



Developing Lexical Resources of Saraiki Verbs: A Corpus Based Study

Research Article

| | | |
|-----------------|--------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| Correspondence: | Muhammad Awais < awaisafzal130@gmail.com > | MPhil Scholar, Department of English, University of Sargodha, Punjab, Pakistan. |
| | Dr. Musarrat Azher < musarratazher@gmail.com > | Associate Professor, Department of English, University of Sargodha, Punjab, Pakistan. |
| | Muhammad Farukh Arslan < farukhgill99@gmail.com > | Lecturer, NUML, PhD Scholar, Department of Applied Linguistics, Government College University, Faisalabad, Punjab, Pakistan. |

Publication Details

Received: July 22, 2023

Accepted: August 25, 2023

Published: September 30, 2023

Abstract

Saraiki is an Indo-Aryan language and is recognized as the fourth most widely spoken language in Pakistan. It is extensively used in Pakistan, especially in south Punjab and Sindh, and is also spoken in some parts of Afghanistan and India. The language holds significant historical and geographical importance. Despite numerous studies emphasizing its distinctiveness, Saraiki remains less explored in terms of its unique linguistic features. The current corpus-based study aims to create synsets of Saraiki verbs by establishing an interface for their synonyms. A corpus of three million words has been developed using literary and non-literary sources. Data collection involved sourcing information from online platforms and scanning hard copies of literary and non-literary works, which were then converted into machine-readable files. From the corpus, one hundred high-frequency verbs were selected and categorized based on Fellbaum's (1993) model, which comprises fifteen files developed according to semantic domains. The verbs falling within these categories were analyzed for their lexico-semantic relations to construct an interface of their synonyms. This study holds significance as it contributes to the development of synsets for verbs, encompassing verb meanings, definitions of associated concepts, example sentences, and lexicosemantic relations. Consequently,



this research proves valuable for students, teachers, and researchers of Saraiki, as well as those engaged in the creation of Wordnet.

Keywords: Lexico-semantic relations, Saraiki verbs, Synset, corpus-based study

1. Introduction

Saraiki, classified as an Indo-Aryan language, holds the distinction of being the fourth most widely spoken language in Pakistan. It serves as the linguistic medium for approximately 25—40 million people worldwide, with significant prevalence in central Pakistan, spanning the southwestern region of the Punjab province, as well as the adjacent districts of Sindh, Baluchistan, and the North-West Frontier Province (Shackle, 2001). Additionally, Saraiki finds resonance in certain areas of Afghanistan and India (Bashir, Connors & Hefright, 2019), contributing to its considerable historical and geographical importance. Over time, Saraiki has proliferated across various regions of Pakistan, leading to the emergence of distinct dialects with individual identities, named after the regions of their use. Shackle (1976), as detailed in Atta (2020), identified six distinct dialects of Saraiki.

1. Southern Saraiki: It represents the variant of Saraiki predominantly spoken in the districts of Rahim Yar Khan, Bahawalpur, Muzaffargarh, and the southern regions of Dera Ghazi Khan (D.G. Khan). In comparison to the central variety, Southern Saraiki covers a relatively compact geographical area. Notably, this particular variant exhibits a subtle influence from Punjabi, resulting in a discernible blend of Punjabi and Saraiki elements.

2. Northern Saraiki: It is predominantly spoken in the districts of Mianwali and Dera Ismail Khan. This variant maintains close affiliations with the Central variety, both in terms of geographical boundaries and shared linguistic features. Given the geographic proximity of the districts of Mianwali and Dera Ismail Khan to the Pashtoon region, Northern Saraiki also demonstrates significant influences from Pashto.

3. Sindhi Saraiki: It is chiefly spoken in Sindh and represents a fusion of the Sindhi and Saraiki languages.

4. Jhangi Saraiki: It, primarily utilized in the Jhang district of Punjab, stands out as a distinct variant within the Saraiki language. Its uniqueness lies in pronounced phonological features, specifically in the utilization of dental and retroflex implosives, as highlighted by Shackle (1976, p. 8). Despite the frequent use of implosives in this variety, it lacks the phonemic contrast between plain stops and dental implosives, as noted by Atta (2020, p. 2).

5. Shahpuri Saraiki: It, spoken in Sargodha and Jhang districts, exhibits shared linguistic features with the Central variety of Saraiki. Notably, Shahpuri Saraiki is linguistically related to Punjabi and is acknowledged as a dialect of the Punjabi language.

6. Central Saraiki: It, also known as Derawali, is prevalent in the districts of Multan, Muzaffargarh, Bahawalpur, and the northern areas of Dera Ghazi Khan. Recognized as the most widely spoken dialect, the prime city for this variety is Multan. In Multan, Saraiki remains the language of choice among locals, utilized informally and as a home language within their communities.

The current study is specifically centered on Central Saraiki and aims to enhance lexical resources by identifying various synsets of Saraiki verbs. A synset, as defined by Kar and Chakrabarti (2011), refers to a group of synonymous words. According to Hassan et al. (2015), a synset represents a distinct sense or concept. The formation of synsets is grounded in sense relationships, and the construction process adheres to three fundamental principles: minimality, coverage, and replaceability, as highlighted by Rattan (2011). To illustrate, words like در، دروازہ، بوہا collectively represent the concept of a door, forming a synset. Similarly, words such as سکول، مدرسہ encapsulate the concept of an institution, constituting another synset.

Several Pakistani languages, including Punjabi and Urdu, have been subject to exploration concerning their morphological and syntactical patterns. Arslan et al. (2023) recently conducted a corpus-based study focusing on the morphology of Shahmukhi Punjabi, with a particular emphasis on nouns and their properties within the Punjabi language. Employing the theoretical framework of distributed morphology (DM) proposed by Harley and Noyer (1999), the researchers concluded that various patterns in Punjabi morphology necessitate consideration in the analysis of higher levels such as syntax, semantics, and discourse.

In a related study, Hassan et al. (2015) conducted research on an online Punjabi Shahmukhi lexical resource, creating a database for Pakistani regional languages to facilitate processing of cross-lingual data, including word sense disambiguation, machine translation, and part-of-speech tagging. The researchers explored various methods for constructing lexical databases for languages worldwide. The overarching goal of this research was to develop a web interface that simplifies the retrieval of updates and query-based results from a lexical database. The system created by the researchers generated grammatically correct sentences that succinctly expressed the concepts of synsets.

Adeeba and Hussain (2011) have contributed significantly to the development of Wordnet in Pakistan, particularly through the creation of Urdu Wordnet derived from Hindi Wordnet. Their approach involved translating the obtained Hindi Wordnet, resulting in the classification of 50,000 words into 28,967 synsets for the Urdu Wordnet project. The study also delved into the notable challenges encountered during the translation process from Hindi to Urdu Wordnet, along with the subsequent manual cleaning procedures.

In the realm of Saraiki language research, scholars have been engaged for several decades, recognizing it as an independent language distinct from Punjabi dialects. Previous studies on Saraiki include a general grammar description (Wagha, 1990), focusing on a variety slightly different from that described in Shackle (1976, 1977), and an exploration of English acquisition by Saraiki speakers (Syed, 2013). A grammar book encompassing Hindko, Punjabi, and Saraiki (Bashir et al., 2019) provides a detailed description of the Saraiki language based on Shackle (1976).

Various researchers have delved into the phonological aspects of Saraiki, including Hussain (2018) with an acoustic analysis of Saraiki stops, and Atta (2019, 2020) examining Saraiki pronunciation. Additional studies cover lexical resources, computational linguistics, syntax, and morphology of Saraiki (Garcia, 2016). Some studies have focused on the identity of Saraiki as a dialect or language (Imtiaz, 2005; Shackle, 2001). The sole study on developing lexical resources for Saraiki language by mapping word senses comes from Gul et al. (2021), who applied an expansion approach to map Urdu word senses (Zafar et al., 2014) onto Saraiki word senses.

Given the significant historical, cultural, and geographical importance of Saraiki, and recognizing the existing gaps in knowledge, this research aims to identify the lexico-semantic properties of Saraiki verbs. The goal is to contribute to the exploration, preservation, dissemination, and understanding of the Saraiki language by uncovering its diverse and unique linguistic features.

Verbs, as the most crucial lexical and syntactical category in any language, provide the relational and semantic framework for constructing sentences (Fellbaum, 1998). Fellbaum's classification of verbs into 15 files revolves largely around semantic criteria, with a division based on conceptual categories defined by Jackendoff (1983, p. 170) and Dowty (1979). These 15 files are further categorized into two groups: verbs denoting events or actions, and verbs denoting states.

The first group, encompassing 14 files, represents verbs associated with semantic domains such as bodily care and functions, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction, and weather verbs (Fellbaum, 1998, p. 41). The second group consists of one file containing state verbs that do not fall under any specific semantic domain. Fellbaum's classification serves to organize verbs without implying theoretical or psychological implications.

Despite the lack of elaboration in these concepts, verbs within the semantic field are derived through semantic relations. In the context of the present research, it is crucial to develop lexical resources for the Saraiki language by investigating the types and lexico-semantic relations of Saraiki verbs. In line with the study's objectives, the following research questions are addressed in the present research:

- What categories of verbs are being used in Saraiki language?
- What are the lexico-semantic relations between verbs?

3. Methods and Procedures

3.1 Data Collection and Development of the Corpus

To conduct this study, a corpus comprising three million words has been meticulously compiled. This corpus consists of 1.8 million words extracted from various online sources, including poetry, novels, dramas, short stories, essays, diaries, columns, history, and geography. Additionally, 0.7 million words were sourced from the Associated Press of Pakistan (AFP) newspaper, and 0.4 million words were obtained from hard-copy books published by Saraiki Adabi Board, Multan. Saraiki data was also extracted from Wikimedia, Saraiki Waseb, and Saraiki literature. The online data and hard-copy materials encompass a diverse range of genres, including news related to economics, international law, weather, and sports.

To convert the text into machine-readable form, the following steps were employed. First, hard-copy data was scanned using an HP DeskJet All-in-One Printer and converted into PDF format using the iLovePDF site. For some data that was not machine-readable, Optical Character Recognition (OCR) was performed using Google Lens, transforming the data into image files. Subsequently, these image files were processed in Google Docs, which read the images and converted them into text form. The text data from both hard-copy and online sources were then combined. After completing these steps,

the researcher saved all the data in UTF-8 format using Notepad++, and it was further processed in Antconc 3.5.9 for analysis.

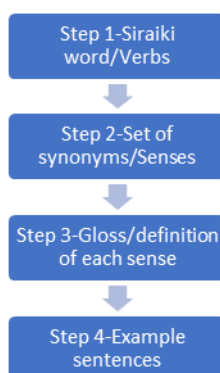
3.2 Development of Verb Synsets

In this study, a selection of one hundred high-frequency verbs was made from the corpus. These verbs were then categorized according to Fellbaum's (1993) model of fifteen files, which were developed based on their semantic domains. The subsequent analysis focused on the lexico-semantic relations of verbs within these categories, aiming to construct an interface of their synonyms.

A hybrid approach was employed for the development of synsets of Saraiki verbs, combining elements of both the merge and expansion approaches. Mapping of verbs involved consultation with native speakers and poets of Saraiki, as well as references to various Saraiki and Punjabi dictionaries. This comprehensive approach aided in capturing the true sense of the words.

The core objective of this study lies in the creation of synsets for Saraiki verbs, a process involving four key steps. Firstly, a verb is selected from the corpus. Secondly, its senses are developed through the use of dictionaries and examination of its occurrences in the corpus. Thirdly, each sense is defined to clarify its meaning. Finally, an example sentence is provided. The components of synsets include the verb itself, its senses, gloss for each sense, example sentences, troponymy, and entailment. A synset is a group of synonymous words, gloss offers a concise definition of a sense, and example sentences demonstrate the use of a sense in a sentence. Troponymy represents the manner relation; for instance, "ہکلاونڙاں" and "بڙبڙاونڙاں" are both manners of "الاونڙاں" and are troponyms of it. Entailment, on the other hand, is a unidirectional propositional relation between two verbs; for example, "خراڻے مارنڙاں" entails "سمنڙاں".

Figure 3.1 Basic Steps involved in Synset Creation



The following steps were undertaken to establish synsets for Saraiki verbs using AntConc 3.5.9, a corpus analysis toolkit. The process involved the following actions:

3.3 Corpus Loading

The corpus was loaded into AntConc by navigating to "File," selecting the corpus, choosing "Wordlist," and initiating the process by clicking the "Start" button. This action generated a Word List containing all words, their ranks, and frequencies within the corpus.

3.4 Utilizing AntConc Functions

AntConc offers various functions for corpus analysis. For instance, by clicking on a specific verb like بنڑا, the researcher accessed a 'concordance' preview. The Concordance view pinpointed the selected word in smaller contexts. The researcher meticulously examined its usages in different contexts to map out the word's senses. If deemed insufficient, the researcher selected the highlighted word to locate it in the corpus under the File view, providing a more extensive context.

3.5 Consulting External Resources

After recording the senses, the researcher sought gloss and example sentences for each sense. This involved consulting dictionaries and Princeton Wordnet in addition to using AntConc. For instance, searching the verb "make" in Princeton Wordnet provided a comprehensive understanding of possible senses for the Saraiki verb بنڑا.

3.6 Troponyms and Hypernyms

The researcher identified and documented troponyms (words specifying manner or mode) and hypernyms (words denoting categories or more general terms) for each verb.

3.7 Repetition for Each Verb

The entire process was repeated for each selected verb in the analysis, ensuring a thorough and systematic approach to creating synsets.

This method combined corpus analysis, external resource consultation, and careful examination of various contexts to capture the diverse senses and relationships associated with Saraiki verbs.

4. Analysis and Discussion

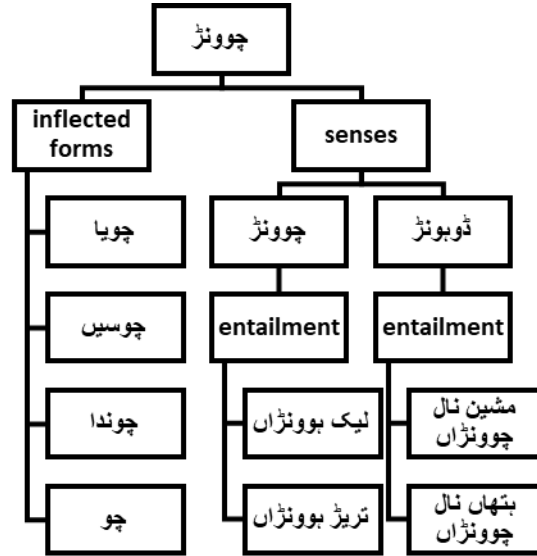
After the careful analysis of the data, 15 types of verbs have been found in the Saraiki corpus. The types of verbs and their lexico-semantic relationships have been discussed in this chapter.

4.1 Verbs of Bodily Function and Care

According to Fellbaum (1998), this category comprises verbs that denote actions or events not under the control of an argument functioning as their subject. These verbs are termed unaccusatives, where their subjects are considered to be underlying objects. Examples from Saraiki include verbs such as ڈرنڑ, سم، جم، and چکراونڑ، ٹھرنڑ، پگھرنڑ.

Table 4.1 The synsets of verb 'سم'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|-------|------|--------|--------------|--------------------------------------------|
| سم | 1 | sleep | verb | سمنڑاں | نیندر کرنڑاں | اساڈی دھرتی دے لوک نیندراں ولہیٹ سم گئن |



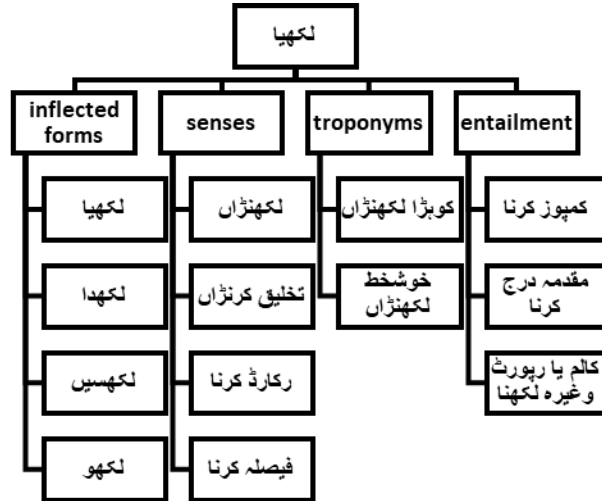
This verb encompasses two senses derived from the corpus. Additionally, each sense involves two entailments. The researcher independently developed glosses and example sentences based on their usage in the corpus.

4.3 Verb of Communication

This semantic category includes verbs related to both verbal and nonverbal communication (gesturing). Verbs of verbal communication encompass speaking and writing. Many language-related communication verbs, such as petition and hail, can be categorized as troponyms of speech act verbs (Austin, 1962). Verbs of speaking inherently involve manner-related verbs like 'بڑبڑاونڑاں، بڑبڑاونڑاں' etc., as well as verbs reflecting the speaker's intention, such as 'حکم ڈیونڑاں، دعا ڈیونڑاں، منگنڑاں' etc.

Table 4.3 The synsets of verb 'لکھیا'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|-------|--------|---------|------|-----------------------|----------------------------------------------------------------|---------------------------------------------------------------------------|
| لکھیا | 1 | Write | Verb | لکھنڑاں | تحریر کرنڑاں | دنیا دا سب کتوں پہلا تصویری رسم الخط سومیر عراق دے قدیم شہر وچ لکھیا ملدے |
| | 2 | Decide | Verb | فیصلہ کرنا | طے کرنا | ہائی کورٹ دا لکھیا سپریم کورٹ وچ چیلنج ہو سگدے |
| | 3 | Compose | Verb | تخلیق کرنا | ادیب، مصنف، تخلیق کار یا شاعر وغیرہ دا کوئی نویں شے تخلیق کرنا | س: سرانیکہ ہیر رانجھا جھنگ دے ربائشی لکھیا، ناں ڈساؤ؟ ج: دمودر داس دمودر |
| | 4 | Record | Verb | رکارڈ ہونا، ثبوت ہونا | ثبوت دے طور تے تحریر کرنا | بنداں کوں چاہیدے جو معاہدہ اونہاں دے درمیان طے ہووے اوکوں لکھ گھننڑاں |



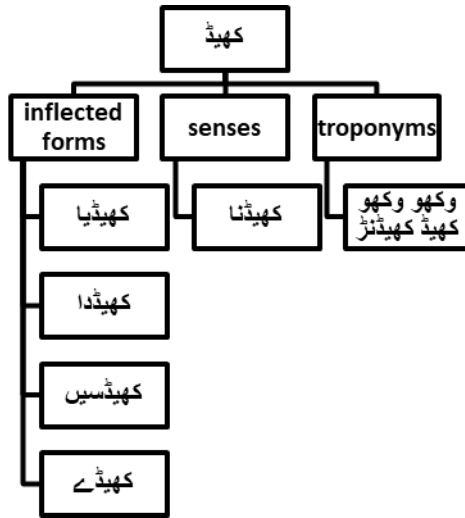
'لکھیا' is classified as a verb of communication as it involves the act of communication through writing. This verb is polysemous, featuring four distinct senses. The first sense has been sourced from online dictionaries, while the remaining senses have been developed with the assistance of the corpus. To enhance their credibility, these senses have been discussed with professionals. Glosses and example sentences have been extracted from the corpus, while troponyms and entailments have been formulated with reference to Princeton Wordnet.

4.4 Competition Verbs

This category encompasses verbs related to sports, games, and warfare, as explained by Fellbaum (1998). She notes that there are many verbs referring to actions specific to games or sports, serving as troponyms of verbs with more general meanings. Examples in Saraiki include 'کھیڈ' and 'نیزنٹ'.

Table 4.4 The synsets of verb 'کھیڈ'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|------|------|----------|---------------------|-------------------------------------------------|
| کھیڈ | 1 | Play | Verb | کھیڈنڑاں | کسے گیم دا کھیڈنڑاں | پاکستان دی ٹیم کراچی اچ 41 ٹیسٹ میچ کھیڈ چکی اے |



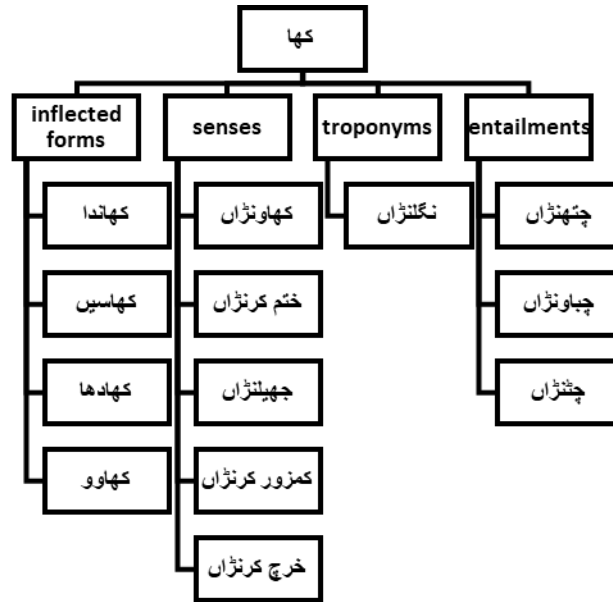
‘کھیڈ’ is categorized under the semantic category of 'competition verb.' It is monosemous, possessing only one sense that has been derived from the corpus. Additionally, an example sentence has been provided from the corpus to illustrate its usage.

4.5 Consumption Verb

This semantic category comprises verbs related to ingesting, using, exploiting, spending, and sharing, as outlined by Fellbaum (1998). It includes unergatives that have only an agent, such as کھاوئڑ تے پیوئڑ etc.

Table 4.5 The synsets of verb ‘کھا’

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|----------------------------|------|-----------------|----------------------------------|----------------------------------------------------------------------------------------|
| کھا | 1 | eat, finish, swallow | Verb | کھاوئڑاں | دندان نال چٹھ کے نگلنڑاں | حلوه کھا تھار کے، مر عمر گزار کے |
| | | | Verb | ختم کرنڑاں | مکا ڈیونڑاں | روز جھاڑی نویں ایجاد پرانڑیں کوں کھا ویندی اے، سب تون وڈا خوش خوراک موبیل فون اے |
| | 2 | Bear | Verb | جھیلنڑاں | برداشت کرنڑاں، تن ورتنڑاں | چوراں پلس کنے اتنڑیں مار کھاہی اے جوں ولا زندگی چ بھیڑا کم نہ کریسن |
| | 3 | Weak | Verb | کمزور کرنڑاں | طاقت یا برداشت اچ کمی ہوونڑاں | سوچاں انسان کوں کھا ویندن، تے اساکوں بنڑ تک وراثت اچ سوچاں ای ملن |



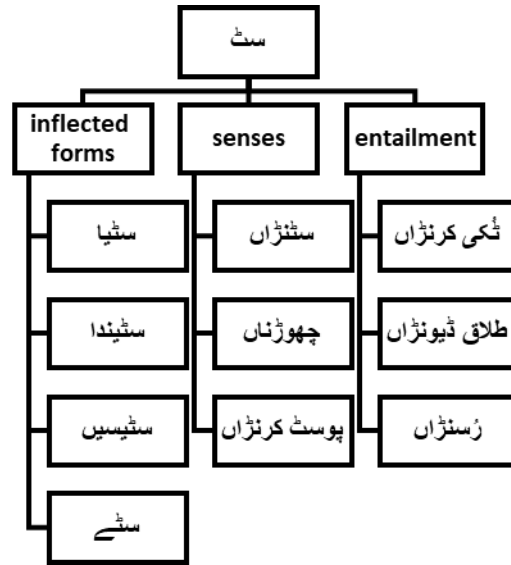
This verb is polysemous, featuring four senses, one troponym, and three entailments. The senses, glosses, and example sentences have been extracted from the corpus, while semantic relations have been developed using English Wordnet.

4.6 Contact Verb

The majority of contact verbs are troponyms of 'base verbs.' Concepts of these verbs include 'بنہڑاں، چھوونڑاں تے کٹنڑاں، ڈھکنڑاں، چھوونڑاں تے کٹنڑاں'. Many verbs are derived through a manner relation that encodes force, intensity, or iteration of the action (Fellbaum, 1998). Some base verbs require instruments or materials to fulfill the action. For instance, the verb 'کٹنڑاں یا کاتنا' requires instruments like 'چاقو، چھری، چاقو' etc. In the context of touching, verbs could be 'مارنا، چھونا، دھک دینا' etc. Examples include 'سٹ، رنگ، مار، گھٹ، چا، کوہونڑ'.

Table 4.6 The synsets of verb 'سٹ'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|--------------|------|-------------|------------------------------|--------------------------------------------------------------------------------------------------|
| سٹ | 1 | throw, spray | Verb | سٹنڑاں | بتھاں نال چھوڑ ڈیونڑ | ملزم سجاد ماء دھی کوں پٹرول سٹ کر ایس بھاء لا ڈتی |
| | 2 | Leave | Verb | چھوڑنا | تعلق توڑ ڈیونڑ | جیہڑے لگدے لیندے ہک دفعہ سٹ چھوڑیوں ول انہاں کراہیں کسے کم وستے ونجنا بے وقوفی نی تے ہور کیا اے؟ |
| | 3 | Post | Verb | پوسٹ کرنڑاں | کسے ڈیجیٹل فورم تے پوسٹ کرنا | جہیں بنیاد تے ریفرینس مسترد یا اگوں بھجوائے گئے ہن انہاں دی وضاحت ویب سائٹ تے سٹ ڈتی گئی |



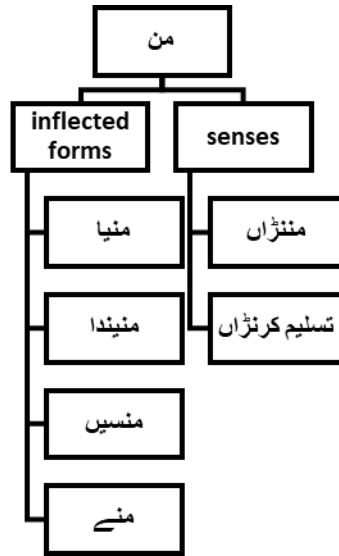
'سٹ' is classified under the semantic category 'Contact Verb.' It is polysemous, featuring three senses. The first sense is obtained from the corpus after consulting dictionaries such as Akhar (2016) and Shabdkosh. The other two senses are retrieved from the corpus. Senses, glosses, and example sentences are derived from the corpus, while entailments are developed with reference to English Wordnet.

4.7 Cognition Verb

This semantic category comprises verbs that express cognitive actions and states, including reasoning, judging, learning, memorizing, understanding, and concluding (Fellbaum, 1998). Troponymy serves as the organizing relation in this file, expressing kinds of reasoning (deduce, induce) or degrees of certainty (infer, guess, assume, suppose). Examples of this semantic category include پڑھ، جانڑدے، سکھ، لگدے، من، and بہل.

Table 4.7 The synsets of verb 'من'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|-------------|------|--------------|------------------------------|----------------------------------------------------------------------------|
| من | 1 | acknowledge | Verb | گل مننڑاں | تسلیم کرنڑاں، قبول کرنڑاں | تیکوں پھل وانگوں پروان چڑھایا بہاویں اج میکوں نگہبان نہ من دربان تاں من |
| | 2 | Unite | Verb | منیونڑاں | روسے مکا ڈیونڑاں | انا پرست بندے ہک وار رس پون تا ساری زندگی نئیں منیندے |



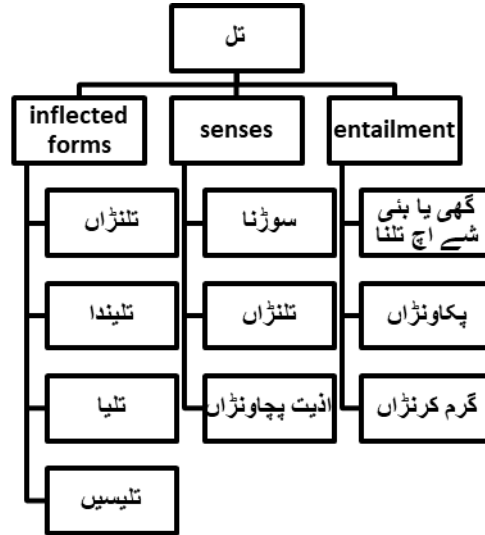
‘من’ is categorized as a cognition verb as it involves mental acknowledgment. It is polysemous, featuring two senses, both present in both the corpus and online dictionaries. The glosses and example sentences are derived from the corpus.

4.8 Creation Verb

Creation verbs can take various forms, depending on the manner of creation, as mentioned by Fellbaum (1998). This may involve creation by a mental act (invent, conceive, etc.), creation by artistic means (engrave, illuminate, print), or creation from raw material (weave, sew, bake). Examples of this semantic category include بنڑاونڑ، تراشنڑ، پک، سی، تل، and گھولی.

Table 4.8 The synsets of verb ‘تل’

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|------|------|-----------------|--------------------------|----------------------------------------------------------------------|
| تل | 1 | Fry | Verb | تلنڑاں | گرم گھی وچ سٹ کے تلنا | جھیں فراق دا ڈکھ ڈیکھنڑاں ہووے اوہ کڑاہ اچ تلیندے پکوڑے ڈیکھ گھنے |
| | 2 | Hurt | Verb | تکلیف پچاونا | اذیت ڈیونڑاں | نیڈیاں بے پروانیاں میڈے ہاں کوں تل چھوڑے |



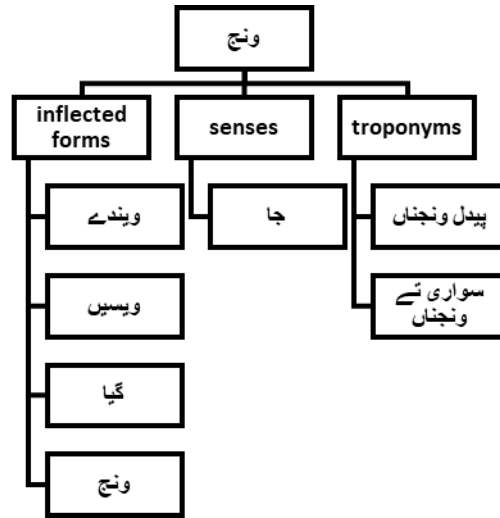
‘تل’ is classified as a creation verb because it involves baking or cooking. It is polysemous, featuring three senses. The first sense is present in both online dictionaries and the corpus, while the other two senses have been developed with the help of native speakers and the corpus. Senses, glosses, and example sentences are derived from the corpus.

4.9 Motion Verb

Verbs related to movement and traveling fall under this category. Miller and Johnson-Laird (1976) distinguish "motion in place," such as shake or twist, and Pinker (1989) refers to it as "continued" motion, from locomotion as in run and crawl (p.529). Troponyms of these verbs, according to Fellbaum (1998), may be on scales such as the speed of locomotion, medium of transportation, the medium in which travel takes place, or other similar scales. Examples of this semantic category include نکل، ونج، لا، چڑھ، ٹر، وڑ، جا، آئی، آندے.

Table 4.9 The synsets of verb ‘ونج’

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|----|------|--------|------------------|----------------------------|
| ونج | 1 | go | Verb | جا | جاون دا حکم دینا | ونج میرے کیتے پانڑیں گھن آ |



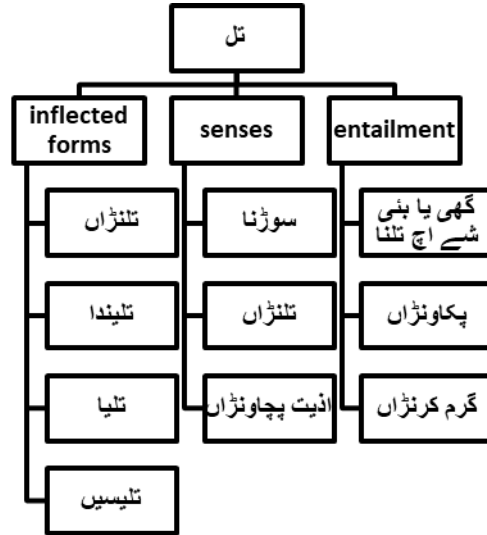
‘ونج’ is categorized as a motion verb because it involves some kind of movement. It is monosemous. The sense, gloss, and example sentence have been taken from the corpus. Semantic relations have been developed through Princeton Wordnet using the expansion approach.

4.10 Emotion or Psych Verb

Verbs expressing any of the emotions (happiness, sadness, adore, love, despise, fear, anger, disgust, etc.) fall under this semantic category. Examples of this semantic category include پٹ، سراہیا، کھل، ہس، رو، and چاہندے.

Table 4.10 The synsets of verb ‘پٹ’

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|-------|------|-------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| پٹ | 1 | Mourn | Verb | پٹنا، ماتم کرنڑاں | واویلہ مچاونڑاں، شور شرابہ کرنڑاں | انہیں کوں وعدے معاہدے یاد کراوے چا تاں او بندا انہاں کوں ملک دشمنڑ لگدے، جے کر زیادہ دھاڑ پٹ کرو تاں او بو پرانڑیں رٹے ہوئے سبق |



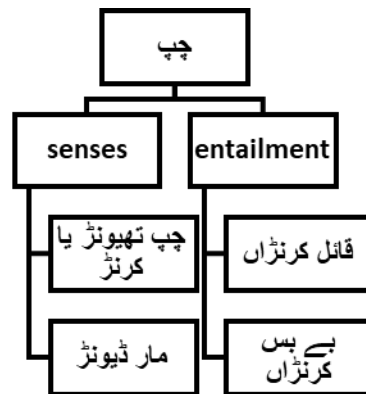
'پٹ' is classified under 'emotion or psych verb' because it involves emotions of sadness. It has four senses, sourced from the online Punjabi dictionary ShabdKosh. The glosses have been manually developed, and example sentences are taken from the corpus.

4.11 Stative Verb

Stative verbs express the state of an object, subject, or argument. These verbs sometimes have non-stative senses; for example, 'پہنچیا' is a stative verb with both stative senses referring to spatial relations and non-stative sense referring to a verb of motion (Fellbaum, 1998). Examples of this semantic category include رہ، پچ، ہووے، پوندے، جیونڑ، and پچ.

Table 4.11 The synsets of verb 'چپ'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|-------|------|-------------------------------|--------------------------|--------------------------------------------------------------------|
| چپ | 1 | quite | Verb | خاموش ہوونڑاں یا کرنڑاں | آواز نہ نکلنڑ ڈیونڑاں | ماڑی توں تاں چپ تھی ونج تیڈی بک بک میرے کنی کوں پکانی کھڑوتی اے |



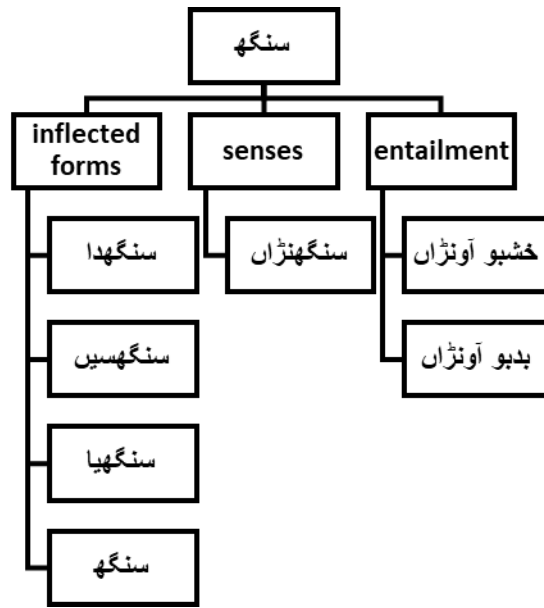
'چپ' falls under the same category as mentioned above. It is polysemous and has two senses, both present in the corpus as well as dictionaries. The example sentences have been taken from the corpus.

4.12 Perception Verb

These verbs refer to perception through the five senses. Their troponyms might differ based on which sense is being used in the phenomenon. For instance, the base verb 'ویکھنڑاں' may have different troponyms like 'سروے کرنڑاں، جسوسی کرنڑاں' based on the intention of the agent; 'جھاتی پاونڑاں، گھورنڑاں' based on the manner of the agent; 'مشاہدہ کرنڑاں، دریافت کرنڑاں' based on the circumstances of the agent (Fellbaum, 1998). Examples of this semantic category include تک، سنگھ، and ڈیکھ.

Table 4.12 The synsets of verb 'سنگھ'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|-------|------|--------|------------------|--------------------------------------------------------------------------------------------|
| سنگھ | 1 | smell | Verb | سنگھنا | نک نال بو لینڑاں | کڈابیں تان ساکوں منہ دے نیڑے کر کے سنگھ ڈیکھ تیکوں پاک محبتاں تے وفاواں دی خنڈبو آسی |



'سنگھ' is a perception verb because it refers to the sense of smelling. It is monosemous. The sense is taken from the Punjabi dictionary Akhar (2016). The gloss has been developed manually by the researcher. The example sentence has been taken from the corpus. The semantic relations have been developed using the expansion approach.

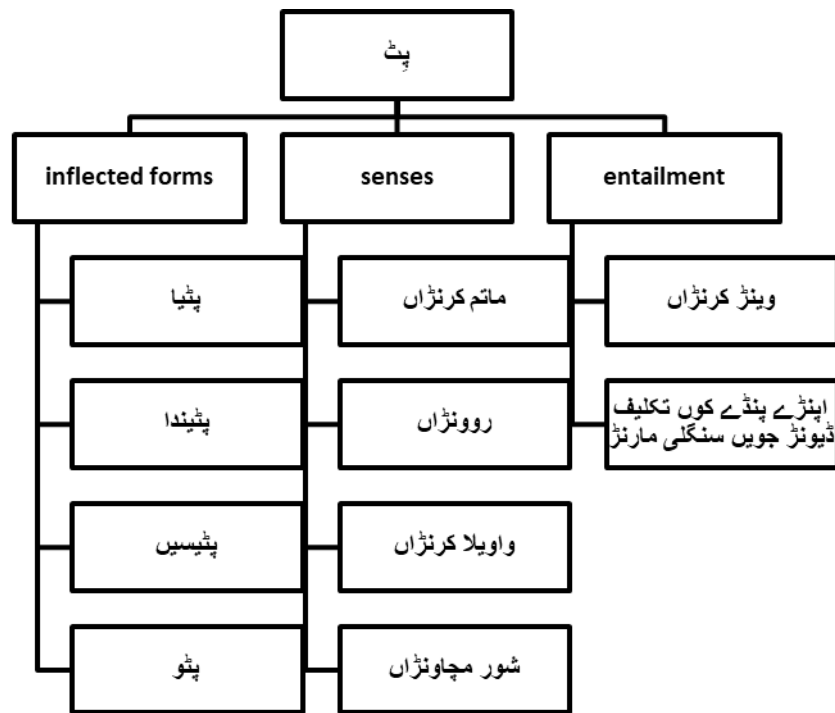
4.13 Verb of Possession

These verbs originate from three fundamental concepts: رکھنڑاں، دیونڑاں، گھلنڑاں {ملکیت ہونڑاں}، حاصل کرنڑاں {لینڑاں}. As Fellbaum (1998) asserts, these verbs signify the "change of possession and its prior or resultant state." Troponyms of these verbs delineate the various ways in which transfer occurs in society: through legal or illegal means such as (چوری، ڈاکہ وغیرہ); or through formal or

informal gifts (بدیہ، بھیک، رشوت، تحفہ وغیرہ)۔ Examples of this semantic category encompass بخش گھدی، لکاونڈ and پاتی، ونڈ، ُطا کرنڈ، رکھا، خریدنڈ، ویچنڈ

Table 4.13 The synsets of verb ‘گھدی’

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|---------|------|-------------|----------------------------|-----------------------------------------------------------------------|
| گھدی | 1 | get | Verb | حاصل کرنا | محنت نال کماوناں | علامہ اقبال میونخ توو ڈاکٹریٹ دی ڈگری گھدی |
| | 2 | Receive | Verb | وصول کرنڈاں | وصول کرنا | اج میں کم تے ناہ آ سگدا ، مالک کنوں میری پچھلی مزدوری وی گھدی آویں |
| | 3 | take | Verb | لے آونا | کسے لئی کوئی شے لے آونا | امیر حسین اوں کنے کجھ کتاباں گھدی آیا |



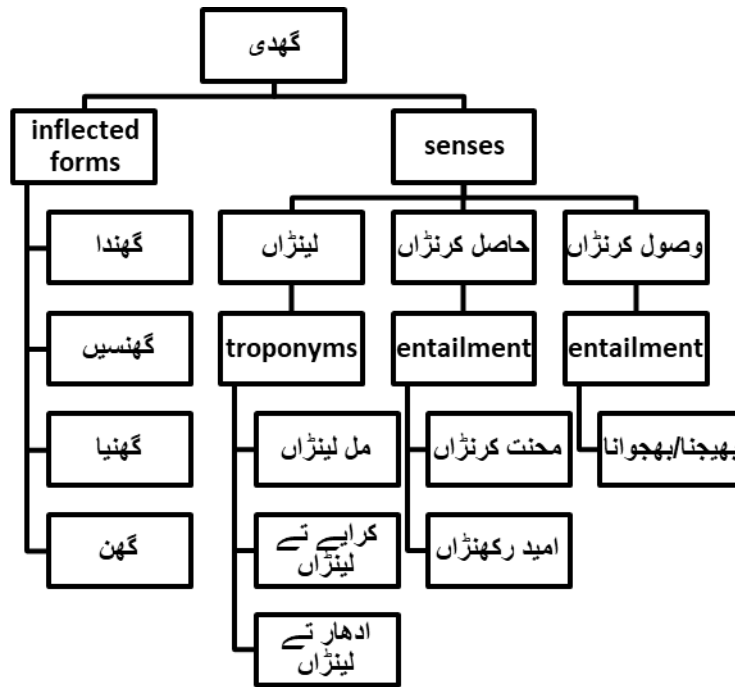
‘گھدی’ is a verb of possession, signifying the act of receiving something. It is polysemous, featuring three senses. The first two senses are drawn from Akhar (2016), while the third sense is extracted from the corpus. The glosses have been developed by the researcher, aligning with their uses in the corpus. Example sentences have been sourced from the Saraiki corpus. Semantic relations have been established using English Wordnet.

4.14 Verb of Social Interaction

Verbs originating from various spheres of social life, including law, politics, economy, education, family, religion, etc. Many possess a specific meaning confined to a particular domain of social life. Examples of this semantic category comprise چھوڑ، چمنڈ، دعوت کرنڈ، دفناونڈ، چھپنڈ، پرناونڈ، مل، and گنڈھنڈ.

Table 4.14 The synsets of verb 'مل'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|-----------------|------|-----------------|------------------------|----------------------------------------------------------------|
| مل | 1 | hug | verb | جپھی پاونڑ | ہک بئے کوں گل لاونڑ | اونکوں جتنا وی گھٹ گھٹ کے ملی ہاں اودر نمی لتھی |
| | 2 | meet | verb | ملنڑاں | ملاقات کرنڑاں | ٹھر گنڈیاں اکھیاں تے ٹھر گیا دل اج میکوں راہ وچ مابی گیا مل |
| | 3 | acquire, get | verb | حاصل ہوونڑاں | کسے شے دا مل جاوناں | میکوں میری محنت دا صلہ مل گئے کہ میں پاس تھی گیاں |



'مل' serves as a verb of social interaction. It is polysemous, featuring three senses. The second sense has been sourced from Shabdkosh and Akhar (2016), while the other senses have been derived from the corpus. Glosses and example sentences have been drawn from the corpus. Semantic relations have been established using English Wordnet.

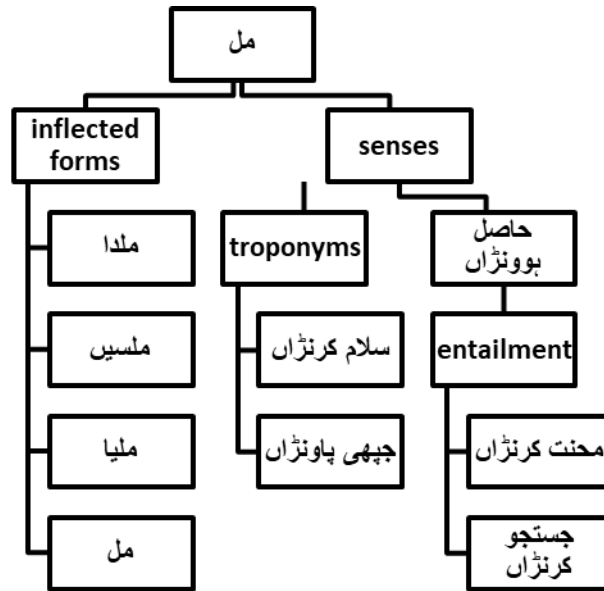
4.15 Weather Verb

This category encompasses verbs associated with weather, including وسنڑ، وسنڑ، وسنڑ، etc.

Table 4.15 The synsets of verb 'وس'

| Word | Synset | ET | GC | Senses | Glosses | Example Sentences |
|------|--------|------|------|---------------|--------------------------------|----------------------------------------------------------------|
| وس | 1 | Rain | Verb | میںہ ہوونا | آسمان توں پانی دے قطرے گرنا | اچا چیت بدلاں دے گجکار شروع تھی گئے اتے منہ زور میںہ وس پیا |

| | | | | | |
|--|---|------|------|-----------------------------------------------|-------------------------------------------------------------|
| | 2 | Live | Verb | وسنڑاں خوشی خوشی رہنڑاں، آباد رہنڑاں | اساں وس پیوسے نال وسدئیں دے نال وسدئیں دے جھوک وسوں تھئی |
|--|---|------|------|-----------------------------------------------|-------------------------------------------------------------|



‘وس’ is classified as a weather verb. It is polysemous, featuring two senses. The first sense is extracted from the Punjabi dictionary Akhar (2016), while the other sense is derived from the corpus. Glosses have been formulated by the researcher, and example sentences have been drawn from the corpus. The researcher has established semantic relations using the expansion approach.

4. Results

In the light of the analysis of the 100 most frequently found verbs in Saraiki corpus, the researchers found out that Saraiki has all the fifteen verb categories that Fellbaum (1993) has mentioned for English verbs. The first category *Verbs of Bodily Function and Care* contains six verbs namely: سم، جم، چوونڑ، ٹرنڑ، چکراونڑ، ٹھرنڑ، پگھرنڑ، ولا and پٹ. The second category *Verb of Change* has eighteen verbs that are: سنوار، کھڑ، اٹھ، پھیر، کھول، مکاونڑ، بدل، توڑ، بہر، ودھاونڑ، تھی، بولی and لکھیا، الاونڑ، روکنڑ، بلا، آکھیا، ڈسایا. The third category *Verb of Communication* has seven verbs that are: کھیڈ and نڈنڑ. The fourth category *Competition Verb* has only two verbs namely: کھا، پی and ہنڈھاونڑ. The fifth category *Consumption Verb* has six verbs namely: سٹ، رنگ and مار، گھٹ، چا، کوہونڑ، من، لگدے، سکھ، تھل، سی، پک، تراشنڑ، بنڑاونڑ، بھل and جانڑدے، پڑھ، ونج، لا، چڑھ، ٹر، وڑ، جا، آئی، آندے، نکل and گھولی. The sixth category *Cognition Verb* has six verbs that are: پٹ، سراہیا، کھل، بس، رو، پچ and چابندے. The seventh category *Emotion or Psych Verb* has six verbs that are: پچ، جیونڑ، پوندے، ہووے، پچ، رہ، ڈیکھ and سنگھ، تک. The eighth category *Creation Verb* has three verbs that are: لک and گھدی، بخش، پاتی، ونڈ، طا کرنڑ، رکھا، خریدنڑ، ویچنڑ، چھوڑ and پرناونڑ، چھینڑ، دفناونڑ، دعوت. The ninth category *Motion Verb* has eight verbs that are: گنڈھنڑ and کرنڑ، چمنڑ، چھوڑ and گجنڑ. The tenth category *Weather Verb* has only two verbs that are گجنڑ and وسنڑ. It has become obvious by analyzing the verbs that Saraiki has verbs that are homographs but have

different pronunciations and different meanings, for example, *Zabar* on first consonant means 'break' while the same verb having *Pesh* on the first consonant means 'fry'.

Unlike English, Saraiki verbs contain inflections of the person, the gender, and the tense. The simple stem / root verb 'پی' may have the inflections, for example, پیساں contains the information of person i.e., first person singular and tense i.e., future thus it means *I will drink*, پیسایں means *We will drink*, پیسیں means *You(singular) will drink*, پیسو means *You(plural) will drink*, پیسیں means *He/She will drink*, and پیسن means *They will drink*. Same verb with constant person, third person singular, for example, may have the following inflections for tenses پیتا meaning by that *drank*, پی means *drink(imperative)*, پیندا means *He drinks*, پیدی *She drinks*, پیسیں means *He/She will drink* etc.

Every verb has four inflected forms, one for each tense and one imperative. Every verb is carefully studied for the analysis. All the verbs are explored on three planes. Firstly, the researcher has tried to investigate all the possible senses and synsets of every verb. Secondly, the researcher has looked for gloss and example sentence of them. Thirdly, lexicosemantic relations of every verb are explored.

In the light of previous chapter, it is obvious that the verbs share two types of relations namely troponymy and entailment. It is difficult to maintain hyponymic relation between verbs, in contrast to that of nouns, because people would not accept that *بھجنڑاں* is *حرکت کرنڑاں* as they accept that *تاہلی* is hyponym of *رکھ* or *رکھ* is hyponym of *جاندار*.

5. Conclusion

A Saraiki corpus comprising 3 million words was developed to establish synsets for Saraiki verbs. From this corpus, a set of one hundred high-frequency verbs was selected for in-depth data analysis. Following the categorization of these verbs into fifteen semantic categories, a hybrid approach was employed to construct synsets for Saraiki verbs.

Upon thorough data analysis, it was revealed that Saraiki exhibits 15 semantic categories for verbs, aligning with Fellbaum's (1993) classification. These categories encompass verbs of bodily care and functions, verbs of change, cognition verbs, communication verbs, competition verbs, consumption verbs, contact verbs, creation verbs, emotion verbs, motion verbs, perception verbs, possession verbs, verbs of social interaction, weather verbs, and stative verbs. The study also identified two predominant types of relations among these verbs, namely troponymy and entailment.

This research lays the foundation for the potential development of synsets for Saraiki adjectives, nouns, and adverbs. The outcomes of this study can be instrumental in creating multilingual and bilingual dictionaries for Saraiki language learners. Moreover, it paves the way for the exploration of lexicosemantic relations among verbs and opens avenues for future research on Saraiki, potentially extending to other Pakistani languages like Sindhi. Furthermore, this research contributes to the eventual creation of a WordNet for the Saraiki language.

Funding: This study was not funded in any shape or form by any party.

Conflict of Interest: The author declares that he has no conflict of interest.

Bio-note:

Muhammad Awais is a student of English Language and Literature. He completed his MPhil from Sargodha University in 2021. His article titled “Code-Switching as a Marker of Identity: A Linguistic Analysis of Pakistani TV Morning Shows” has been published in RJLS. Besides, he is a firefighter in Punjab Emergency Services Department and a poet. His interests include corpus linguistics, morphology and discourse analysis.

Dr. Musarrat Azher is an accomplished Associate Professor in the Department of English at the University of Sargodha. With a keen academic focus on linguistics, her expertise lies in the fields of Corpus Linguistics, WordNet of Saraiki Language, and Language Variation.

Muhammad Farukh Arslan is a lecturer at National University of Modern Languages, Faisalabad Campus, Pakistan. He also serves as a visiting lecturer in Government College University Faisalabad (GCUF) and is enrolled in the Ph.D. program in the Department of Applied Linguistics, GCUF, Pakistan.

References

- Adeeba, F., & Hussain, S. (2011). Experiences in Building Urdu WordNet. *Proceedings of the 9th Workshop on Asian Language Resources*, pp. 31-25. Retrieved from <https://aclanthology.org/W11-3406>
- Arslan, M. F., Mahmood, M. A., Shoaib, M., Sana, I., & Zunaira, T. (2023). Morphological Description of Nouns in Shahmukhi Punjabi; A Corpus Based Study. *Journal of Positive School Psychology*, 7(3), P. 1259-1269. <https://www.researchgate.net/publication/372549062>
- Fellbaum, C. (1990). English Verbs as a Semantic Net. *International Journal of Lexicography*, 3(4), pp. 278–301.
- Fellbaum, C., Gross, D., & Miller, K. (1993). Adjectives in WordNet. *International Journal of Lexicography*, pp. 26-39
- Hasan, E., Iqbal, M., Azeemi, Q., & Javeed, A., (2015). An Online Punjabi Shahmukhi Lexical Resource. *Science International*, 25(3), pp. 2529-2535.
- Hashmi, R. S., & Majeed, G. (2014). Saraiki Ethnic Identity: Genesis of Conflict with State. *Journal of Political Studies*, 21(1), 79-101
- Kar, S., & Chakrabarty, A. (2011). Expansion of the First Hindi-Nepali Word-Net based Bi-Lingual Dictionary and the advancement of the Human-Machine Interface. *International Journal of Computer Applications*, (0975 – 8887) on Electronics, Information and Communication Engineering, pp. 8-11
- Kaur, R., Sharma, R. K., Preet, S., & Bhatia, P. (2010). Punjabi WordNet Relations and Categorization of Synsets.

- Garcia, M. I. M. (2016). Saraiki: Language or Dialect?. *Eurasian Journal of Humanities*, 1, pp. 40-53.
- Mehta, D. (2021). Part of Speech Tagging – POS Tagging in NLP | byteiota. byteiota | From Bits to Bytes. Retrieved 12 March 2021, from <https://byteiota.com/pos-tagging/>.
- Miller, G. (1995). WordNet. *Communications of the ACM*, 38(11), pp.39-41.
- Miller, G. A. (1993). Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography*, 3(4), pp. 245–264, Retrieved from <https://doi.org/10.1093/ijl/3.4.245>
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K., (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4), pp. 235–244.
- Moldovan, D. and Novischi, A., (2004). Word sense disambiguation of WordNet glosses. *Computer Speech & Language*, 18(3), pp. 301-317.
- Mushtaq, M., & Shaheen, M. (2017). The Siraiki Province Movement in Punjab, Pakistan: Prospects and Challenges. *Journal of the Punjab University Historical Society*, 30(2), pp. 139-150.
- Prabhu, V., Desai, S., Redkar, H., Prabhugaonkar, N., Nagvenkar, A., & Karmali, R., (2012). An Efficient Database Design for IndoWordNet Development Using Hybrid Approach, pp. 229–236.
- Rattna, R. (2011). Creation of Punjabi Word-Net and Punjabi Hindi Bi-Lingual Dictionary.
- Raza, G. (2016). Etymology of the Saraiki language name. *Journal of Linguistics & Literature*, 1(1), pp. 61-81
- Gul, S., Azher, M., & Sana, N. (2021). Development of Saraiki WordNet by Mapping of Word Senses: A Corpus-based Approach. *Linguistics and Literature Review*, 7(2), pp. 47-66. <https://doi.org/10.32350/llr.72>
- Shackle, C. (1977). Siraiki: A Language Movement in Pakistan. *Modern Asian Studies*, 11(3), pp. 379-403. Retrieved April 19, 2021, from <http://www.jstor.org/stable/311504>
- Shackle, C. (2015). Siraiki language. *Encyclopedia Britannica*. <https://www.britannica.com/topic/Siraiki-language>