Ilkogretim Online - Elementary Education Online, Year; Vol 20 (Issue 5): pp. 1100-1108
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.05.119

Analyzing the Lexico-Semantic Relationships of Nouns Used in The Saraiki Newspaper: A Corpus Based Study

Noreen Zamir, Visiting lecturer, University of Sahiwal, Punjab, Pakistan**,** noreenzameer@uos.edu.pk
Benish, Fast National University, Lahore, Pakistan.
Jahanzeb Jahan, Lecturer, Department of English Language and Literature, University of Education, Lahore, Pakistan
Bilal Asmat Cheema, Lecturer, Division of Arts and social sciences, University of Education, Lahore, Pakistan

Abstract- The current study is an effort to analyze the lexico-semantic relationships of nouns in Saraiki newspaper. Although spoken as a first language by almost 20 million people in Pakistan, Saraiki has a limited Corpus in the form of ijunoonSaraiki dictionary (2017). A functional corpus for Saraiki would be helpful in developing a WordNet through which the digital applications of Saraiki can be run. Of the four open class categories included in the WordNet the present study explores the semantic relations found among the nouns of Saraiki. A corpus of 2 million words was created from the Saraiki newspaper Jhoke after POS tagging a list of 1500 nouns was generated and the nouns were semantically categorized for identifying the lexical relationships among them using machine readable dictionaries. Each lexical and semantic relation was quantitatively analyzed using Antconc 3.5.7. The 3A model of corpus linguistics was used for annotation and analysis of Saraiki nouns (Wallis & Nelson, 2001). The results revealed ten lexico-semantic relationships frequently found among the nouns in the newspaper Jhoke, which are essential for the development of a lexical data base such as WordNet. The lexico-semantic relationship analysis showed that the singular / plural relationship was the most frequent among the Saraiki nouns while, synonymy was the least frequent relationship. Identifying the noun relationships of Saraiki are a step towards the development of a Saraiki WordNet that would support NLP and other digital applications.

Key words: Lexico-semantics, Nouns, Corpus linguistics, Saraiki, Newspaper.

I. INTRODUCTION

Pakistan is a multilingual country and a honeycomb of linguistic heritage with over 74 spoken languages. The national language is Urdu, spoken by a small portion of the population as their mother tongue, while English is the official language and used in the educational institutes as the medium of instruction (Saini, 2003). In addition to Urdu and English, the regional languages include: Punjabi, Pashto, Sindhi, Saraiki, and Balochi spoken by a large number of speakers (Rahman, 2009). These languages are an offshoot of the Indo-Iranian branch with sub-branches belonging to the Indo Aryan and Iranian subgroups, spoken within a specific geographical area and connected through dialect continuums (Renschler & Kleiner, 2013).

Among all the regional languages Punjabi is the most widely spoken language in the Punjab which the largest province of Pakistan. The dialects include Majhi, Hinko, Rachnavi, Dogri, Dhani, and Saraiki distinguished by the varying linguistic features that makes them stand apart from each other (Javaid, 2004). However, *Majhi* is the prestige dialect of Punjabi acts as a binding force to other dialects and is also used in the textbooks of Punjabi (Shackle, 2003). Although a majority of people report Punjabi as their first language however, only 20.68% use the Saraiki dialect as their mother tongue. (Pakistan Bureau of Statistics, 2017)

Spoken by almost 20 million people in various parts of Pakistan with majority belonging to south Punjab,(Samina et al., 2020) Saraiki ranks 60th out of the hundred languages spoken world-wide based on the percentage-fraction of world population of native speakers of this language (Gillani, 2013). The absence of etymological dictionaries and the dearth of written resources are considered to be some of the reasons for Saraiki as a dialect while the proponents of Saraiki movement consider it to be a language (Hussain &Khan, 2016). Saraiki lacks in Natural language processing and a limited amount of work is done in its morphology, syntax, and semantics (Maldonado, 2015). In order to develop encyclopedias, online dictionaries and web translations of a language, corpus development is considered to be the first step of Natural Language Processing (NLP) and the non-availability of any corpus a language into resource poor which results a slow move towards ultimate decline (Saini, 2018).

However, Lexical-semantic classifications have proved useful in supporting various natural language processing (NLP) tasks such as word net data bases and online dictionaries (Korhonen & Briscoe, 2004). The Lexical relations are one of the most important semantic relations in exploring the meanings of words in any language. They are mainly used to classify the lexical items in terms of their relations to each other within sentences. They vary according to the kind of the relation that a lexical item may have with another word or words (Mukund & Srihari, 2009). Lexical-semantic classes which aim to capture the close relationship between the syntax and semantics of verbs have attracted considerable interest in both linguistics and computational linguistics (Maziarz et al., 2011).

All the applications in Natural Language processing work with the back end digital sources and Word Net is one of the reliable digital forms of language thesaurus. In word net lexical data base of nouns, verbs and adjectives are combined to from cognitive synonyms which are further interlinked by semantic and lexical relations. These relations have been developed on the basis of word meanings rather than word forms (L'Homme & Cormier, 2014). While the word net deals with the semantically related lexical items whose meanings are mutually interdependent after classifying and analyzing their relationship to sentence meaning and syntax (Wynne, 2005).

Corpus semantics has a significant role in language structure, connected etymology, interpretation, stylistics and lexicology (Thompson & Hunston, 2000). The principle of corpus linguistics has been used in research for over a century and it has also helped in digitalizing languages. The Lexicographers have been collecting examples of language in order to define the words accurately. Corpus linguistics broadly looks at the patterns that are associated with lexical and grammatical features (Bennett, 2010). However, it is a study of language which is based on extensive collections of the "real life" language in use known as corpora. In order to develop semantic nets and lexico-semantic relation among a language a real-life corpus needs to be developed (Stubs, 2001). The newspaper language being a functional language is considered to be useful in developing a corpus and the lexical relations play an imperative role in the formation of a thesaurus (Vandana & Dash, 2018).

The newspaper language being a functional language is considered to be useful in developing a corpus and the lexical relations play an imperative role in the formation of a thesaurus (Vandana& Dash, 2018). The language of newspapers depicts the language of its readers as it utilizes the lexicon which the readership uses in their spoken discourse. Therefore, for the development of a specialized copora newspaper, books and articles are the most suitable data as it serves different kind of linguistic variation (Khokhlova, 2014). Lexical data bases contain a specialized corpus (used for a specific research area) and work as a thesaurus which is an alternate to the machine-readable alphabetical dictionaries and indigenous languages lack such online lexical databases and thus traditional lexicographic information (Dash, 2005).

1.1 Statement of the problem

A limited thesaurus for Saraiki is available in the form of (ijunoonSaraiki dictionary, 2016) and lacks semantic relations of the lexical units. Therefore, the study intends to develop a more functional corpus for Saraiki language by analyzing the lexical semantic relationship among its nouns used in Jhoke newspaper.

1.2 Research questions

The study responded to the following research question

1. What type of lexcio-semantic relationships are frequently found in the use of nouns in Saraiki language?

1.3 Significance of the study

This study will be useful in developing a WordNet for Saraiki language through which digital applications of Saraiki can be run, thus improving the digital standards of Saraiki language. It will develop the Saraiki corpus and provide awareness about lexicosemantic relationship of nouns used in Saraiki language.

II. LITERATURE REVIEW

Corpus linguistics is referred to as the corpus-based studies in applied linguistics (Tognini -Bonelli, 2001). Before the development of modern digital corpus, a conventional corpus existed in thirteenth century that was limited only to the corcordance of Holy books. Corpus linguistics emerged as an academic discipline after the development of computational technology after the third industrial revolution in twentieth century. Corpus analysis is directly linked to the digital tools and web content because corpus analysis is directly linked to the digital tools and web content which became available after the third industrial revolution in twentieth century. Moreover, the use of a corpus and linguistic issues along with statistical software are the most vital elements of corpus linguistics as argued by (Boulton & Tom, 2017).

(Bennet, 2006) is of the view that corpus linguistics assists in the research that deals in language variation and its use through factual and empirical probe as a research paradigm. This paradigm has become more feasible over the past few decades as it has started providing more validity and clichéd results which were not considered feasible. He further adds that it is a methodology which incorporates a generalized text comprised of a real-life corpus. (Biber, Rappen, & Friginal 2002) have however, viewed corpus as more than just a methodological approach because researchers are now able to find answers of different fundamental questions using corpus.

Modern corpus is however, based on computer-assisted analysis of text, which evolved in late 1950s. However, in the beginning, it was opposed by Noam Chomsky, who prioritized rationalism over empiricism connected with corpus-rooted approaches. (Wilson, 2001). The usefulness of copora has wider spectrum in linguistics especially in grammatical description and lexicography. Modern corpus was initially used for English language only; however, it is now applied to different languages. On the other hand, Neo-Firthian, another approach, considers collocation in corpora, study of words and phraseology as basis of linguistic theory viewed by (Mc Enery &Wilson, 2001).

However, the use of corpora in semantic research is a rapidly developing method. The range of quantitative techniques employed in the field can make it difficult for the non-specialist to keep abreast with the methodological development. Lexical semantics is a subfield of semantics in which the meaning of words in relation to the sentence structure or semantics is formed as explained by (Cruse, 2001). It is concerned with the organized study of word meanings. The two most important questions which lexical semantics addresses is that (a) what is the procedure to for the description of the meaning of words (b) how the contextual scenario changes the meaning of a word within a sentence. Both the questions having a connection support each other on the basis of the ability to interpret and the variation which occur among the meaning of words. These relationships include antonomy which deals with the opposite meanings of a word, hyponymy deals with specific meanings which (Chen, 2017) described in their study.

(Miller, 1990) points out that Word net is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and its concept superficially resembles a thesaurus, in that its groups words together based on their meanings. However, there are some important distinctions. First, Word Net interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, Word Net labels the semantic relations among words, whereas the groupings of words in a thesaurus do not follow any explicit pattern other than meaning similarity.

(Fellbaum 1998) while discussing the lexical information provided by the dictionaries gives information about dictionary entries which according to him are devised for the betterment and readability of human beings not for the online software and Artificial intelligence. (Bhattacharyya 2010) provided a Hindi word net which was inspired by English and German word net in which lexical information was given on the Hindi words. This data base is devised by research methodology in which the words or phrases such as verbs, adjectives, nouns and adverbs are combined to form semantic word net through which words of a language are encoded into abstract information.

According to (Charles, 1988) concordances which are specifically designed to identify and give a research basis for the studies which are empirical in nature and contain a lexical database that provides identification for lexical semantic relations found in corpus. He further adds that the concordances which include a corpus in the form of text and the lexical items (which might be a phrase or word) are properly linked to provide information about the whole sentence. In brown corpus which is one of the earliest corpuses the words are connected to form a Word net which gives understanding of the meaning of the word. However, in order to have an indicative sense of polysemous word there is need to generate a semantic concordance which is specialized enough to provide samples of representative texts in a thesaurus.

(Saini, 2019) is of the opinion that Saraiki language lacks in Natural language processing and a written or spoken corpus of Saraiki is not available therefore this language is not digitalized which makes it a resource poor language. He listed Saraiki NLP research as an example and classified all the work into six categories which according to him will help in the development of corpus and other digital applications and software's these categories include, script, Evolution, phonetics, Machine Translation System, corpus and Social-Science. Along with providing possible future aspects this study gave research domains for Saraiki Natural Language processing (Nadiem, 2005).

III. RESEARCH METHODOLOGY

The present study follows a corpus-based quantitative research design. It is exploratory in nature as it explores the lexical semantic relations as well and there were some predefined areas which were to be verified/ studied with the help of the corpus Baker, (2010).

3.1 Theoretical framework

The theoretical framework used for the annotation and analysis of the Saraiki corpus is based on the 3A Model introduced by Wallis and Nelson (2001), who claimed the three key stages of Annotation, Abstraction and Analysis as a process to study corpus linguistics. Each stage is characterized by a progression between the original text and the assessment of research questions.

3.2 Sample

The study used purposive sampling technique to select the sample from the Saraiki daily Jhoke newspaper, the only Saraiki newspaper being published from Multan. Homogenous Purposive sampling was considered suitable for the study that allowed for collecting a specific set and systematic inclusion of 180 issues of the daily newspaper Jhoke from July 2020 to December 2020.

3.2 Corpus creation

A corpus of 2 million words was created in the inpage 2012 software. The corpus text was selected from front and back page of Jhoke newspaper excluding the Urdu and Punjabi text thus making it a monolingual and written corpus, as the main aim of the corpus was to analyze the lexico-semantic relations of Saraiki.

3.3 Text processing and annotation

The corpus created in inpage 2012, software was taken to the notepad file with UTF-8 encoding and was POS tagged with CLE POS tagger to generate the list of nouns. Owing to the unavailability of a Saraiki tagger and data handling errors a few nouns showed their ambiguous representation which was removed manually.

IV. RESULTS

1500 Nouns were extracted out of the 2 million corpus of Saraiki language newspaper, generated from a POS tagger. Due to the unavailability of NLP tools for Saraiki language the lexical relations were developed manually with the help of machine-readable Saraiki language dictionaries. The analysis of nouns was performed using 3A Model for Corpus linguistics. Keywords for the list of nouns were generated which came

to 1500 in number. WordNet exhibits ten lexical and semantic relations for nouns including the relation of synonymy, antonymy, meronymy, holonomy, hyponymy, hypernymy, masculine, feminine, singular and plural. Antconc 3.5.7 version was used to calculate the frequencies for each relation. Subsequently, meanings for each of the target noun concurring to the lexico-semantic relationships were retrieved by means of investigating online resources and utilizing the assistance of the Saraiki language native speakers.

4.1 Relation of Synonymy

Table 1: Frequency and percentage of synonymy in Jhoke newspaper

Lexico-semantic relationship	Frequency	Percentage
Synonymy	254	16.93%

Table 1 shows the frequency and percentage of the relationship of synonymy foundamong the nouns of Jhoke newspaper. As the table illustrates that only 245 nouns showedthe relation of synonymy with a total percentage of 16.93 percent which implies that mostof nouns in newspaper Jhoke did not have alternative words to be replaced with the targetword. After the analysis of list of Jhoke newspaper it wasfound that the remining 83% nouns which did nouns show any synonym include name ofjobs, countries, cities, a few eatables and certain loan words of English, Urdu and Punjabidid not have their synonyms. Moreover, the name of caste, and the words related to thelanguage used in politics along with random words suggesting that Saraiki language isnot a diverse and rich language and has borrowed a number of words from English, Urduand Punjabi language.

4.2 Relation of Antonymy

After the analysis of nouns four different kinds of the relation of antonymy werefound among the list of nouns of which was created from the Jhoke newspaper corpus.These antonyms included Gradable opposites, Near opposites, Complementary oppositeand Relational opposites. Antonymy is the opposition relationship, which is mostly found among the adjectives and adverbs but rare amongst nouns however, in Jhoke corpus 445nouns showed the relation of antonymy with a frequency of 29.66%. Results showed thatthe loan words, name of games, certain nouns used in the politics and a few randomwords had no antonyms.

4.2.1 Gradable opposites

Only 121 nouns were found as gradable antonyms of each other which made 12.60percent of total antonyms in the corpus (See appendix). The results showthat a few Saraiki nouns express grading quality among them. which means that the Saraikilanguage has less inherent compatible binary relationship and its words lack continuousspectrum due to unavailability of opposite pairs.

4.2.2 Near Opposites

The number of nouns, having relation of near opposites which have been found in theJhoke newspaper corpus were only 8.73% of the total corpus and 27.19% of the totalantonyms respectively (see Appendix). The table also indicates the frequency of Saraikinouns as near opposites came out to be 132 out of 445 antonyms. There are a variety ofinstances such as hot and cold ڈاڈ ہےی/رف ب where the terms do not appear to be in relation toreal opposition, but they seem like near opposites It is evident from the results that thisrelation is not recurrent among the nouns of Saraiki language.

4.2.3 Complementary opposites

The nouns which exhibited complementary relationship among them had a frequencyof 103 and their percentage comprised of 23% of the total antonyms and were 6% of thecomplete list of nouns the complementary antonyms used in Jhoke Newspaper (seeappendix). As it is characteristic of nouns of lexical items that the denial of the oneimplies the assertion of the other and the assertion of the one implies the denial of theother. However, it is considered that there is not any middle word between which cancomplement this lexical category of complementary opposites but results showed thatSaraiki nouns used in Jhoke corpus contained nouns which had the characteristic ofmiddle words to be used between two complementary opposites and were quite close togradable opposites.

4.2.4 Relational opposites

The relational opposites were not prominent among the relation of antonymy and isamong the least frequent kind of antonyms with a frequency of only 89 which made the23.30 percent of the total antonyms and 5.93% of the complete list of nouns generatedfrom the Jhoke newspaper corpus (See appendix).The low frequency of this relationspecifies that the language used in the newspaper Jhoke was politics centered and hadless words related to social discourse.

4.3 Relationship of Meronymy and Holonomy

Table 2: Frequency and percentage of meronymy and holonomy in Jhoke newspaper

Lexico-semantic relationship	Frequency	Percentage
Meronymy	289	19.26%
Holonomy	389	25.93%

Table 2 presents the frequency and percentage of meronymy and holonomy in Jhoke newspaper, the statistical result in the table shows that out of 1500 the nouns from the corpus of Jhoke newspaper 389 nouns had the relation of meronymy which is 23.93% of the total list of nouns. The table also shows that only 289 nouns existed as holonyms with a percentage of 19.26% respectively. The results indicate that almost 80 percent of the nouns in Jhoke newspaper did not have their representative parts and around 74% nouns were individual entities with no whole. However, the part and whole relationship together makes up 45.2% of the nouns generated from the corpus of newspaper Jhoke.

4.4 Relationship of Hypernym and Hyponym

Table 3: Frequency and percentage of Hypernymy and Hyponymy in Jhoke newspaper

Lexico-semantic relationship	Frequency	Percentage
Hypernymy	345	25.20%
Hyponymy	378	25.20%

Table 3 shows the relationship of hypernym and hyponym in Jhoke newspaper, according to which 345 nouns showed the quality of hypernymy which made 23% of the total nouns generated out of Jhoke newspaper corpus. Whereas, 37.80% nouns were hyponyms and their frequency came out to be 567 respectively. The relative analysis of both the relations show that percentage of hyponyms is higher than that of hypernyms and they together makeup 60.8% of the complete list of Saraiki nouns.

A few abstract nouns did not possess this relation andthe nouns specifically used for newspaper, animates, name of jobs, name of week daysand seasons also did not show the relationship of hypernym and hyponym.

4.5 Relationship of Singular and Plural

Table 4: Frequency and percentage of Singular and Plural in Jhoke newspaper

Lexico-semantic relationship	Frequency	Percentage
Singular	656	43.73%
Plural	523	34.86%
Singular/plural	321	21.3%

Table 4 presents the frequency and occurrence of relationship of singular and plural nouns in Jhoke newspaper. The table indicates the occurrence of 656 plural nouns out of 1500 nouns generated from the Jhoke newspaper corpus, which made the 43.73% of the list of Saraiki nouns extracted out of the target corpus. Singular nouns came out to be 523 with a frequency of 34.86% respectively. As the relationship of singular and plural is necessary and common part of every dictionary and wordnet as well therefore, the collective percentage of singular and plural relationship was 78.6% and their frequency was 1,179 in total list of Saraiki nouns. However, the table also shows that 321 Saraiki nouns showed attributes of both singular and plural with a percentage of 21.3% out of the 1500 nouns. The results indicate that most of the Saraiki nouns showed the relationship of singular and plural among them except few loan words used in the newspaper Jhoke.

4.6 Relationship of Masculine and Feminine

Table 5: Frequency and percentage of Masculine and Feminine in Jhoke newspaper

Lexico-semantic relationship	Frequency	Percentage
Masculine	434	28.93%
Feminine	567	37.80%
Masculine/Feminine	499	33.26%

Table 5 shows a unique ratio of masculine and feminine relationship among the nouns of Jhoke newspaper. The percentage of feminine nouns was more than the masculine nouns, as the frequency of feminine nouns is shown as 567 with a percentage of 37.80% according to which most of the nouns in Saraiki are feminine. According to the table the percentage for masculine relation is 28.93% with a frequency of 434 nouns. A few species and social relations of Saraiki language nouns were frequent in exhibiting their respective masculine and feminine relationships. The remaining 71% of the nouns were not related to masculine gender. However, the results in table 5 highlight the fact that 499 nouns in the Jhoke newspaper corpus with a frequency of 33.26% were neither masculine nor feminine nouns as their gender was not clear. The results also indicate that newspaper Jhoke also contained nouns that were neither feminine nor masculine those among those nouns most of them were related to politics, counting, name of cities and some administrative posts which could not be categorized as masculine or feminine.

4.7 Comparative analysis of the relationship of nouns in Jhoke newspaper

A comparative analysis of the lexico-semantic relationship of Saraiki nouns estimates the relative presence of each of the relationship.

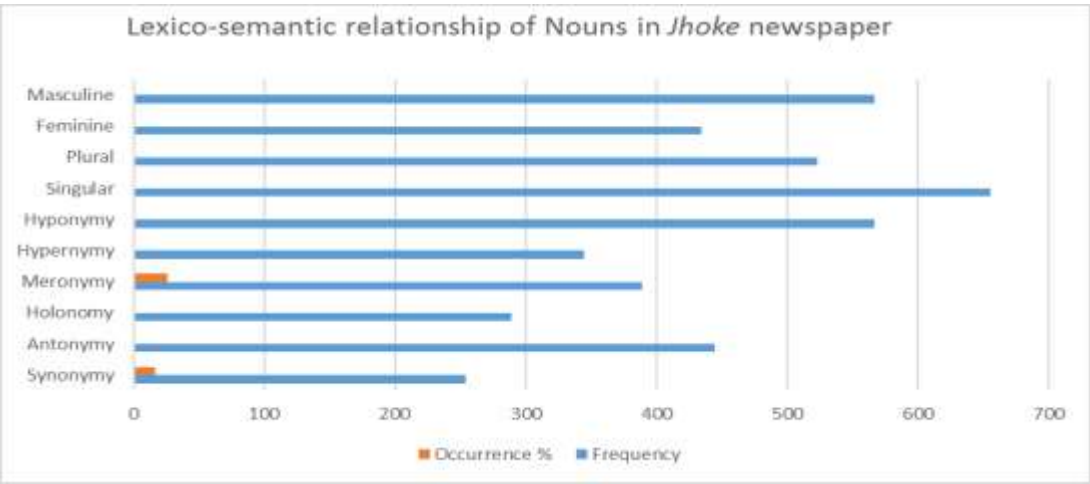

Figure 1

Figure 1, reveals the comparative frequency and percentage of the occurrence of each lexical and semantic relation as found among the nouns in the newspaper Jhoke. The most frequent relation evident in the list of nouns formed out of the newspaper corpus was the relation of singular and plural with the graph highlighting the occurrence of 656 Singular nouns. The figure also indicates that the relation of synonymy with the lowest frequency of 254 as compared to other relations found among the nouns used in the newspaper Jhoke. This implies that most of the nouns used in the newspaper did not show considerable alternative words to be replaced with the target nouns and the least lexical and semantic relation found while analyzing the nouns was the relation of synonymy.

V. Results and discussion

In the light of the results, it is concluded that the lexico-semantic relationship of nouns used in Saraiki newspaper Jhoke can help in the development of a Saraiki WordNet. The nouns of the Jhoke newspaper showed ten lexical and semantic relationships which are essential for the development of the digital thesauruses and exhibited by Princeton WordNet (Miller, 1990).

The research findings also indicate that the Saraiki language is losing its diversity and purity as lexical borrowing is evident in its functional language due to which the frequency of synonyms and antonyms is less. The Jhoke newspaper corpus can also be utilized for developing Natural language processing tools like machine translation, named entity recognition and morphological parsers will be a part of the Saraiki WordNet to be designed on the basis of Princeton WordNet which is considered to be a solid lexical database and is supposed to be important for digitally young languages such as Saraiki language.

Although Saraiki language has a grammar and a dictionary but research done into its syntax and morphology is limited. Through scientific investigation it can be proved that Saraiki is a language with its own standard and non-standard dialects. Moreover, a diachronic analysis of Saraiki language is recommended to verify and examine the lexical changes and influence of other languages on Saraiki that took place over the decades.

References

1. Bennet, R.G. (2010). Using corpora in the language learning classroom: Corpus linguistics for teachers. Michigan ELT, 34-48.
2. Biber, D., Conrad, S., Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, Cambridge.
3. Baker, P. (2010). Sociolinguistics and corpus linguistics. Edinburgh, Texas: Edinburgh
4. O'Hammoran (eds.) Applying English Grammar. Functional and Corpus Approaches.eachers. Michigan ELT, 34-48.
5. Boulton, A. & Tom, C. 2017. Corpus use in language learning: A meta-analysis. Language Learning 67(2): 348-393.
6. Chen, C. (2017). Science mapping: A systematic review of the literature, Journal of Data and Information Science 2(2): 1-40
7. Charles, W. G. (1988). The categorization of sentential contexts. J. Psycholinguistic Res, 403--411.
8. Cruse, D. A. (2001). Aspects of the micro-structure of word meanings. In Ravin andLeacock, 30–51.
9. Dash, S.N. (2005) Corpus Linguistics and Language Technology: With Reference to Indian Languages, New Delhi: Mittal Publications.
10. Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. The MIT Press.
11. Gilani, M.H. (2013). Historical background of Saraiki language. Pakistan journal of social sciences, 33(1).
12. García, M.I.M. (2015). A Corpus Based Quantitative Survey of the Persian and Arabic Elements in the Basic Vocabulary of Urdu Language. Pakistan Vision, 16 (1), 63.
13. Hussain S.S., Khan B.M. (2016). Integrated Linguistic Links of Pushto and Saraiki Languages, Pakistan Journal of Social Sciences, 36 (2) 653-663.
14. IjunoonSaraiki dictionary (1sted). (2019). Retrieved fromhttps://www.ijunoon.com/saraiki
15. Javaid, U. (2004). Politics of Bahawalpur: From State to Region: Lahore: Classic.

16. Korhonen, A. & Briscoe, T. (2004). Extended Lexical-Semantic Classification of English Verbs. 10.3115/1596431.1596437.
17. Khokhlova, N. (2014). Understanding of abstract nouns in linguistic disciplines. Computer science, 26(2). DOI:10.1016/j.sbspro.2014.05.278
18. L'Homme, M. C., & Cormier, M. C. (2014). Dictionaries and the digital revolution: A focus on users and lexical databases. International Journal of Lexicography, 27(4), 331–340. https://doi.org/10.1093/ijl/ecu023
19. Miller, G. A. (1990). Nouns in WordNet: a lexical inheritance system. International journal of Lexicography, 3(4), 245-264.
20. McEnery, A.M., & Wilson, A. (2001). Corpus Lin-guistics (second edition). Edinburgh University Press, Edinburgh
21. Maziarz, M., Piasecki, M., Szpakowicz, S., & Rabiega-Wiśniewska, J. (2011). Semantic relations among nouns in Polish WordNet grounded in lexicographic and semantic tradition. Cognitive Studies| Études cognitives, (11).
22. Mukund, S., & Srihari, R. K. (2009). NE tagging for Urdu based on bootstrap POS learning. Presented at the Third International Workshop on Cross Lingual Information Access.
23. Nadiem, I. H. (2005). Punjab: Land, history, people. al-FaisalNashran.
24. Pakistan Bureau of Statistics (1981). Census report of Pakistan. Retrieved from http://www.pbs.gov.pk/sites/default/files/other/pocket_book2006/2.pdf
25. Renschler, I. & Kleiner, B. (2013). Considering Dialect in Survey Research. Ongoing Research, 51-59.
26. Saini, J.R. (2018). An Exhaustive Meta-analytical Study of the History, Evolution and Development of 'Saraiki NLP'.INFOCOMP, 17(2), p. 41-5.
27. Samina. K & Uzma. N & Muhammad. I & Younas. M (2020). THE STUDY OF ORTHOGRAPHICAL DIFFERENCE BETWEEN PUNJABI LANGUAGE AND SIRAIKI DIALECT IN PUNJAB PROVINCE. Hamdard Islamicus: quarterly journal of the Hamdard National Foundation, Pakistan. 43. 175-193.
28. Shackle, C. (2003). Punjabi in G, Cadona& D. Jain (Eds). The Indo Aryan Languages(4581-621).
29. Thompson, G. & Hunston, S. (2000). Evaluation: An introduction. Evaluation in Text: Authorial Stance and the Construction of Discourse. 1-27.
30. Tognini-Bonelli, E., 2001. Corpus Linguistics at Work. John Benjamins, Amsterdam and Philadelphia.
31. Vandana& Dash, S.N. (2018). Creation and Compilation of Hindi Newspaper Text Corpus. Language in India, 18(2).
32. Wynne, M. (2005). Developing Linguistic Corpora: a Guide to Good practice. Oxford: Oxbow Books. Available online from http://ota.ox.ac.uk/documents/creating/dlc/
33. Wallis, S. A. & G. Nelson. (2001). "Knowledge discovery in grammatically analyzed corpora". Data Mining and Knowledge Discovery 5(4): 305-336.