Analysis of Lecxico-Semantic Relations of Punjabi Shahmukhi Nouns: A Corpus Based Study

Muhammad Ahmad Hashmi[1], Muhammad Asim Mahmood[2] & Muhammad Ilyas Mahmood[3]

[1] SESE English, School Education Department, Okara, Punjab, Pakistan

[2] Dean, faculty of Arts and Social Sciences, Government College University, Faisalabad, Punjab, Pakistan

[3] Department of English, University of Okara, Punjab, Pakistan/Faculty of Education, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Correspondence: Muhammad Ahmad Hashmi, SESE English, School Education Department, Okara, Punjab, Pakistan. E-mail: muhammadahmadhashmi@gmail.com

Abstract

The current study is an effort in the development of Lecxico-semantic relations among Punjabi Shahmukhi nouns. Semantic relations are those nets, which are found among nouns on the bases of word meanings. Development of semantic nets is taken as a key part while developing WordNet of any language. The WordNet of Punjabi Shahmukhi is not developed yet. The digital exposure and progress of Punjabi Shahmukhi is very slow in comparison to other languages of the world. The present study explores the kind of semantic relations found among the nouns of Punjabi Shahmukhi. WordNet organizes words on the basis of word meanings rather than word forms. WordNet of English includes four open class categories including; nouns, verbs, adverbs and adjectives, but present study is limited to the analysis of nouns. A corpus of 2 million words of Punjabi Shahmukhi was taken from different sources. Then, it was POS tagged and a list of 846 nouns was generated. Then, each noun was analyzed individually to develop its Lecxico-semantic relations including: synonymy, antonymy, meronyms, holonymy, hyponymy, hypernymy, singular, plural, masculine, feminine and HAS a part. The present research is significant and useful in the development of WordNet for Punjabi Shahmukhi. With the development of WordNet, it will be possible to run digital applications in Punjabi Shahmukhi including: machine translation, information retrieval, querying archive and report generation to automatic speech recognition, data mining, read aloud, robotics and many more. On the other hand, WordNet will help to maintain an international status for Punjabi Shahmukhi.

Keywords: semantic relations, WordNet, Punjabi Shahmukhi, nouns

1. Introduction

Pakistan is a multilingual country having six main languages and fifty-nine minor languages. In major languages: Punjabi Shahmukhi is spoken by 46.15% of population; Pashto is spoken by 13.42%; Sindhi speakers are14.10%; Saraiki speakers are 10.53%; Urdu is spoken by 7.57% and Balochi speakers are 3.57% (Census of Pakistan, 2017). Punjabi Shahmukhi is spoken by the majority of population of Pakistan including the 10.53% of the Saraiki, which is also treated as southern dialect of Punjabi (Nadiem, 2005). If we exclude the Saraiki language, Punjabi Shahmukhi still persists as a major language and has got a significant number of speakers in Pakistan. It has its significant literary history, territory and culture. This language has its two scripts: Gurmukhi is a Hindi script of Punjabi, which is being spoken in Eastern Punjab of India while Shahmukhi is an Urdu script of Punjabi, which is spoken in Western Punjab of Pakistan. A notable work has been done in Gurmukhi Punjabi in India as Lehal and Singh (2000) developed a scheme for segmentation. Optical Character Recognition system for Gurmukhi has been developed by Lehal, Singh and Lehal (2001), which serves as a system for the detection of hand written characters of Gurmukhi (Sharma & Jhajj, 2010). The hand written recognition of Gurmukhi script has been further refined by Verma and Sharma in 2017 by taking this as a popular script of northern part of India. Unfortunately, Punjabi Shahmukhi is lagging behind in the race of Natural Language Processing (NLP). Punjabi Shahmukhi does not have any reliable digital corpus, thesaurus or dictionary for its digital processing. Digital form of language serves as a base to run the applications of NLP including Automatic Summarization, Computational Lexicography, Machine Translation, Automatic Morphological analyzer, Named Entity

Recognition, Optical Character Recognition, Digital documentation, Parsing, POS-Tagging, Sentence Breaking, Text Mining, Sentiment analysis, Text to Speech, Automatic speech recognition, Speech to Text, Speech to Speech, Information Retrieval, Speech Identification and many more (Kumar & Paul, 2016).

All these applications work with the backend digital sources of language and WordNet is one of the reliable digital form of language. The present study investigates the kinds of semantic relations among the nouns of Punjabi Shahmukhi. According to L'Homme and Cormier (2014) words have been organized in semantic nets, which are linked with different relations in WordNet. These relations have been developed on the basis of word meanings rather than word forms. There are four open class categories in Princeton WordNet of English including; nouns, verbs, adverbs and adjectives, but the current study is limited to develop semantic nets only among nouns. The kind of relations, which are considered the part of nouns are synonymy, for example, synonym of روز is دیہاڑی. Similarly, the synonym of دُکھ is اذیت and جہان is سنسار. Here, the changing of original word with its synonym does not change the context. Then comes the relation of opposition among nouns as سنگی/ دشمن and امیر/غریب, دُنیا/ آخرت, دُکھ/ سُکھ are antonyms of each other. Another relation among nouns is part-whole relation, we call it meronymy/holonymy, as 'deck' is meronymy of 'ship' and itself is the holonymy. It holds among synsets as in Punjabi روڑا is a meronym of اٹ and اٹ itself is holonymy. Moreover, the semantic relation of hypernym and hyponym, which are not common in dictionaries but are significant part of WordNet. Augenstein, Riedel, Vikraman and McCallum (2017), define hypernym is finding out the upper class of a given noun in the particular organization of natural hierarchy or can be called upper to its superordinate as in 'لکڑ, the hypernym of میز، کرسی تے تختہ. Here "میز، کرسی، تختہ are made from its upper class source لکڑ. Hyponym displays the one class lower category. e.g., ٹالی is a hyponym of رُکھ, while رُکھ is a hyponym of بوٹا. As hyponym of پکھو is منڈا. and انسان is کاں and منڈا. These above and some other semantic relations including singular/plural, masculine/feminine and Has a Part that are related to nouns. In this way, the purpose of present study is to develop semantic nets among nouns of Punjabi Shahmukhi.

1.1 Significance of the Study

The current study is significant as it will be useful in the development of WordNet for Punjabi Shahmukhi Language. It will help in the translation of Punjabi Shahmukhi language into other languages and vice versa. This study will serve as a mile stone in digitizing the Punjabi Shahmukhi. The most significant part of this study is to develop confidence in the native speakers of Punjabi Shahmukhi language to promote Shahmukhi as a first language in Pakistan and all over the world.

1.2 Research Question

What are those are the semantic relations which found among the nouns of Punjabi Shahmukhi?

2. Methodology

2.1 Data Collection

A corpus of 2 million words of Punjabi Shahmukhi language was collected from the various sources including: Newspapers, News items, Novels, Published Books, Poetry, Short stories and Articles. Corpus was taken in soft form with UTF-8 encoding in Notepad file. Then the corpus was transliterated in Gurmukhi script to tag it with a software name as Akhar and after it again transliterated in Shahmukhi script. This was because POS tagger for Punjabi Shahmukhi is not developed yet. The obtained data was checked manually to remove machine error and to correct tags, which had been changed during transliteration processes. After the correction of tags, the list of nouns has been generated using Laurence Anthony's antconc 3.5.7: a tool to handle corpus.

2.2 Data Analysis

The final list of nouns was analyzed manually in the available online sources including Punjabi dictionary https://www.ijunoon.com/punjabi_dic/, Punjabi Wikipedia https://pnb.wiktionary.org/wiki/, which has more than 9000 words in its database and another Punjabi Wikipedia https://pnb.wikipedia.org/wiki/, which has a huge data of 46546 articles. None of these resources inform about the semantic relations of synonymy, antonymy, meronyms, holonymy, hyponymy, hypernymy, singular, plural, masculine, feminine and HAS a part. These sources have been used to verify the existence of a particular word in Punjabi Shahmukhi language. The relations have been found by asking the native speakers and finally verified by a Punjabi expert form the Department of Punjabi in Government College University, Faisalabad.

During the data analysis diacritical marks have been given to the words to specify their particular sense because many words in Punjabi language have more than one meaning. The current study is unique because it a breakthrough in digitizing Punjabi language by using diacritical marks.

3. Results and Interpretations

3.1 Analysis of Nouns

From the total corpus of 2 million Punjabi Shahmukhi words, the list of nouns was extracted as mentioned earlier in data collection. Analysis showed that there was a repetition of nouns. A significant number of nouns showed their ambiguous representation because of machine translation and data handling errors. This happened due to the unavailability of NLP tools for Punjabi Shahmukhi. The generated list was analyzed and a final list of 846 nouns was developed. Many of the nouns were not in their correct orthography. The meaning of each target noun has been found using online resources of Punjabi Shahmukhi.

3.2 Interpretations of Lecxico-Semantic Relations of Nouns

The above analyzed data has been shown in the following bar-graph, which informs about the frequency of every lexical and semantic relation with its comparative graphical representation.

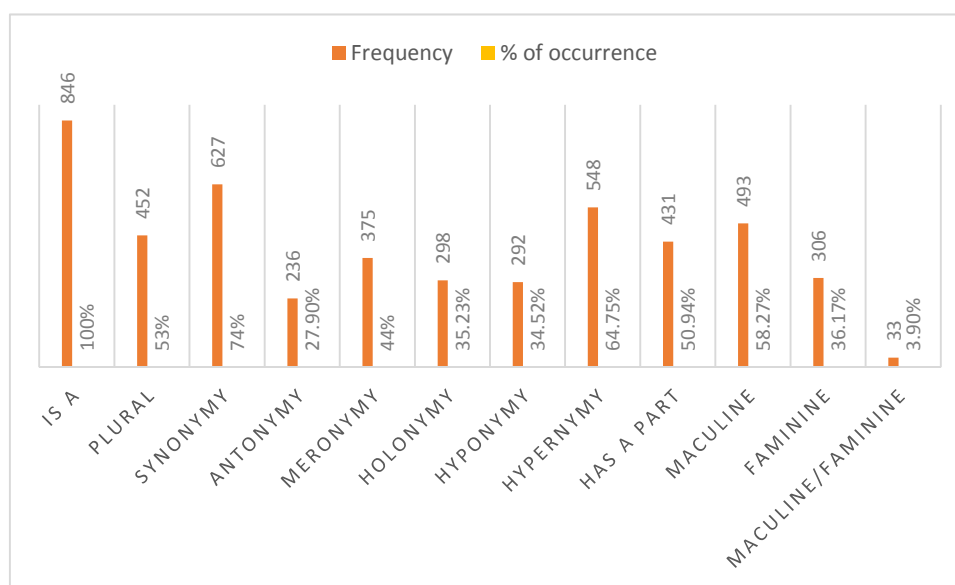In this way, the following graphical representation shows the frequency of nouns for each semantic relation.

Figure 1. Comprative representation of semantic relations of Punjabi shahmukhi nouns

3.2.1 Synonymous Relation

The most frequent lexical relation found in the analysis of nouns is synonymy. Two words are synonym of each other, if the replacement of one word with other does not change the context (Fellbaum, 2010). The longest bar of synonymous relation in bar graph helps us to understand the diversity and richness of language, where a number of words can be replaced with the target word.

Many words in the corpus are very unique and do not have their replacements. The analysis of those nouns shows that these nouns include the names of animals (طوطا ,بطخ برن مور، بلبل، چڑابلی، مکڑی ,تتر , گدھ ,کیوی ,چیتا ,گدھ) and other (.etc ... آڑو , کیلا,ساگ ,ادرک ,مٹر , ٹِنڈا, بھنڈی, کدو, آلو،مولی,مٹر), fruits, vegetables (.etc,شیر ,کُتا ,بلا ,بکری) naturally occurring materials (.etc ... وَتّا ,بوا ,پتھر ,ریت).

I have analyzed that the names of countries and cities (مدین,روم , یمن , جھنگ ,گلگت، روس، تُرکی،) ...etc.) are rare in showing their synonyms relations. Furthermore, the name of games which are borrowed from English culture also do not have their replacement in Punjabi Shahmukhi, for example فُٹ بال ,کرِکٹ and باکی.

Many man-made things like وان ,بل ,میز ,وردی ,کِل ,بیڑا ,کُرسی and توپ are rare in showing synonymous relation. Moreover, the analysis shows that Punjabi Shahmukhi has borrowed a range of words from English like مشین, and, چرس the name of narcotics like ...صوفہ, بلب ,بینڈ ,زونگ ,کلپ ,ڈِش ,لوشن ,سِتار ,کلب ,کرین ,ٹینک ,نٹ ,موٹر and افیم too are not having their substitutions in Punjabi Shahmukhi.

3.2.2 Relation of Antonymy

Antonymy is the relation of opposition, which is found frequent among the adjectives and adverbs (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) but rare among the nouns, so the bar of antonyms does not get a significant height in the bar-graph. Antonyms are of different kinds according to their nature of opposition and their comparative occurrence in the selected corpus is as given below.
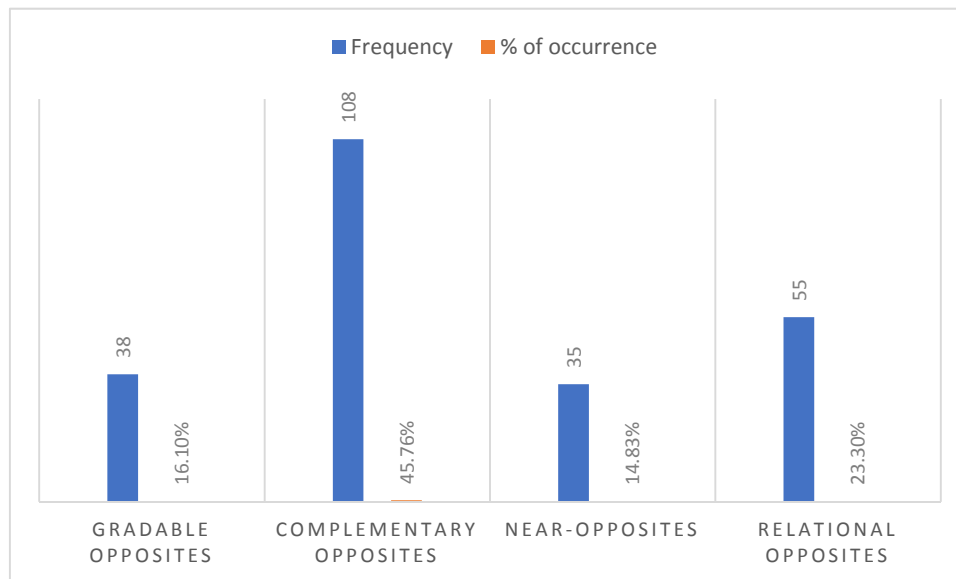

Figure 2. Comparative representation of kind of oppostion relation

3.2.2.1 Antonyms as Gradable Opposites

Many antonyms show gradable quality and are classified under category of adjectives (Fellbaum, 2010) as لمبا/چھوٹا, اچھا/برا, ٹھنڈا/گرم and so on. These pairs of opposites add grading quality of 'more and less'. As ٹھنڈا and گرم both are gradable opposites and we can say that A is colder than B and C is very cold. Some of the nouns also showed quality in grades as in محبت/نفرت and in دوست/دشمن but this quality was not very common in nouns.

The above graph shows that only 38 nouns were found gradable antonyms, which made 16.10 percent of total antonyms. Some examples of nouns showing gradable opposition from the current corpus are کوڑ/سچ, نڈر/ڈر, گرمی/جاڑا, فرمابردار/باغی, میٹھا/کوڑا, بماری/صحت, غربت/دولت, مایوسی/ہمت, رونا/ہاسا...etc.

3.2.2.2 Complementary Opposites

Antonyms are also complementary opposites, which are quite similar to gradable opposites (Vetulani, 2012). These are associated to the opposite conceptual compartments and cannot fall under same concept. For example, if something is not right, it must be wrong and if something is not dead, it must be alive. According to Cruse (1986), there is not any middle way in these opposite categories as found in gradable opposite. Nouns like (ماننا/ناماننا) and ذندہ/مرد,صیح/غلط and adjectives like (صیح/غلط) are examples of complementary antonyms. However, a question of grading complementary opposites is answered by Boholm (2017) that these can be graded rather than using their main sense. For example, when we say that someone is alive not dead but we can say that someone is energetic or lively. As the gradation of دن/رات can be صبح،دوپہر or شام which is not the representation of main sense but itself these are other nouns.

The analysis of antonyms represents the longest bar of complementary opposites in bar-graph with the highest frequency of 108 out of 236 antonyms. It also means that nouns frequently show complementary opposition, for example: حقیقت/خیال, دوزخ/جنت, پانی/اگ, جنگ/امن, زنانی/آدمی, خشکی/آبی, تریاق/زبر, بندی/بنده, سوال/جواب, دُکھ/سُکھ, زندگی/موت, کُڑی/مُنڈا, فتح/شکشت, آسمان/زمین … and so on.

3.2.2.3 Relational Opposites

Gagné and L'Homme (2016) define relational opposites which are also called relative and conversive terms, include the opposite pairs as اوپر/نیچے, آگے/پیچھے, باپ/بچہ and استاد/شاگرد. Egan (1984) further explains that in

relational opposites 'one' cannot be described without suggesting the 'other one'. In this way relational opposites are taken as a subclass of directional opposites e.g., in آگے/پیچھے the description is spatial, while some of opposites showed social roles, for example ڈاکٹر/مریض, باپ/بچہ, استاد/شاگرد or the kinship relations in nouns as ماں/باپ.

Relational opposites are prominent as the second most frequent category of antonyms in the bar graph, which made the 23.30 percent of the total antonyms. Examples of relational opposites are: طلاق/ویاہ, پین/ویر, ابا/امّاں, آقا/سیوک, مربم/زخم, طوطی/طوطا, مریض/حکیم, حوا/آدم, دھی/پُتر, باں/لت, نوکرانی/رانی, جسم/جان, کھسم/بیوی ...etc.

3.2.2.4 Near-Opposites

There are number of cases, where the words do not seem to be in relation of real opposition, but they look like near opposites. Cruse (1986) names them "impure" antonyms, for example the word دیو is between بڑا and چھوٹا likewise, ملک/قصبہ and کام/کھیل, سرگوشی and چیخ etc. The number of nouns, which have been categorized in near-opposites has made 14.84% of the total antonyms. Some examples of near-opposites from the current analysis are شرٹ/پینٹ, سوٹی/سوٹا, گوشت/بڈی, تنا/جڑ, شہر/صوبہ, نثر/شعر, پانڑیں/برف, پیر/بتھ, جُھگّیٰ/محل, تڑکا/شام ...etc. کولی/تھالی.

The overall position of antonym bar in graph shows that the relation of antonym is not frequent in nouns. Only 27.90% of the total corpus showed the antonymy relation as Miller (1990) discussed that this relation of opposition is not a fundamental organizational principle for nouns, although it exists among nouns. It serves as basis for the organization of adjectives in WordNet or other dictionaries.

3.2.3 Meronymy/ Holonymy (Part-Whole Relation)

Meronymy/Holonymy are called part-whole semantic relations. If the part of any whole represents the whole entity, it is known as its meronym (Dave & Jaswal, 2015). The analysis of nouns highlighted that 375 nouns out of 846 had the relation of meronymy e.g., عقل is a meronym of سوچ. When someone talks about عقل, the capability of thinking سوچ comes in our mind, so it can be said that عقل is a representation of سوچ. On the same lines کھلاڑی is a meronym of کھیڈ and یاری is a meronym of یار. The wholes of which the meronyms are part called holonyms. In the examples above سوچ, کھیڈ and یار are the holonyms.

Another important piece of information, which the graph represents is the difference in number of meronyms and holonyms in the given corpus. The number of meronyms were 375, while holonyms were 298 in number. It describes the nature of nouns in the corpus that 35.23% of total corpus showed the relation of holonymy. Further, it explains that 64.73% nouns from the corpus were not the part of any whole but those were individual entities.

While the presence of 46% of meronyms represented that these nouns have their representative parts. Moreover, 54% of the total nouns do not have their representative parts e.g. جادو, امن, پیار, سچ, خبر, حُسن... etc. The possible reason behind it can be the vague or more generalize quality of these nouns. Further, the social relations like (ویر, رانی, بیوی, صدر, میاں, دِھی, پتی... etc.); nouns related to administrative posts (جج, افسر, مالک...etc.); natural materials (تیل, مِٹی, وادی, پتھر, وَٹّا... etc.); inborn feelings (پیار, مزہ, دشمن... etc.) and many general abstractions (امن, شکست, جادو, خیال... etc.) were rare in having meronym relation.

Likewise, name of animals (مور, کاں, وچّھا, گڈا, کتّا, گڈا, لیلا... etc.); a number of man-made things (صوفہ, گولا, کھڈّی ... etc.); name of cultivations (بھنڈی, اناج, باجرہ... etc.) are not the part of any wholes, so these nouns also lack the relation of holonyms.

3.2.4 Relation of Hypernym/Hyponym

Sigman and Cecchi (2002) explain the relation of hypernym/holonymy as a relation of superordinate and subordinate. Here, the target word is defined by looking at the higher category (hypernym) and lower category (hyponym). In the corpus 292 words, which made 34.52% of total nouns showed their hyponym relation. For example, the subordinate category of آنڈا is زردی and hyponym of محل is کمرا, موتی is بار and so on.

Likewise, hypernym of آنڈا is پکھو, ریت is ماده, صوبہ is ضلع, کھنڈ is گڑ and so on. The percentage of hypernym relation found in the corpus of noun was 64.75%. Comparative analysis of both these relations exposed that occurrence of hyponyms was frequent than the occurrence of hypernyms. The reason behind this fact is the nature of nouns in the corpus as مزہ, زبر, گڑھ, ڈیرہ, فن, تصویر, روگ, حُسن, یار, کھیڈ, ولی, سوچ and so on. These type of nouns do not have their superordinate category, so the comparative occurrence of hypernym relation is less.`

The name of games, which are not from Punjabi culture were in lack of showing their subordinate categories, for example وکٹ, باکی and کرکٹ. The names of: animals and birds (باز, بیل, چڑا, بُلبل, طوطا, بطخ, بلد, مچھر, مکّھی, أَلو... etc.), specific instruments (چھج, ٹیکا, کرین etc.); social relations (ویر, امی, بیوی, میاں, پُتر, بہن, پتی, ابو, ماسی, پاپا ...etc.); abstract nouns (پیار, توپ, کنڈا, تیر, تیسی, نیزہ, چابی ... etc.); body parts (بتھ, قدم, دِل, کان, گل, نین, مک, متھا ... etc.); ...

(بمت and a number of other nouns (چوٹ,سوٹا , عینک, دیگ, جیب, جال, پتھر, پگ, زبر, تخت…etc.), do not have their subordinate categories because these are already the final nodes.

The nouns of social positions (صدر and افسر); abstract entities which cannot be measured like (سوچ, حُسن,فن…etc.); nouns which stand as individual (دولت,جیل, تخت, مسجد, محل … etc.); the final unchangeable realities (جان, خواب, جہنم, حُور, وقت, موت, جنت, عمر…etc.) and some of the natural entities (پہاڑ, اگ, خون, چانن, صحت…etc.) did not have their super-ordinate categories.

3.2.5 Has a Part Relation

The presence of Has a part relation is another quality of WordNet, which helps in the physiological representation of any entity in human mind and allows to have a clear picture of the target word (Banerjee & Pedersen, 2003). This relation is also not frequent in common dictionaries. The occurrence of Has a Part relation was 50.94%, which make 431 nouns out of 846 nouns. It means remaining 415 nouns are in lack of showing their parts e.g. خبر, سچ, فن, پیار, ڈر, مزہ, بمت, بانی, خطا, خیال and so on. The analysis of these nouns has showed that many of these nouns belong to the category of abstract nouns. As the other nouns like جُگَّت, دری, دوا, ران, شکر, پتھر, لاش and so on are already so compact, complete and concise, that their parts are not visible to describe.

3.2.6 Relation of Singular/Plural

According to Giunchiglia, Maltese, Farazi and Dutta (2010) the relation of singular/plural is the necessary and common part of every dictionary and of WordNet as well. The nouns of the target corpus have already been converted in their singular form under the heading of correct ontology. As graph highlights the occurrence of plurals: 493 nouns out of 846 nouns, which made the 58.27% of the target corpus.

The nouns like سُکھ, لہو, وار, کاں, یار, سچ, مال, ویاہ, ڈر and کیس do not have their plurals. Moreover, these are used equally for singular and plural forms. There exists no plural relation for feelings and emotions (بمت, ڈر, جیون, غُصّہ, پیار…etc.); vegetables and fruits (آلو, آڑو, ساگ, کدو…etc.); a number of animals and birds (گِدڑ, باز, بُلبل, کرکٹ…etc.); sports (کپاہ, دھان, مکئ…etc.); countries (روس, روم, تُرکی…etc.); cities (جھنگ, جہلم, یمن…etc.); crops (باکی…etc.); directions (مشرق, مغرب, شمال…etc.) and metals (سٹیل, سِلور, لوبا…etc.).

A number of nouns in the corpus belonged to the kind of collective nouns (ڈار, نیوز, ٹبر, آئین, جگ, وچار…etc.) and some of the nouns did not show their plurals including: آٹا,تیل, اگ, مِٹی, دولت, شراب, گھی…etc.

3.2.7 Relation of Masculine and Feminine

The results of the analysis for the relation of masculine/feminine have been shown in the following bar graph.
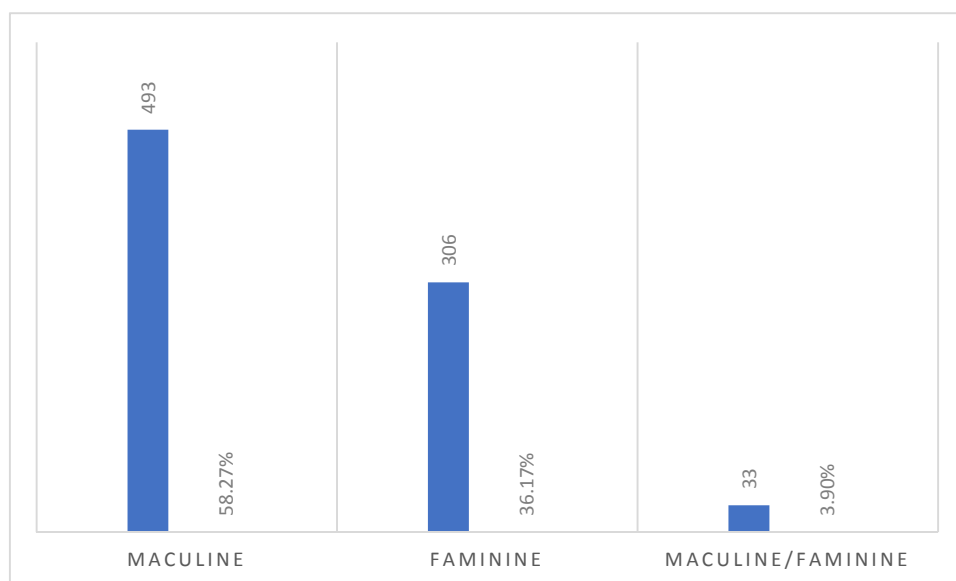
Figure 3. Representation of the relation of Masculine and Feminine

According to Ordan and Wintner (2005) the relations of masculine and feminine for nouns are also very frequent in common dictionaries and in WordNet as well. The ratio of masculine and feminine relations in comparison with whole nouns was 58.27% and 36.17% respectively. Generally, the relations of masculine and feminine show the presence of species and social relations in the society.

The rules for defining gender to inanimate objects is arbitrary and there exist no hard and fast rules to define gender, but a pattern can be established, as the words which end with ای are taken as feminine and the words which end with ا and آ are taken as masculine.

The results also highlighted that 306 nouns out of 846 nouns represents feminine gender, which made 36.17 % of the total corpus. 25 nouns were found purely feminine and their respective masculine gender has been found. Some examples of pure feminine gender are بکری ,گُڈّی ,لڑکی ,کُڑی ,ماں ,مائی ,عورت ,رانی ,امی … etc. Excluding these pure feminine gender, the remaining 281 nouns were identified as feminine following the general pattern rule of آرمی ,موتی ,تصویر ,شام ,خبر ,جنگ ,کتاب ,شے ,سوچ ,اِٹ ,as the ending part of the noun or predecessor of it e.g. ای بجلی ,صدی and so on.

The results of analysis identified the presence of 493 nouns as a masculine in the given corpus. From 493 nouns the 50 nouns were found pure masculine gender, representing any species or any social position including: مور, بندہ ,آدمی ,ألو ,بچہ ,بیگ ,جِن ,مالک ,مرد ,میاں … etc. Excluding the 50 nouns from the total masculine nouns of 493, the remaining 442 nouns were identified as masculine following the general pattern rule of ا and آ as the ending part of the noun or can occupy the position of predecessor. Some of the examples from the given corpus are وار, پیر ,آنڈا ,روگ ,آٹا ,تارا ,حقّا ,خط ,خیال … etc.

After categorizing 493 masculine nouns and 305 feminine nouns the remaining 35 nouns were identified as general nouns, which included both of the relations: masculine and feminine. These are general descriptions, for example تاجر ,گابک ,ملوک ,جابل ,بشر ,مُسلم ,مکین ,مریض … etc. 8 out of 35 nouns showed professions, for example قاصد ,مخبر ,کلرک ,ٹیچر ,ڈاکٹر ,صدر ,and جج. These are used equally for the males and females. These titles are dynamic and not gender specific. Furthermore, 10 nouns represented the species of different birds and animals in the corpus, as بُلبل ,ڈنگر ,کیوی ,گِدھ ,پسُو ,مچھر ,باز ,مچھی and عُقاب. These are the general categories for gender which are equally used for males and females. Some of these type of nouns become specific by adding prefixes, for example نر مچھی and ماده مچھی. There found 33 nouns, which are equally used for males and females. They become specific, when are used with particular gender, as جابل ,بشر ,ضعیف ,ملوک ,مُسلم ,گابک ,مکین ,مریض ,ناک, لوگ ,کہرل ,دوست ,کھیڈَ and لوک.

4. Conclusion

The Present research was set out to explore the lexicosemantic relations of nouns with a future aim to create Punjabi Shahmukhi WordNet. WordNet is taken as a reliable resource to run digital applications of language which ultimately create valid results. Punjabi Shahmukhi has a least exposure in digital world because the digitization process of this language is very slow so, it is aimed to develop its WordNet. This project was undertaken to determine the possible semantic nets of Punjabi Shahmukhi nouns. Punjabi is a rich and diverse language showing a significant number of sematic relations.

The lexicosemantic relations have been found out for 846 nouns. After the relation of masculine and feminine, the second most frequent lexical relation among Punjabi Shahmukhi nouns is relation of synonymy. This is because of a huge number of speakers of this language. As the speaker of any language increases, the diversity in that particular language also increases. The results also help us to conclude that nouns are very rare in showing antonymous relation as Miller (1990) mentions that the relation of antonymy is not a fundamental principle to categorize nouns.

Finally, these lexical and semantic relations of nouns will be the part of Punjabi Shahmukhi WordNet, which is aimed to develope in near future.

References

Augenstein, I., Das, M., Riedel, S., Vikraman, L., & McCallum, A. (2017). *Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications*. arXiv preprint arXiv:1704.02853. https://doi.org/10.18653/v1/S17-2091

Banerjee, S., & Pedersen, T. (2003, August). Extended gloss overlaps as a measure of semantic relatedness. *Ijcai*, *3*, 805–810.

Boholm, M. (2017). The semantic field of risk. *Safety Science*, *92*, 205–216. https://doi.org/10.1016/j.ssci.2016.10.011

Dave, H., & Jaswal, S. (2015, September). *Multiple text document summarization system using hybrid summarization technique* (pp. 804–808). In 2015 1st International Conference on Next Generation Computing Technologies (NGCT). IEEE. https://doi.org/10.1109/NGCT.2015.7375231

Fellbaum, C. (2010). *WordNet* (pp. 231–243). In Theory and applications of ontology: computer applications. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-8847-5_10

Gagné, A. M., & L'Homme, M. C. (2016). Opposite relationships in terminology. Terminology. *International Journal of Theoretical and Applied Issues in Specialized Communication*, *22*(1), 30–51. https://doi.org/10.1075/term.22.1.02gag

Giunchiglia, F., Maltese, V., Farazi, F., & Dutta, B. (2010, May). *GeoWordNet: a resource for geo-spatial applications* (pp. 121–136). In Extended Semantic Web Conference. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13486-9_9

Khokhlova, L. V. (2014). *2. Majority language death*. University of Hawai'i Press.

Kumar, A., & Paul, A. (2016). *Mastering Text Mining with R*. Packt Publishing Ltd.

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

Lehal, G. S., & Singh, C. (2000). *A Gurmukhi script recognition system* (Vol. 2, pp. 557–560). In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. IEEE. https://doi.org/10.1109/ICPR.2000.906135

Lehal, G. S., Singh, C., & Lehal, R. (2001). *A shape based post processor for Gurmukhi OCR* (pp. 1105–1109). In Proceedings of Sixth International Conference on Document Analysis and Recognition. IEEE. https://doi.org/10.1109/ICDAR.2001.953957

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235–244. https://doi.org/10.1093/ijl/3.4.235

Nadiem, I. H. (2005). *Punjab: Land, history, people*. al-Faisal Nashran.

Ordan, N., & Wintner, S. (2005). Representing natural gender in multilingual databases. *International Journal of Lexicography*, *18*(3), 357–370. https://doi.org/10.1093/ijl/eci032

Rose F., & Egan. (1984). *Survey of the history of English synonymy* (pp. 5a–25a). Webster's New Dictionary of Synonyms.

Sharma, D., & Jhajj, P. (2010). Recognition of isolated handwritten characters in Gurmukhi script. *International Journal of Computer Applications*, *4*(8), 9–17. https://doi.org/10.5120/850-1188

Sigman, M., & Cecchi, G. A. (2002). Global organization of the Wordnet lexicon. *Proceedings of the National Academy of Sciences*, *99*(3), 1742–1747. https://doi.org/10.1073/pnas.022341799

Verma, K., & Sharma, R. K. (2017). Comparison of HMM-and SVM-based stroke classifiers for Gurmukhi script. *Neural Computing and Applications*, *28*(1), 51–63. https://doi.org/10.1007/s00521-016-2309-5

Vetulani, Z. (2012). Wordnet and Gödel's Completeness Theorem. *Studies in Logic, Grammar and Rhetoric*, *27*(40), 213–221.