

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

<https://llrjournal.com/index.php/11>

**MANEUVERING ROLE OF ADJECTIVES IN PUNJABI
GRAMMAR: A CORPUS BASED COMPUTATIONAL
GRAMMATICAL ANALYSIS**



Sarwat Sohail^{*1}, Dr. Ghulam Ali², Ansar Ali³

^{*1} *PhD Scholar, CeLTS, Iqbal Open University
Islamabad, PAKISTAN*

² *Associate Prof. CeLTS, Iqbal Open University
Islamabad, PAKISTAN*

³ *Lecturer CeLTS, Iqbal Open University Islamabad,
PAKISTAN*

^{*1}*sarwatsohail7@gmail-com, ²g.ali@aiou.edu.pk,
³ansar.ali@aiou.edu.pk*

Abstract

This study presents a corpus-based computational analysis of Punjabi prose written in Shahmukhi script, aiming to explore grammatical patterns employed in this language. Specifically, a unique appearance of adjective is noted, traced and described in the study. Usage-Based Model of Grammar, specifically Bybee (2001) and Langacker (1987), is used here for descriptive analysis of linguistic components of Punjabi. A corpus of approximately 3 million words is compiled from contemporary Punjabi fiction to ensure representativeness and diversity of styles. The research integrates computational methods with corpus linguistics to examine how grammatical structures emerge from actual language use. The methodology involves systematic preprocessing, including text normalization and tokenization, followed by the development and application of a Part-of-Speech (POS) tagging model adapted for Shahmukhi Punjabi. The findings are expected to reveal dominant grammatical constructions, high-frequency lexical classes, and patterns of morphological productivity, as is set through framework, specific to Punjabi fiction. Specifically, the role of adjective determining the gender and number is noted. As adjectives like 'patla', 'lamman', 'chota', 'چھوٹا', 'المان', 'پتلا', and many more, display an interesting role and show gender and number differences. This research contributes to the limited insights into Punjabi grammatical organization and by developing foundational tools for future natural language processing. Study has brought out striking features of the native language and it is a stepping stone for researches in computational analysis of minor neglected languages.

Key Words: *Adjective, corpus-based, computational analysis, Punjabi fiction, Shahmukhi script, Usage-Based Model, descriptive analysis, representativeness, corpus linguistics, preprocessing, text normalization, tokenization, POS tagging, natural language processing.*

Introduction

Punjabi is an indigenous language in Punjab: that makes almost 45% of the whole Pakistan. (Crystal, 2000) argues, “When a language disappears, cultural identity, historical memory, and traditional knowledge disappear with it”. He adds “Language loss is not only linguistic loss; it is cultural and intellectual extinction.” Like many other local languages, it has also been neglected in the wake of urbanization, modernization advancement and literacy. As a Sociolinguist, (Nettle, Daniel & Romaine, Suzanne, 2000) connect language extinction with colonialism, globalization, and political domination. And opine, “Indigenous languages suffer because dominant languages gain institutional power.” In the article “Postcolonial Linguistic Marginalization: The Kalasha Language in Pakistan’s National Discourse.” Urooj, Nadia, Rafaqat Hussain Shah, & Zafar Iqbal Bhatti.(Nadia,2025) (academia.edu.pk) portray how Indigenous languages in Pakistan suffer

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

because of: educational exclusion, lack of media representation, state neglect, dominance of Urdu and English. The article quotes, “Indigenous language decline is often caused by political structures...” (Bromham, 2023), (al., 2021), (Akhunzada, 2026), (Issa, 2023), (Younus, 2023), (Literature. 1986), (Ngũgĩ, 1986).

(Phillipson, 1992), (Unesco), and UNDRIP Text plead for the revival of endangered indigenous languages in their own ways respectively.

Punjabi poets of the past are very well known all over the world. Their poetry is rich, both in content and style. Contemporary writers have also added much in literature. Research work is mostly literature oriented. Whatever is done, is related to history and text. Linguistic part is neglected in this language.

Gone are the days when people used to follow traditional ways of research in language. Modern communication depends fully on computational techniques of analyzing language and linguistical implications. These ways and methods have improved the speed, accuracy, security, efficiency and presentation of data and its analysis. AI, NLP, Machine Learning, Data Compression Techniques, Encryption and Cyber-security, Cloud Computing and Big Data Analytics are a few to mention and avail globally and very effectively.

Computational Grammatical Analysis of Punjabi (Shahmukhi) Fiction in a corpus of contemporary language data and narrowing down the topic, role of adjective in determining the gender and number position of the noun is focused here. Very interestingly adjective in Punjabi displays this behavior and intriguingly runs through the fabric of text. Previously researches on adjectives in Punjabi show how are they derived from nouns by use of suffixes but this very aspect of their existence has never been brought to light. Adjectives keep on changing with gender and number of the nouns following them.

Research Questions

1. What is Corpus Compilation?
2. What is preprocessing?
3. What is tokenization?
4. How to do Multi-level Annotation
5. Theoretical Interpretation
6. How to do Corpus based descriptive analysis of units of grammar/adjectives, with NLP tools.

Literature Review

Punjabi displays all the characteristics of a mature culturally rich, having very established poets and writers, with universal thoughts and scholarly literary works. In the past. Linguistically, it is being worked upon by a few, recently. Very basic ideas of grammar and morphology have been explored and described by some researchers. Researchers like Gill, 1969; Malik’s 2006; Meenu, 2007; Rupinderdeep, 2010 have found out that there are 65 phonemes in Punjabi including 49 consonants and 16 vowels. Arslan and Mahmood (2021)’s study focuses on the idea of sound shift in Punjabi language and highlights number of sounds. Islam (2011) conducted study on the morphology of loanwords in Urdu. Hussain, (2018) has studied nominal markers of Punjabi nouns. Adjectives are derived from nouns generally. In inflectional morphological operation the morphemes such as /i/, /ia/, /vi/, /ae/ and /la/ are added at the end of the words like ‘Azad’ آزاد and Udaas, اداس the masculine nouns and as a result adjectives are formed as ‘Azadi’, آزادی and ‘Udaasi’ اداسی .

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

Hussain(2018) continues to tell, “The second pattern which is observed during the analysis that adjectives are also derived from nouns by joining the morpheme *vi*, ‘وی’ to make adjective like the words *afsana* ، افسانه ، is changed in to افسانوی *afsanvi* and رومانی *romani* to رومانوی *romanvi* which is adjectives. It shows the pattern that *vi*, وی’ is attached to derive the adjectives.” Researches on Shahmukhi Punjabi linguistics are few and far between, can be counted on finger tips. But still are substantially very significant in initiating these studies though mostly in traditional way.

Punjabi is also the native language in Indian Punjab, with Gurmukhi script. It has travelled with immigrants to other parts of the world, most prominently to United Kingdom, Canada and Australia, it has been strengthened there has been enjoying elite status there, because Punjabi community is using it confidently.

Recent statistics shows that 38% of the population in Pakistan speaks Punjabi. A rich literature both ancient and modern, full of wisdom and morality is found in this language

Gradually now it has to do struggle for survival. It is in a combat with national and foreign languages. Linguistically it has been less attended to. Researches are seen using traditional approaches and some using corpus techniques. But computational works, AI techniques and NLP is rarely done. This study explores Punjabi using machine and NLP and highlights one of the exclusive pattern of adjectives

Grammar blue prints a language and mirrors the structural system to communicate effectively and consistently without confusion and ambiguity. Linguists are of the opinion that grammar is not just a set of rules but a framework that reflects how language functions in human cognition and society. (Fromkin, 2018), (Biber, 1999), (Halliday, 2014), (Chomsky, 1957) and many more. Describe the grammar and establish its significance. Grammar constitute word classes and Punjabi has 11-word classes like most of other languages. One of them: adjectives enrich expression by adding descriptive and emotional depth to discourse.” Traditional Grammar, Structural Linguistics, Generative Grammar, Functional Grammar, and Discourse Analysis, are the linguistic perspectives that can provide theoretical framework about adjectives along with other word classes. Traditional grammar tells of adjectives that modify or describe nouns and pronouns and their function is to provide additional information about qualities, quantities, sizes, colors, emotions, and states. ‘...adjectives are essential because they enrich nominal expressions and make communication more precise.” (Quirk, 1973). In Structural Linguistic Theory (Bloomfield, 1933) describe attributive and predicative position of adjective with noun. In Noam Chomsky’s Generative Grammar, adjectives are analyzed as part of the deep and surface structure of sentences and serve as determiners. (Halliday, Halliday’s Introduction to Functional Grammar (4th ed.). , 2014) mention many functions that are performed by adjectives in the text. The significance of this class of words is not less in Semantic Theory of Adjectives and Discourse Framework. Only adjectives portray cultural values, gender attitudes, political ideology, emotional stance. The color and richness of any discourse lie in the adjectives found in it. “.. Adjective choice can influence readers’ perceptions and reinforce power relations.” (Fairclough, 1995).

This research shows/intends to show methodological gap along with it the gap of idea. Bybee's Dynamic Usage-based framework has greatly inspired usage-based models of language. Bybee’s model explains synchronic, diachronic and typological patterns within languages, such as which class of variant will occur in which contexts, what forms they will take, and about their diachronic end results by implementing the linguistic phenomenon of splits means when a word starts to show subtle polysemy, and morphological possibilities for the originally single form. Bybee proves that

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

even irreducibly, smaller, irregular word-forms are seen to be non-arbitrary when the context it occurs in is taken into consideration. In the same way, she presents that even apparently regular allomorph is context-sensitive. Linguistic forms, act in context. They alter shapes and forms. They cannot be studied as isolated entities, but rather in relation to the strength of their attachment to other entities.

Methodology and Corpus Development

Punjabi corpus has been developed for the purpose. Punjabi novels and short stories are collected. And are dug out from the real published works to convert into machine readable data.

Only one novel Sarr Gai Dharti having 16385 words produced the results given in table 1.

NLP methods are used for the preprocessing. For finding word frequency, tokenization and POS tagging are the basic preliminary actions performed. Python codes have helped finding an adjective that varies with the context showing number and gender condition of the subject in the sentences, from the corpus. Frequency analysis is then employed to identify recurrent grammatical categories, lexical distributions, and morpho-syntactic patterns. By focusing on frequency and distribution, the study highlights the relationship between usage and grammatical structure, aligning with usage-based theories of language tasks, including parsing, machine translation/ morphosyntactic features and language modeling.

Findings

The data received is shown in the tables below:

Table 1.

Word frequency from the corpus:

8 : سڑ	شاہنامہ : 8	خیالی : 1	کے : 13	ر	بنتر : 1
گئی : 54	دا : 199	، : 608	کے : 132	ایس : 6	ستار ہوئی : 1
دھرتی : 2	حوالہ : 1	تاریخی : 1	پرو : 1	جہاں : 1	عرب : 1
ترجمہ : 1	دتا : 35	واقعی : 1	نے : 1	دے : 216	فتح : 4
کار : 8	گیا : 133	بڑے : 8	نے : 115	جنم : 1	تک : 26
ثروت : 1	اے : 292	سوہنے : 1	ایہدے : 5	توں : 459	ایران : 1
				لے : 36	اوبدے : 82

Table 2.

POS FREQUENCY -----

VERB : 2055	PART : 476
AUX : 680	ADV : 44
NOUN : 2347	DET : 50
PROPN : 6190	CCONJ : 25
ADJ : 375	SCONJ : 97
PUNCT : 2645	NUM : 139
ADP : 484	INTJ : s3
PRON : 774	

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

ANALYSIS RESULT =====

TOTAL SENTENCES: 1088

TOTAL WORD COUNT: 16385

UNIQUE WORD COUNT: 2236

----- WORD POS TAGGING -----

سڑ → VERB	وڈھی → PROPN	کتاب → PROPN
گئی → AUX	تے → PROPN	شابنامہ → PROPN
دھرتی → NOUN	مشہور → ADJ	دا → PROPN
ترجمہ → NOUN	فارسی → PROPN	حوالہ → NOUN
کار → PROPN	زبان → NOUN	دنا → VERB
ثروت → PROPN	وچ → PROPN	گیا → AUX
سہیل → PROPN	فردوسی → PROPN	اے → AUX
سڑ → PROPN	دی → PROPN	- → PUNCT
گئی → PROPN	لکھت → PROPN	مشہور → ADJ
دھرتی → PROPN	لمی → PROPN	فارسی → NOUN
وچ → PROPN	تعریفی → ADJ	نظم → NOUN
یارھوین → ADJ	نظم → NOUN	مہا → PROPN
یارھوین → ADJ	مہا → PROPN	کاوی → PROPN
صدی → PROPN	کاوی → PROPN	بادشاہواں → PROPN
دی → PROPN	لمی → PROPN	دی → PROPN
	تعریفی → ADJ	

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

Patley, “rukawat lei patley rod lagey sun.”

“Thin rodes were fixed for stop signal.”

پتلیاں - "پنڈ دے اندر پتلیاں گلیاں وچ لوک بھجے جاندے سن۔"

‘patlian’ , “pind wich lok patlian galian wich bhjdy Jandy sn.”

People were running in the narrow streets of the village.

پتلیاں

Conclusion:

The recurrent word patterns in the data, the grammatical features and parsing are all traced and picked from the Punjabi Corpus with the help of NLP techniques. Ever changing role of adjectives have traced and noted from the corpus. Modern techniques have been practiced on Punjabi corpus successfully. The research leads to exploration of grammatical aspects of the indigenous languages and can play key role in revitalization and reservation of local ignored languages.

REFERENCES

- Akhunzada, F. “. (2026). “Language Endangerment: A Case Study of Gawarbari, Kalasha and Yadgha Languages of Chitral.” (2026). . (*wahacademia.com*).
- al., B. e. (2021). Bromham et al. “Global Predictors of Language Endangerment and the Future of Linguistic Diversity.” *Nature Ecology & Evolution* (2021). (Nature). *nature*.
- Biber, D. J. (1999). . *Longman Grammar of Spoken and Written English*. . London: Longman.
- Bloomfield, L. (1933). *Language. Structural Linguistic Theory*. New York: Holt.
- Bromham, L. (2023). “Language Endangerment: Cambridge University Press (2023) .
- Chomsky, N. (. (1957). *Syntactic Structures*. . Mouton: The Hague.
- Crystal, D. (2000). *Language Death. Cambridge University Press*. London: Cambridge University Press.
- Fairclough, N. (. (1995). *Critical Discourse Analysis. London: Longman*. London : Longman.
- Fromkin, V. R. (2018). . *An Introduction to Language (11th ed.)*. . Boston: Boston: Cengage Learning.
- Halliday, M. A. (2014). *Halliday’s Introduction to Functional Grammar (4th ed.)*. London: Routledge. London: Routledge.
- Halliday, M. A. (2014). *Halliday’s Introduction to Functional Grammar (4th ed.)*. . London: Routledge.
- Issa, M. e. (2023). Language Erosion: An Overview of Declining Status of Indigenous Languages of Gilgit-Baltistan, Pakistan.” . (*ijmejournal.org*).
- Literature., N. w. (1986). *Decolonising the Mind: The Politics of Language in African Literature*.
- Nettle, D. &. (2000). *Vanishing Voices: The Extinction of the World’s Languages*. . London: Oxford University Press.
- Phillipson, R. (1992). *Linguistic Imperialism. Oxford University Press*. Oxford University Press.
- Quirk, R. &. (1973). *A University Grammar of English*. . London: Longman.
- TEXT, U. O. (n.d.). United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP),. *Articles, Articles 13–16*.
- Unesco. (n.d.). . *Indigenous Knowledge and Identity. UNESCO Language Atlas*.

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

- Urooj, N. R. (2025). *Postcolonial Linguistic Marginalization: The Kalasha Language in Pakistan's National Discourse.* . (academia.edu.pk).
- Younus, J. e. (2023). "Analyzing the Factors Involvement in Declining Kalasha Language." (2023). (journals.internationalrasd.org). (*journals.internationalrasd.org*).
- Bauer, L. (2003). *Introducing Linguistic Morphology.* Edinburgh: Edinburgh University Press.
- Gill, H., & Gleason, H. (1969). *A Reference Grammar to Punjabi*". Revised Edition, Dept. of Linguistics, Punjabi University, Patiala.
- Hussain, K. G. (2018). Nominal Markers in Punjabi. *Hayatian Journal of Linguistics and Literature*, 2(1), 46-57.
- Yule, G. (1996). *The study of language.* Cambridge: Cambridge University Press.