aa
ISSN E: 2709-8273
ISSN P:2709-8265
JALT
JOURNAL OF APPLIED LINGUISTICS AND TESOL

JOURNAL OF APPLIED LINGUISTICS AND TESOL

Vol.8. No.3.2025

DISTRIBUTIONAL SEMANTICS AND TAGSET OF NOUNS TAXONOMIES: A CORPUS-DRIVEN STUDY OF SARAIKI FOLK SONGS

Dr. Iram Amjad
Assistant Professor, English
FAST National University of Computer and Emerging Sciences, Lahore, Pakistan
Email: iram.amjad@nu.edu.pk & iramamjad7@gmail.com

Dr. Zahida Mansoor
Assistant Professor, English
FAST National University of Computer and Emerging Sciences, Lahore, Pakistan
Email: zahida.mansoor@nu.edu.pk

Abstract

This study investigates the distribution of semantic networks of nouns in Saraiki folk songs through a corpus-driven approach, utilizing 3A Model for annotation and analysis. Saraiki, spoken by nearly 20 million people in Pakistan, has limited digital resources, with the iJunoon Saraiki dictionary (2017) being one of the few available corpora. To address the need for a structured lexical resource, this study aims to uncover the underlying semantic relationships among Saraiki nouns, which could contribute to the development of a Saraiki WordNet. A specialized corpus of 0.25 million words was compiled from the folk songs, *and nouns were tagged, categorized, and manually annotated using UAM Corpus Tool. Unlike traditional approaches, a discourse analysis framework was employed to examine the contextual use of nouns and identify recurring patterns in semantic relations. Specifically, relational semantics was used to analyze connections such as hyponymy, meronymy, and part-whole relations. The findings revealed that meronymic relationships were the most frequent, while antonymy was the least common occurrence. This study highlights the importance of corpus-driven methods for understanding the semantic structures of under-resourced languages and suggests that insights gained from the Saraiki noun relations could be central to creating digital tools for Natural Language Processing (NLP) and computational linguistics applications.*

Keywords: *Corpus-Driven approach, Noun Taxonomies, Semantic Networks, Saraiki Folk Songs, Word Formation.*

Introduction

Pakistan, a linguistically diverse country, is home to multiple languages, regional customs, and traditions, all creating a colorful yet unique montage of cultural identity. With the origins of Saraiki dating back to over 4500 years in the Indus civilization and considered the second largest language of Pakistan, its simple words, allow people of all provinces understand and communicate easily, attracting writers and mystic poets as well. Traditional stories play an important role in the formation and development of the lexical system of the language, representing distinctive features of informal speech, literary language, and the precise meaning of the word. These tales contain the symbolic and emotionally-evaluative tones, important for the life and history of any language at different times (Davronovich, 2022).

The Saraiki-speaking region, in particular, is rich in folklore and music, carrying a legacy that not only entertains but also embodies the cultural essence of the Saraiki people, the fourth largest ethnic group in Pakistan (Bureau of Statistics, 2017). Central to this cultural heritage is folk music, a powerful expression of regional identity and values that preserves the historical and social narratives of local communities. Folk music in Pakistan is deeply connected to regional culture and reflects the lives, struggles, and beliefs of the people, conveying emotions and stories through traditional linguistic structures (Fatima, 2024; Hemani, 2017).

Folk songs often illustrate social pictures and the unwritten history of the society generally unknown to the current generations (Khandoker, 2024; Amjad 2017), representing not only the culture of times past but that of the contemporary society as well. Sung during happy times and during sad times; they often tell a story about human condition, about finding or losing love, deceit, war, and natural disasters; the status and form of folk songs in traditional culture is constantly evolving (Dinç, 2020; Novik, 2020), preserving the essence of earlier cultures as the long-term records of history, human experience, and language.

Saraiki folk songs are linguistically rich, relying heavily on noun classifications to depict places, objects, people, and emotions, thereby offering insights into the cultural worldview (Gul al., 2021; Bashir et al., 2019). A corpus-driven approach to studying Saraiki folk songs offers a systematic way to uncover the unique linguistic features and semantic relationships within this oral tradition (Malik et al., 2022). Given that Saraiki is an under-documented and low-resource language in terms of digital linguistic resources (Hussain, 2016), this research also aims to contribute to the enhancement of Saraiki WordNet and advance Natural Language Processing (NLP) in underrepresented languages. By examining noun taxonomies through a distributional semantics framework, this study aims to explore the preservation of the linguistic structures and cultural meaning through Saraiki folksongs.

Research Problem

Saraiki folk songs, a rich store of cultural heritage and linguistic expression, remain underexplored in terms of their linguistic structures and semantic relationships. The lack of corpus-based studies with detailed analysis of noun taxonomies and their cultural significance impedes efforts in the preservation of the oral traditions of the resource-poor Saraiki language. This study addresses the gap by employing a corpus-driven approach to understanding noun use and semantic relations by analyzing the Saraiki folksongs.

Research Aims and Objectives

This study aims

1. To analyze the distributional semantics within the noun taxonomies of Saraiki folk songs.
2. To develop a tagset and taxonomy for Saraiki nouns based on corpus data.
3. To explore cultural, thematic, and syntactic patterns in folk song narratives.

Research Questions

a. What are the dominant noun categories present in Saraiki folk songs?
b. Can Saraiki noun taxonomy be represented using a domain-relevant tagset?
c. What distributional patterns reflect cultural and semantic associations in the Saraiki folk songs?

Significance of the Study

This study contributes to the growing literature on regional linguistics by focusing on Saraiki folk songs as a corpus for understanding noun taxonomies through distributional semantics. Documenting regional songs would not only preserve the language but also enrich linguistic resources by recognizing semantic structures that convey cultural knowledge and social values, crucial for cultural and linguistic studies. By integrating a culturally specific tagset for analyzing noun classifications, this research addresses the underrepresentation of the low-resource Saraiki language within corpus linguistics and NLP.

Literature Review

Research in distributional semantics and noun taxonomies in folk songs provides a foundation for understanding the linguistic patterns within specific communities. Distributional semantics, studies word meanings based on their contextual distribution in large text corpora, and is particularly useful for uncovering the meanings and relationships between nouns within folk songs. Based on the ‘distributional hypothesis’ that words used in similar contexts have similar meanings allows researchers to classify and examine cultural and emotional undertones within a language (Hussain, 2016).

As, research suggests that preservation of ethnic cultural identities, is important allowing communities to protect links to their origins, strengthening their sense of self and contributing to a stable social structure (Shah & Sahito, 2024). Folk music, globally recognized as a vessel for cultural expression, has been widely studied for its linguistic richness. Research into folk songs from diverse traditions, such as Irish, Mexican, and Native American, has shown how folk songs function as cultural narratives that reflect collective experiences and values. These songs often revolve around universal themes of joy, sorrow, and struggle offering insights through oral storytelling and vivid descriptions (Onwuegbuna, 2012). From a linguistic standpoint, Saraiki folk songs have drawn scholarly attention for their complex syntactic structures, rich metaphors, and cultural codes (Nazeer et al., 2024; Malik, 2023).

Historically, folk songs and noun taxonomies have been explored mainly within the Western languages. Early studies predominantly used qualitative methods to manually classify folk songs and their linguistic features, while recent technological advances have allowed more quantitative, data-driven analyses. Research on folk songs now increasingly recognizes them as tools of cultural resistance and identity preservation, moving beyond mere entertainment value to view them as expressions of social struggles and collective consciousness. This shift is significant in understanding how linguistic frameworks within folk songs help preserve cultural identity, a vital concern for languages at risk.

Nasir and Asif (2022) argue that the Saraiki language functions as a unifying force for southern Punjab, promoting a distinctive traditional identity within the sociopolitical struggles. However, this linguistic identity does not merely create a cultural group, but a political tool in the movement for a separate province, challenging the central Punjabi narratives. Critics, however, question the authenticity of the linguistic movement, as Saraiki is spoken across multiple regions in Pakistan, suggesting that the central demand remains socioeconomic and administrative freedom rather than linguistic separation (Iqbal, 2021).

Despite opposition from Pakistan’s ruling elite—who dismiss the movement as constitutionally disruptive or linguistically superficial—Saraiki poets and writers continue to resist through their work. Their poetry asserts cultural pride, critiques marginalization, and reinforces the demand for political recognition, making literature a key site of resistance. Folk songs also embody "folk taxonomies," classification systems shaped by cultural perspectives, which differ from scientific classifications and represent local knowledge and belief systems. Such classifications, often centered on nouns that denote people, places, and objects, are pivotal in understanding how communities view their world.

Research studies have recurrently highlighted that, although Pakistani music faces challenges due to societal deviations, thematic shifts in music have the potential to bring improvement and enable songs to compete in the modern era (Yusuf, 2024). While Nezeer et al., (2024) pointed out that this gap in methodologies could be tailored to focus on resource-scarce languages, which could address these challenges. In studying Saraiki folk songs, distributional semantics provides a framework to identify noun taxonomies specific to Saraiki culture. However, research remains limited, with some studies employing the corpus-driven approaches to analyze languages. Despite its widespread cultural use, Saraiki has limited digital linguistic resources, and most linguistic analyses rely on a relatively small corpus due to resource constraints (Gul et al., 2021). While large datasets are traditionally favored in computational linguistics, some argue that small datasets may still capture cultural meanings adequately, particularly in rich, culturally embedded corpora like folk songs (Krishnaiah, 2018). Recent computational advances position folk songs especially understudied traditions like Saraiki as vital corpora for distributional semantics. Interdisciplinary collaboration between linguists, folklorists, and NLP researchers is key to unlocking these taxonomies.

Methods and Materials

The Data

The selection of Saraiki folk songs was guided by the theoretical frameworks of distributional semantics (Lenci & Sahlgren, 2023) and the 3A Model (Wallis & Nelson, 2001). The corpus includes Saraiki folk songs obtained from classical Sufi poetry, and modern adaptations in the form of poetic compositions, and oral traditions of folk performances.

A purposive sample of 30 Saraiki folk songs (*Ek Phul Mootiye Da, Pyaar Nal Nah Sahi, Dil Mangda milay dhola, Gulshan may aashiaan hai, Qasam khuda di dar dilbar da, Na chan rol meikuun tey ker cha wafa, mai saari umraan panjh haari, sun meidi dua mola! Meikun yaar mila mola! Tere Ishq Nachaya Kar Thia Thia, Bulleya Ki Jaana Main Kaun, Ni Main Jana Jogi De Naal, Mera Piya Ghar Aaya, Ae Raatan Saanu Jaan na Dendi, Sajan Mera Makhna, Medah Ishq Vi Toon, Chal Bulleya Chal Othay Chaliye, Jithay Khairan Na Puchhn Walay, Rohi Tu Ratian Kithay Guzaariyan, Fareed Da Meharban Dhola, Toon Saadi Saans Saans Vich Vasda, Ajj Kaleyan Mendi Dhola Nay, Zindagi Da Ki Baney, Dukhda Apna Das Ke Rowan, Mein Vi Pagal, Tu Vi Pagal, Wey Lagda Na Ji Dil Mera, Duniya Da Dukh Sajda Rihnda, Kaheen Vi Dard Na Hove, Main Saraiki Bolna Aan, Mein Mar Jawan Tey Rola Paween),* written by *Bulleh Shah, Khawaja Ghulam Fareed* and *Shakir Shujah Abadi* were selected based on their relevance to the research questions and theoretical framework. Although these texts differ in authorship and are written in diverse historical periods, they exhibit the hallmark features of the Saraiki folk song genre which renders them appropriate for a corpus-based investigation of distributional semantics and hierarchical structures of nouns.

The poets represent a strong tradition of Saraiki folklore well-known for its linguistic complexity, cultural relevance, and thematic variety. The inclusion of these poets covers a significant temporal range, offering historical insights into the development of Saraiki folk poetry from the late 1600s to the present. Bulleh Shah's poetry dates back to the late 17th early 18th century, a time of socio cultural and spiritual changes, while Khawaja Ghulam Fareed's work towards the end of 19th century when the subcontinent was under colonialism and the cultures were changing. These works provide classical historical accounts while Akram Niazi and Shakir Shujah Abadi's poetry from late twentieth century provides historical perspectives of modern socio-politics, combining traditional poetic forms with contemporary sensibilities. This range makes it possible for a thorough investigation of processes of continuity and change in all aspects of the Saraiki folk song tradition. Each poet has his share in terms of linguistic complexity and artistry. Their folk songs capturing the intricacies of the Saraiki language are appropriate for the study of the distributional semantics and tagset of nouns taxonomies. The differences in their vocabulary, idioms, and stylistic features contribute in making the corpus interesting and suitable to address the linguistic features of Saraiki folk poetry.

The inclusion criteria of the Saraiki folk songs included temporal range, linguistic richness, cultural and thematic relevance, representation of Saraiki dialects, availability of authentic texts and recordings, and length and complexity. Folk songs from different time periods were included to depict the changes over time in the semantics and taxonomy of the nouns. Selection of songs with authentic lyrics or recordings helped in the transcription accuracy ensuring consistent semantic annotation development. Songs of moderate length and complexity were prioritized to balance the depth of analysis with the feasibility of manual and computational coding techniques. The selected songs went beyond simple noun use to appreciate the song context ensuring that the dataset offered sufficient semantic diversity for the study of noun taxonomies, and included primary Saraiki cultural aspects of love, identity, spirituality and folklore. This criterion supported in make certain that the songs represented the relevant cultural aspects essential in the study of contextual semantics. The folksongs also included variations from the different Saraiki dialects, representing linguistic diversity within the Saraiki-speaking regions.

Steps for Data Analysis

A specialized corpus collection of Saraiki folk songs was created by gathering lyrics through oral performances and written collections to have actual representation of Saraiki cultural diversity and dialectal variation. The songs were transcribed into a machine-readable format when they were still in oral form and not already in text form. Typically, this meant settling transcription inconsistencies and labeling linguistic levels in the transcription. According to Lenci and Sahlgren (2023), coding techniques were developed mainly at three levels: lexical annotation, contextual coding, and 3A Model application. First, in lexical annotation, the nouns were located and tagged with the predetermined Saraiki tagset based on linguistics work done in several regions. Second, in contextual coding, distributional semantic methods were implemented such as co-occurrence analytical measures and vectors based modeling were used to study the semantic vectors. Last, for 3A Model application, the annotative, algorithmic and analytic methods were used to classify the nouns into taxonomies depending on distributional characteristics. To ensure accuracy and rigor, annotated data was reviewed by some linguists of the Saraiki language while the coding techniques were iteratively refined based on their feedback (see Bashir & Conners, 2019; Wallis & Nelson, 2001).

It is the consistent method of selection and analysis of songs that keeps this investigation simple, elaborate and ethnically relevant as well as contributes to further development of understanding the language of Saraiki folk songs.

Conceptual Model for Computing Tagset for Siraiki Nouns and Semantic Network Analysis

The present descriptive work is grounded in distributional semantics as proposed by Lenci and Sahlgren (2023) and the 3A Model of Wallis and Nelson (2001). This theoretical framework explains the meanings of Saraiki nouns from distributions of contextual co-occurring patterns. To this effect, it is argued that a word is endowed with connotation through the co-occurring and the distributional elements surrounding the word in a linguistic context. It is of the same notion as the theoretical approach of Tognini-Bonelli (2001) inductive linguistics, explaining where one learns linguistic rules/generalizations from language in actual use. As a result, corpus-driven research into semantic relations within the networks of Saraiki nouns was carried out. Therefore, the tagset for Saraiki nouns was constructed following conceptual-lexical structure theories which advocate for the use of relational semantics (for instance: hyponymy, meronymy, antonymy, synonymy, and nominal metaphors.) and grammatical categories (singular, plural, masculine, and feminine) that stress morpho-syntactic constructions. Using these theoretical lenses, the study highlights some of the patterns present as noun usage that can be said to be coherent with the structure of Saraiki lexical classification, which serves the theory of linguistics and has practical uses in NLP for low resource languages. This results-driven method of analysis uncovers underlying meaning without preconceived structure.

In addition, the 3A Model consists of annotation, abstraction and analysis, which enables effective specialized corpus annotation and processing. In the annotation stage, nouns are labeled and categories are imposed on the nouns, while in the abstraction stage, the information is arranged to suit analysis using different models such as tables or graphs that explain different aspects of the research. Finally, in the analysis phase, the developed hypotheses are tested and out of the Saraiki noun tagset some trends are revealed. These frameworks integrate other strategies that are useful in understanding the noun taxonomies of Saraiki and their semantics.

Further, this study utilizes aspects of relational semantics to explicate the interrelations between Saraiki nouns, thereby extending the frontiers in the field of Lexical Semantics. The analysis is corpus-driven and as such it does not begin with given categories but is data driven increasing the probability that the tagset for Saraiki nouns is based on how those nouns are actually used in oral tradition discourse. The objective with the investigation of co-occurrence relationships in the folk songs corpus, a specialized corpus is to determine the semantic networks of Saraiki nouns and how the nouns interrelate within discourse. This is in contrast to the classical lexicography approach of doing which underlines the advantages of distributional semantics in the building of lexical resources for under-resourced languages like Saraiki.

The understanding derived from both relations and distributional analysis of Saraiki nouns facilitated in realization of the wider objective of constructing a Saraiki WordNet, a systematic lexical database suitable for application in different areas of NLP. Focusing on the tagging and categorization of nouns employing UAM CorpusTool, the study is in line with the efforts of developing corpus-based tools for underrepresented languages which can later be used for improving natural language processing tools such as machine translation, information retrieval and automatic summarization for Saraiki.

This study integrates multiple theories on distributional semantics to depict the networks and the meanings of the Saraiki nominal forms appearing in folk songs. The study offers both theoretical comprehension of the lexical and semantic relations of Saraiki as well as practical assistance in creation of computational resources through corpus and discourse approaches.

Linguistic Features in Saraiki Folk Songs

This study of folk songs in regional language of Pakistan provides a rich lexical resource giving insights into Saraiki culture, way of life, and language features itself. Saraiki, which is an Indo-Aryan language, is widely used in various parts of Punjab, Pakistan. There are numerous folk songs composed in Saraiki language, which makes an emphasis on how nouns are emotively used through relational semantics. This specialized corpus-driven study aims at investigating and elucidating the distributional nature of noun classes in Saraiki folk songs. Figure 1 visually showcases a taxonomy created with Tableau in four quadrants. It charts the distribution of noun classifications based on grammatical gender and number (singular/plural and masculine/feminine) as well as semantic dimensions (relational semantics and morpho-syntactic stress). The visualization integrates the corpus results and exemplifies how language conveys emotion and culture systematically.

Figure 1. Semantic Noun Taxonomy Quadrants in Saraiki Folk Songs

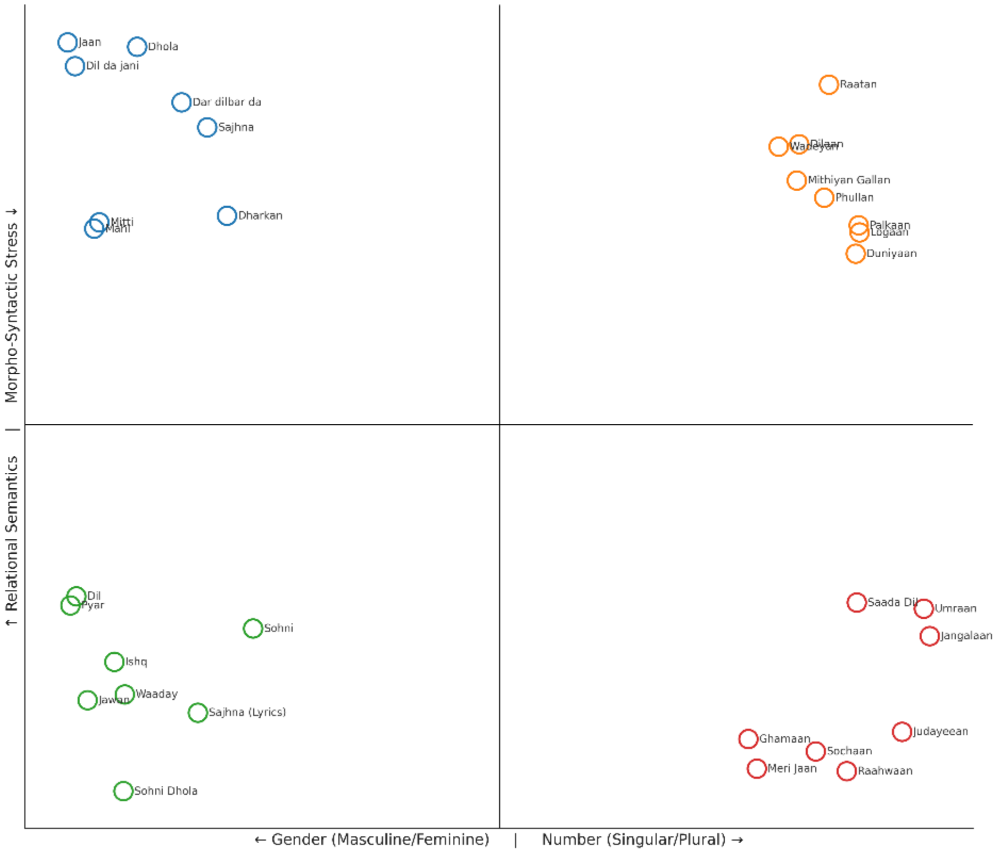

Table 1 presents relational semantics of hyponymy, meronymy, antonymy, synonymy and nominal metaphors in Saraiki folk songs.

Table 1: Relational Semantics in Saraiki Folk Songs

Relation Type	Tag	Examples from Folk Songs	Description
Hyponymy	**REL-HYP**	**Hypernym:** *Dharti* (earth) **Hyponyms:** *Chann* (moon), *Phul* (flower) **Hypernym:** *Gulzaar* (garden) **Hyponyms:** *Gul* (Flower) **Hypernym:** *Dil* (Heart) **Hyponyms:** *Dharkan* (Heart beat) **Hypernym:** *Phal* (fruit) **Hyponyms:** *Angoor* (grape), *Aam* (mango)	General to specific relationships used to describe elements of nature or love. Specific fruits mentioned as part of larger cultural metaphors for abundance and sweetness.
Meronymy	**REL-MER**	**Whole:** *Raat* (night) **Parts:** *Chann* (moon), *Tara* (star) **Whole:** *Zameen* (Earth) **Parts:** *Mitti* (Soil) **Whole:** *Duniya* (world) **Parts:** *Dil* (heart) **Whole:** *Dil* (heart) **Parts:** *Dhadkan* (heartbeat), *Khoon* (blood)	Parts of nature (like the moon and stars) often used to describe the entirety of night. Mitti (soil) is a part of Zameen (earth/land), symbolizing attachment to one's homeland. Dil (heart) is a part of a larger metaphorical concept of the inner emotional world, signifying that the heart (part) is used to signify the entire emotional world (whole). Parts of the body, particularly the heart, used to convey deep emotions.
Antonymy	**REL-ANT**	*Khushi* (happiness) vs. *Gham* (sorrow) *ro-ro* (crying) and *hasde aan* (laughing) *asaan* (easy) vs *mushkil* (difficult) in *bicharna hai asaan milna mushkil* *ojriian* (devastated) vs *vasaa* (establish) in *hayee wasatt panj tan da ojriian kuun vasa mola!* *ghussa* (anger) and *sukh* (peace/comfort) *pyaar* (love) and *ghussa* (anger) in *Pyaar nal* *Milna* (meet) vs. *Judai* (separation) *Dil da jani* (close to heart)/ *unjaane* (unknown/stranger) in *jaan saanjaa kay dil da jani ban gaya ajj unjaane* *Beemaaraan* (Sick) and *Shaafa* (Healing) *Fulaad* (Steel) and *maum* (wax)	Opposites used to reflect the contrasts of emotions and life experiences. Emotional contrasts between sadness and happiness. Reinforces the healing power of love through repetitive antonyms
Synonymy	**REL-SYN**	*sukh* (peace) and *qarar* (comfort) *Sohna* (beautiful), *Haseen* (pretty) *Sajjan* (beloved), *saathi* (companion) in *zindagi da hei pandan ookha nahi koi sajjan saathi* *Kaali raat andhaari* (black night, dark and terrible night)	Different words with similar meanings are used to amplify the emotional and aesthetic effect. Share close conceptual meaning in the context of emotional relief. Describing elements of nature
Nominal Metaphor	**REL-MET**	*Chann* (moon) as metaphor for the beloved *Pyar* (love) as a river *Waaday* (Promise) as a *Maum* (wax) in *maum kay waaday pighal rahay hain* *Chandar* (Moon) in kaali raat andhari hai koi chandar charha de mola!	Metaphors drawn from nature or everyday life to express complex emotions or relationships. Metaphors drawn from natural surroundings to express hope/miracle.

ISSN E: 2709-8273
aa
ISSN P:2709-8265
JALT
JOURNAL OF APPLIED LINGUISTICS AND TESOL

It is evident from Table 1 that there is a connection between language and various aspects of nature, emotions, and social life. For example, in hyponymy, general phrases are used in the songs like earth and love and are then followed by specific phrases like moon and flower. The term *Dharti* (earth) is represented with hyponyms such as *Chann* (moon) and *Phul* (flower). This suggests a greater world of nature and life, and particular items within that world. Likewise, *Dil* (heart) has *Dharkan* (heartbeat) as its metaphorical hyponym, illustrating emotional intensity. *Meronymy* also plays a significant role in describing the world, where parts symbolize the whole. An example is, while discussing the night (*Raat*), moon (*Chann*) and star (*Tara*) can be used to represent the whole. Smaller elements are used to reflect larger notions/concepts related to nature and other emotional states. The merging of whole/part relationships enhance the metaphorical meanings of the folk songs highlighting emotional intensity of lover and beloved (see also Amjad, 2017).

In addition, these folk songs also serve to amplify the emotions manifested in their lyrics through the use of antonymy, intersecting between *Khushi* (happiness) and *Gham* (sorrow) or *Pyaar* (love) and *Ghussa* (anger). These actions exemplify the range of feelings one is able to experience and helps reinforce themes like conflict, yearning, and resolution. The songs' intensity is further highlighted by the emotional divergence rendered through the use of words like *Pyar* (love) and *Mohabbat* (affection) or *Ishq* (devotion). Moreover, the imagery of *Chann* (moon) as the beloved and *Pyar* (love) as a river showcases how ordinary life, as well as nature, are metaphorically used to convey intricate feelings of yearning and devotion. As a result of the metaphorical language, the vivid imagination of nature and things that life brings makes these emotions deep transforming simple sights into unimaginable art that exhibits human connection and longing. Saraiki folk songs have a rich tapestry of the grammatical categories of number (singular/plural) manifesting emotional states (see Table 2).

Table 2: Grammatical Categories: Number (Singular/Plural)

Category	Tag	Examples from Folk Songs	Description
Number (Singular/ Plural)	**NP.SG**	*Multan (a single city)*	Singular proper noun denoting a specific cultural and geographic entity.
	NP.PL	*Saharanpur (multiple cities)*	Plural proper noun indicating collective geographic locations.
	NPC.SG	*Dil* (heart)/ *Dilaan* (hearts) in *Ṭuṭṭe hue dilaan noon qaraar mil jaande nen*	Singular form expresses personal emotions; plural form often represents shared or collective feelings.
	NPC.PL	*mithiyan gallan (sweet conversation)* in *dhola karda mithiyan mithiyan gallan*	
	NADJ.SG	*Log* (person)/ *Logaan* (people)	
	NADJ.PL	*jihyaan* (yours) in *Tere jihyaan howe hamsafar*	Indicates singular or plural forms in relation to social or collective actions in folk songs.
		Phul (flower)/ *Phullan* (flowers)	
		Mithha (sweet - singular)	
		Mithhe (sweet - plural)	Singular and plural use in nature references, especially in metaphorical expressions of beauty.

The grammatical difference in singular and plural forms in Saraiki folk songs, as shown in Table 2, works as a marker of individual versus collective emotions within the distributional semantics and the 3A Model (Wallis & Nelson, 2001). Singular proper nouns (e.g., *Multan)* and plural proper noun (e.g., *Saharanpur*) indicate a specific cultural and geographic entity or collective geographic locations. Nouns in singular form such as *dil* (heart) operate at the domain of personal, self-reflective expressions while its pluralized form *dilaan* (hearts) evokes collective emotions which help in sustaining social bonds. Relational semantics, especially the representation of meronymic and hyponymic relations in folk songs, also resonates with this pattern. Additionally, *mithiyan gallan* (sweet conversations) is another abstraction whose pluralization demonstrates the morpho-syntactic strategy of intensification, which is common in oral cultures where meaning is reinforced through repetition and plurality. The shift from *phul* (flower) to *phullan* (flowers) similarly illustrates the semantic change from appreciation of singular beauty to a metaphor of abundance, which fits folk taxonomy theories. This study has adopted a corpus driven approach combined with cognitive linguistics to emphasize how the grammatical number in Saraiki marks not only quantification but also constructs narratives of emotion and culture pivotal to folk songs discourse. Saraiki folk songs expands on the grammatical categories involving gender (masculine/feminine) as an important category (see Table 3).

Table 3: Grammatical Categories: Gender (Masculine/Feminine)

Category	**Tag**	**Examples from Folk Songs**	**Description**
Gender (Masculine/ Feminine)	**NP.M** **NP.F**	*maida* in *maida wei manda rang roop saarh sattayiaa aayee, Dhola* (beloved man) in *mera dhola chail chabila* and *dhola saraiki changay rakhda ae shoq, Dil da jani* in *bole hans hans dil da jani Mahi* in *raati dihin mahi di apne,* *Sajhnaa* (beloved woman) in *ek teri sajhnaa lor ae*	Gendered nouns used to convey gender roles and relationships in folk songs.
	NP.MF	*dar dilbar da* (Dear beloved or O beloved heart) in *qasam khuda di dar dilbar da hargiz chorh na wesoon* *Jawan* (young man)/*Sohni* (beautiful woman)	Masculine terms often relate to strength, while feminine terms focus on beauty and affection.

Gendered nouns are used to depict rigidly defined roles and their associated sentiments in Saraiki folk songs, as shown in Table 3. Within distributional semantics (Lenci & Sahlgren, 2023) and the 3A Model (Wallis & Nelson, 2001), the presence of masculine and feminine nouns reflects pronounced patterns of discourse in gendered folk discourse. Co-occurrence of *Dhola* (beloved man), *Mahi* and *Dil da jani* (heart's beloved) with terms signifying power, appreciation, and action, nourish the prevailing masculine identities in the cultural narratives. On the other hand, female terms like *Sajhnaa* (beloved woman) and *Sohni* (beautiful woman) appear in frames pertaining to beauty, emotional responsiveness, and affection that corresponds to the analysis of hyponymy and meronymy in folk discourse on the relational semantics framework. The recurrent use of *dar dilbar da* (dear beloved heart) shows where subtle semantic meanings are mapped onto metaphorical language concerning a distribution, whereby emotional closeness is expressed using metaphors. This links gendered language with broader culturological notions of love, desire, and dedication.

Within cognitive linguistics, such classifications of gendered nouns help in the development of folk taxonomy where masculine terms prevail in action-centered relational semantics, and feminine nouns fall within the affective and aesthetic realms. Employing specialized corpus, this study uncovers systematic gendered lexical patterns that exist within the cultural practices as well as the morpho-syntactic prominence of emotion in the Saraiki oral tradition. Table 4 presents the significant morpho-syntactic stress on noun constructions appearing in these folk songs showcasing affection and emotional intimacy.

Table 4: Morpho-Syntactic Stress on Noun Constructions

Construction Type	Tag	Examples from Folk Songs	Description
Possessive Construction	**N.POSS**	*Meri jaan* (my life) *Teḍa deedaar channaan* (your sight is like moonlight) *Saada dil* (my heart) in *wei saada dil ban kay turr gayain ae*	Expresses possession and relational intimacy.
Gendered Nouns	**N.GEN**	*Sohni dhola* (beautiful beloved)	Gender distinctions add nuance to the role of characters or relationships in the songs.
Plural Construction	**N.PL** **N.PL.EMO**	*Logan diyan gallan* (talk of people) *Pahaaṛaan* (mountains) *Jangalaan* (jungles) *Raahwaan* (paths) *judayeeaan* (separation) *taqdeeraan* (luck) in *judayeeaan taqdeeraan de naal* *sochaan* (thinking) in *har ik raat sochaan tey weindhi whaa* *ghamaan* (sorrow) in *aehuu haa ghamaan da pau yehsaan* *umraan* (long life) in *mai tehdiaan aasharian the umraan nabhayee ae* and *mai saari umraan panjh haari*	Plural form used to emphasize the collective or societal context of the action. Plural form used to describe expansive natural landscapes. Plural nouns are used to express shared emotional states and cumulative experiences in relationships.

Table 4 outlines the morpho-syntactic patterns of Saraiki folk songs, indicating particularly the way noun phrases carry cultural, emotional, and relational meanings. Phrases like *Meri jaan* (my life) and *Saada dil* (my heart) show possessive constructions that signify affection on a deep personal level, strengthening the narratives of ownership and loyalty. Such constructions belong to the relational semantics domain where having something goes beyond mere possession to involve deep emotional connections. Classifying nouns and verbs like *Sohni dhola* (beautiful beloved) reveal the ethnocentric constructions of sense of gender, which adds deeper meaning through choice of words. The plural forms *judayeeaan* (separations), *sochaan* (thoughts), and *ghamaan* (sorrows) embody singular emotions to mark a shared feeling, moving beyond individualistic peripheries to societal parameters. This fits the distributional semantics and relational morphology theory, which postulates that patterns in language expose collective culture. Using the 3A Model (Wallis & Nelson, 2001) allows one to study folk songs within a systematic framework of marking, categorizing, and analyzing nouns provided the specialized corpus of folk songs. Morpho-syntactic stress in folk songs serves to highlight the function of language structures in embedding cultural and emotional memory, which is beneficial for computational linguistics, such as constructing a Saraiki WordNet.

Conclusion

This study offered a corpus-based examination of the distributional semantics of nouns in the folk songs of Saraiki, contributing to the sociolinguistic reality of this neglected language. Applying the 3A Model and relational semantics, this study classified noun taxonomies in a systematic way so that patterns of semantics noticed in hyponymy, meronymy, antonymy, synonymy, and nominal metaphor were identified. The findings demonstrate the importance of noun constructions in aspects related to culture and emotions encoded in folk songs, particularly those resulting from morpho-syntactic stress, such as the possessive adaptations, plural forms, and gendering of nouns. Such linguistic phenomena represent and construct personal and social memories, as well as the mythology of Saraiki folk ethnos.

This study has noteworthy implications from the linguistic, cultural, and computational aspects. This analysis enhances our understanding of the interfaces between language and emotion or culture, and underscores the value of folk songs as a source of language heritage. Additionally, this study assists in building a Saraiki WordNet, which is useful for Natural Language Processing (NLP) of lesser-known languages. The methodological framework employed - corpus-driven annotation juxtaposed with quantitative research - provides a template for the study of other regional languages with scanty digital presence.

Our study expands the theoretical frontiers of distributional semantics and noun taxonomy under folk songs through the contribution of classifying them as emerging traditions. It is also methodological due to the application of corpus-driven annotation to an under described language, hence, creating a framework for annotation and analysis. From a computational perspective, the study will serve as a foundation for developing NLP resources aimed at improving the electronic representation of the Saraiki language. In the sociocultural context, it emphasizes the role of folk songs in maintaining the linguistic identity of a people and advocates for their recognition as linguistic heritage.

The study has certain limitations too. Even though the corpus is likely extensive, it does not seem to include the dialectal differences of Saraiki. There is also a problem of an assumption with any kind of 'counting' done with UAM CorpusTool because of the reliance on classic manual annotation. And even though the study does trace some linguistic constituents and structures of the folk songs, it refrains from ‘digging deep’ into syntactic parsing that could shed more light into morphosyntactic variation. At last, the computational resources available to the researcher capped the full adoption of the NLP tools which is unfortunate because it is an area that needs further exploration in other studies.

Multi-faceted research can attempt to create a more representative corpus by including more dialects of Saraiki as well as historical documents. Creating a Saraiki WordNet is important, as is developing other NLP tools for the language, and will be a focus in one of the subsequent studies. Besides looking into patterning of other South Asian languages, these studies may rely on more advanced methods of syntax parsing and artificial intelligence based language models to perform a more thorough analysis of the linguistics. At the same time, sociolinguistic studies can examine the impact of folk song narratives on the current language of the Saraiki speakers.

To conclude, this study highlights the importance of Saraiki folk songs as a verbal and cultural relic, which shows the interconnection between language, people, and traditions. It also serves as a starting point for future work on the safeguarding and modern technology application of neglected languages through the inclusion of corpus linguistics, computer science, and cultural studies.

References

Amjad, I. (2017). A Multimodal Analysis of Qawwali: From Ecstasy to Exotic Trance. *Linguistics and Literature Review (LLR), 3*(1), pp. 12-25. https://journals.umt.edu.pk/index.php/llr/article/download/263/259

Bashir, E., Conners, T. J., & Hefright, B. (2019). A descriptive grammar of Hindko, Panjabi, and Saraiki. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9781614512257

Davronovich, J. R. (2022). Lexical Features of Folklore. *EPRA International Journal of Research and Development (IJRD),7*(2), pp. 56-58. https://eprajournals.com/IJSR/article/6615

Dinç, M. (2020). What happened to folk songs: About folk songs changed for political-ideological reasons. *Folklor/Edebiyat*, *26*(103), pp. 483–508. https://doi.org/10.22559/folklor.1246

Fatima, P. (2024, February 1). Importance of folk music for the culture of Pakistan. Republic Policy. Retrieved July 6, 2025, from https://republicpolicy.com/importance-of-folk-music-for-the-culture-of-pakistan/

Gul, S., Azher, M., & Nawaz, S. (2021). Development of Saraiki WordNet by mapping of word senses: A corpus-based approach. *Linguistics and Literature Review*, *7*(2), pp. 46–66. https://doi.org/10.32350/llr.72/04

Hemani, S. (2017). Music: Pakistan. In S. Joseph (ed.), *Encyclopedia of Women & Islamic Cultures Online*. Brill. https://doi.org/10.1163/1872-5309_ewic_COM_002126

Hussain, S. S. (2016). The Growth of Saraiki Language. *Pakistan Journal of Social Sciences*, *36*(1), 387-396. https://pjss.bzu.edu.pk/index.php/pjss/article/view/424

iJunoon. (2017). Retrieved July 6, 2025, from https://www.ijunoon.com/saraiki/

Iqbal, S. F. (2021). Dual aspects of demand of province of South Punjab: Redefining federalism. *Sir Syed Journal of Education & Social Research,* *4*(3), pp. 18–27. https://doi.org/10.36902/sjesr-vol4-iss3-2021(18-27)

Khandoker, N. (2024). becoming-woman: exploring decolonial feminist possibilities with Bhawaiya folk songs of Bengal. *Feminist Review, 138*(1), pp. 20-40. https://doi.org/10.1177/01417789241280499

Krishnaiah, K. (2018). The role of folk songs in social movements: A case study on the separate Telangana state formation in India. *European Journal of Multidisciplinary Studies, 3*(2), pp. 124–133. https://doi.org/10.26417/ejms.v7i2.p124-133

Lenci, A., & Sahlgren, M. (2023). *Distributional Semantics*. Cambridge University Press. https://doi.org/10.1017/9780511783692

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*(II), 34–63. https://doi.org/10.52015/numljci.v21iII.261

Malik, M.H., Ghous, H., Ahsan, I., & Ismail, M. (2022). Saraiki language hybrid stemmer using rule-based and LSTM-based sequence-to-sequence model approach. *Innovative Computing Review*, *2*(2). pp. 18-40 https://doi.org/10.32350/icr.0202.02

Nasir S, & Asif S I. (2022). Rewriting History of Saraiki Region: A Socio-cognitive Critique of Aslam Javed and Ashraf Javed Malik's Poetry. *Research Journal of Language and Literary Studies*, *2*(1), 32-52. https://www.researchgate.net/publication/384438517_Rewriting_History_of_Saraiki_Region_A_Socio-cognitive_Critique_of_Aslam_Javed_and_Ashraf_Javed_Malik%27s_Poetry

Nazeer, M., Musarrat Azher, Azhar Pervaiz, & Iqra Yasmeen. (2024). Developing Lexico-Semantic Relations of Saraiki Nouns: A Corpus-Based Study. *University of Chitral Journal of Linguistics and Literature*, *8*(I), 162-182. https://doi.org/10.33195/

Novik, A. (2020). Gjirokastra Folklore Festival as the main ritual event in Albanian cultural life at the beginning of the 21st century. *Yearbook of Balkan and Baltic Studies*, *3*(1), pp. 157–182. https://doi.org/10.7592/ybbs3.08

Onwuegbuna, I. E. (2012). Music as an embodiment of culture and philosophy: A survey of Nigerian folk songs. In C. K. Ikwuemesi (Ed.), Astride Memory and Desire: Peoples, Cultures and Development in Nigeria (pp. 291–309). ABIC Books.

Pakistan Bureau of Statistics. (2017, May 28). Census 2017 language data (Report No. P70-140). https://defence.pk/pdf/threads/census-2017-language-data.560777/

Saleemi, F. J., Asghar, M. N., Iqbal, S., Chaudhry, M. U., Yasir, M., Bazai, S. U., & Khan, M. Q. (2021). A basic parts of speech (POS) tagset for morphological, syntactic and lexical annotations of Saraiki language. *Journal of Applied and Emerging Sciences*, *11*(1), pp. 77–88. https://doi.org/10.36785/jaes.111459

Shah, M. A., & Sahito, M. S. (2024). Cultural Institutions in Pakistan: Promoting Cultural and National Identity. *Pakistan Languages and Humanities Review*, *8*(2), pp. 377–391. https://doi.org/10.47205/plhr.2024(8-II)33

Tognini-Bonelli, E. (2001). Corpus linguistics at work. Philadelphia: John Benjamins. https://doi.org/10.1075/scl.6

Wallis, S., & Nelson, G. (2001). Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery*, *5*, 305–335. doi:10.1023/A:1011453128373

Yusuf, A. (2024, December 31). From Vital Signs to Young Stunners: The ever-evolving sound of Pakistan. The Gazelle. https://www.thegazelle.org/issue/267/ever-evolving-sound-of-pakistan