International Conference on Information and Communication Technologies (ICICT 2014)

Role of Semantic Relations in Hindi Word Sense Disambiguation

Satyendr Singh[a],*, Tanveer J. Siddiqui[a]

[a]Department of Electronics & Communication, University of Allahabad, Allahabad 211002, India

Abstract

Semantic relations play an important role in resolving the ambiguity of a polysemous word. This paper investigates the role of hypernym, hyponym, holonym and meronym relations in Hindi Word Sense Disambiguation. In this work, we have considered five different cases: all relations, hypernym and hyponym, hypernym, hyponym and holonym. No semantic relations are used in the baseline. We obtained an overall improvement of 12.09% in precision over the baseline for the case when all relations are considered. The maximum improvement for single semantic relation was obtained using hyponym which resulted in an overall improvement of 9.86% in precision.

1. Introduction

Natural languages contain polysemous words that bear different meanings in different context. For example, the Hindi word 'हल' (hal) as noun has three different meanings (senses) listed in Hindi WordNet [1] as shown in figure 1.

> 1.समाधान, निबटारा, निपटारा, हल, निराकरण, अपाकरण; सोच-समझकर ठीक निर्णय करने या परिणाम निकालने की क्रिया; "मेरी समस्या का समाधान हो गया"
> {samadhan, nibtara, nibtara, hal, nirakaran, upakaran,; soch-samajkar theek nirnay karne ya

* Corresponding author. Tel.: +91-9452-523410
  E-mail address: satyendr@gmail.com

parinam nikalne ki kriya; "meri samasya ka samadhan ho gaya"}
(solution, settlement, redress, thoughtful act of taking right decisions or results " My problem is solved.")
2. हल, लाँगल, लांगल, सीर, कुंतल, कुन्तल, गाकील, नागल, नाँगल; जमीन जोतने का एक उपकरण; "किसान खेत में हल चला रहा है"
{hal, langal, langal, seer, kuntal, kuntal, gaakil, nangal, nangal; jameen jotne ka ek upkaran; "kisaan khet me hal chala raha hai"}
(hal; ploughing instrument; Farmer is ploughing the field.)
3.हल,संवर्तन, संवर्तन; एक अस्त्र; "बलराम का अस्त्र हल था जिसके कारण उन्हें हलधर भी कहते हैं"
{hal, savartan, savartan; ek astra; " balram ka astra hal tha jiske karan unhe haldhar bhi kahte hai"}
(hal,; a weapon; "The weapon of balram was hal because of which he is also called as haldhar")

Fig. 1. Senses of 'हल' (hal) in Hindi WordNet

In a particular context only one of these senses applies. For example, the most appropriate sense of 'हल' (hal) in the context "वो हमेशा समस्या का तुरंत हल चाहता है" {wo hamesha samasya ka turant hal chahta hai} (He always wants quick solution to a problem) is sense 1 (solution sense). Human beings can easily arrive at the correct meaning of the word using the surrounding context, usually on the basis of short context with five words generally proving sufficient [2] . However, computational identification of the correct sense of the word is a difficult task. The task of automatic identification of the correct sense of a word in a given context is known as Word Sense Disambiguation (WSD). It is one of the most difficult problems in Natural Language Processing (NLP) and is often described as "Artificial Intelligence (AI) - complete". A lot of published literature exists on English WSD [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. These works make use of context, dictionary definitions, syntactic, semantic and domain information for disambiguation. However, there are only few reported works on Hindi WSD [12, 13, 14].

One of the pioneer works on automatic WSD using dictionary definition was done by Lesk [8]. He used direct overlap between the context and the glosses in dictionary to predict the correct sense of a word. Following Lesk's work a number of overlap-based algorithms were proposed [4, 5]. Instead of using direct overlap Resnik [10] introduced a measure of relatedness based on information content. He used the information content of concepts along with their positions in the noun is-a hierarchy of the WordNet [15] to disambiguate a word. In [3] a measure of conceptual distance among the concepts is used for disambiguation. Navigli and Velardi [9] proposed a method called Structural Semantic Interconnections (SSI). Structural specifications of the possible senses of each word in the context were created and a grammar describing relations between sense specifications was used to select the most appropriate sense. The method utilized WordNet and a small amount of tagged corpora to draw syntactic generalization for disambiguation.

Semantic and lexical relations have also been utilized for WSD [4, 5, 6, 7, 11]. Banerjee and Pederson [4] adapted Lesk algorithm for WSD using WordNet. They assigned a score to all possible tag sequence in a small window containing target word in the middle. The glosses of the synset, hypernym, hyponym, holonym, meronym, troponym and attribute of each word were used for computing overlap. The sense of target word in candidate combination having the highest score was assigned as the winner sense. In [5], Banerjee and Pederson proposed a new measure of semantic relatedness between concepts based on the number of overlaps in glosses. The extension was achieved by including the glosses of   other   concepts to which a concept is related in the WordNet concept hierarchy. Vasilescu et al. [11] compared variants of Lesk's algorithm in which simplified Lesk algorithm was found significantly better than the original Lesk's algorithm. In [6], both the sense definitions and the context vector were extended by including the hypernyms of nouns and verbs appearing in the sense definition and the context respectively. Instead of using direct overlap weighted overlap was used. The weights were assigned to words on the basis of the depth of the associated synset in the WordNet taxonomy. Leacock et al. [7] exploited lexical relations to locate training examples in a general text corpus. They used statistical classifier that used topical context and the local clues to perform WSD. However, these results cannot be generalized for Hindi or other Indian languages without experimental evidences.

Earlier works on Hindi language includes [12, 13, 14]. Sinha et al. [14] utilized Hindi WordNet to construct sense definition context. They used overlap between context of a polysemous word and extended sense definitions to disambiguate a polysemous Hindi noun. They extracted words from synonyms, glosses, example sentences, hypernyms, glosses of hypernyms, example sentences of hypernyms, hyponyms, glosses of hypernyms, example sentences of hypernyms, meronyms, glosses of meronyms and example sentences of meronyms for creating sense

definition. The context of polysemous word were taken as a list of words in the context of target polysemous word. The sense with maximum overlap was assigned as the winner sense. Singh and Siddiqui [13] evaluated the effects of context window size, stemming and stop word removal on Hindi WSD. Khapra et al. [12] studied domain specific WSD for nouns, adjectives and adverbs in a trilingual setting of English, Hindi and Marathi. They used dominant senses of words in specific domains for performing disambiguation. First monosemous words were identified and then bi, tri and polysemous words were disambiguated iteratively. Information in WordNet Graph structure and corpus biases for senses was used to arrive at sense decisions. Evaluation was performed on Tourism and Health domains. They obtained an accuracy level of 65% on F1-Score for all the 3 languages. None of these work report on the role of semantic relations on Hindi WSD. The proposed work aims at studying the role of individual semantic relations and their combinations for Hindi WSD.

This paper presents an adaptation of Lesk's dictionary-based algorithm for Hindi WSD. Instead of using dictionary definitions, Hindi WordNet is employed. Hindi WordNet offers a rich hierarchy of semantic relations which our algorithm uses to enrich the glosses. This improves chances of match between contextual terms and extended sense definitions (glosses). We experiment with five different cases: all relations, hypernym and hyponym, hypernym, hyponym and holonym. The contributions of these relations, individually or in combination, on Hindi WSD are not studied earlier. This paper aims to study the role of semantic relations individually and their combinations in a Lesk-based setting for Hindi WSD task. The results show that the right combination of relations can lead to significant performance improvement in overlap based WSD.

The rest of the paper is structured as follows: In section 2, relations in Hindi WordNet and WSD algorithm used in this work are discussed. Section 3 provides details of the data set and experiments conducted. Results are discussed in section 4. Finally, we conclude in section 5.

2. Our Approach

We use the direct overlap between the sense definitions and the surrounding context of a polysemous word to disambiguate it. In order to increase the chances of overlap, the sense definition is extended by adding example sentences and synonyms to its gloss. To study the role of various semantic relations, we additionally utilize glosses and example sentences of these relations. The underlying assumption is that the words related to a sense will be better indicator of that sense and will improve the chances of match with the context. The relations and their combinations studied in this paper are shown in Table 1.

Table 1. Combinations of semantic relations used in the experiment.

Combination	Semantic relations used
All relations	Synonym
	Hypernym
	Hyponym
	Holonym
	Meronym
Holonym	Synonym
	Holonym
Hypernym and Hyponym	Synonym
	Hypernym
	Hyponym
Hypernym	Synonym
	Hypernym
Hyponym	Synonym
	Hyponym

Baseline   Synonym

2.1. Hindi WordNet and Semantic Relations

Hindi WordNet is among the most popular lexical resource used in WSD research for Hindi language. It organizes words in terms of set of synonym sets, called synset. Synonym sets are the words bearing similar meanings. The total number of unique words in Hindi WordNet is 98651 and total numbers of synsets is 38687. It captures various semantic and lexical relations between synsets and words such as is-a, kind-of and part-whole relations. These relations can be utilized in WSD to provide useful clues for disambiguating a polysemous word. The relations in Hindi WordNet includes synonymy, hyponymy, hypernymy, meronymy, holonymy, entailment, troponymy, antonymy, gradation, causative, ability link, capability link, function link, attribute, modifies noun, modifies verb and derived from. We study the effects of hypernym, hyponym, holonym and meronym.

Hypernym and hyponym capture is-a or kind-of relation between synsets. Figure 2 shows an example of hypernym and hyponym.

The synsets {घर, गृह, मकान} has {सदन, शाला, आलय, धाम, निकेतन, आश्रय} as hypernym. These synsets are involved in superset relation with {घर, गृह, मकान} synsets. The synsets {घर, गृह, मकान} has {सराय, द्यूतगृह, जुआघर, आश्रम, कुटिया, गुरुकुल, अतिथि-गृह, शवदाहगृह, हवाईघर, हवेली, कोठी, कुटिया, झोंपड़ी, बंगला, घरौंदा, मढ़ी, आशियाना, खण्डहर, खोली, चुंगीघर} as hyponym, which is subset of {घर, गृह, मकान}.

Fig. 2. Example of hypernym and hyponym

Meronym and holonym capture part - whole relation between synsets. If two synsets A and B are related in the manner such that A is a part of B or constituent of B, then A is meronym of B, and B is holonym of A. Holonym relation is reverse of meronym. Example of meronym and holonym is shown in figure 3.

{जड़, मूल, सोर} is a part of {पेड़, वृक्ष} and hence {जड़, मूल, सोर} is meronym of {पेड़, वृक्ष} and {पेड़, वृक्ष} is holonym of {जड़, मूल, सोर}.

Fig. 3. Example of meronym and holonym

2.2. WSD Algorithm

A Lesk-based algorithm has been used in this work for evaluating the effects of semantic relations on Hindi WSD task. The steps in the WSD algorithm are given in figure 4. The context vector of the target word is created by extracting words appearing in $\pm$ n window size, keeping the target word in the middle. For a window size of n, the size of context vector is 2n+1. The sense definition vector is created by extracting words occurring in sense definitions (including synsets, glosses, example sentences and semantic relations, their glosses and example sentences) of the target polysemous word. The stop words are dropped from both sense definitions as well as test instances before constructing sense definition vector and context vector. The context vector is matched with the sense definitions vector and a numerical value is assigned using frequency of the matching words. Target word is not counted in matching. The sense maximizing the score is assigned as winner sense.

1.w ← target polysemous word to be disambiguated
2. N ← number of senses of w
3. Remove stop words from the sense definitions and from the test instances of w
4. Create context vector (C) by extracting all the words to the left and right of w in the proximity of n words
5. for i = 1 to N do

```
        Create sense definition vector for sense i (Sᵢ) of the target word w
        Scoreᵢ← similarity (C, Sᵢ )
6. Return Sᵢ for which score is maximum.

Computing score:
Similarity (C, S) // C is context vector and S is sense definition vector
sense_score← 0
    for each word x in C
        wcount← frequency of x in S
        sense_score = sense_score + wcount

    return sense_score
```

Fig. 4. WSD Algorithm

3. Data Set and Experiments

3.1. Data Set

For evaluation, we manually created a sense annotated dataset consisting of 60 polysemous Hindi nouns, for Hindi lexical sample task. The sense inventory was derived from Hindi WordNet. Hindi WordNet contains very fine grained sense listing. Some of the senses having very fine grained sense distinctions were merged on a subject based evaluation as subjects were not able to discriminate among those senses manually. The subjects were native speakers of Hindi language and included three students (two doctoral and one post graduate) of University of Allahabad, Allahabad, India. For example, the sense listing of Hindi noun 'ग्राम' (gram) in Hindi WordNet is shown in figure 5. When we presented instances of sense 1 and sense 3 to three subjects, all of them marked sense 1 and sense 3 as similar. Hence we merged sense 1 and sense 3 for 'ग्राम' (gram). Some of the senses not commonly used have been dropped because instances involving those senses were rare and could not be found.

1. गाँव, ग्राम, गाम, देहात, दिहात, अवसथ, आवसथ; बहुत छोटी बस्ती; "भारत की अधिकांश आबादी गाँवों में निवास करती है"

2. ग्राम; वजन मापने का एक मापक; "उसने आठ सौ ग्राम आटा खरीदा"

3. गाँव, ग्राम, गाम; किसी गाँव में रहनेवाले लोग; "शोर सुनते ही पूरा गाँव इकट्ठा हो गया"

Fig. 5. Senses of Hindi noun 'ग्राम' (gram)

The lexical sample approach is used for creating sense annotated dataset. We first identified 60 polysemous Hindi nouns. Their instances were collected from Hindi corpus [16], a raw corpus available at Centre for Indian Language Technology (CFILT), Indian Institute of Technology (IIT) Bombay. Instances were also collected by performing search using sense definitions as query from www.khoj.com and www.google.co.in. The retrieved documents were used to create context. The context (instances) covered domains like news, medical, literature, sports, stories, medicine etc. Our dataset has a total of 7506 instances. The average number of instances per word is 125.1, the average number of instances per sense is 49.70 and the average number of senses per word is 2.51. The translation, transliteration and statistics of dataset are given in Figure A1 in Appendix A.

For performance evaluation precision and recall is computed. Precision is defined as the ratio of the correctly answered instances and the total number of instances answered for the target word by the system. Recall is defined as the ratio of the correctly answered instances and the total number of instances to be answered for the target word by the system.

3.2. Experiment

Table 1 shows the five combinations of semantic relations considered in this work. Meronym is not taken alone as it is not defined for 45 words in our dataset.

For evaluating the proposed algorithm, we conducted test runs by varying context window size from 5 to 25 in steps of 5. In the baseline the synonyms of the target word along with its glosses and example sentences is taken.

Average precision and recall over the context window size is computed for all the words. The mean average precision for 60 words and percentage improvement in precision over the baseline is shown in Table 2. Table 3 shows the mean average recall over 60 words and percentage improvement in recall over the baseline.

Table 2. Mean Average Precision (over 60 words) and percentage improvement in Precision over the baseline

Combination	Mean Average Precision (over 60 words)	Percentage Improvement
All relations	0.56767	12.09%
Holonym	0.51699	2.08%
Hypernym and Hyponym	0.56393	11.35%
Hypernym	0.51910	2.50%
Hyponym	0.55637	9.86%
Baseline	0.50641	

Table 3. Mean Average Recall (over 60 words) and percentage improvement in Recall over the baseline.

Combination	Mean Average Recall (over 60 words)	Percentage Improvement
All relations	0.51325	11.33%
Holonym	0.46868	1.67%
Hypernym and Hyponym	0.51130	10.91%
Hypernym	0.47293	2.59%
Hyponym	0.50425	9.38%
Baseline	0.46098	

4. Results and Discussion

As shown in Tables 2 and 3, the mean average precision for combinations of all relations, hypernym + hyponym, holonym, hypernym and hyponym are 0.56767, 0.56393, 0.51699, 0.51910 and 0.55637. The mean average precision for the baseline is 0.50641. The mean average recall for all relations, hypernym + hyponym, holonym, hypernym and hyponym are 0.51325, 0.51130, 0.46868, 0.47293 and 0.50425. The mean average recall for the baseline is 0.46098.

We obtained 12.09% improvement in the precision for the combination of all relations over the baseline. Semantic relations provides rich hierarchy of is-a, kind-of and part-whole relations between synsets resulting in better content words in sense definitions. This improves disambiguation accuracy.

The combination of hypernym and hyponym results in an improvement of 11.35% in precision which is quite close to all relations case. These relations provide semantic synsets in the superset and subset of the target word's synset accounting for major contribution of disambiguation accuracy.

We obtained a minor improvement of 2.08% and 2.50% in precision using holonym and hypernym alone respectively. Hypernym is equivalent to going up in the hierarchy to root level resulting in inclusion of more general words in sense definitions, their glosses and example sentences. The general words are not much useful for discriminating among senses.

The maximum increase of 9.86% in precision by a single relation was observed using hyponym. Hyponym results in inclusion of subset synsets of the target word. This is equivalent to moving down in the synset hierarchy

leading to more specific words in sense definitions. This helps in discriminating among various senses of target polysemous word. This accounts for the maximum improvement observed in precision with the use of hyponym.

5. Conclusions and Future Work

In this paper, we evaluated the effects of four semantic relations and their combinations on Hindi WSD. We obtained maximum improvement in precision over the baseline using all semantic relations. The major contributor in this improvement is hyponym. Hyponym provides subset synsets of the synsets of the target word. Hence more specific content words are included in the sense definitions accounting for increase in the disambiguation accuracy. The dataset used is a lexical-sample of 60 nouns. In future, we would like to confirm our findings on larger dataset and for all-word disambiguation task. Lack of gold standards, like SENSEVAL, in Hindi prevents us from generalizing our results. Creation of such standards will help in expediting research in Hindi WSD.

There are a number of directions in which this work may be extended in future:

- We have used direct overlap in order to measure the similarity and extended sense definitions. A number of measures of relatedness and similarity have been proposed for English WSD. These measures can be explored for Hindi WSD.
- Like sense definitions, the context representation can be enriched by involving related words.
- We have used only direct relations; another possible extension of this work is to use indirect relations, e.g., hyponyms of hyponyms.

References

1. Hindi WordNet http://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php
2. Choueka, Y. and Lusignan, S. Disambiguation by Short Contexts. *Computers and the Humanities*, 1985, 19:3, p. 147 – 157.
3. Agirre, E. and Rigau, G. Word sense disambiguation using conceptual density. In: *Proceedings of the 16th Conference on Computational Linguistics (COLING'96)*, Copenhagen, Denmark, 1996; p. 16-22.
4. Banerjee, S. and Pederson, T. An Adapted Lesk Algorithm for Word Sense Disambiguation using WordNet. In: *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '02)*, 2002; p. 136-145.
5. Banerjee, S. and Pederson, T. Extended gloss overlaps as a measure of semantic relatedness. In: *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico, 2003; p. 805-810.
6. Fragos, K., Maistros Y., and Skourlas C. Word Sense Disambiguation using WORDNET relations. In: *Proceedings of the First Balkan Conference in Informatics*, Thessaloniki, 2003
7. Leacock, C., Chodorow, M. and Miller, G. A. Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, 1998, 24:1, p. 147-165.
8. Lesk, M. Automatic Sense Disambiguation using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In: *Proceedings of the 5th annual International Conference on Systems documentation (SIGDOC'86)*, Toronto, Ontario, 1986; p. 24 – 26.
9. Navigli, R. and Velardi, P. Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27:7, p. 1075-1086.
10. Resnik, P. Using information content to evaluate semantic similarity. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCIA)*, Montreal, Canada, 1995; p. 448 – 453.
11. Vasilescu, F., Langlasi, P. and Lapalme G. Evaluating Variants of the Lesk Approach for Disambiguating Words. In: *Proceedings of the Language Resources and Evaluation {LREC}*, 2004; p. 633 – 636.
12. Khapra, M., Bhattacharyya, P., Chauhan, S., Nair, S. and Sharma A. Domain Specific Iterative Word Sense Disambiguation in a Multilingual Setting. In: *Proceedings of International Conference on NLP (ICON 08)*, Pune India, 2008.
13. Singh, S. and Siddiqui, T. J. Evaluating Effect of Context Window Size, Stemming and Stop Word Removal on Hindi Word Sense Disambiguation. In: *Proceedings of the International Conference on Information Retrieval and Knowledge Management (CAMP'12)*, Malaysia, 2012; p. 1-5.
14. Sinha, M., Kumar, M., Pande, P., Kashyap, L. and Bhattacharyya, P. Hindi Word Sense Disambiguation. In: *International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems*, Delhi, India, 2004.
15. WordNet http://wordnet.princeton.edu/
16. Hindi Corpus http://www.cfilt.iitb.ac.in/Downloads.html

Appendix A.

Word	Sense Number : Translation of senses in English (Number of Instances)	Word	Sense Number : Translation of senses in English (Number of Instances)
अंग (ang)	Sense1 : Any part or organ of human body (88) Sense2 : component (30) Sense3 : Part of a community, organization or unit (105)	ताव (tao)	Sense1 :torrid (18) Sense2 :ream of paper (8)
अंश (ansh)	Sense1 :Numerator in math in Hindi (42) Sense2 :component (36) Sense3 :Degree, measurement of angle (53)	तिल (til)	Sense1 :Sesame, a plant from which oil is extracted from its seeds (41) Sense2 :mole (263)
अचल (achal)	Sense1 :immovable (12) Sense2 :Person's name (34) Sense3:Immovable property (27)	तीर (teer)	Sense1 :arrow (103) Sense2 :shore of river or sea (39)
अशोक (ashok)	Sense1 :Name of a tree in India (33) Sense2 :Name of an Indian King (21)	तुलसी (tulsi)	Sense1 :Basil, A plant which is considered holy and medicinal (193) Sense2 :A saint who was follower of God Ram and who wrote Ramayana (81)
उत्तर (uttar)	Sense1 :Answer (30) Sense2 :North direction (79) Sense3 :A person's name (36)	तेल (tel)	Sense1 :oil (128) Sense2 :crude oil obtained from mines (53) Sense3 :A ceremony performed in Indian marriages (14)
कदम (kadam)	Sense1 :Initiative (16) Sense2 :foot (13) Sense3 :step (11)	थान (thhaan)	Sense1 :roll of cloth, bolt (21) Sense2 :A place where domestic animals are tied (9) Sense3 :Place of Indian God or Goddess (8)
कमान (kaman)	Sense1 :Bow , curved piece of resilient wood with taut cord to propel arrows (28) Sense2 :Command (35) Sense3 :An special army (e.g., Navy) (33)	दक्ष (daksh)	Sense1 :A king in Indian mythology who was father of sati and father in law of Lord Shiva (64) Sense2 :Qualified, efficient, skilled (15)
कलम (kalam)	Sense1 :Pen, quill (67) Sense2 :cutting of a tree (69) Sense3 :Style of painting of a particular place (66) Sense4 :Place near ear and cheeks , where there are hairs (26)	दर (dar)	Sense1 :Standard cost, rate (147) Sense2 :door (67)
कांड (kaand)	Sense1 :Part of religious literature (43) Sense2 :Negative event or happening (29)	दाद (daad)	Sense1 :To praise someone, accolade (27) Sense2 :skin disease, ringworm (51)
कुंभ (kumbh)	Sense1 :Water pot made of mud (65) Sense2 :A Sun Sign (Aquarius) in Hindi (58) Sense3 :A Holy event happing every 12 years in India (64)	दाम (daam)	Sense1 :Cost, price (61) Sense2 :Type of strategy or policy (20)
कोटा (kota)	Sense1 :Reservation, quota (70) Sense2 :Name of a district in Rajasthan in India (64)	धन (dhan)	Sense1 :Money , Wealth (126) Sense2 :Sign of addition in mathematics in Hindi,+ (16)
क्रिया (kriya)	Sense1 :Verb in Hindi grammar (116) Sense2 :Activity, action (71)	धारा (dhaaraa)	Sense1 :Law charges for crime in Indian constitution, section (44) Sense2 :River's flow, stream (67) Sense3 :flow of speech, thought or events (50) Sense4 :Electric Current (67)
क्वार्टर (quarter)	Sense1 :A place allotted to live for temporary period (26) Sense2 :A quantity of wine (14) Sense3 :A match , in which after winning, a player or team reaches semifinal (12)	धुन (dhun)	Sense1 :Music tune (84) Sense2 :cult, flakiness, mania (10)
खान (khan)	Sense1 :mine (60) Sense2 :Vast storage of subject knowledge or quality (13) Sense3 :Surname of a Muslim community in India (65)	फल (phal)	Sense1 :fruit (90) Sense2 :result (79) Sense3 :Front sharp part of arrow or spear (11)
गल्ला (galla)	Sense1 :Food grains (wheat, corn, cereal) (41) Sense2 :Penny bank, piggy bank (29)	बाल (baal)	Sense1 :hair (111) Sense2 :child (47)

Word	Senses	Word	Senses
गुना (guna)	Sense1 :number of times (22) Sense2 :Name of a district in Madhya Pradesh in India (21)	मत (mat)	Sense1 :Religious community (41) Sense2 :Opinion, thought, idea (31) Sense3 :vote (92)
गुरु (guru)	Sense 1:teacher (89) Sense 2:Jupiter (name of a planet) (60)	माँग (maang)	Sense1 :Requirement, need, (13) Sense2 :Parting of hairs on head where married Hindu woman put vermilion as a sign of marriage (33)
ग्राम (gram)	Sense1 :village (169) Sense2 :A unit of measurement, gram (77)	मात्रा (matra)	Sense1 :Quantity , amount , volume (41) Sense2 :some time period in music (8) Sense3 :Vowel sound in Hindi speech (39)
घटना (ghatna)	Sense1 :event (65) Sense2 :Lowering of water level, subside (14)	मूल (mool)	Sense1 :root of plant (6) Sense2 :Basic reason, fundamental (49) Sense3 :Time for a type of star (97) Sense4 :Capital/Principal money (40)
चंदा (chanda)	Sense1 :moon (82) Sense2 :Financial contribution, subscription (75)	लाल (lal)	Sense1 :red color (129) Sense2 :Son, child (26)
चरण (charan)	Sense1 :Stage, phase (72) Sense2 :foot (49) Sense3 :Quarter part of anthology (78)	वचन (vachan)	Sense1 :whatever one speaks or says, saying (23) Sense2 :Promise, commitment (27) Sense3 :Agent in Hindi grammar to denote singular or plural (23)
चारा (chara)	Sense1 :Domestic animal's food, provender, forage (100) Sense2 :option (21)	वर्ग (varg)	Sense1 :Community, category, class (90) Sense2 :square object (15) Sense3 :square of number, unit of measurement of area(e.g., square feet) (129)
चाल (chaal)	Sense1 :speed (13) Sense2 :move to be taken in chess or similar games (97) Sense3 :A place where people stay, tenement house (11) Sense4 :behavior (37) Sense5 :Strategy in game, trick (26)	विधि (vidhi)	Sense1 :Way or process of doing something (72) Sense2 :law (69)
जीना (jeena)	Sense1 :To live, survive (39) Sense2 :staircase (33)	शेर (sher)	Sense1 :Tiger, lion (166) Sense2 :type of Urdu poetry (41)
जेठ (jeth)	Sense1 :Name of a month in Hindi (10) Sense2 :Husband's elder brother, brother in law (20)	संक्रमण (sankraman)	Sense1 :Process of sun's transition from one star-sign to another (28) Sense2 :Process of disease infection (60) Sense3 :Process of transition from one place or state to another place or state (22)
टीका (tika)	Sense1 :A sign on forehead using sandalwood (15) Sense2 :vaccination (22) Sense3 :To write about something in detail (24) Sense4 :A ceremony to confirm marriage in India, Engagement Ceremony (10) Sense5 :A jewelry which is worn by Indian bride on forehead (24)	संबंध (sambandh)	Sense1 :relation (23) Sense2 :Agent in Hindi grammar that shows relation between two words (33) Sense3 :marriage (8)
डब्बा (dabba)	Sense1 :Box , made up of plastic, wood or metal, bin (21) Sense2 :Coach of train which carries passengers (24)	सीमा (seema)	Sense1 :Limit, threshold (28) Sense2 :boundary, border (23)
डाक (dak)	Sense1 :Bid, bidding (60) Sense2 :post, postal system (59)	सोना (sona)	Sense1 :gold (65) Sense2 :sleep (24)
ढाल (dhal)	Sense1 :Sloping or sliding land (31) Sense2 :A protective covering used for saving attack of sword, armor (28)	हल (hal)	Sense1 :solution (26) Sense2 :ploughing instrument, plough (76)
तान (taan)	Sense1 :Process of stretching (14) Sense2 :Music tone (19)	हार (haar)	Sense1 :defeat (33) Sense2 :necklace, garland (63)

Fig. A1. Translation, Transliteration and Statistics of Dataset