



*Asesorías y Tutorías para la Investigación Científica en la Educación Puig-Salabarría S.C.
José María Pino Suárez 400-2 esq a Lerdo de Tejada, Toluca, Estado de México. 7223898475*

RFC: ATI120618V12

Revista Dilemas Contemporáneos: Educación, Política y Valores.

<http://www.dilemascontemporaneoseduccionpoliticayvalores.com/>

Año: VI

Número: Edición Especial

Artículo no.:38

Período: Agosto, 2019

TÍTULO: Desarrollo de nombres en Punjabi y relaciones léxico-semánticas.

AUTORES:

1. M.Phil. Nosheen Akhter.
2. Ph.D. Muhammad Asim Mahmood.
3. M.Phil. Muhammad Tahir Nadeem.

RESUMEN: Este estudio tiene como objetivo el desarrollo de conjuntos sintéticos del sustantivo Punjabi Language (PL) en Shahmukhi. El estudio ha desarrollado y utilizado un corpus de 2 millones de palabras del Punjabi en escritura Shahmukhi. Se ha tomado una lista de 1000 nombres en inglés de fuentes de aprendizaje en línea del Punjabi. La lista de sustantivos se ha traducido del inglés al gurmukhi y al shahmukhi utilizando el software: Akhar 2016. Este estudio ha desarrollado conjuntos de sustantivos que siguen los diccionarios de WordNet y Punjabi de Princeton, y su metodología se ha ideado de varios términos. Este estudio concluye que los sentidos tomados de Princeton WordNet variaron de la cultura punjabi.

PALABRAS CLAVES: Síntesis de sustantivo, lenguaje punjabí, corpus punjabí, traducción, relaciones léxico-semánticas.

TITLE: Development of Punjabi Noun Synsets and Lexico-Semantic relations.

AUTHORS:

1. M.Phil. Nosheen Akhter.
2. Ph.D. Muhammad Asim Mahmood.
3. M.Phil. Muhammad Tahir Nadeem.

ABSTRACT: This study aims developing synsets of noun of Punjabi Language (PL) in Shahmukhi. This is a corpus-based study. The study has developed a corpus of 2 million words of Punjabi Language in Shahmukhi script for selecting nouns to develop their synsets. The corpus of Punjabi Language was POS tagged and processed through software: AntConc.3.4.4.0 for getting nouns. The list of 1000 nouns was retrieved in English with respect to semantic categories. The developed list of nouns has been further translated from English to Gurmukhi to Shahmukhi using software: Akhar 2016. For the purpose of assistance, Princeton WordNet and Punjabi dictionaries were used in the construction of Punjabi noun synsets. As a result, 5000 synsets of nouns have been developed in term of identity number, word, grammatical category, synsets of noun, number of senses and sentence examples.

KEY WORDS: synsets of noun, Punjabi language, Punjabi corpus, translation, lexicon-semantic relations.

INTRODUCTION.

There are number of local or regional languages prevailing in Pakistan such as Punjab, Pashto, Saraiki, Sindhi, Balochi and Urdu (Bhurgri & Abdul-Majid, 2006). It has been noted that these languages do not have alphabets in standard form. In the present research, Shahmukhi Punjabi is considered as the main consideration. The present study focuses on the development of synsets of nouns of Punjabi Language in Shahmukhi script.

English WordNet is the first work that contains 57,000 noun word forms which have been organized into 48,800-word meanings, named as synsets. This is so-called Princeton WordNet. In case of Punjabi WordNet, no solid attempt has been made on the development of Punjabi Shahmukhi synsets and its WordNet. The present study is the first refined attempt in Pakistan that is about the development of synsets of Punjabi nouns using mixed approach which is a combination of merge and expansion approaches as practiced by Kaur, Sharma, Preet and Bhatia (2010).

The basics of WordNet are viewed as a large lexical source of English, Punjabi, Urdu and etc. The development of WordNet is constituted with the ingredients: nouns, verbs, adjectives and adverbs in form of group as a set of cognitive synsets which express a different concept. The term *synsets* is explained as they are interconnected under the concepts of lexico-semantic relations. The structure of WordNet is a helpful source for Natural Language Processing (NLP) and Computational Linguistics (CL). This study comes under NLP.

The focus of the present study is made on the constituent of WordNet such as noun synsets of PL. This time 5000 synsets of Punjabi nouns have been developed following the Princeton WordNet, Akhar (2016) and online dictionaries such as “ijunoon” and سبقدوش. The list of nouns has been used for the development of synsets of PL.

The important ingredients for the development of synsets of Punjabi nouns are as follows:

- 1) Identification number (ID) of Punjabi word from the database.
- 2) Number of senses of the word taken from Princeton WordNet, Akhar 2016 (a software), Punjabi dictionaries and online dictionary (Hypernym - Free English online dictionary based on WordNet 3.0, ijunoon dic.com and سبقدوش) and students of Punjabi discipline.
- 3) Synonyms of an individual senses.
- 4) Gloss for the synset.
- 5) Example from the developed Punjabi corpus and above-mentioned sources.

There are number of approaches being used by the scholars in the construction of WordNet such as, merge approach, expansion approach, parts of speech (POS) tagging approach, transliteration approach and etc. In this connection, the present study has practiced mixed approach which is combination of merge, expansion, POS tagging and machine transliteration approach.

Using expansion approach, this approach is helpful where there is already work has been done. This approach is well guided in the process of building WordNet of the source language. The semantic relations can also be borrowed from the source language being used in target language. The advantage of this approach is helping in borrowing semantic relation of the given WordNet (Bhattacharyya Puspak, 2010).

This expansion approach was used for development of Urdu WordNet (Ahmed, 2010; Adeeba & Hussain, 2011). In the same way, Kaur et al., (2010) have used an expansion approach for the development of Punjabi WordNet relations and the classification of synsets using Hindi WordNet; Rattan and Bhatia (2011) used this approach for the creation of Punjabi WordNet and Punjabi Hindi Bilingual Dictionary.

The contribution of works on the development of Punjabi nouns is viewed as: Nouns can be divided into different groups based on their inflection which are mentioned by Bhatia (1993, p. 164-166) and Shackle (2003, p. 600-601).

Hasan, Iqbal, Azeemi and Javeed (2015) recently carried a study on online Punjabi Shahmukhi lexical resource. The study constructed a database for Pakistani regional languages in order to process cross lingual information in form of word sense disambiguation, machine translation and part of speech. In the last, according to Humayoun and Ranta (2010) the study was devised in order to develop corpus, building of a lexicon for PL and described implementation of morphology. This study was considered as source of development of Punjabi corpus in Shahmukhi script as developed in the present study.

The present study was conducted to probe into the following speculated research questions:

- 1) What are the synsets of the Punjabi nouns in Shahmukhi?
- 2) What are the problems encountered by the researcher during developing synsets of the Punjabi nouns in Shahmukhi?

DEVELOPMENT.

Literature review.

WordNet is a lexical resource of language, in which words are packed together in accordance with the similar meaning. WordNet's language contains compounds, collocations, phrasal verbs, and idiomatic phrases, the word is the basic unit (Fellbaum, 1998).

The design of WordNet looks like a thesaurus in which building block is a synset containing of all the words that express a given concepts (Miller, 1991). Similarly, WordNet is defined as: WordNet resembles a traditional dictionary; for example, WordNet gives definitions and sample sentences for most of its synsets. A definition is valid for all the synonyms in a synset, since it expresses the meanings of the concept. The sample sentences may not be felicitous for all synonyms, and often different sentences are given for different members of the synset (p. 9).

PakWordNet was constructed for major Pakistani languages. In a nutshell, WordNets for Pakistani languages are being developed following expansion approach using Urdu WordNet which has been built at University of Engineering and Technology (UET), Lahore.

Approaches for the WordNet creation.

In construction of WordNet, there are two main approaches which can be helpful for the construction of WordNet and the following approaches are merge approach and expansion approach.

Merge approach.

In case of merge approach, different senses in which a word can be used is first recorded (Puspak, 2010). In this regard, the lexicographers build a synset for each of the sense, following the three principles of minimality, coverage and replaceability for synset creation.

In merge approach, there is no distracting influence of another language, which particularly will happen when the lexicographer will encounter cultural and regional concepts of the source language. The quality of the WordNet will be good, if the synset maker has a good knowledge of the language. But the disadvantage for this process is that is very time consuming.

Expansion Approach.

In expansion approach, the synsets of a source language SL are provided. The taken synset is sensibly studied for its meaning. Therefore, target language's words which are representing that meaning are collected and kept together in a set in repeatedly order. This approach is helpful where there is already work has been done.

This approach is well guided in the process of building WordNet of the source language. The semantic relations can also be borrowed from the source language in order to be used in target language. The advantage of this approach is helping in borrowing semantic relation of the given WordNet (Puspak, 2010). As the current study is quite similar with respect to borrowing sysnets relation from English WordNet using transliteration method.

The expansion approach for the creation of WordNet implemented in the development of EuroWordNet was also the principal methodology experienced in the construction of IndoWordNet. This expansion approach was used for development of Urdu WordNet (Ahmed, 2010). Similarly, the same approach was practiced by Adeeba and Hussain (2011) experiencing in building the Urdu WordNet. The whole Punjabi WordNet was developed using expansion approach. In the same way, Kaur et al. (2010) employed an expansion approach for the development of Punjabi WordNet

relations and the classification of synsets using Hindi WordNet. Likewise, this approach was used in creation of Punjabi WordNet and Punjabi Hindi Bilingual Dictionary by Rattan and Bhatia (2011). As these studies practiced the same approach in developing of WordNet the present study practiced the expansion approach. In the end, if both source and target languages belong to the same language family, the role of expansion approach becomes more attractive, as cancelling influences of regional and cultural concept is supposed to be minimal in this scenario.

Constituent Elements of WordNet.

WordNet deals four syntactic categories separately such as Noun, Verb, Adjective and Adverb. The separation is because of the semantic differences among the relations that connect words and concepts from the four syntactic categories (Miller, 1995). First of all, WordNet has the potential of 80,000 noun word forms which have been organized into some 60,000 lexicalized concepts. Miller views as many of these nouns are collocations; a few artificial collocations invented for the convenience of categorization. In this study, the ultimate focus is on noun word forms which are used to build synsets of Punjabi language in Shahmukhi script.

Noun in WordNet.

The focus of the present study is focusing on the construction of synsets of noun. Noun is one of the constituents which are involved in the development of WordNet.

Punjabi nouns are inflected in number and case. They also have an inherent gender that can be masculine or feminine. Morphologically speaking, it has five cases such as direct, oblique, vocative, ablative, and locative/instrumental; the last two cases are vestigial (Shackle, 2003, p. 599). Since locative/instrumental case is extremely rare to find, we omit it and define just the first four cases in our implementation.

Nouns can be divided into different groups based on their inflection which are mentioned by Bhatia (1993, p. 164-166) and Shackle (2003, p. 600-601). Many past works were seen on nouns and its functions in WordNet.

Similarly, Haq (1987) distributes nouns in two major classes such as proper noun and common noun. He also divides proper nouns further into four sub-groups such as attributive name, alias, title and nom-de-plume, and common nouns are further categorized into four sub-groups such as state, collective, locative and instrumental nouns. In contrast, Siddique (1971) categorizes nouns on the basis of structure and nature. According to nature as noun, it has been into substantive, quality nouns and pronouns.

As far as noun is concerned, noun in WordNet is an information about hyponymic relations between nouns is given in the definitional phrases of conventional dictionaries (Amsler, 1980). The function of noun in WordNet is forming a lexical inheritance system which is systematic struggle has been done to link hyponyms with their hypernyms and vice versa.

In order to understand the construction of WordNet, a number of studies were carried out of which “Nouns in WordNet” conducted by George Miller (1993), “Modifiers in WordNet” carried by Katherine Miller (1993). Similarly, Turney and Littman (2003) and Turney (2005) used paraphrases as features to analyze noun-modifier relations. Paraphrases expressed more overtly the semantic relation between a noun and its modifier.

Hasan et al. (2015) recently carried a study on online Punjabi Shahmukhi lexical resource. The study constructed a database for Pakistani regional languages in order to process cross lingual information in form of word sense disambiguation, machine translation and POS. This study was recently conducted but the researchers did not expose the detail of list of noun words. The present study was basically was devised in order to extend work on Shahmukhi PL.

The significant study on computation of gender of Urdu nouns under morphology has been carried by Akbar and Abid (2009). This study has been made for interested people who know the correct gender of Urdu nouns for automatic POS tagging.

Synset and Concept.

Synset is a group of synonyms words. A synset shows a unique different sense or also called concept (Hasan et al., 2015). For example, (مدرسه, اسکول) shows a complete concept of (Institution) as synset. Different types of words are seen in WordNet as:

Polysemous.

In which words have multiple senses are recognized as polysemous. In case of WordNet, it is perceived as each word has many synsets as it has senses. For example, کرسی word shares five senses such as کسے وی چیز دی نی یا کرسی, لوکاں د سردار یا لیڈر, کسے اوچے بندے دا رتبہ, ڈھووالی چوکی, پیڑھی, موڑھا and خاندان دی لڑی یا پیڑھی

Monosemous.

In which words can have only one sense are called as monosemous. For instance, چیتا word has one sense such as تیز طرار وڈی بلی

Compound Words.

Because of single words, WordNet synsets also sometimes contain compound words which are made up of two or more words but are treated like single words in all respects. For Example, پٹواری word is in English (land record officer).

Gloss.

Gloss is used to describe the concept. It is having two components: text definition and example sentence (Hasan et al., 2015). In order to understand this phenomenon, the given below example

has shown five senses of کرسی word. Each sense was properly defined according to the sense such as خاندان دی; کیسے وی چیز دی نی یا کرسی; لوکاں د سردار یا لیڈر; کیسے اچے بندے دارتبه; ایک بندے دی بین لی چوکی یا پیڑھی لڑی یا پیڑھی

Text Definition.

It is a primary part which is used to explain the concept as highlighted by the synset (Narang & Bhatia, 2012; Adeeba & Hussain, 2012). For example, the word کرسی has text definition as ٹولا کے بیٹھن دی شے

Example Sentence.

It is secondary part of gloss which is used to give the usage of the words in the example sentence (Narang & Bhatia, 2012; Adeeba & Hussain, 2012). For example, کرسی word is exemplified in sentence as او نے اپنا کوٹ کرسی دے پیچھے ٹنگیا نے بے گیا

Recent advancement in Pakistan on WordNet by Adeeba and Hussain (2011), the researchers reported on the development of Urdu WordNet from Hindi WordNet. The technique they followed that was to translate downloaded Hindi WordNet into Urdu WordNet taking 50000 unique words and organized into 28967 synsets.

Another study was carried by Kaur et al. (2010), the study described the development of Punjabi WordNet by employing an extension approach in Hindi WordNet under Indradhanush WordNet Project. In development of Punjabi WordNet contained the characteristics of origin, symbols, morphological and syntactic were presented in the study. The lexical semantic relations were elaborated in Punjabi WordNet. In conclusion of this study, approximately 35000 synsets were identified in Hindi WordNet was exactly the same in categorization in synsets in relation with Punjabi WordNet.

According to Humayoun and Ranta (2010), the study was devised in order to develop corpus, building of a lexicon for PL and described implementation of morphology. In this study, the developers found some transliteration systems to exchange “Shahmukhi” and “Gurumukhi” with partial solution.

Research methodology.

This section deals methodology for the current study. The present study is about the development of noun synsets of PL in Shahmukhi.

Before creating noun synsets relation, this study has followed few steps: firstly, the corpus of 2 million words of Shahmukhi has been developed and used in this study. Secondly, two list of nouns have been retrieved from online PL source. Thirdly, a machine transliteration system has been used in this study. Fourthly, these lists of nouns have been compared with the developed corpus of PL. Fifthly, a corpus of Punjabi Language has been POS tagged manually. Next, the total outcome of nouns in 2 million words of Shahmukhi has been observed.

The most frequent 1000 noun words of Punjabi corpus have been semantically categorized and finalized to create Punjabi synsets of noun using Princeton WordNet and Punjabi dictionaries; moreover, the experiences of developing synsets of nouns in Shahmukhi have been discussed in detail. These above-mentioned steps are the significant components of the present work and they have been discussed below.

Development of Corpus.

The corpus size of PL was 2 million words. The data was extracted from online websites (as discussed in coming section) in form of online news items, novels, poetry, short stories, drama and columns. Some data was taken from newspapers such as “Khabrain Lokae” from Missal Publishers, Aminpur Bazar, Faisalabad, and from open source set by Dr. Hamayoun (2010).

Process of Retrieval of Corpus.

The corpus for this study was developed following the certain steps. First, after going to online sources all PL data in Shahmukhi scripts were manually downloaded from online sources in the form of categorizations such as novels, poetry, columns, drama and online news in PL. Secondly, all data were kept in form of classification.

Development of lists of nouns.

For this study, two lists of noun words of PL involved the following steps. Initially, a list of Punjabi words was taken from two major sources: online lexical Punjabi learning and grade 5 English textbook. These lists of words were in English translated into Punjabi using online dictionaries.

1500 nouns were taken from online source: lexical Punjabi learning.

Table 1. Detail List of Semantic Categories of Lexical Punjabi Learning.

Sr. No.	Semantic categories	Number of Nouns	Sr. No.	Semantic categories	Number of Nouns
1	Agriculture	15	25	Hotel	27
2	Animals	48	26	Insects	23
3	Articles in Common daily use	47	27	Institutions, building and their parts	81
4	Birds	27	28	In the restaurant	25
5	Bank	37	29	In the shop	22
6	Collective Noun	22	30	Industry	15
7	Colours	36	31	Material Set	64
8	Clothes and Dresses	71	32	Metals	11
9	Crime and theft	15	33	Minerals	14
10	Days	07	34	Nature and Natural things	33
11	Dinner Menu	22	35	Ornament and Jewellery	27
12	Directions	15	36	Parts of body	74
13	Diseases	79	37	Parts of the Car	16
14	Domestic Articles	42	38	People	20
15	Drink Set	16	39	Places	34
16	Edibles	25	40	Print Media	20
17	Education	25	41	Professional	82
18	Feelings	18	42	Public Transportation	15
19	Flowers	09	43	Relations	52
20	Foods	40	44	Sports	49
21	Fruits	22	45	Stationary	45
22	Groceries	19	46	Vegetables	28
23	Grains and Crops	23	47	Tools	28
24	Hospital	14	48	Weather	17

Table 2. Detail List of Nouns from Kindergarten to Grade 5th Books.

Kindergarten	1 st Grade	2 nd Grade	3 rd Grade	4 th Grade	5 th Grade
43	91	90	62	83	86

Table 3. List of Sources of Nouns.

Lexical Punjabi learning	Grade 5 th English Textbook	Accumulative Words of Nouns
1500	455	1755

Machine Transliteration.

The machine transliteration system was employed in order to translate lists of nouns words from English language to Shahmukhi. In this connection, the lists of noun words in English were made in corpus expressions (as given in Appendix) and then were translated into Punjabi (i.e. Gurmukhi) using ‘Google translator’ (See Figure 1 below). After that, the translated Punjabi (i.e. Gurmukhi) corpus expressions were placed in software: Akhar (2016), a Punjabi translator converted Gurmukhi into Shahmukhi. Finally, lists of Punjabi words were translated in Shahmukhi.

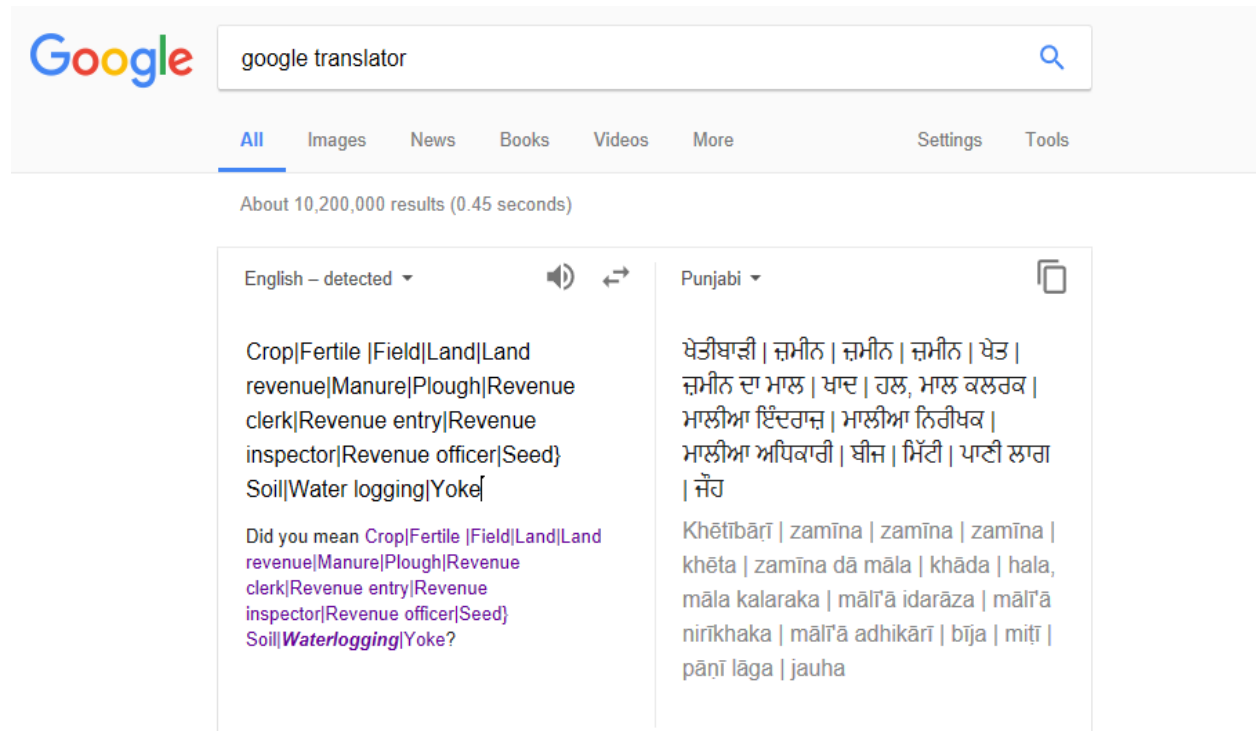


Figure 1. Google Translation from English to Gurmukhi.

After having converted from English to Gurmukhi, such Gurmukhi expressions were placed in a software: Akhar; a Punjabi translator that converted Gurmukhi into Shahmukhi (See Figure 2 below).

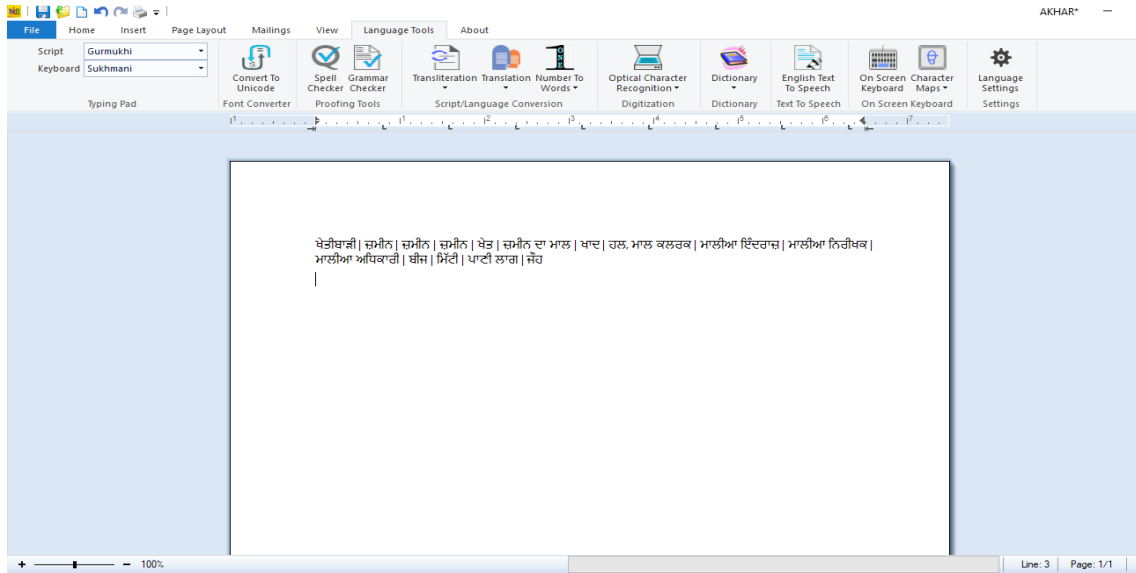


Figure 2. Akhar Translation from Gurmukhi to Shahmukhi.

After having translated data in Gurmukhi from Google translator, the data was again translated from Gurmukhi to Shahmukhi through software: Akhar, a Punjabi translator. See figure 3 below.

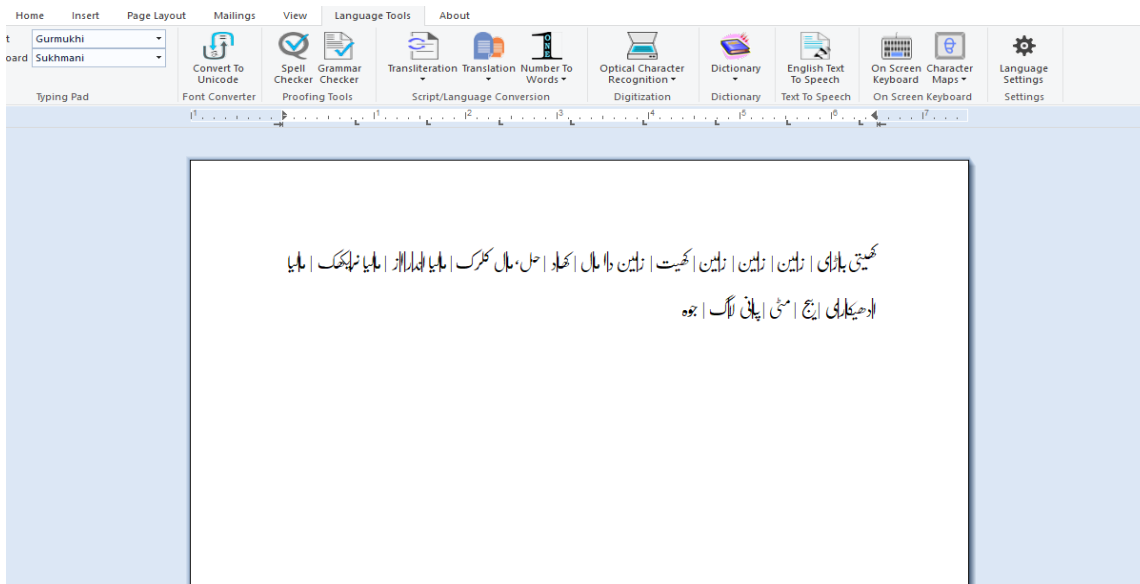


Figure 3. Translation in Shahmukhi.

For tagging Punjabi corpus, a number of steps were taken. The detail has been given below.

Process of Tagging Punjabi Corpus.

This is very significant process of data tagging. This process of tagging includes 5 steps. First, for doing this step an online software (i.e. Online Shahmukhi to Gurmukhi transliterator) which converts Shahmukhi script into Gurmukhi script. This step is manually completed with great care. This step is called transliteration. Secondly, the converted Gurmukhi script is manually tagged in accordance with parts of speech (POS). This step is regarded as POS tagging. At third, the tagged Gurmukhi data is converted back to Shahmukhi script. This process is known as back transliteration. While translating from Gurmukhi to Shahmukhi, a number of problems in form of formatting, missing words and unrecognizing words. Fourthly, all these problems are manually cleaned. This way of cleansing is known as manual cleansing. At last, the possible solutions are directed to the identified problems. This final step is normally perceived as problem solving. The graphical representation of process of tagging is given below.

Table 4. Process of Tagging

Step 1	Machine Transliteration	Conversion of Shahmukhi to Gurmukhi Script
Step 2	POS tagging	Gurmukhi script is tagged
Step 3	Back Machine Transliteration	Conversion of tagged Gurmukhi to Shahmukhi
Step 4	Manual Cleansing	Identifying unknown words, tags and etc.
Step 5	Problem solutions	Removing unknown tags, manipulating tags and etc.

Step 1: Machine Transliteration.

At first, go to online converter of Shahmukhi to Gurmukhi script “Online Shahmukhi to Gurmukhi Transliterator”. After pasting shahmukhi script in left side as mentioned in this snapshot, translate into Gurmukhi and copy that converted stuff into notepad files.

ਪੰਜਾਬੀ ↔ ਪੇਸ਼ਾਬੀ Punjabi – Gurmukhi ↔ Shahmukhi Transliteration

Home | PMT Home

Transliterate Text

Enter your Text

جنت وی جانده دوزخ وی جانده
اقرا وی جانده عوزے کوثر وی جانده
اے نیت دی گل اے عمل دی نیسی
اسی عشق قلندر مرنا وی جانده
بہار عشق دے نالوں سانوں موت حولی
اسی چک لہا اپنا دفتانا وی جانده
گل عشق پا کے اسی رونا کیوں
سادا دکھاں دا شوق اخیراے اسی نچ کے پار منانا وی جانده

Shahmukhi - Punjabi -> Gurmukhi - Punjabi Transliterate

Suggest a better Transliteration

Transliterated Text

ਜੰਤ ਵੀ ਜਾਂਦੇ ਦੋਖ ਵੀ ਜਾਂਦੇ
ਅਕਰਾ ਵੀ ਜਾਂਦੇ ਖੋਰੇ ਕੋਸਰ ਵੀ ਜਾਂਦੇ
ਏ ਨੇਤ ਦੀ ਗਲ ਏ ਅਮਲ ਦੀ ਨੇ
ਅਸੀ ਅਸ਼ਕ ਕੁਲੰਦਰ ਮਰਨਾ ਵੀ ਜਾਂਦੇ
ਭਾਰ ਅਸ਼ਕ ਦੇ ਨਾਲੋਂ ਸਾਨੋਂ ਮੋਤ ਹੋਲੇ
ਅਸੀ ਚੁਕ ਲਾਭਾ ਅਪਨਾ ਦਫ਼ਨਾਨਾ ਵੀ ਜਾਂਦੇ
ਗਲ ਅਸ਼ਕ ਪਾ ਕੇ ਅਸੀ ਰੋਨਾ ਕਿਉ
ਸਾਦਾ ਦੁਖਾਨਾ ਦਾ ਸ਼ੁਕ ਅਖੀਰਾਏ ਅਸੀ ਨਚ ਕੇ ਪਾਰ ਮਨਾਨਾ ਵੀ ਜਾਂਦੇ

Figure 4. Visualization of Machine Transliteration from Gurmukhi to Shahmukhi.

Step 2: Parts of Speech (POS) Tagging.

After having machine transliteration, the next stage is called POS tagging in which retrieved Gurmukhi converted script is pasted in the upper box and get tagged data at lower box through online software “Online Punjabi Part of Speech Tagger”. This way of tagging is known as online POS tagging.

Research Centre for Punjabi Language Technology
Punjabi University, Patiala

Punjabi Part of Speech Tagger

Input Punjabi Text

ਆਖ ਨੀ ਮਾਏ ਆਖ ਮੇਰਾ ਹਾਲ ਸਾਈਂ ਔਗੇ ਆਖ ਨੀ!
ਪਰੇਮ ਦੇ ਧਾਗੇ ਅੰਤਰ ਲਾਗੇ ਸੁਲਾਂ ਸੇਤੀ ਮਾਸ ਨੇ
ਨਿਜ ਜਨੌਂਦਿਓ ਭੋਲਏ ਮਾਏ ਜਨ ਕਰ ਲਾਈਓ ਪਾਪ ਨੀ!
ਕਰੇ ਹੁਸੋਂ ਫ਼ਕਰ ਨਮਾਨਾਂ ਜਾਂ ਦਾ ਮੇਲਾ ਆਪ ਨੇ

Tagged Output

ਆਖ\N_NNP ਨੀ\N_NNP ਮਾਏ\PSP ਆਖ\PSP ਮੇਰਾ\DM_DMMD
ਹਾਲ\N_NN ਸਾਈਂ\PSP ਔਗੇ\N_NN ਆਖ\PSP ਨੀ\N_NN
ਪਰੇਮ\N_NN ਦੇ\PSP ਧਾਗੇ\N_NN ਅੰਤਰ\PSP ਲਾਗੇ\N_NN
ਸੁਲਾਂ\N_NNP ਸੇਤੀ\N_NNP ਮਾਸ\N_NN ਨੇ\PSP ਨਿਜ\N_NN
ਜਨੌਂਦਿਓ\PSP ਭੋਲਏ\N_NN ਮਾਏ\PSP ਜਨ\N_NN ਕਰ\N_NN VM
ਲਾਈਓ\PSP ਪਾਪ\N_NN ਨੀ\N_NN ਕਰੇ\N_NN VM_VF
ਹੁਸੋਂ\N_NN VAUX ਫ਼ਕਰ\UnKnown ਨਮਾਨਾਂ\UnKnown

ਖਾਲੀ ਥਾਂ ਲੁ ਇ ਈ

Rule Based Statistical

Tag the text

View Punjabi POS Tag Set

Figure 5. Visualization of POS Tagging.

Step 3: Back Machine Transliteration System.

As completing online tagging, the retrieved tagged data is again translated through software “Online Shahmukhi to Gurmukhi Transliterator”. Actually, this software translates Shahmukhi to Gurmukhi and vice versa. This conversion of tagged Shahmukhi data into Gurmukhi script is called back machine transliteration system.

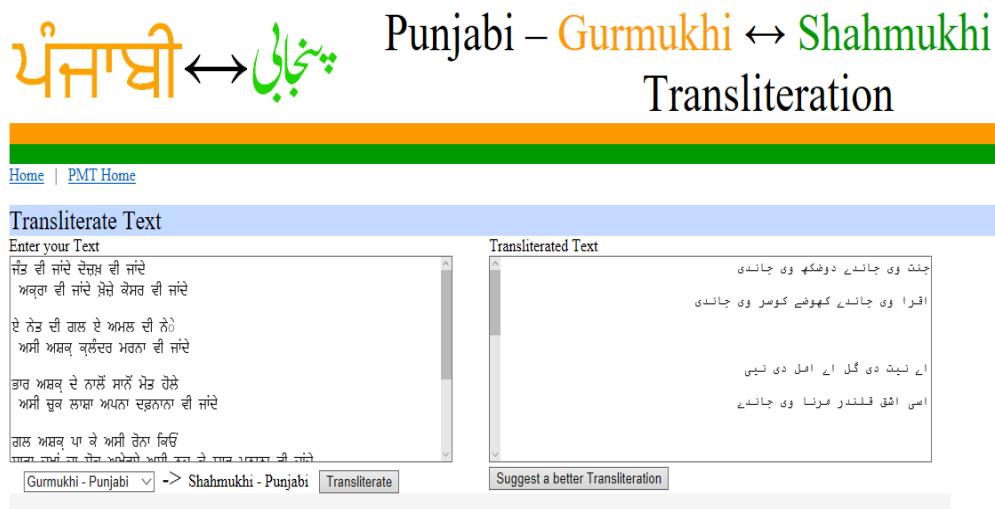


Figure 6. Visualization of Back Machine Transliteration System.

Step 4: Manual Cleansing.

After having completed this tagging process, the ultimate retrieved data is manually observed, and the differences are noted. Due to phonetic differences, while translating some consonants variation are identified.

Comparison of Data.

After completion of data collection, the lists of nouns were devised in form of corpus expressions. Afterward, those expressions were compared and matched with the target source: extracted nouns from tagged Punjabi corpus using software: AntConc.3.4.4.0. The most frequent words which were matched with corpus expressions. They were considered for further analysis. The detail of words has been given in Appendix.

Semantic Categorization and Finalization of Lists of Noun Words.

The most frequent noun words after execution of matched comparison between lists of nouns from online sources and the list of nouns from the developed corpus, the finalized data of 1000 noun words was categorized semantically as above mentioned in Table 4. The detail lists of noun words as categorization have been given in Appendix.

Development of Noun Synsets of Punjabi Language.

The main and significant purpose of this study is about the development of synsets of noun of Punjabi Language. In order to create synsets of noun, the following components of the development of noun synsets were devised in form of identification number (i.e. ID of noun words), list of noun words, numbers of senses, synsets of noun, synonyms, gloss (i.e. definition of word), example sentence (i.e. providing example sentence from developed corpus, Princeton WordNet and Compleat Lexical Tutor). For this purpose, the excel sheet for this activity was used instead lexical matrix. The detail of process of creation of synsets has been given below.

Semantic Category.

This was the first step which was actually providing the semantic category of noun words during the creation of synsets of noun. It was already discussed in above section that 48 semantic categories of noun words were taken in case of creation of synsets of noun. For example, Agriculture was one of the semantic categories of noun words. The purpose of categorization of noun words was keeping data in a comprehensive way for the readers.

Identification Number.

The second step of creating noun synsets was giving cardinal number to the noun words for identification. This way helped in searching words out of data; for example, ID no. 1.

List of Noun Words.

In this section, the finalized noun words were taken in this column of “list of words”. These 1000 noun words were considered for synsets of noun. For example, کرسی is one of the lists of noun words. The detail of noun words has been given below in Appendix.

Numbers of Senses.

This section showed the actual senses of noun word which were existed as per different contexts. This is a convenient source of recording senses of noun word while making synsets of noun. For example, کرسی word shared five senses in numbers such as sense 1, sense 2, sense 3, sense 4, and sense 5.

Grammatical Category.

In order to develop synsets, there are number of syntactic categories such as verb, adjective, noun, and adverb which are the components of WordNet. In this relation, the present study focused on noun synsets. The only grammatical category noun was used for creating synsets of noun words.

Synsets of Noun.

This section has great importance because this study is based on this section. This present work studied synsets or senses of noun words. The purpose of this activity was giving more senses of noun contextually. For this purpose, senses of noun words were extracted from developed corpus of Punjabi, Princeton WordNet and online dictionaries. The more meanings were considered as synsets or senses of noun words. For example, کرسی word shared five senses such as:

خانداں دی and کیے وی چیز دی نی یا کرسی , لوکاں د سردار یا لیڈر , کیے اچے بندے دا رتبہ , ڈھووالی چوکی, پیڑھی, موڑھا
لڑی یا پیڑھی

Gloss.

Gloss means definition of word. This section is also an important because a single word has multiple senses and each sense is defined as per contextual meaning. This study provided definition of every sense of noun word. In order to understand this phenomenon, the given below table has shown five senses of *ਕਰਸੀ* word. Each sense was properly defined according to the sense.

Sentence Example.

The last component of creating synset is sentence example. This component was done after having assistance from developed corpus of Punjabi language, Princeton WordNet, Complete Lexical Tutor and Punjabi dictionaries. This study extracted sentences of searched word and its senses from the said sources. In this connection, sentence examples were brought into Shahmukhi using software: Akhar 2016 and online Gurmukhi to Shahmukhi Translator. For example, *ਕਰਸੀ* word has five senses. Each sense has been exemplified through sentences.

Table 5. Visual Representation of Semantic Category of Noun Words.

Semantic Category	Word ID	List of Words	No. of Senses	Grammatical Category	Synset/Senses	Gloss/Definition	Sentence Example
In the Restaurant	1	ਕਰਸੀ	Sense 1	Noun	ਠਹੋਰੀ ਚੋਕੀ, ਪਿਠੀ, ਮੋਠਾ	ਇਕ ਬੰਦੇ ਦੀ ਬਿਨ ਲੀ ਚੋਕੀ ਯਾ ਪਿਠੀ	ਓਨੇ ਅਨਾਕੋਟ ਕਰਸੀ ਦੇ ਪਿਠੇ ਠੰਗੀ ਟੇ ਗੇ ਗਿਆ
			Sense 2	Noun	ਰੱਬੇ, ਮਕਾਮ	ਕਿਸੇ ਅਚੇ ਬੰਦੇ ਦਾ ਰੱਬੇ	ਅਨੋ ਮੇਅਸ਼ੀਅਤ ਓਚ ਅਚੇ ਡਰੇ ਡੀ ਕਰਸੀ ਲਬ ਗੀ
			Sense 3	Noun	ਅਫਸਰ, ਅਕਤਦਾਰ, ਲੀਡਰ	ਲੋਕਾਨ ਦ ਸਰਦਾਰ ਯਾ ਲੀਡਰ	ਅਪਨੇ ਖੀਅਲਾਟ ਨੋ ਕਰਸੀ ਟਕ ਪਿਠੋ
			Sense 4	Noun	ਨੀ, ਬਨੀਅਦ	ਕਿਸੇ ਓ ਚੀਜ਼ ਡੀ ਨੀ ਯਾ ਕਰਸੀ	ਮਿਰੇ ਗੇਹਰ ਡੀ ਕਰਸੀ ਮੀ ਨਲ ਖਰਾਬ ਬੋਨੰਦੀ ਜਾਨੰਦੀ
			Sense 5	Noun	ਖਾਨੰਦਾਨ, ਪਿਠੀ	ਖਾਨੰਦਾਨ ਡੀ ਲੜੀ ਯਾ ਪਿਠੀ	ਓਡੇ ਖਾਨੰਦਾਨ ਡੀ ਕਰਸੀ ਮਗਲਾਨ ਨਾਲ ਮਲੰਦੀ ਅ

Developing Template of Synset.

After defining each above-mentioned step, in the end of creation the researcher has developed a template viewing a complete picture of synsets. The synset entry format in file is shown.

Table 6. Synset Entry Format.

Sr. No.	Tags	Descriptions
1	WORD CAT	The semantic category of noun word
2	ID	The synset identifier.
3	WORD	Search word
4	SYNSET-PUNJABI	It gives the set of synonyms for the sense in the Punjabi language
5	CAT	The syntactic category of the sense.
6	CONCEPT	It explains the concept represented by the synset.
7	EXAMPLE	It gives the usage of the words of the synsets in the sentence

Tools for this Research.

In this study, multiple tools were brought into use for the convenience of the researcher. The following software as tools were used to reduce the consumption of time, such as:

- 1) Online Shahmukhi to Gurmukhi Transliterator.
- 2) PakInPage to Unicode Converter.
- 3) Notepad.
- 4) Part of Speech tagger system.
- 5) AntConc 3.4.4.0

Results.

The lists of nouns after comparing with developed corpus, these finalized words have been discussed in upcoming table. As for as results are concerned, 5000 Punjabi synsets have been developed using 1000 noun words of Shahmukhi Punjabi.

The numerical results of the lists of nouns have been explained in Appendix. These words have been supposed for data analysis and discussion. The two major sources: from Kindergarten to Grade 5th books and a semantic category of lexical Punjabi learning have been considered for data analysis. The first source has been discussed through Punjabi synsets.

Discussion.

This section deals synsets of Shahmukhi Punjabi nouns which have been developed and analyzed through exemplifications. The given below tables are the examples of data taken from the data analysis of Punjabi nouns of Shahmukhi. As for as discussion of synsets is concerned, the experiences and the problems of developing synsets of Punjabi nouns have been encountered. The detail of semantic categories has been discussed ahead.

Semantic Category: Agriculture.

Table 7. Synsets of Punjabi Noun فصل (Crop)

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Agriculture	1	فصل	Sense 1	Noun	فصل, پکی ہوئی فصل	کاشت کردہ زمین دی پیداوار	پاکستان فصل دیے اعتبار نال زرعی ملک ا
			Sense 2	Noun	بوٹا, پودے	کاشت کیتا ہویا بوٹا	ساڈے ملک وچ لوکی اپنے باغاں وچ فصل بیجنا شروع ہو گئے
			Sense 3	Noun	کٹھ, ہجوم	لوکاں دا کٹھ	پنج پرا اک فصل انگو کٹھے جوان ہوئے
			Sense 4	Noun	انسان دی بنائی ہوئی شے, مصنوعی	موسمی فصل	آج کل لوک بے موسمی فصل اگان لگ پے
			Sense 5	Noun	ذخیرہ, مال	غلہ یا پیداوار	فصل دا ذخیرہ کرن والی لوک مجرم ہوندے
			Sense 6	Noun	اوجھڑی, جانور دا معدہ, پوٹا	جانور دی اوجھڑی جس وچ اناج ہوندا	ہر جانور دیے پیٹ وچ اک فصل نا پری اوجھڑی ہوندی

Semantic Category: Animals.

Table 8. Synsets of Punjabi Noun بھینڈ (Sheep).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Animals	10	بھینڈ	Sense 1	Noun	جانور, میما	بکرے دا بچہ	وڈی عید لئی اسی بھینڈ خرید لے
			Sense 2	Noun	بندہ, پولا بندہ	پولا بندہ جیرا آرام نل پھس جاوے	شریف آدمی آج دیے دور وچ اک بھینڈ سمجھیا جاندا
			Sense 3	Noun	بندہ, پیروکار	گل منن والا	ساڈے ملک وچ ڈاکٹراں دی بھینڈ چال چلی ہوئی آجکل

Semantic Category: Articles in Common Daily Use.

Table 9. Synsets of Punjabi Noun پٹی (Strap).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Articles in Common Daily Use	39	پٹی	Sense 1	Noun	فیتہ، تسمہ، ڈوری، سیبا	چیز نوں بنن والی شے	مزدور نے بوبت مضبوطی نال ڈبہ پٹی نال بن دیتا
			Sense 2	Noun	پٹی	کسی چیز دا ٹکڑا	لکڑ دی پٹی آگ بالن لئی استعمال ہوندی آ
			Sense 3	Noun	بیلٹ، ڈوری	کسی چیز نوں بنن والی بیلٹ	پینٹ نال پٹی بننا ضروری سمجھیا جاندا اے
			Sense 4	Noun	بینڈز، دو لمکدیاں ہوٹیاں پٹیاں جو پادری اڑے وکیل کالر دوآلے پاؤندے ہن؛	گلے وچ پان والی پٹی	وکیل عدالت وچ کالے رنگ دی پٹی لائے اندے آ
			Sense 5	Noun	ہتھکڑیاں، بیڑیاں	ہاتھ بنن لئی سنگلیاں	پولیس مجرمان دے ہتھوں نوں پٹی بن کے عدالت وچ لے جاندی آ

Semantic Category: Bank.

Table 10. Synsets of Punjabi Noun قرضہ (Debt).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Bank	52	قرضہ	Sense 1	Noun	ادھار، قرض	ادھار لئی ہوئی شے	قرضہ اک بیماری دی طرح جان نہی چھڈ دا
			Sense 2	Noun	دینداری، لین دین	احسان ہونا	میں اپنے دوست دا قرضہ ادا نہی کر سکدا

Semantic Category: Birds.

Table 11. Synsets of Punjabi Noun طوطا (Parrot).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Birds	59	طوطا	Sense 1	Noun	طوطا، پرندہ	سبز رنگ دا اڑن والا پرندہ	بچیاں نوں طوطے پالن دا بوبت شوق پندا آ
			Sense 2	Noun	رٹے باز، طوطے وانگ کوئی گل رٹ لین والا آدمی	طوطے وانگ کوئی گل رٹ لین والا آدمی	ساڈے اسکول سسٹم وچ بچیاں نوں طوطے وانگو یاد کرن نوں کہندے آ

Semantic Category: Clothes and Dresses.

Table 12. Synsets of Punjabi Noun کوٹ (Coat).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Clothes and Dresses	67	کوٹ	Sense 1	Noun	کوٹ پہناوا	کپڑیاں دے اتو پان والی شے	دولہے نے شادی تے پینٹ کوٹ پایا ہویا سی
			Sense 2	Noun	پرط، تہہ، ملما، لیپ، پوچا	زمین یا کوئی وی شے تے رنگ کرنا	سلمہ نے کندے رنگ دا کوٹ پہیریا
			Sense 3	Noun	چھلکا، چھل،	کسی کھان والی شے دا غلاف	کیلے دے کوٹ دا رنگ پیلا ہوندا آ
			Sense 4	Noun	وردی	کسی جگہ تے پان والی ضروری وردی	بچے نے اسکول دی وردی نال کوٹ پایا ہویا سی

Semantic Category: Crime and Theft.

Table 13. Synsets of Punjabi Noun حملہ (Attack).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Crime and Theft	85	حملہ	Sense 1	Noun	وار	کسی چیز دا وار	جنگل وچ شیر نے ہرن تے حملہ کر دیتا
			Sense 2	Noun	بیماری دا دورہ	کسی بیماری دا حملہ	مریض نوں شوگر دے حملے تے معذور کر دیتا
			Sense 3	Noun	حملہ، چڑھائی،	دشمن تے جنگ وچ چڑھائی کرنا	پاکستانی فوج نے بھارت حملے دا پورا جواب دیتا

Semantic Category: Dinner Menu.

Table 14. Synsets of Punjabi Noun پھل (Fruit).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Dinner Menu	95	پھل	Sense 1	Noun	پھل، میوہ	درخت دا پھل	میرے ابا جی روز گھر پھل لے کہ اندھے آ
			Sense 2	Noun	اولاد، نسل	اللہ دی دیتی ہوئی نعمت اولاد	اللہ نے انسان نوں سکون دین لئی اولاد دی شکل وچ پھل نہ نوازی
			Sense 3	Noun	انعام، بدلہ	کسی چیز دا بدلہ	عثمان نے امتحان وچ پہلی پوزیشن لے کہ محنت دا پھل پا لیا
			Sense 4	Noun	نتیجہ	محنت دا صلہ	محنت دا پھل کدی رایگاں نہی جانداں

Semantic Category: Diseases.

Table 15. Synsets of Punjabi Noun خون (Blood).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Diseases	107	خون	Sense 1	Noun	رت، لہو، خون	نساں وچ چلن والا خون	گہرے زخم ناں بوہت خون وگ گیا سی
			Sense 2	Noun	ساک، رشتہ، نسل، اولاد	خونی رشتہ	ساڈا خون دھوکہ نہی دے سکدا
			Sense 3	Noun	فساد، قتل	کسی نو جان توں مار دینا	مجرم نے خون کر کے خودکشی کر لی
			Sense 4	Noun	فطرت	کیسے دے خون وچ شامل خوبی	انصاف کرنا اودھے خون وچ شامل ا

Semantic Category: Domestic Articles.

Table 16. Synsets of Punjabi Noun رسی (Rope).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Domestic Articles	138	رسی	Sense 1	Noun	ڈوری، تاگا	کپڑے سین لئی دھاگہ	مزدور نے کپڑے بنن لئی رسی سیدھی کیتی
			Sense 2	Noun	جالا، تار	بجلی دیاں تاراں	پتنگ دی رسی نے بچے دی گردن نوں زخمی کر دیتا
			Sense 3	Noun	پھندا،	رسی دا بنیا جال	مچھوارے نے شکار لئی رسی نوں دریا وچ سٹیا

Semantic Category: Drink Set.

Table 17. Synsets of Punjabi Noun چاہ (Tea).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Drink Set	165	چاہ	Sense 1	Noun	چاہ-پتی، پتی؛	پتی جیڑی چاہ وچ پیندی	چاہ نے حکومت بوبت ٹیکس لائی جاندی آ
			Sense 2	Noun	چاہ	دودھ دی بنی ہوئی چاہ	آج کل ہر کوئی چاہ دا شدھائی ہويا پیا اے
			Sense 3	Noun	قہوہ	پانی نے پتی دا بنیا قہوہ	پٹھان سردیاں نوں چاہ پیندے نے

Semantic Category: Education.

Table 18. Synsets of Punjabi Noun کتاب (Book)

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Education	169	کتاب	Sense 1	Noun	کتاب	کوئی علم دی لیکھ	پیر کامل اک بڑی اچھی کتاب وا
			Sense 2	Noun	کاپی، نوٹ-بک، رجسٹر	لکھن لئی کاپی	دکاندار نے ادھار دیاں چیزاں نو کتاب وچ لیکھ لیا
			Sense 3	Noun	قصہ، کہانی، واقعہ	کسی واقعہ دا خلاصہ	میری زندگی دی کتاب بند ہون والی آ
			Sense 4	Noun	الہامی کتاب	اللہ دی کتاب	قرآن اللہ دی کتاب آ

Semantic Category: Edible Items.

Table 19. Synsets of Punjabi Noun مکھن (Butter).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Edible Items	187	مکھن	Sense 1	Noun	مکھن، مکھنی؛	دودھ دا بنیا مکھن	ماں اپنے نیانے نوں مکھن نال روٹی دیندی آ
			Sense 2	Noun	روغن، تیل	مکھن دا بنیا تیل	بانڈی وچ مکھن نال سالن دا ذائقہ ودھ جاندا اے
			Sense 3	Noun	چکنیاں چوڑیاں، گلاں، چاپلوسی، خوشامد	مٹھیاں گلاں	دھوکے باز بندہ اپنی مکھن ورگیاں گلاں نال غریب لوکاں نوں لٹ ریا سی
			Sense 4	Noun	سوهنا بندہ	بوہت سوہنا بندا	عاشق لئی محبوب مکھن ورگا ہندا اے

Semantic Category: Feelings.

Table 20. Synsets of Punjabi Noun پیار (Love)

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Feelings	201	پیار	Sense 1	Noun	پیار، محبت	پیار والا احساس	وڈیاں نوں نیانیاں نال پیار کرنا چاہی دا اے
			Sense 2	Noun	چاہت، پسند	کسی نوں بوہت پسند کرنا	مینوں اپنے رب نال بہت پیار وا
			Sense 3	Noun	معشوقہ، محبوبہ	کسی نال عشق ہو جانا	نوی نسل پیار دے جھانسه وچ پاگل ہوئی پئی اے
			Sense 4	Noun	رونقی بندہ، سوہنی چیز	چیڑا ماحول نوں خوش گوار بنا دے	چارلی چیپلن ایک بوہت پیار والا بندا سی
			Sense 5	Noun	میل جول	دعا سلام ودھانا	پوری قوم نوں آپس وچ پیار نال رہنا چاہی دا اے

Semantic Category: Food.

Table 21. Synsets of Punjabi Noun شراب (Alcohol).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Food	217	شراب	Sense 1	Noun	انگور دا شراب، انگوری وائین	انگور دا رس	انگور دا رس نال شراب بندی آ
			Sense 2	Noun	پھلاں توں بنی شراب	سارے پھلاں دا مرکب رس	کوکتیل شراب سارے پھلاں دی رس نوں ملا کے بنائی جاندی آ
			Sense 3	Noun	سوبا رنگ	لال رنگ	پنڈ وچ رہن والے لوکاں دا رنگ شراب ونگوں سوبا ہوندا اے

Semantic Category: Professionals.

Table 22. Synsets of Punjabi Noun مالک (Owner).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Professionals	510	مالک	Sense 1	Noun	مالک، والی-وارث	کسی چیز دا مالک	باپ دی موت توں بعد پتر جائیداد دا مالک ہندا اے
			Sense 2	Noun	سرمایہ دار ہندا	جیڑے بندے کول بوہت پیسہ ہووے	ساڈی ٹیم نوں جتن دی خوشی وچ شہر دی مشہور کمپنی دی مالک نے دعوت دیتی آ
			Sense 3	Noun	سردار، مالک، حاکم	کسی علاقے دا سردار	ساڈے پنڈ دا مالک بڑا ظالم سی
			Sense 4	Noun	زمیندار، جاگیردار	بوہت زیادہ زمین جائیداد والا ہندا	سندھ وچ آج وی مالک راج کردے نے

Semantic Category: Weather.

Table 23. Synsets of Punjabi Noun ہوا (Air).

Semantic Category	Sr. No.	List of Words	No. of Senses	Grammatical Category	Synset/Sense	Concept/Gloss	Example
Weather	616	ہوا	sense 1	Noun	وا، ہوا	ہوا دا چلنا	ٹھنڈی ٹھنڈی ہوا لوکاں دی منہ نے خوشی لے آندی آ
			sense 2	Noun	موسم	موسمی ہوا	موسمی ہوا نال بیماریاں وچ اضافہ ہو جاندی اے
			sense 3	Noun	رنگ ڈھنگ	طور طریقے	پیسے نال لوکاں دی ہوا خراب ہو جاندی اے
			sense 4	Noun	نخرہ	ادا	خوبصورت بندے دی ہوا ہمیشہ اچی رہندی آ

CONCLUSIONS.

This study concludes that the development of nouns of PL and the problems experienced during the analysis of the data have been reported. This study answers that 5000 noun synsets of Punjabi language have been developed.

During the development of noun synsets, a list of 1000 nouns has been finalized for making noun synsets. After the development of the list of nouns, a mixed approach has been practiced. A mixed approach is a combination of both merge and expansion approach. For the development of Punjabi synsets of nouns, a merge approach has been used for developing gloss and sentence examples of synsets of Punjabi nouns.

An expansion approach has been employed for constructing senses of Punjabi nouns. In this study, POS tagging, machine transliteration system and back transliteration system have been used for developing IDs for noun words, list of nouns and senses of Punjabi nouns.

The study also deals the problems experienced during the construction of noun synsets of Punjabi language. Such as the collection of list of nouns has been developed after having a great fatigue as mentioned in chapter 3; the senses of nouns have been developed using expansion approach with great efforts using Princeton WordNet, Akhar (2016) and online dictionaries: ijunoon and سبقدوش; lastly, the gloss and sentence examples have been constructed manually after taking assistance from the students of Punjabi discipline and Punjabi dictionaries.

BIBLIOGRAPHIC REFERENCES.

1. AbdulJaleel, N., & Larkey, L. S. (2003). Statistical transliteration for English-Arabic cross language information retrieval. In Proceedings of the twelfth international conference on Information and knowledge management (pp. 139-146). ACM.
2. Adeeba, F., & Hussain, S. (2011). Experiences in building the Urdu WordNet. Asian Language Resources collocated with IJCNLP 2011, 31.

3. Arbabi, M., Fischthal, S. M., Cheng, V. C., & Bart, E. (1994). Algorithms for Arabic names transliteration. *IBM Journal of Research and Development*, 38(2).
4. Akhar, (2016): <http://akhariwp.com/>
5. Princeton WordNet: <https://wordnet.princeton.edu/>
6. Bhatia, T. K. (1993). *Punjabi: a cognitive-descriptive grammar*. Routledge, ISBN 9780415003209.
7. Bhattacharya, P. (2010) IndoWordNet. In: lexical resources engineering conference, Malta Narang, Ashish, R. K. Sharma, and Parteek Kumar. "Development of Punjabi WordNet." *CSI transactions on ICT 1.4* (2013): 349-354.
8. Bhurgri, A. M. (2006). *Enabling Pakistani Languages through Unicode*. Microsoft Corporation white paper at <http://download.microsoft.com/download/1/4/2/142aef9f-1a74-4a24-b1f4782d48d41a6d/PakLang.pdf>
9. Haq, M. Abdul. (1987) “اردو صرف و نحو”, *Amjuman-e-Taraqqi Urdu* (Hind).
10. Hasan, E., Iqbal, M. M., Azeemi, Q. R., & Javeed, A. (2015). *An online Punjabi Shahmukhi Lexical Resource*.
11. Humayoun, M., & Ranta, A. (2010). *Developing Punjabi Morphology, Corpus and Lexicon*. In *PACLIC* (pp. 163-172).
12. Knight, K., & Graehl, J. (1998). Machine transliteration. *Computational Linguistics*, 24(4), 599-612.
13. Kaur, R., Sharma, R. K., Preet, S., & Bhatia, P. (2010). Punjabi WordNet relations and categorization of synsets. In 3rd national workshop on IndoWordNet under the aegis of the 8th international conference on natural language processing (ICON 2010), Kharagpur, India.
14. Khan, S. A., & Khan, M. A. (2009). *Computation of Gender of Urdu Nouns*.

15. Malik, M. G. (2006). Punjabi machine transliteration. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics.
16. Nastase, V., Sayyad-Shirabad, J., Sokolova, M., & Szpakowicz, S. (2006). Learning noun-modifier semantic relations with corpus-based and WordNet-based features. In AAAI (pp. 781-787). Online Dictionary: http://www.ijunoon.com/punjabi_dic/
17. Online Dictionary: <http://urduthesaurus.com/index>
18. Rosario, B., Hearst, M., & Fillmore, C. (2002). The descent of hierarchy, and selection in relational semantics. In Proc. of ACL 2002, 417–424.
19. Rosario, B., & Hearst, M. (2001). Classifying the semantic relations in noun-compounds via a domain specific hierarchy. In Proc. of EMNLP 2001, 82–90.
20. Rattan, R., & Bhatia, P. G. (2011). Creation of Punjabi WordNet and Punjabi Hindi Bilingual Dictionary (Doctoral dissertation).
21. Siddiqi, Dr. Abu-ul-Lais. (1971). “جامع القواعد”, Markazi Urdu Board, Lahore. 67.
22. Sharma, D. V. (2011). An Analysis of Difficulties in Punjabi Language Automation due to Non-standardization of Fonts.
23. Shackle, C. (2003). Panjabi, in Cardona, George; Jain, Dhanesh, The Indo-Aryan Languages, Routledge, 581–621, ISBN 9780415772945.
24. Turney, P. (2005). Measuring semantic similarity by latent relational analysis. In Proc. of IJCAI 2005, 1136–1141.
25. Turney, P., & Littman, M. (2003). Learning analogies and semantic relations. Technical Report Technical Report ERB-1103. (NRC#46488), National Research Council, Institute for Information Technology.

26. Wan, S., & Verspoor, C. M. (1998). Automatic English-Chinese name transliteration for development of multilingual resources. In Proceedings of the 17th international conference on Computational linguistics; Volume 2, 1352-1356. Association for Computational Linguistics.

DATA OF THE AUTHORS.

1. **Nosheen Akhter.** Visiting Lecturer in English, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan.
2. **Muhammad Asim Mahmood.** Dean Faculty of Social Sciences, Government College University, Faisalabad, Pakistan. Email: masimrai@gmail.com
3. **Muhammad Tahir Nadeem.** Primary School Teacher, Government Primary School, MC 22, Gujranwala, Pakistan. Email: dtetahirnr@gmail.com

RECIBIDO: 2 de julio del 2019.

APROBADO: 13 de julio del 2019.