

Marathi WordNet Development

Naik Ramesh Ram¹, C. Namrata Mahender²

¹Dept. Of .Computer Science and Information Technology
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India,
ramesh.naik31@yahoo.com

²Dept. Of .Computer Science and Information Technology
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India,
nam.mah@gmail.com

Abstract: WordNet is a dictionary of word meanings/concepts. Hence there must be a standard representation of the concepts in order to simulate a lexical matrix on a machine. Marathi is an Indo-Aryan language spoken by about 71 million people mainly in the Indian state of Maharashtra and neighboring states. In WordNet, which is basically a semantic network, the different lexical categories of words (nouns, verbs,..) are organized into 'Synsets' (sets of synonyms). Each synset represents a lexical concept and they can be linked by different types of relation (Hypernymy, antonym, etc.). The WordNet for Marathi 36842 unique words grouped in more than 26988 Synsets, linked synset 24398 [1]. We would like to increase unique words, with that we want to give special efforts on words which are oriented towards Marathi culture only for example culture aspects of words like that Lawani, Abhang, Fugaddi, Ringan etc.

Keywords: WordNet, Synset, Lexical, Marathi WordNet.

1. Introduction

WordNet is a machine readable lexical database organized by semantics. WordNet is a dictionary plus a thesaurus and much more. Building a functional WordNet for a language is no easy task. It requires lexicographers and computer scientists to work jointly to create such an online lexical resource. This gives rise to a need to develop a wide range of software tools to create, maintain and make potential use of the WordNet [2].

The Marathi WordNet is a system for bringing together different lexical and semantic relations between the Marathi words. It organizes the lexical information in terms of word meanings and can be termed as a lexicon based on psycholinguistic principles. The design of the Marathi WordNet is inspired by the famous English WordNet. In the Marathi WordNet the words are grouped together according to their similarity of meanings. Two words that can be interchanged in a context are synonymous in that context. For each word there is a synonym set, or synset, in the Marathi WordNet, representing one lexical concept. This is done to remove ambiguity in cases where a single word has multiple meanings. Synsets are the basic building blocks of WordNet.

2. Synset

It is assumed that, the person reading the synset already has the knowledge of the concept and the synset is merely a representation of the concept. For example, consider the word

chair. This word could either mean a seat for the person or a person who presides at a meeting. To distinguish the two senses of the word chair, the word forms which also represent the same word meaning can be used along with the word chair. Hence the representation {chair} and {chair, chairman, chairperson} helps us to distinguish between the two different sense of chair. However, sometimes if the appropriate synonym is not available, the polysemy can be resolved with the help of a small gloss (a definition which explains the concept) and an example. This combination of the synonymous words and the gloss is called as a Synset. Often the word syntactic category or the part-of-speech information is also included in the synset.

3. Parts of speech

3.1 Noun (नाम):-

Learning the Marathi Nouns is very important because its structure is used in every day conversation. The more you master it the more you get closer to mastering the Marathi language. But first we need to know what the role of Nouns is in the structure of the grammar in Marathi. Marathi nouns are words used to name a person, animal, place, thing, or abstract ideas. Nouns are usually the most important part of vocabulary.

3.2 Adjective (विशेषण):-

Learning the Adjectives is very important because its structure is used in every day conversation. The more you master it the

more you get closer to mastering the Marathi language. But first we need to know what the role of Adjectives is in the structure of the grammar in Marathi. Marathi Adjectives are words that describe or modify another person or thing in the sentence.

Grammar Tips:

While in English an adjective doesn't change when the noun changes, in Marathi an adjective should agree in gender and plural with the noun. For example:

a) Masculine to feminine example:

Ha majha chhota mulaga aahe (this is my little son) becomes: He majhi chhoti mulagi aahe (this is my little daughter)

As you can see from the example above, the adjective comes before the noun and also takes the feminine form.

b) Singular to plural example:

He pandhari manjar majhi aahe (this is my white cat) becomes: Hya pandharya manjari majhya aahet (these are my white cats). As you can see from the example above, the adjective comes after the noun and also takes the plural form.

3.3 Verbs (क्रियापद):-

Learning the Verbs is very important because its structure is used in every day conversation. The more you master it the more you get closer to mastering the Marathi language. But first we need to know what the role of Verbs is in the structure of the grammar in Marathi. Verbs are words that convey action (bring, read, walk, run), or a state of being (exist, stand). In most languages a verb may agree with the person, gender, and/or number of some of its arguments, such as its subject, or object.

Grammar Tips:-

- Present Tense

In Marathi, verbs take the following endings to form the present tense:

Aahe, To, Te

I am playing cricket: Me cricket khelate aahe. (Feminine)

Me cricket khelato aahe (masculine)

- Past Tense

In Marathi as well as in English the simple past tense (imperfect) is used to describe past events. The endings for the past tense verbs are: Hoto, hote, O, le, li. Example (feminine/masculine): gele/gelo/geli, khelat hoto/hote, gadi sutali

I played cricket: Me cricket khelale (feminine)

Me cricket khelalo (masculine)

I was playing cricket: Me cricket khelat hoto/hote.

- Future Tense

To form the future in Marathi it's very easy, just use the whole infinitive verb plus these following ending: een, eel Example: jaen, karen, yesheel

I will play tomorrow: Me udyaa kheleen.

3.4 Tables

Learning the Adverbs is very important because its structure is used in every day conversation. The more you master it the more you get closer to mastering the Marathi language. But first we need to know what the role of Adverbs is in the structure of the grammar in Marathi.

Marathi adverbs are part of speech. Generally they're words that modify any part of language other than a noun. Adverbs can modify verbs, adjectives (including numbers), clauses, sentences and other adverbs.

Grammar Tips:

While in English adverbs are usually formed by adding (-ly) to adjectives. In Marathi adverbs can be formed different ways. Some of the are formed by repeating the word or by adding (-ch).

halu (slow) becomes haluch/haluhalu (slowly)

Yogya (perfect) becomes yogya (perfectly)

However that's not always the case. Some words are adverbs by nature. For example:

aatta (now), khara/kharach (really), and lavakar (soon) are all Marathi adverbs.

Here the words present in the system

Table 1: Lexical Categories of words [3]

Noun नाम	Adjective विशेषण	Verbs क्रियापद	Adverbs क्रियाविशेषण
सहमति: अनुज्ञा,समंती, सवलत,मोकळी क	खालावेल:मीकरणे, कमीहोणे	फूस : लाचदेणे	इंतभूत: बाकीसर्व, इत्यादी, अनिसर्व.
पावटा:माप, प्रमाण	जीवघेणी: त्रासदायक,भयंकर, जबर, खूप, अतिशय, प्रचंड,भयावह	वेटोळे : गुंडाळणे, दाबूननिटबांध णे, परिभ्रमणकरणे	
ऋणानुबंध : बंधुत्व,बंधुप्रेम	शेंडेफळ: अपकार, अपाय, खोडकर, टवाळी, खोडी, बदमाश.	आभासी: माहितीनसले ला, अनभिज्ञ	

3.5 Semantic Relations

A semantic relation is a relation between word meanings. As meanings are represented through a synset, these relations are pointers between Synsets. If two synset S1 and S2 are related through a relation R, then the individual word forms belonging to these Synsets are also related by the relation R [2]. There are various relations between the synset in a WordNet. Some of the important relations are Synonymy, Hyponymy/Hypernymy, Meronymy/Holonymy.

Synonymy

Synonymy is the most important relation between two word forms. Two word forms are synonymous if the substitution of one for the other does not alter the meaning of a sentence in which the substitution is made [2]. This forms the basis for the creation of synset in the WordNet. The replace ability of synonymous words also makes it necessary to partition the WordNet into various syntactic categories like nouns, verbs, adjectives and adverbs. This is because only words in different categories cannot be synonyms and hence are not interchangeable. Synonymy is a symmetric relation. If W1 is a synonym of W2 then W2 is also a synonym of W1.

Hyponymy and Hypernymy (is a kind of): Hypernymy is a semantic relation between two Synsets to capture super-set

hood. Similarly, hyponymy is a semantic relation between two Synsets to capture sub-set hood. The hyponymy relation is transitive and asymmetrical. Hypernymy is the reverse of hyponymy.

Meronymy and Holonymy (Part-whole relation): It is a semantic relation between two synset. If the concepts A and B are related in such a manner that A is one of the constituent of B, then A is the Meronymy of B and B is the Holonymy of A. The Meronymy relation is transitive and asymmetrical. Holonymy is the reverse of Meronymy. It is used to construct a part-of hierarchy.

Ontology:- Ontology is a hierarchical organization of concepts. All the concepts are classified into categories, sub-categories, and then sub-sub-categories and so on. The highest level categories correspond to syntactic categories. All these categories with links to other categories are stored in a file in a pre-defined format.

4. Wordnet database

WordNet database: The database created consist more than 10,000 Marathi words with its noun,adjective,verb,adverb,semantic relations like synonymy,Hypernymy,Holonymy ontology in the form of multiple tables which are in relationship. The required data type to store Marathi font is nvarchar is set in the table definition.

5. CONCLUSION

The Marathi WordNet deals with the content words, or open class category of words. Thus, the Marathi WordNet contains the following category of words- Noun, Verb, Adjective and Adverb.

Our work mainly focuses on developing the WordNet for Marathi language. WordNet consists of a database. The WordNet database consists the synonym set (synset) of a word. Synsets are interconnected with other Synsets via a number of lexical and semantic relations.

References

- [1] "Hindi and marthi wordnet," <http://www.cfilt.iitb.ac.in/>.
- [2] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, "Five papers on WordNet" in WordNet: An Electronic Lexical Database, Ed. MIT Press, May 1998.
- [3] Naik Ramesh R, C.Namrata Mahender, "Development of Wordnet for Marathi Language", (NCAC'2013), 05-06 march 2013.
- [4] Iftikar Hussain, "Development of Assamese wordnet" for M.tech Dissertation, Department of Computer Science and Engineering Tezpur University, Napaam Tezpur-784028 Assam, INDIA June, 2010.