corpus	Primary Function
1	-وݨ /-w-n̲/	98	17.8%	Derivational (Causative) & Inflectional (Purposive)
2	-یسی /-ī-sī/	64	11.6%	3rd Person Singular/Plural Present Agreement
3	-ییسن /-īs-n/	54	9.8%	3rd Person Plural Present Agreement
4	-دا /-da/	43	7.8%	Masculine Singular Present/Habitual
5	-w-ī-sī	36	6.5%	3rd Person Present Tense

4.3 Analysis of Main Morphological Structures

4.3.1 Analysis of Polyfunctional Suffix -وݨ /-w-n̲/

The most common pattern in the corpus, accounting for 17.8% of tokens, is the suffix -وݨ /-w-n̲/. Its frequency is indicative of its polyfunctionality as both a derivational causative and an inflectional purposive participle. Because of this complexity, Manzoor et al. (2025) identify this suffix as a 'Level 1' instructional priority, using 'Causative Construction Mapping' to help students bridge the gap between this single Saraiki suffix and English periphrastic constructions like *make*, *have*, and *let*.

In the results section, the function and distribution of this suffix to morphemes were visible, represented in a concordance plot in Figure 4.1, providing evidence of its operation across the textual contexts of the corpus. The plot of the concordance to the use of -وݨ /-w-n̲/ suffix and relevant exemplifications are provided in Figure 4.1.

Each vertical line represents one use of the suffix to indicate use across the high frequency and category. In this representation, we can see the remarkable patterns of how the suffix is widely offered and consistently represented in the text across contexts. Visually demonstrating one more layer of evidence for thinking of it as a foundational morphological pattern in the language typology and corpus data.

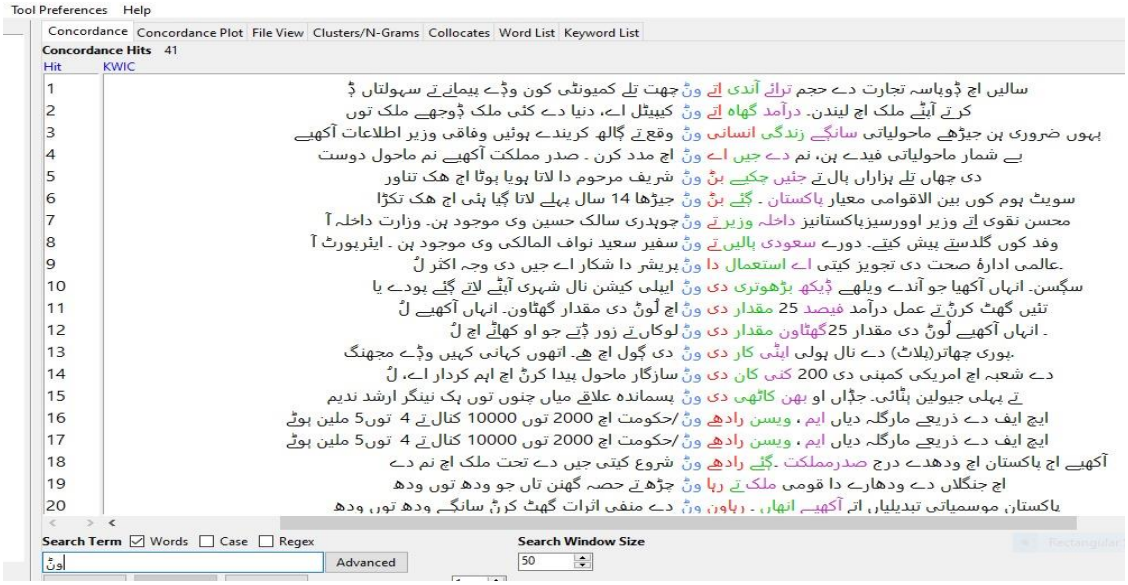

Figure 4.1: Concordance Plot showing the distribution of the suffix '- وݨ ' in the corpus texts.

The twofold properties of the suffix can be specified as follows:

- As a Derivational Causative Marker, it can be applied to a verb root, thereby increasing the valency of a verb, which in turn results in emergent causative verbs that introduce an agent participant causing the action. For example, the intransitive bhāg ('to flee') becomes transitive bhāg-w-n̲ ('to cause to flee' or 'to make someone flee'). This was the highly productive derivational morphological process found in the language.
- As an Inflectional Purposive Participle, the suffix may also contribute to the formation of purposive participles, expressing intention or purpose (e.g., "for the purpose of doing X" or "in order to do X") in other syntactic contexts. Here, it operates as part of the inflectional paradigm of the verb.

Together, these functions provide rich fodder for examples of polysemy that serve as evidence against a neat, rigid binary distinction between inflection on the one hand and derivation on the other. In theoretical terms (in particular within Construction Morphology (Booij, 2010)), the frequency and polysemy of -وݨ /-w-

n̲/ suggest it is a highly entrenched and malleable schematic construction. This is problematic for models that rely on a strict categorical separation and supports a usage-based perspective where the distinction between creating a new word (derivation) and modifying a word for grammatical purposes (inflection) is a cline.

However, for computational linguistics, the polyfunctionality of -وݨ /-w-n̲/ poses an important challenge in disambiguation. Any automated morphological analyzer for Saraiki must be sensitive enough to know to analyze -وݨ /-w-n̲/ as derivational (causative) when it is used as a causative, and as inflectional (purposive) when it is used as purposive, making it a proof of concept for developing advanced NLP tools for the language.

4.3.2 Core Agreement Suffixes

The frequency of -یسی /-ī-sī/ and -پیسن /-īs-n/ indicates the role of person and number agreement within Saraiki grammar.

- -پیسن /-īs-n/ is a clear exemplar of marking 3rd person plural agreement in the present tense. Its frequency indicates the value for marking a plural subject, particularly in the narrative contexts more prevalent in the texts within the corpus.
- -یسی /-ī-sī/ is a more flexible agreement suffix for third person agreement in both singular and plural instances based on dialect and phonological environment.

4.3.3 The Long Tail-Low-Frequency Patterns and Morphological Diversity

The core patterns discussed above may provide the framework of the Saraiki verbal morphology, but the "long tail" of the Zipfian distribution is far from empty. This fringe is made up of 23 distinct patterns that occur fewer than five times in the corpus and provide vast morphological richness, pointing toward dialectal variation and special functions. Such low-frequency forms are not merely regarded as anomalies; they can be placed into functional categories:

1. Several special derivational suffixes that often express very specific functions, for instance, intensity or reflexivity, or others that express subtle changes to the core meaning of the verb, for instance, -یجیاں /yajiān/, probably exerting a special causative or intensive function, occur only once in the corpus.
2. The marked morphological distinction within the Saraiki dialects thus produces dialectal and regional variants. Some low-frequency suffixes mark a particular dialect region less represented in the corpus but legitimate in the larger Saraiki linguistic system.
3. **Morphophonemic Variants:** Some suffixes appear in a unique form, depending on the phonological environment created by the corresponding verb root to which they attach, forming rare yet predictable variants. Such rare manifestations are captured in Table 4.3, presenting selected examples of those low-frequency patterns determined in the corpus to indicate their morphological nature and their rarity.

Table 4.3:

Suffix (Transcription)	Morphological Nature	Frequency	Notes
یجیاں (yajiān)	Derivational (causative/intensive)	1	Rare, specialized derivation

Suffix (Transcription)	Morphological Nature	Frequency	Notes
یجی (yajī)	Inflectional (habitual)	3	Low-frequency inflectional variant
واوے (wāwe)	Dialectal variant	6	Likely reflects regional usage

The analysis of such rare patterns is important; they illustrate the expressive depth of Saraiki with respect to highly subtle semantic distinctions, historical forms, and regional identities that high-frequency patterns do not capture. Computational linguistics would need to include such patterns in robust morphological analyzers that can take care of exceptions and ambiguity, moving beyond productive-core boundary modeling of the language at large. 4.4 Interaction of Tense, Aspect, and Agreement While the preliminary analysis indicates an 81% inflectional character of the verbal morphology in Saraiki, a more thorough examination of the corpus data will reveal the precise interdependencies of tense, aspect, person, number, and gender in Saraiki. Such grammatical categories are thus distributed not uniformly, but according to clear patterns among them and what is still being unfolded in the language as an ongoing, shaping process.

4.4. The Interrelation of Tense, Aspect, and Agreement

From a preliminary analysis, it follows that the Saraiki verbal morphology is indeed highly inflectional (81%), and further investigation into the corpus would provide more insight into how exactly tense, aspect, person, number, and gender interrelate. The not-so-even distribution of the grammatical categories is likely to show recognizable patterns of usage and of continuing linguistic change.

4.4.1 Precedence of the Present Perfect and Third Person Forms

The analysis of the corpus reveals an apparent inclination towards specific classes of tense and person, as is the case with the analyzed texts, which are rich in narrative and description. Figure 4.2 depicts that the Present Habitual/Progressive aspect is the most commonly used, with 320 occurrences. This suggests that the discursive practice in the corpus primarily revolves around regular or ongoing actions. The Past Tense, meanwhile, is the second most frequent category, but the distribution is overwhelmingly dominated by the Third Person, which represents about 88% of all annotated tokens. Such a concentration indicates a stable core of person marking, whereas first- and second-person forms remain sporadic in these registers.

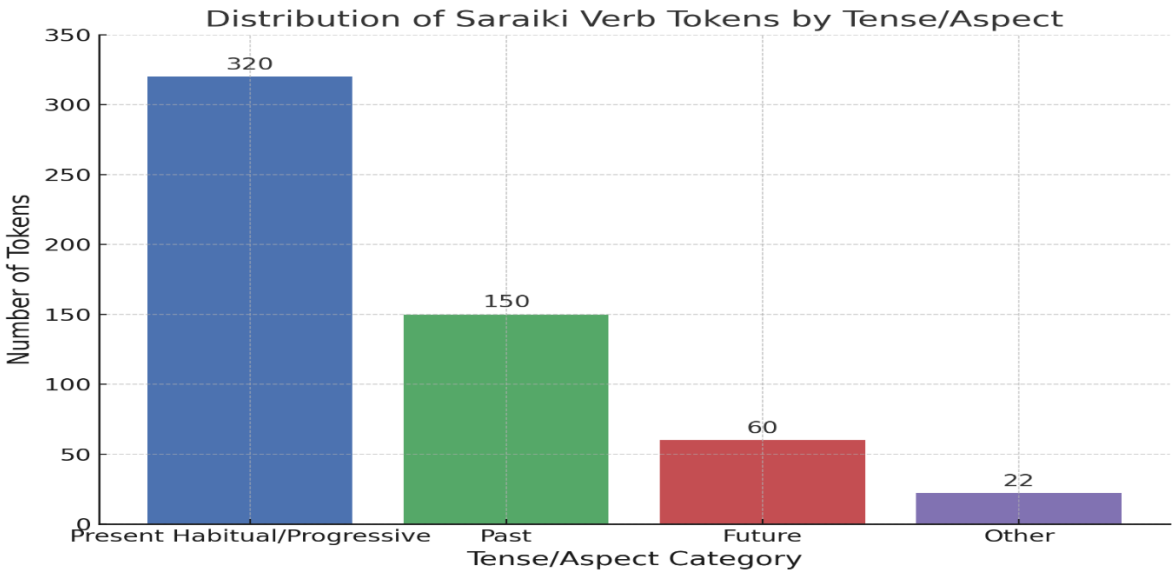

Figure 4.2: Distribution of Saraiki Verb Tokens by Tense/Aspect

This bar chart shows the distribution of verb tokens belonging to Saraiki in the corpus by their tense/aspect categories:

- The present habitual/progressive is the most frequently used, indicating its relevance in conversational communication.
- The past forms are infrequent, but they have been attested.
- Future forms occur according to corpus data and, therefore, infrequently.
- The last category, 'Other,' consists of infrequent or unverifiable tense/aspect forms. In addition, the third person is predominant over other verb forms' distribution, dominating about 88 percent of all annotated tokens. This is manifestly seen in the analysis of person, number, and gender agreement illustrated in the figure.

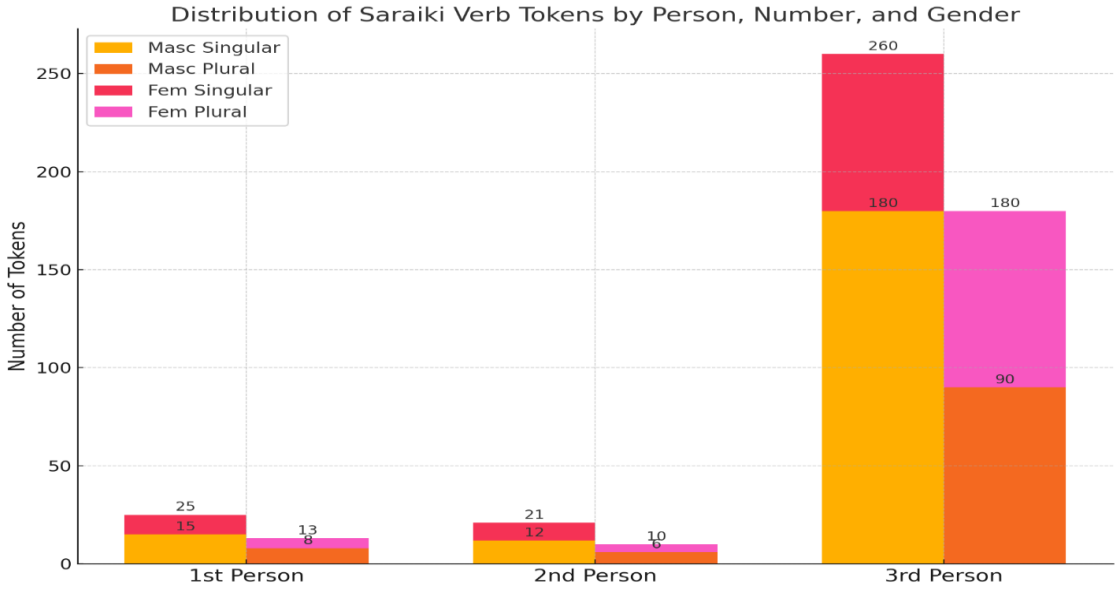

Figure 4.3: Saraiki Verb Tokens Distribution by Person, Number, and Gender

Gender:

1. In the 3rd person, masculine singular and plural dominate.
2. Singular form is preserved by gender. Distinction into masculine and feminine is present.
3. In plural, a near equal count of masculine and feminine denotes gender neutralization.
4. Very few tokens from the first and second person categories are recorded.

The graphical representation provides enriched insight into the distribution trends in the Saraiki verb morphology.

4.4.2 Gender Neutralization in the Plural

The corpus data reveal that there is a constant change in Saraiki that is already taking place: the elimination of gender differences in plural verb forms. Figure 4.3 shows that gender is very strongly marked and recognized in the singular forms, where the masculine and feminine tokens are very clearly differentiated. In the plural category, gender distinction becomes weak, signaling a process of linguistic simplification through syncretism. As noted by Manzoor et al. (2025), this transition provides a strategic opening for 'Agreement System Reduction' training, where learners are taught to consciously strip away complex gender-marking expectations to align with the more minimal agreement system of English verbs.

Singular Specificity: The "-دا /-da/" suffix with 43 occurrences is assigned solely for the third-person masculine singular, whereas its counterpart "-دی /-dī/" is feminine singular.

Plural Syncretism: The very active plural ending "-ییسن /-īs-n/" with 54 occurrences, on the opposite end, effectively serves as a non-binary gender identification marker, applying to both male and female subjects without any change in form.

This syncretism not only indicates that the gender agreement system of the Saraiki language is experiencing a transition, but also that it is becoming less complicated in regard to the plural paradigm. This result places the Saraiki language in the company of the evolutionary trends common to other Indo-Aryan languages.

4.4.3 Tense and Other Person Marking

Present and third-person have a clear dominance; however, the corpus contains markers for all other tense and person systems. For instance, the future tense is actually marked while being relatively low in frequency; the suffix -یساں /-īsə̃/ does unequivocally mark the first-person future (e.g., تھیساں /thīsə̃/ 'I will become').

Such a fine-grained quantitative analysis moves beyond the claim made in the literature regarding Saraiki being "inflectional". It shows in detail how the grammar of the language gets systematically encoded.

4.5 Interaction of Grammatical Categories

4.5.1 Person, Number, and Gender

The analysis shows a system in which person, number, and gender are layered onto each other within verb suffixes.

1. **Person:** There is a strong bias toward third-person forms, which make up nearly 88% of the annotated tokens. This is as expected, considering the narrative-descriptive nature of the corpus. First- and second-person forms are significantly rarer.
2. **Number:** Strongly differentiating between singular and plural is ensured. Singular is marked by -da, while plural is clearly indicated with the suffix -īs-n.
3. **Gender:** An important finding is the marked asymmetry regarding gender. In the singular forms, it is consistently opposed between masculine (e.g., -da) and feminine (e.g., -dī): there is little room for gendering in plural forms, and plurals such as those with the agreement marker -īs-n are commonly neutral in gender. This is known as syncretism, and it indicates an ongoing diachronic simplification of the gender system.

4.5.2 Tense and Aspect

The data from the corpus illustrates a marked inclination towards specific tense-aspect categories.

1. Present habitual/progressive ranks as the most frequent category, typified by suffixes -da (masculine singular) and -dī (feminine singular), which is not surprising given its general value in describing events carried on often or regularly.
2. Past tense is second in frequency, marked either by suffixes such as -yā or by employing alternates to the stem.
3. Future tense forms possessed the lowest frequency, marked as they are by suffixes such as -īsa̰ (1st person) and -wa̰ (3rd person), although this may be a reflection of future tenses being less frequently encountered in writing styles for the source texts in this corpus.

4.6 The Long Tail: Inferring Low-frequent Patterns and Morphological Diversity

A restricted range of pervasive patterns mainly constitutes the Saraiki verbal morphology. Yet the rare forms—particularly the 23 we identified with fewer than five occurrences each—are significant for understanding the depth and richness of the language's morphological system. Such rare patterns, often called the "long tail" of the Zipfian distribution, still play a small role in the language's overall complexity. The infrequent patterns might represent:

1. **Dedicated Derivational Suffixes:** Markers of intensity, reflexivity, or other nuanced shifts in meaning backgrounded in the comparatively more justifications towards their usages to causatives, these derivational suffixes will have subtle meaning distinctions.
2. **Regional and Dialectal Variations:** Suffixes that have prevalence for specific dialects but are infrequent, if anywhere used, within a somewhat standard qualitative Saraiki found in written literature and news.
3. **Archaic Forms:** Older local morphological conventions that exist in written art, literature, and poetry but are not part of any current communicative conventions.

5. Discussion

This research's outcomes are significant because they not only change but also emphasize the very nature of typological classification of Saraiki; they also provide information about its relations to larger morphological theories and about language interactions and transformations.

5.1 Morphological Typology: An Agglutinative-Fusional Continuum

The findings show that Saraiki is not at a polar extreme, but instead exists on a continuum between an agglutinative language and a fusional language. Its agglutinative nature is evidenced by its proclivity for stacking discrete suffixes with recognizable grammatical functions (e.g., root-tense-agreement), yet there is ample evidence of a fusional system. For example, the suffix -da encodes third person, singular, masculine, and present habitual all at once; this compound suffix is a classic feature of fusional languages. Additionally, the polyfunctionality of suffixes like -w-n̲/ and case-syncretism in plural gender marking are clear departures from an agglutinative prototype. Again, this hybrid or mixed nature is typical of many languages in South Asia, and it demonstrates the usefulness of a continuum-based typological model.

5.2 The Inflection-Derivation Interface

The dual function of the suffix -w-n̲/ has strong empirical evidence towards modern morphological theory, which contradicts the traditional, strict binary of inflection and derivation. Inflection and derivation do not appear as two distinct modules of grammar but as flexible categories whose boundaries can be crossed by a single morpheme based on syntactic and semantic context. This finding argues for gradient modeling, which sees morphemes as "more inflectional" or "more derivational," as capturing more linguistic reality in Saraiki

than other models. Other languages exhibited similar phenomena, but the current one shows robust, quantitative data from a newly understudied language family.

5.3 Syncretism and Diachronic Change

The data collected in this study give a strong diachronic result: the systematic neutralization of gender opposition within the plural forms of Saraiki verbs. The gender is very strictly preserved in the singular paradigm—the verbs exactly agree with either male or female subject—but plural verbs often combine or merge the distinction into a single, gender-neutral form.
The phenomenon, syncretism, is thus regarded as a major sign that morphological simplification is taking place. The data further imply that the Saraiki agreement system is being altered in such a way as to move from the original complex gender-marking to a simpler plural paradigm. This development is not exclusive to Saraiki; rather, it is a typologically similar change that has occurred in the entire Indo-Aryan language family, with rapid inflectional complexity reduction as the major feature. Thus, by way of the two-million-word corpus, this research not only documents this shift but also takes an "ongoing change" from its empirical capture; thereby, it shows how the language inevitably progresses to grammatical economy.

5.4 The Comparative Areal Perspective

The morphological patterns apparently present in Saraiki find much heavier resonance in its genetic associates, Punjabi and Sindhi. Commonalities, such as cognate agreement markers (e.g., forms related to -īs-n) or a causative suffix cognate to -w-n̲/, strongly network Saraiki in the Northwestern-Indo-Aryan family. Though Saraiki has a whole host of truly independent innovations, such as morphophonemic alternations of its own and the very specific application of gender syncretism, it earns its right to be seen as a language equal in status with Punjabi and Sindhi. These areal patterns bear traces of the inheritance of ancient traits, independent innovations, and diffusion through contact over centuries in the South Asian linguistic area.

5.5 Limitations

Like any scientific investigation, this study is limited in its design. For an under-resourced language, the corpus is undeniably large; however, it is biased towards written, formal registers. Conversation, as it happens, casual digital communication, and peripheral dialects are still under-represented, and hence some morphological variation must have slipped through. Inter-annotator reliability studies could also serve to validate the extent of the interpretive frameworks used in the careful manual annotation of multifunctional suffixes used in the corpus.

5.6 Future Research

These limitations now suggest avenues of future research.

1. **Corpus Expansion:** There is a pressing need to construct a more balanced corpus encompassing a wider variety of genres, especially spontaneous spoken data and social media texts from non-homogeneous dialectal regions.
2. **Diachronic and Longitudinal Studies:** Gender-neutralization trends need a history and longitudinal studies that can allow, on one hand, a tracing of the trajectory of this change over time through older texts and, on the other hand, its tracing within contemporary speech.
3. **Computational Modeling:** Future work should concentrate on constructing robust morphological analyzers for Saraiki. A promising hybrid model combines the rule-based finite state transducer to model regular processes, and statistical methods trained on annotated data to handle ambiguity and low-frequency patterns.

4. **Psycholinguistic Research:** Experimental studies can test how native speakers process morphologically complex and ambiguous forms, providing cognitive grounding for the theoretical claims made in this paper.

6. Conclusion

This has been one of the first comprehensive corpus-based quantitative analyses of verbal morphology in Saraiki, an important and under-resourced Northwestern Indo-Aryan language. An analysis of 552 verb tokens from a multi-genre corpus allowed us to identify 40 morphological patterns that suggest a mainly inflectional system accounting for 81 percent of verb tokens, thus empirically confirming its utterly suffixal nature. Pattern frequency distribution strongly conforms to a Zipfian curve, showing that usage is largely governed by a small set of highly productive "core" suffixes, while the rest, being rare, form a "long tail" contributing to the morphological richness of the language. The investigation into the Saraiki verbal morphology has shown a pronounced polyfunctionality of the suffix /-w-n̲/, which operates simultaneously as a derivational causative marker and an inflectional purposive participle.

These empirical facts raise questions about the traditional theoretical differences between inflection and derivation, which have been considered rigid. The data throw light on a gradient, continuum-based model of morphology rather than a fixed binary, implying that morphemes occupy a position on a scale between new word creation and grammatical modification.Our analysis additionally revealed a diachronic shift in the gender agreement system of Saraiki: the singular forms clearly distinguish masculine from feminine, whereas this very distinction has become largely neutralized in the plural agreement markers whose development shows evidence of syncretizing tendencies, likely providing insights into the ongoing trend towards morphological simplification within the plural class and providing evidence for the ever-altering identity of grammatical systems within the Indo-Aryan family.

The quantitative insights from this study fill a vital empirical gap in the linguistic description of Saraiki. This work enriches Indo-Aryan typology and offers critical data for NLP tools, such as morphological analyzers and machine translation systems. Furthermore, the successful integration of these findings into the pedagogical framework developed by Manzoor et al. (2025) confirms that a quantitative understanding of Saraiki morphology is a prerequisite for both technological advancement and effective, culturally responsive language education.

The study has never before furnished this insight, but still, there is a need for further investigations. An extension of the corpus to incorporate a wider variety of genres, especially data including conversation and social media, will provide a fairer picture of Saraiki verbal usage across registers and dialects.

Finally, in addition to such research, analyses via comparison with South Asian neighbouring Indo-Aryan languages using the same quantitative framework would help to further test common trajectories of evolution and typological divergence of the said South Asian verbal morphologies. This research serves as a crucial step towards elevating the scientific understanding and computational utility of the Saraiki language.

References

Bailey, T. G. (1924). *A Dictionary of the Punjabi Language*. London: Royal Asiatic Society.

Bhatia, T. K. (1993). *Punjabi: A Cognitive-Descriptive Grammar*. London: Routledge.

Booij, G. (2010). *Construction morphology*. Oxford University Press.

Butt, M. (1995). *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, M., & King, T. H. (2004). The Status of Case. In V. Dayal & A. Mahajan (Eds.), *Clause Structure in South Asian Languages* (pp. 153–198). Dordrecht: Kluwer.

Comrie, B. (1989). *Language Universals and Linguistic Typology* (2nd ed.). Chicago: University of Chicago Press.

Grierson, G. A. (1919). *Lahnda (Western Panjabi)*. Oxford University Press.

Halle, M., & Marantz, A. (1993). Distributed Morphology and the Pieces of Inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111-176). MIT Press.

Hankamer, J., & Mikkelsen, L. (2018). Structure, architecture, and blocking. *Linguistic Inquiry*, *49*(1), 61-84.

Haspelmath, M. (1994). Morphological typology: An overview. In R. E. Asher & J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics* (pp. 227–231). Oxford: Pergamon Press.

Manzoor, L., Bano, N., Majeed, Z., & Naeem, R. (2025). A Corpus-Driven Pedagogical Framework for Teaching English Verbs to Saraiki Speakers: Leveraging Morphological Patterns from a 2-Million Word Corpus Analysis. *Qualitative Research Journal for Social Studie*s, 2(4), 20–39.

Masica, C. P. (1991). *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Najam, U. (2015). A Comparative Study of Saraiki and Punjabi Verb Morphology. *Pakistan Journal of Language Studies*, 3(2), 45–66.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130.

Rahman, T. (1995). *Language and Politics in Pakistan*. Karachi: Oxford University Press.

Shackle, C. (1976a). *Saraiki: The language of the Multan area of Pakistan*. University of London Press.

Shackle, C. (1976b). *The Siraiki Language of Central Pakistan: A Reference Grammar*. London: SOAS.

Shackle, C. (2003). *Siraiki: A Language Movement in Pakistan*. Islamabad: National Institute of Pakistan Studies.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Zeldes, A. (2017). *Multilinear Grammar: Ranks and Interpretations in the Hierarchy of Meaning*. Berlin: Language Science Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Harvard University Press.