

Bhojpuri WordNet: Problems in Translating Hindi Synsets into Bhojpuri

Imran Ali

Banaras Hindu University

Varanasi, India

aravimran@bhu.ac.in

Praveen Gatra

Banaras Hindu University

Varanasi, India

praveengatra@bhu.ac.in

Abstract

Today, artificial intelligence systems are incredibly intelligent, however, they lack the human-like capacity for understanding. In this context, sense-based lexical resources become a requirement to develop artificial intelligent machines. Lexical resources like Wordnets have received scholarly attention because they are considered crucial sense-based resources in the field of natural language understanding. They can help the machines in knowing the intended meaning of the communicated texts, as they are focused on the concept rather than the words. Wordnets are available only for 18 Indian languages. Keeping this in mind, we have initiated the development of a comprehensive wordnet for Bhojpuri. The present paper describes the creation of the synsets of Bhojpuri and discusses the problems that we faced while translating Hindi synsets into Bhojpuri. Some of the challenges are lexical anomalies, lexical mismatch words, synthesized forms, lack of technical words, etc. Nearly 4000 Hindi synsets were mapped for their equivalent synsets in Bhojpuri by following the expansion approach. We have also worked on the language-specific synsets, which are unique to Bhojpuri. This resource is useful in machine translation, sentiment analysis, word sense disambiguation, cross-lingual references among Indian languages, and Bhojpuri language teaching and learning.

1 Introduction

Today's era is one of science and technology. People have been communicating using the Internet and social media and enjoying different forms of media and entertainment. For this, they require accessible resources in their own languages; however, we Indians are forced to depend on the tools that are available, either in English or only in a few major Indian languages. The creation of linguistic resources in a language, particularly in a low-resourced language, is a very challenging task. To understand the intended meaning of a communicated text, one needs knowledge of the world along with competency in

the language, which cannot be captured with any traditional resources as meaning resides not in the words but in the minds of the people using them (Nida, 1979). We need a very comprehensive and intelligent tool to understand a text like a human. In recent years, wordnets have been considered a very crucial tool in the field of natural language processing. WordNet is an online lexical resource and a semantic network (Bhattacharyya, 2010). It is constituted of synsets. Each synset expresses a distinct concept. So synsets are the basic building blocks of WordNet (Bhattacharyya et al., 2006). WordNet's design is inspired by the current psycholinguistic theories of human lexical memory (Miller, 1998). WordNet stores lexical items in ontological that are used to represent IS-A-KIND-OF, IS-A-PART-OF and other relations such as the hypernymy-hyponymy and holonymy-meronymy. Wordnets have been developed for more than 200 languages (Rebele et al., 2016) because wordnets are considered to be the most important lexical resource available for natural language processing tasks like word sense disambiguation, information retrieval, machine translation, sentiment analysis, and as well as for language learning and teaching. Wordnets have been developed for 18 Indian languages (Bhattacharyya, 2010). Except Hindi, all Indian language wordnets have been developed following the expansion approach, and Hindi has been considered as their source language (Bhattacharyya, 2010). Bhojpuri is a spoken by millions of people in India, as well as in several countries such as Mauritius, Nepal, and others throughout the world. There are fewer efforts have been made in the realm of digitization and the development of lexical resources for this language. This is the motivation behind the creation of Bhojpuri WordNet. The main goal of this paper is to create synsets for Bhojpuri. We have discussed the creation of Bhojpuri Synsets, considering Hindi Synsets as its source language. We followed the expansion approach to create Bhojpuri synsets.

The paper is divided into six sections. Section 2 discusses the related research on wordnets, particularly in the Indian context. Section 3 briefly explains the Bhojpuri WordNet, its methodology, and the statistics of the Bhojpuri Synsets. Section 4 covers the problems and difficulties encountered while translating Hindi synsets into Bhojpuri. Section 5 explores Bhojpuri language-specific synsets, and we conclude the paper in the last section.

2 Review of literature

The first WordNet was developed for the English language at Princeton University in 1985 by G. A. Miller. It contains only content words. It doesn't give pronunciation, etymology, usage notes, or pictorial illustrations. The current structure of WordNet was inspired by Levin's works English Verb Classes and Alternations (Miller, 1995). Levin tried to organize more than 3,000 English verbs into categories based on their common behavior and meaning (Levin, 1993). WordNet is structured in lexical hierarchies in the form of synsets. Synsets are a set of synonyms. Minimality, coverage, and replaceability (MCR) principles govern the creation of the synsets (Bhattacharyya, 2010). Minimality means the synonyms must have minimal differences from other synonyms, coverage is that the synonyms must cover the concept, and replaceability is the synonyms that could be substituted in most cases without changing the meaning of a concept. Here, in WordNet, the focus shifts from words to concepts (Dash et al., 2017). Later wordnets for European languages were developed under an umbrella project for 8 European languages like Dutch, Spanish, Italian, German, French, Czech, and Estonian (Vossen, 2002). It was named Euro WordNet and developed under the headship of P. Vossen from 1996 to 1999 (Vossen, 2002). Each concept was linked to the closest synset in Princeton's WordNet. So it allows cross-language information retrieval from one language to another. In recent, there were efforts to develop lexical resources for low resource languages like KangleiWordnet. It was developed at IIIT, Manipur. For its development, both the linkage approach and the expansion approach were applied to (Nongmeikapam, 2023). It is an integrated wordnet of 5 major local languages of Manipur, viz., Manipuri, Tankhul, Thadou, Mao, and Kabui wordnets. For KhagleiWordNet, the linked language is Manipuri instead of Hindi and English is used as the sec-

ondary language. Apart from it, (M, 2017) worked for Tirukkural WordNet. He used the expansion approach, but Tamil as the pivot language.

2.1 Indian Language WordNets

Hindi WordNet was the first wordnet and was started in 2000 and developed in 2006 at IIT Bombay. Since then, wordnets for a number of Indian languages have been developed, in parallel with Hindi WordNet (Narayan et al., 2002). Hindi WordNet is a system for bringing together different lexical and semantic relations between Hindi words. The design of the Hindi WordNet is inspired by the famous English WordNet. It was developed using the merge approach and further linked with English WordNet for cross-lingual references. No attempt was made for compound and conjunct verbs. Each synset was mapped onto some places in the ontological structure of wordnet with a specific synset ID number. Linkages between nominal and verbal, adjectival and adverbial concepts like ability link, capability link, and functional, or derived from, modified nouns have been additionally added (Narayan et al., 2002).

IndoWordNet: IndoWordNet is a project similar to EuroWordNet. It is a linked lexical resource for 18 Indian languages' wordnets (Dash et al., 2017). However, Hindi has been their pivot language, and they followed the expansion approach (Bhattacharyya, 2010). In the expansion approach, the lexicographers translate the source synsets in the target language. It allows to add or drop synonyms in the synset depending upon the language richness. Unlike Hindi WordNet, it covers typical complex Indian language phenomena like complex predicates and causative verbs (Dash et al., 2017). Due to the morphological richness and different cultural traits of Indian languages, a linkage approach was also adopted (Dash et al., 2017).

Assamese WordNet (AWN): Assamese WordNet was developed at Guwahati University. (Moromi, 2019) dealt with the design and development of the AWN. She followed the expansion approach. Problems, challenges, and complexities faced in the development of the AWN have been briefly discussed in her Ph.D dissertation. This work also classifies Assamese text by utilizing AWN.

Bangla WordNet: Dash, N.S., and his team worked for the development of Bangla WordNet at ISI Kolkata, IIT Kharagpur, and Jadavpur University (Dash, 2017b). They followed the expansion

approach and used Hindi as a source language. The encountered challenges are paradigmatic lexical gaps in wage terms, reordering of phrases, differences in flora and fauna, lexical mismatches, and false cognates during the synset creation for Bengali.

Gujarati WordNet: DDU Gujarat worked for Gujarati WordNet. According to Bhattacharyya (Bhatt et al., 2017), synsets of Hindi were translated into Gujarati following the expansion approach. Sources of translation were Bhagvat and Mandal (Patel, 1958) and the Gujarati Lexicon (Chandariya, 2005). Till 2017, 108 Gujarati language-specific synsets have been recorded.

Kashmiri WordNet: The University of Kashmir developed Kashmiri WordNet and compiled 29469 synsets for Kashmiri (CFILT, 2023). It also used Hindi as a pivot language and followed the expansion approach (Kak et al., 2017). The authors talk about language-specific synsets (LSS) for Kashmiri.

Konkani WordNet: Amrita University started working for Konkani in 2009, and till 2023, approximately 32370 synsets (CFILT, 2023) have been developed following the expansion approach. (Desai et al., 2017) classifies two types of challenges. They are discrepancies and issues in the source language, and challenges due to differences in the source and target languages.

Marathi WordNet: Bhattacharya and his team at IIT Bombay worked on the Marathi WordNet, which was created utilizing the expansion approach from the Hindi WordNet (HWN) (Popale and Bhattacharyya, 2017). The lexicographer’s experience is that Hindi and Marathi are close members of the same family, as many Hindi words have the same meaning in Marathi. However, they also find it difficult to find a single word to express the concepts of HWN, lack color concepts, and have borrowed some words from Hindi. The developers think that there is a need for LSS for Marathi.

Odia WordNet: The University of Hyderabad has worked for Odia WordNet by following the expansion method. It is an interlingual WordNet in Odia (Mohanty et al., 2017). The authors identify some gaps that were encountered in kinds of wages, derivation of nouns from nouns or adjectives, complex kinship in Hindi, and the absence of some Hindi concepts in Odia. They think that there is a need to create an LSS for some new or unique expression of Odia.

Punjabi WordNet: Thapar University and Punjabi University worked for Punjabi WordNet. Rattan (2011) used the expansion approach and used Hindi as a source language for Punjabi. The author developed a web application for the Punjabi-Hindi bilingual and Punjabi-Hindi-English trilingual dictionaries. The IL-MultiDict tool has been used for the creation of Punjabi WordNet (Rattan and Bhattia, 2011). The authors observe a lower number of synonyms in Punjabi in comparison with the Hindi.

Sanskrit WordNet: Kulkarni and his team worked for Sanskrit WordNet at IIT Bombay. Sanskrit WordNet was developed using the Synskarta tool (Kulkarni et al., 2010). It is an online interface for synset creation following the expansion approach specific to Sanskrit. However, it has additional information like etymology, references, and expectancy for the words. (Nair, 2011) worked for the most celebrated thesaurus in Sanskrit. This work is a web application for the Sanskrit ontological representation of each word in Amarakosha named ‘Amarakośajñānājālam’.

Tamil WordNet: Tamil University worked for the Tamil WordNet. 25419 Tamil synsets (CFILT, 2023) have been made using the Hindi synsets (Dash et al., 2017). (Rajendran et al., 2002) claim that the majority of co-synonyms listed under a synset in Hindi are deceptive since they group terms together with diverse meanings. They advised that it would be better if an independent wordnet was made for Tamil.

Telugu WordNet: Dravidian University worked for Telugu WordNet, and 21091 synsets (CFILT, 2023) have been developed using the expansion approach. (Arulmozi and Kesava Murty, 2017) have discussed the problems, challenges, and complexities faced in the development of the Telugu WordNet. For many kinship terms, particularly in gender terms and younger-elder issues, it is a problem to have their equivalent in Hindi.

Urdu WordNet: Urdu WordNet was developed at Jawaharlal Nehru University, New Delhi. (Rahman et al., 2017) list technical difficulties, cultural inadequacy, and synset linking issues while creating the synsets of Urdu from Hindi by following the expansion approach. They suggested translation, transliteration, derivation, neologism, multi-words, and explanation to tackle the issues.

In our survey, we find no work has been done towards the synsets creation in favor of Bhojpuri till 2021. We assume that there is also a need for

lexical resources in Bhojpuri, as wordnets have emerged as a crucial resource developed for NLP applications. So we started working on the development of synsets for Bhojpuri.

3 The Bhojpuri WordNet (BWN)

The Bhojpuri WordNet is a sense-based lexical resource for the Bhojpuri. It has been developed following the expansion approach and has used Hindi as its source language. Bhojpuri WordNet interface enlists synset ID, synonyms, gloss, examples, and word categories and represents the concepts in MCR principles of WordNet. Since the Bhojpuri WordNet uses Hindi as its pivot language, many indigenous concepts practiced by the Bhojpuri community are not listed in the Hindi WordNet. So this WordNet also includes the Bhojpuri Language-specific synsets (BLSS) for total inclusion of the indigenous knowledge of the community on the technical front.

3.1 Methodology of Bhojpuri WordNet

Many a time, the source language and the target language have a strong kinship relationship. In such a case, the expansion approach becomes all the more attractive since the distracting influences of cultural and region-specific concepts are minimal (Sharma and Kumar, 2017). 17 Indian languages' wordnets were developed following the expansion approach and used Hindi as a source language (Dash et al., 2017). Since Bhojpuri is closely related to Hindi, we used Hindi synsets as a source resource and developed the BWN using the expansion approach. We are using the IL-MultiDict synset creation tool to record equivalent Bhojpuri synsets in parallel to Hindi synsets.

First, we look at the Hindi synsets that appear in the IL-MultiDict synset creation tool, and then we look for concepts in Bhojpuri; if the concept is available in Bhojpuri, we find out equivalent synonyms. Translated synsets are validated based on a bilingual **Bhojpuri-Hindi Shabdkosh**¹ and a multilingual **Bhojpuri-Hindi-English Shabdkosh**² offline dictionaries. We also used online dictionaries like **Glosbe**³ and **Jogira**⁴. We checked words' frequency in the Bhojpuri Language Technological Resources (BHLTR) corpus (Ojha, 2019).

¹Bhojpuri-Hindi Shabdkosh by Tiwari, A. (2019)

²Bhojpuri-Hindi-English Shabdkosh by Neeren, A. (2018)

³<https://hi.glosbe.com/>

⁴<https://jogira.com/bhojpuri-hindi-dictionary-and-translation/>

At last, we got validations of the concepts and their frequency to maintain the MCR principles via 5 native speakers and 2 experts in Bhojpuri. After the validation, we add or include synonyms available in Bhojpuri, save them into the database, and proceed to the next synset. So far, out of 4000 Hindi synsets, we could find only 3267 equivalent Bhojpuri synsets. Figure 1 shows the IL-MultiDict tool used for the development of Bhojpuri synsets. The tool's left panel shows Hindi synsets (source language), and the right panel shows Bhojpuri synsets (target language). The given concept *nīmana*: {nīmana, āchā, baḍhimyā, baḍhimā, bhālā, nika, nimana, sajjanagood} of Bhojpuri is equivalent to 'good' in English. The figure 2 depicts the complete architecture of the Bhojpuri synset creation methods.

3.2 The Bhojpuri WordNet and Synset Statistics

The Bhojpuri WordNet consists of 3267 synsets following the expansion approach, nearly 4000 Hindi synsets were taken into account and mapped for their equivalent translation or for their near counterparts in Bhojpuri. It lowered the quantity of concepts because of the linguistic lacunarity. Only 3267 Hindi synsets could be translated into Bhojpuri, 311 Hindi synsets could not be identified in Bhojpuri; 190 proper names were ignored; and there are still synsets that need to be resolved. To ensure the reliability and consistency of the synonyms of the language, they were cross checked against the Bhojpuri-Hindi bilingual online or offline dictionaries, the **BHLTR corpus**⁵ (Ojha, 2019), and other resources like an online website **Jogira**⁶. The POS statistics of the study are as follows: 2720 nouns, 119 adjectives, 385 verbs, and 43 adverbs.

4 Issues and Challenges

Since we are using Hindi synsets to create Bhojpuri synsets following the expansion approach, we have to translate the Hindi synsets into Bhojpuri. While translating from one language to another, we encountered many lexical and semantic gaps due to the socio-cultural differences, morphological richness of the languages, and so on. Therefore, we also faced many difficulties while creating synsets for Bhojpuri. In this section, we are going to dis-

⁵<https://github.com/shashwatup9k/bho-resources>

⁶<https://jogira.com/>

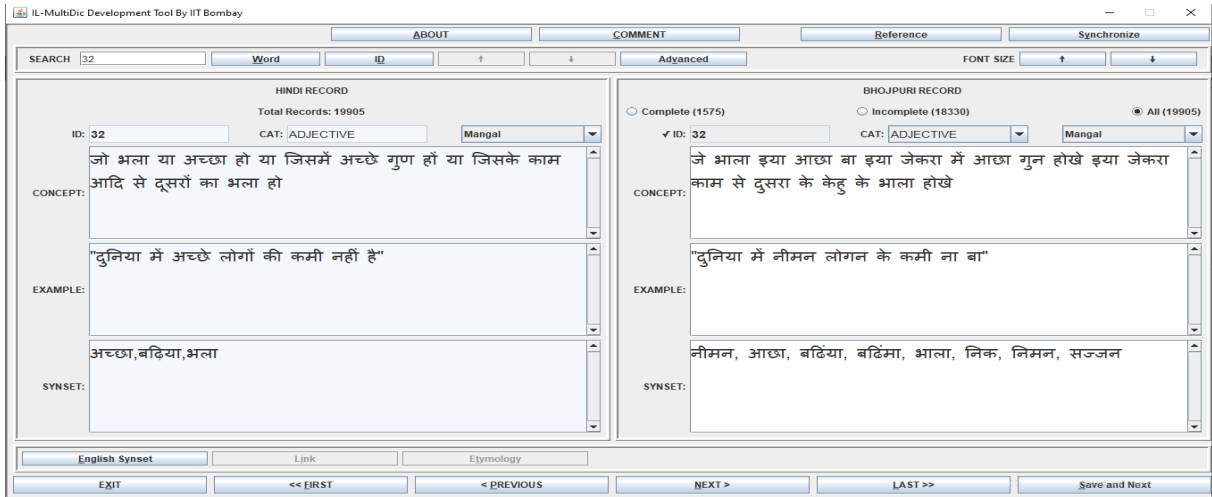


Figure 1: The IL-MultiDic tool showing Hindi and Bhojpuri synsets

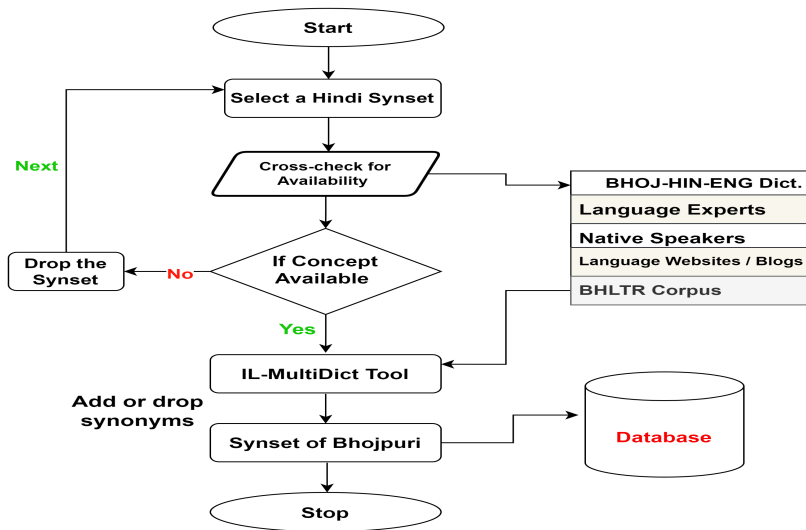


Figure 2: Architecture of Bhojpuri Synset creation

cuss some of the challenges and their solutions in order to fill in the gaps.

4.1 Lexical Anomalies: Equivalent Concept is Not Found

Because of its own cultural practices and distinguishing features, a linguistic community usually differs from its adjacent community. Therefore, there is a potential that a Language A may not have a concept that a Language B does; in this instance, the concept of language A will not have any equivalents in the target language B. In Hindi WordNet, the concept of Hindi does not find any equivalence in Bhojpuri. For example: Consider the Table 1.

Synset ID	Synset	Gloss
19913	mallārī	'A kind of rāgīnī'
19945	saurāṭī	'A kind of rāga'
19958	pum̃gariyā	'An ornament'

Table 1: Non-availability of Bhojpuri equivalents for Hindi synsets

4.2 Lexical Mismatch: False Cognate

Many concepts enlisted in the Hindi synsets look identical to Bhojpuri concepts but they differ in sense denotation. These types of words are called false cognates because learners might be confused by looking at them at the first sight. They might entertain as the equivalent concept. Some of the examples are given in the Table 2.

Syn. Id	Synset	Gloss HIN	Gloss BHOJ
4741	maidāna	‘field’	‘going to toilet’
7171	bādāma	‘peanut’	‘Almond’

Table 2: Semantic mismatches between Hindi-Bhojpuri

4.3 Synthesized Form or Direct Borrowings

Today, the world has become a global village because of the modern developments. As a result, these contemporary concepts have given rise to numerous words. Due to the lack of the sound or sound patterns, the majority of modern words in Bhojpuri have either been directly borrowed or synthesized by the community. Native speakers typically simplify the consonant clusters by the insertion of epenthetic vowel like ‘a’ or ‘i’ between or before clusters. This process breaks the syllable so that it can aid up in pronunciation. For Example, let’s consider the Table 3.

Synset ID	Synset	Gloss	BHOJ Syms
260	svara	Vowel	sovara
7512	voṭa	vote	bhoṭa
2000	pradhāna	Prime	paradhāna

Table 3: Nativized or simplified Bhojpuri equivalents for Hindi Synsets

4.4 Lack of Technical /Scientific Word

Hindi has been the medium of instruction in formal education in Bhojpuri region. So Bhojpuri has not developed technical jargon for scientific and technical concepts. Even though these concepts are there in the language but no word has been coined yet so users continue practicing Hindi terms. Some examples have been listed in the Table 4.

Synset Id	Synset	Gloss
112	ubhayacara	‘Amphibian’
4035	sampreṣaṇa	‘Communication’
338	kaṣerukī	‘Vertebrate’

Table 4: Direct borrowed or transliterated Bhojpuri equivalents for Hindi synsets

To overcome the problem, we follow with some strategies either we should use the transliterated version or go with the direct borrowings of the expression in Bhojpuri. Otherwise we have to coin new equivalent terms in Bhojpuri. However, Bho-

jpuri speakers either go with the explanatory expression or direct borrowings.

4.5 Concept is Available, but with a Reduced Number of Synonyms

The most essential aspect of the Expansion approach is that it allows us to add or drop synonyms based on the available synonyms in the language. We have also noticed that whereas Bhojpuri has fewer synonyms for a notion, Hindi has a greater number of them. For instance, consider the synset ID 2186; Sun, given in Table 5, the concept of the sun has 102 synonyms in Hindi but Bhojpuri hardly enlists a dozen synonymous words for the sun. Likewise, the concept of *śiva* has up to 53 synonyms in Hindi but Bhojpuri enlists only 12 to 15 synonymous words.

Hindi Synsets	BHOJ Synsets
sūrya, sūraja, bhānu, divākara, bhāskara, prabhākara, dinakara, ravi, āditya, dīneśa, āphatāba, aphaṭāba	suruja, sūraja, aragadeva, adita deva, dēva, adita, dinakara, bhāskara, ravi, dīneśa, divākara, aruna
And so on.	

Table 5: Bhojpuri synset with a reduced number of synonyms

4.6 Lexical Gaps

The lexical gap in a language is when the meaning of a word of a particular language does not fit into the meaning of the other language which exhibits a difference in the meaning (Dash, 2017a). Likewise, in certain contexts, Bhojpuri speakers practice more concepts however Hindi enlists less numbers of terms for that kind of concept. The concept of *cāvala* ‘rice’ is used for both cooked and uncooked rice in Hindi, Where as two different words *cāura* for uncooked rice and *bhāta* for cooked rice are used.

5 Bhojpuri Language-specific Synsets

Language specific synsets refers to unique concepts which are available only in the particular language and no conceptual match is find in other languages (Buitelaar and Sacaleanu, 2001). Every language has some concepts or ideas which are unique to only that language. Since Bhojpuri WordNet is being developed by using Hindi Synsets so here, there is potential that many indigenous concepts

specific to Bhojpuri might have not been listed in the Hindi synsets. So there is a need of language specific synsets for Bhojpuri. These language specific words are called *thethee* (deseee). Sometimes, it is better to call it regional specific synsets instead of Language specific synsets (Dash, 2017a).

To create Bhojpuri language-specific synsets, we first collect and assemble a list of LSS for Bhojpuri and provide a complete description of the LSS and examples of its usage in sentences, with a pictorial depiction if possible. We do comparison and validation across languages. We study these LSSs carefully to determine whether they are really monolingual in nature, or originated in the language, and fit the LSS principles (Dash, 2017a). If the concept appears unique to Bhojpuri, we consider it as Bhojpuri Language-specific Synsets (BLSS) otherwise the LSS is dropped. Following confirmations, we approve and augment them in the Bhojpuri LSS database (in blss.acddb). Till now, we have recorded 100 language specific synsets for Bhojpuri. Some of those have been listed in the table 6 and 7.

ID	18
CAT	NOUN
CONCEPT	khānā khilā ke bāda javana kucha baratana meṃ baca ke sukhā jālā
GLOSS	What is left after eating meal and after drying up it hardens
EXAMPLE	“baratana meṃ kharakaṭala jama gila ha ”
SYNSET	kharakaṭala, kharakaṭa

Table 6: Language specific synsets of Bhojpuri -1

ID	6
CAT	ADJ
CONCEPT	u ādamī je jarūrata se jādā bolata hokhe ā bākī oke kuchahu jānakārī na bā basa khālī ṭara ṭara bakavāsa kare lā
GLOSS	The man or boy who talks too much even though he doesn't know anything just talks nonsense
EXAMPLE	“dīpā māṃjhī tejasvī yādava ke labarā kaha dehanī”
SYNSET	labarā, labariyāha, labariyāha

Table 7: Language specific synsets of Bhojpuri -2

6 Conclusion

In this paper, we have tried to delve into the issues and challenges that have occurred during the creation of Bhojpuri synsets. As Bhojpuri is considered closely related to Hindi. So Hindi has been made the source synset for its development. The Bhojpuri WordNet follows the expansion approach and MCR principles of WordNet. What we have experienced during this research that there are several issues like no equivalents found, less derived abstract nouns and adjectives, reduced number of synonyms in comparison to Hindi, and lack of modern scientific technical words. These challenges look for serious involvement at the time of synset creation for Bhojpuri. Synthesized forms, direct borrowings with some sort of simplification, and nativization processes are ways to sort out the complexities. We have presented only some sample cases to explicit the problems and challenges that we faced in the development. Since our work depends on the Hindi synsets and IL-MultiDict offline tool for the Bhojpuri synsets, we found many indigenous concepts or ideas have not been incorporated into the Hindi synsets. This gap requires the creation of BLSS as a part of the Bhojpuri WordNet. We find Bhojpuri is more synthetic than Hindi. However, the Bhojpuri community simplifies the consonant clusters and nativizes some of the borrowed sounds while pronouncing. The Bhojpuri WordNet as a lexical resource could contribute to machine translation, sentiment analysis, word sense disambiguation, and cross-lingual references among Indian languages. The future scope of Bhojpuri WordNet (BWN), a lexical database for the Bhojpuri language, holds immense potential for further development and application. Here are some potential areas where BWN can be expanded and utilized; Education and Language Learning, Visual WordNet, a bilingual Hindi-Bhojpuri dictionary, and in Hindi-Bhojpuri translation applications. These advancements would enable BWN to play a vital role in various applications in the field of natural language processing.

Acknowledgement: We acknowledge NLP team, IIT Bombay for providing the IL-MultiDict tool. We also acknowledge Prof. R.B. Misra and Prof. S. Shahi for their help in the discussion of the creation of Bhojpuri synsets.

References

- S Arulmozi and MC Kesava Murty. 2017. Building telugu wordnet using expansion approach. *The WordNet in Indian Languages*, pages 201–208.
- Brijesh S Bhatt, CK Bhensdadia, Pushpak Bhattacharyya, Dinesh Chauhan, and Kirit Patel. 2017. Gujarati wordnet: a profile of the indowordnet. *The WordNet in Indian Languages*, pages 167–174.
- Pushpak Bhattacharyya. 2010. Indowordnet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
- Pushpak Bhattacharyya, Debasri Chakrabarti, and Vijayanthi M Sarma. 2006. Complex predicates in indian languages and wordnets. *Language Resources and Evaluation*, 40:331–355.
- Paul Buitelaar and Bogdan Sacaleanu. 2001. Ranking and selecting synsets by domain relevance. In *Proceedings of WordNet and Other Lexical Resources: Applications, Extensions and Customizations, NAACL 2001 Workshop*, pages 119–124.
- CFILT. 2023. [Indowordnet statistics](#).
- Niladri Sekhar Dash. 2017a. Defining language-specific synsets in indowordnet: Some theoretical and practical issues. *The WordNet in Indian Languages*, pages 45–63.
- Niladri Sekhar Dash. 2017b. Problems in translating hindi synsets into the bangla wordnet. *The WordNet in Indian Languages*, pages 65–82.
- Niladri Sekhar Dash, Pushpak Bhattacharyya, and Jyoti D Pawar. 2017. *The WordNet in Indian Languages*. Springer.
- Shilpa N Desai, Shantaram W Walawalikar, Ramdas N Karmali, and Jyoti D Pawar. 2017. Insights on the konkani wordnet development process. *The WordNet in Indian Languages*, pages 101–117.
- Aadil Amin Kak, Farooq Ahmad, Nazima Mehdi, Mansoor Farooq, and Muneera Hakim. 2017. Challenges, problems, and issues faced in language-specific synset creation and linkage in the kashmiri wordnet. *The WordNet in Indian Languages*, pages 209–220.
- Malhar Kulkarni, Chaitali Dangarikar, Irawati Kulkarni, Abhishek Nanda, and Pushpak Bhattacharyya. 2010. Introducing sanskrit wordnet. In *Proceedings on the 5th global wordnet conference (GWC 2010), Narosa, Mumbai*, pages 287–294.
- Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*. University of Chicago press.
- Mahesh M. 2017. *Wordnet for tirukkural*. Ph.D. thesis, Annamalai University.
- George A Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.
- George A Miller. 1998. *WordNet: An electronic lexical database*. MIT press.
- Panchanan Mohanty, Ramesh C Malik, and Bhimasena Bhol. 2017. Issues in the creation of synsets in odia wordnet. *The WordNet in Indian Languages*, pages 175–200.
- Gogoi Moromi. 2019. *Design and Development of Assamese WordNet along with Document Classification using wordnet*. Ph.D. thesis, Gauhati University.
- Sivaja S Nair. 2011. *The Knowledge Structure in Amarakosha*. Ph.D. thesis, Department of Sanskrit Studies, University of Hyderabad.
- Dipak Narayan, Debasri Chakrabarti, Prabhakar Pande, and Pushpak Bhattacharyya. 2002. An experience in building the indo wordnet-a wordnet for hindi. In *First international conference on global WordNet, Mysore, India*, volume 24.
- Eugene A Nida. 1979. *A componential analysis of meaning: An introduction to semantic structures*. De Gruyter.
- Kishorjit Nongmeikapam. 2023. [kangleiwordnet](#).
- Atul Kr Ojha. 2019. English-bhojpuri smt system: Insights from the karaka model. *arXiv preprint arXiv:1905.02239*.
- Lata Popale and Pushpak Bhattacharyya. 2017. *Creating Marathi WordNet*, pages 147–166. Springer.
- Rizwanur Rahman, Mazhar Mehdi Hussain, and Niladri Sekhar Dash. 2017. Language-specific synsets and challenges in synset linkage in urdu wordnet. *The WordNet in Indian Languages*, pages 221–229.
- Sankaravelayuthan Rajendran, Selvaraj Arulmozi, B Kumara Shanmugam, S Baskaran, and S Thiagarajan. 2002. Tamil wordnet. In *Proceedings of the first international global WordNet conference. Mysore*, volume 152, pages 271–274.
- Rekha Rattan and Parteek Bhatia. 2011. Creation of punjabi wordnet and punjabi hindi bilingual dictionary.
- Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. Yago: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *The Semantic Web–ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II 15*, pages 177–185. Springer.
- RK Sharma and Parteek Kumar. 2017. *Development of Punjabi WordNet, Bilingual Dictionaries, Lexical Relations Creation, and Its Challenges*, pages 83–99. Springer.

Piek Vossen. 2002. Wordnet, eurowordnet and global wordnet. *Revue française de linguistique appliquée*, 7(1):27–38.