An Analytical Study of Synonymy in Assamese Language Using WorldNet: Classification and Structure

Shikhar Kr. Sarma
Gauhati University
Guwahati,Assam,India.
sks001@gmail.com

Himadri Bharali
Gauhati University
Guwahati,Assam,India.
himadri0001@gmail.com

Mayashree Mahanta
Gauhati University
Guwahati,Assam,India.
mayashreemahanta@gmail.com

Utpal Saikia
Gauhati University
Guwahati,Assam,India.
utpal.sk@gmail.com

Dibyajyoti Sarmah
Gauhati University
Guwahati,Assam,India.
dibyasarmah@gmail.com

Abstract

The present paper aims to categorize different types of synonymous words and also to highlight their synonymic pattern as well as grammatical categories found in Wordnet of Assamese language. Synonymy is an important component of vocabulary of the language. It establishes lexical relation between words. In fact, the term 'synonymy' is applied to the two or more words which share the same semantic features. WorldNet is a lexical database consisting of synsets. A synset is constructed by assembling a set of synonyms that together define a unique sense and synset is the basic foundation of Wordnet. Assamese language is rich in synonyms. In Assamese WorldNet, more than 20,000 synsets are entered under the categories of Noun, Verb, Adverb and Adjective. These synsets can of different types according to their semantic similarity, connotation, denotation, stylistic variations etc.

1 Introduction

Synonym is an important feature of the vocabulary of any language. But it is very difficult to give a clear, precise and correct definition of synonymy. There are various approaches with numerous definitions of synonym and types of synonyms. Linguistically, two or more words in the same language with very closely related meaning are called synonyms. It is to be mentioned here that synonyms does not mean the 'sameness of meaning' as there is no two terms with completely identical meaning. It is generally accepted that complete synonym is rare in natural language. The discussion of synonyms comes under the study of lexical relation. Lexical relation analyses the meaning of the words in the language which have related meanings. The idea of synonym is not only applied to lexical items, but also idioms, larger expressions, of course. A lexicographer builds a synonym dictionary depending on the words which share the same semantic features in a given language.

The present paper deals with lexical synonyms of the same word class, not with the phrasal synonyms. We will categorize the synonymous words considering the semantic features of the words they share based on Assamese Wordnet. Besides, it is also an attempt to point out the synonymic pattern and the grammatical categories of synsets in Wordnet.

2 A rapid sketch of Assamese Language

Assamese is the easternmost New Indo-Aryan language spoken in the Brahmaputra valley comprising at present six districts with Lakhimpur in the extreme east and Goalpara in the west. Tibeto-Burman and the Khasi are the important ones. According to the 1991 census report, the number of speakers of the language is almost 100000 billions. However, it is spoken as a second language by a considerable number of speakers of Tibeto-Burman languages like Bodo, Mising and Karbi. Traditionally, it has served as the lingua-franca or pidgin in the neighbouring states of Nagaland and Arunachal Pradesh.

The word 'Assamese' is an English one based on the anglicized form 'Assam' from the native word 'Asam'. The word Assam was connected with the Shan invaders of the Brahmaputra valley during the 13th century. In modern Assamese, Shan invaders of the 13th century are termed as 'Ahoms'.

Presence of Assamese language dated back to the literatures of Charyapadas, written by Budhist scholars. The Assamese language present in charyapadas reflects its evolutionary stages in initial state. Literatures with distinct Assamese language are found from the Kavyas of the pre Sankari era. This was in 13th century AD. From that time onwards pure Assamese language with its structured forms evolved (Goswami, 1983). Assamese script is derived from Brahmi script. It played a vital role in the evolution of the Indian script. The rock inscription and copper plate from 5th to 9th century showed the evolution of Assamese script. There are eight vowel phonemes in Assamese. There are twenty-one consonant and two semi-vowel phonemes in the Standard Colloquial Assamese (Kakati, 2008).

2 A Brief Discussion of Assamese Vocabulary

The scope of Assamese vocabulary is very vast. It consists of words of Sanskrit origin, Non-Aryan words, dialect oriented words. Besides Assamese socio-cultural influences are also perceived in the vocabulary of the language. It is to be noted here that Assamese still lacks a common vocabulary dictionary in the language. Moreover, no dictionary was found in the early and the middle ages. The selected modern dictionaries are – '*A Dictionary in Assamese and English by Miles Bronson*' (1867); 'Hemkosh' (1900), by Hemchandra Barua and later it is compiled by Debananda Barua (the 14th edition) which included 1, 54,428 words; '*Chandrakanta Abhidhan*' (2004, 3rd edition) '*Adhunik Asomiya Sabdakosh*' (2007, 9th edition), '*Asamiya Jatiya Abhidhan*' (2010) and many other vocabulary dictionaries are available in Assamese language. No common standard vocabulary dictionary has been made till today. Many critics have prepared vocabulary lists in their own way. Earlier philologists like Kaliram Medhi and Banikanta kakati had classified vocabulary list in their own style.

Kaliram Medhi in 'Asomiya Byakaran aru Bhasatattva' has provided a classification Assamese vocabulary such as 'tatsama', 'tatbhava'

and 'desaja'. But his 'Desaja' words are shown as loan words in which maximum words are Perso-Arabic words (Pathak, 2004). Therefore, his vocabulary classification cannot be taken as valid. On the other hand, though Banikanta Kakati's classification of Assamese vocabulary covers almost all the aspects, yet his classification also cannot be regarded as valid one.

It is interesting to note here that there are a large amount of loan words in Assamese language. In day-to-day life these loan words have been used extremely to express feelings, ideas etc. Moreover, it is seen that Perso-Arabic words have been used in Assamese language. These words occupy a significant status in Assamese language.

Assamese vocabulary can be divided into the following heads (Sarma et al., 2012):

1. Aryan or words of Sanskrit origin
   a. Tatsama
   b. Semi-tatsama
   c. Tadbhava
2. Non-Aryan words
   a. Austro-Asiatic
   b. Tibeto-Burmese
   c. Tai-Ahom
   d. Dravidian
3. Loan words
   a. Words coming from N.I.A. languages
   b. Foreign Words
   i.Persian
   ii.Arabic
   iii.Portugese
   iv.English
   c. Loan translations
      i.Translated words
      ii.Terminology
4. Unclassified words
   a. Hybrid
   b. Onomatopoetic
   c. Compound

3 Wordnet and Synonym Sets Building in Assamese Language

Wordnet is a repository of words of a language. It is basically a synonymous lexical database. The words are classed together according to their similarity of meanings. Vocabulary plays a main role in building Wordnet. The task of Assamese Wordnet building is almost ready to provide us with all the lexical words. Though Assamese wordnet tries to cover all the Assamese word

forms, yet there are still many words in the language those need to be entered (Sarma et al., 2010)

3.1 Classification of Synonymy in Assamese

Assamese language is rich in synonyms. We can classify synonyms under the following three heads:

1. Absolute synonymy:
Words can be called absolutely synonymous if they share the complete semantic features in all contexts of occurrences. However, it is generally recognized that absolute synonyms are almost non-existent. Though it is very rare, it certainly exists in Assamese languages. It is limited mostly to dialectical variations and technical or institutional terms. For example:

বিদ্যালয় 'bidyaaloi' (school): পঢ়াশালী, পাঠশালা 'parhaashaalii, haathshaalaa'

খবৰ 'khabar' (news): বাতৰি, সংবাদ, সংবাদ পত্র, বাতৰি কাকত 'baatari, sangbaad, sangbaad hatra, baatari kaakati'

2. Stylistic synonymy:
Stylistic synonyms are words conveying the same concept but differing in stylistic connotations. Stylistic synonymy is very common in Assamese language. For example-

মৃত্যু 'mrityu' (death):

মৰণ, প্ৰয়াণ, প্ৰাণত্যাগ, মহাপ্ৰয়াণ, বৈকুণ্ঠ প্ৰয়াণ, তিৰোধান, তিৰোভাৱ, কাল, লোকান্তৰ,কালগ্ৰাম, দেহাৱসান 'maran, prayaan, praantyaag, mahaaprayaan,boikuntha prayaan, tirodhaan, tirobhaabh, kaal, lokaantar, kaalagraam, dehaawasaan'

সুন্দৰ 'sundar' (beautiful): ধুনীয়া,দেখনিয়াৰ,ৰূপহ, মোহনীয়,নয়নাভিৰাম,চকুত লগা, চকু জুৰোৱা, নয়ন জুৰোৱা, বিতোপন,চকুত চমক লগোৱা 'dhuniaa, dekhaniyaar, rupah, mohaniiya, nayanaabhiraam, sakut lagaa, saku juruwaa, nayan juruwaa, sakut samak lagowaa'

3. Ideographic synonymy:
Ideographic synonyms convey the same concept but differ in denotations. It is also called denotation based synonymy. For example-

টুকুৰা 'tukuraa' (a piece): চকল, ডোখৰ 'chakal, dokhar'

খং 'khong' (anger): ক্ৰোধ, ৰাগ, কোপ, ক্ৰোধাগ্নি 'krodh, raag, kop, krodhaagni'

Apart from these, we can have the following more synonym types in Assamese language depending on its resemblance of meaning, distribution, style, form etc.

3.2 Near Synonymy

Near synonyms are those words whose meaning is relatively close or more or less similar, but not fully intersubstitutable. They vary in terms of their shades of denotation, connotation, implicature, emphasis or register. Near synonyms are extensively found in Assamese. For example:

ভাল 'bhaal' (good): সজ্জন,সৎ 'sajjan, sat'

All these words denote the quality of goodness. But they differ from one another in respect to their denotational meaning. The word ভাল 'bhaal' is a generic term, whereas সজ্জন 'sajjan' is more particular applicable only to human being. Besides, সৎ 'sat' conforms to both animate and inanimate things. The usages of these synsets are shown below-

ভাল 'bhaal' - ভাল ব্যক্তি/ কাম/ কিতাপ 'bhaal byakti/kaam/kitaap' (good person/work/book)

সজ্জন 'sajjan'- সজ্জন ব্যক্তি/ ✳কাম/ ✳কিতাপ

সৎ 'sat'- সৎ ব্যক্তি/ কাম/ ✳কিতাপ

Near-synonyms can vary as follows-

Type of variation	Examples
Collocational	কৰ্ম, চাকৰি *'karma, chaakari'* (work)
Stylistic, formality	সন্মানীয়, মাননীয়, মান্যবৰ *'sanmaniiya, maananiiya, maanyabar'* (honourable)
Stylistic, forced	ধ্বংস, পতন *'dhansha, patan'* (destruction)
Expressed attitude	ক্ষীণ, লাহী, শুকান *'khin, laahii, sukaan'* (thin)
Emotive	মা, আই, মাতৃ *'maa, aai, matri'* (mother)
Continuousness	নিগৰা, বোৱা *'nigaraa, bowaa'* (to drip, to flow)
Fuzzy boundary	বননি, বন, হাবি, জংঘল, অৰণ্য *'banani, ban, haabi, janghal, aranya'* (wood)

Table 1: Type of variation

The first column in the table 1 represents the various classifications of Near-synonyms and in the next column, the examples of respective Near-synonym types are given accordingly. The above mentioned Near-synonym variations are seemed to be almost near in their meanings, but most of them differ in their distributions. The distribution of the first type of variation of Near-synonym in the Table 1 is shown below:

Collocational: কৰ্ম স্থান 'karma sthaan) (work place)

চাকৰি *স্থান 'chaakari sthaan' (Work place)

3.3 Connotation Based Synonymy

More modern approach to classify synonyms may be based on definition of synonymous words differing in connotation. The scope of connotation based synonyms is very vast one. Connotation based synonyms in Assamese language are categorized in the following types:

a. Connotation of degree or intensity: আচৰিত, অবাক, স্তম্ভিত 'aacharit, abaak, stambhita' (surprise)

b. Connotation of duration: জুপি চোৱা, ভূমুকিয়াই চোৱা 'jupi chowaa, jumi chowaa, bhuumukiyaai chowaa' (to peep)

c. Emotive Connotation: অকলশৰীয়া, শূন্যতা 'akalshariiyaa, shunyataa' (loneliness)

d. Evaluative Connotative: It conveys speaker's attitude as good or bad. For example: প্ৰখ্যাত, জনাজাত, বিখ্যাত 'pryakhyaat, janaajaat, bikhyaat' (famous)

e. Causative Connotation: পকা, পৰিপক্ক হোৱা 'pakaa, paripakka howaa' (to ripe)

f. Connotation of manner: সোনকালে, ততালিকে, খৰকৈ, শীঘ্ৰে 'sonkaale, tataalike, kharkoi, shiighre' (fast)

3.4 Cognitive Synonymy

Cognitive synonymy is also known as descriptive, propositional or referential synonymy. Cognitive synonym is sometimes described as incomplete synonymy (Lyons, 1981), or non-absolute or partial synonymy (Lyons, 1996). Cognitive synonymy highlights the fact that though not all speakers of a language will necessarily use, yet they may understand it well. Cognitive synonymy is also termed as denotative synonymy (Stanojević, 2009) It analyzes sense and denotation. Examples of Cognitive synonyms-

গোপন 'gopan' (secret): অপ্ৰকাশ্য, গুপ্ত, গুপুত 'aprakaashya, gupta, guput'

দেউতা 'deutaa' (father): পিতা, পাপা, পিতাই, আতা, বাবা, পিতৃ 'pitaa, paapaa, pitai, aataa, baabaa, pitri'

3.5 Euphemism Synonymy

Euphemism is the substitution of words of mild or vague connotations for expressions rough, unpleasant. These kinds of synonyms are important linguistic tools that are inherent in our language. Most of the people like to use in day to day conversation. The use of such words is both social and emotional. Euphemism deals with the touchy or taboo subjects (like sex, personal appearance or religion) without hurting or upsetting others (Radulović, 2012). As a matter of fact, euphemism can be of two types: (a) Positive euphemisms increase acceptability such as, domesticity, institutional, economical etc., and (b) negative euphemisms decrease negative values that are associated with negative phenomena such as, war, drunkenness, crime, poverty (Rawson, 1981). For example:

Positive euphemism: স্তন 'stan' (breast): পিয়াহ, পয়োধৰ, পয়োভাৰ, কুচকুন্ত, ওহাৰ, বাত '

Negative euphemism: বেশ্যা 'beshyaa' (prostitute): গণিকা, ৰেণ্ডী, দেহোপজীৱী, পতিতা, নটিনী 'ganikaa, rendii, dehohajiibii, patitaa, natinii'

4 Synonymic Pattern and Grammatical Catagories in Assamese Wordnet

There is no fixed pattern of synonymous words in a synset in Assamese wordnet. Sometimes only one word is provided for one concept in the Wordnet. In certain concepts, it covers up to 38 synonymous words. Here, we can take the following example:

Concept: AG: ঘেহুঁ আদিৰ বিশেষ প্ৰকাৰৰ খহটা চূৰ্ণ

EG: milled product of durum wheat (or other hard wheat) used in pasta

Synset: চূজি 'suji'

Concept: AG: ধৰ্ম গ্ৰন্থৰ দ্বাৰা স্বীকৃত এক সৰ্বোচ্চ সত্তা, যি সৃষ্টিৰ গৰাকী

EG: the supernatural being conceived as the perfect and omnipotent and omniscient originator and ruler of the universe; the object of worship in monotheistic religions

Synset: ভগৱান, ঈশ্বৰ, প্ৰভু, বিধাতা, পৰমপিতা, দয়াময়, সৃষ্টিকৰ্তা, পৰমব্ৰহ্ম, জগদীশ, অন্তৰ্যামী, ভুৱনেশ্বৰ, কৰুণাময়, নেদেখাজনা, ওপৰৰজনা, জগজীৱ, মংগলময়, সৰ্বমংগলময়, সনাতন, বিভু, ধাতা, বিধাতাপুৰুষ, জগদীশ্বৰ, জগৎপিতা, জগৎপতি, জগতকৰ্তা, জগতস্ৰষ্টা, জগজীউ, পৰমেশ্বৰ, পৰাৎপৰ, ইচ্ছাময়, পৰমাত্মা, ত্ৰিজগতপতি, ত্ৰিলোকপতি, পৰমানন্দ, নিয়ন্তা, চিন্তামণি, ভবেশ, নিৰাকাৰ

'*bhagawaan, iishwar, prabhu, bidhaataa, parampitaa, dayaamoy, sristikartaa, parambrahma, jagadiish, antarjaamii, bhuwaneswar, karunaamoy, nedekhaajanaa, opararjanaa, jagajiiwa, mangalmoy, sarbamangalmoy, sanaatan, bibhu, dhaataa, bidhaataapurush, jagadiishwar, jagatpitaa, jagatpati, jagatkartaa, jagatsrastaa, jagajiiu, parameshwar, paraatpar, issaamoy, paramaatmaa, trijagatpati, trilokpati, haramaananda, niyantaa, chintaamoni, bhabesh, niraakaar*'

Apart from these, Assamese Wordnet considers only Noun, Verb, Adverb and Adjective class. But there are evidences of synonymous words in closed classes also like preposition, conjunction and Interjection etc. It may be the reason that we can find large amount synonymous words from the open classes and also can be compared with the other languages easily.

Examples of Synsets according to grammatical categories are given below:

Noun: কাগজ, কাকত, তুলাপাত, পেপাৰ 'kaagaj, kaakat, tulaapaat, pepaar' (paper)

Verb: নচা, নৃৎয কৰা 'nachaa, nritya karaa' (to dance)

Adverb: ওচৰতে, কাষতে, সমীপতে, গুৰিতে, অদূৰতে '*osarate, kaashate, samiipate, gurite, aduurate*' (near)

Adjective: অধম, প্ৰবলতাহীন, কম, অপ্ৰবল, বেয়া, নিকৃষ্ট 'adham, prabalataahiin, kam, aprabal, beyaa, nikrista' (bad)

5 Conclusion

Synonymy plays a vital role in the field of lexical study. It paves the way for wordnet building in any natural language. Synonyms in Assamese wordnet cover a large amount of lexical words coprising the grammatical categories, such as noun, verb, adverb, adjectives. Accordingly, we classify synonyms into certain types in Assamese language.

It is to be mentioned here that while building synonym sets in Assamese wordnet, dialectical forms are not considered. Besides, though borrowed words are included in synset building in Assamese, but the numbers are very limited. Yet, there are many foreign words which we use them as native words in day to day communication. This kind of discussion will be dealt later sometime. Yet, wordnet with all its synsets have succeeded in representing Assamese language in a very systematic and novel way.

References

Hemchandra Barua. 1900. *Hemkosh*, ed. and published by Debananda Barua.

Miles Bronson. 1867. *Dictionary in Assamese and English*

Sumanta Chaliha.2007. *Adhunik Asomiya Sabdakosh*, Bani Mandir, Ghy

Golock C. Goswami. 1983. *Structure of Assamese*, Gauhati University, Assam Guwahati,Assam

Banikanta Kakati. 2008. *Assamese: Its Formation and Development*, Lawyers Book Stall, Guwahati , Assam

John Lyons. 1977. *Semantics*. Vol.1. Cambridge: Cambridge University Press.

John Lyons. 1981. *Language and Linguistics: An Introduction*, CUP, Cambridge.

John Lyons. 1996. *Linguistic Semantics*, CUP, Cambridge.

Maheswar Neog and Upendranath Goswami ed., 2004, C*handrakanta Abhidhan,* Publication Deptt. Gauhati University

Ramesh Pathak. 2004. *Studies in Assamese Vocabulary*, Anita Pathak, Guwahati, Assam

Milica Radulović: *Expressing Values in Positive and Negative Euphemism*. Facta Universitatis, Series: Linguistics and Literature Vol. 10, No 1, 2012, pp. 19 – 28

Hugh Rawson.1981. *A Dictionary of Euphemisms and Other Doubletalk*. New York: Crown Publishers, Inc.

Debabrat Sarma. 2010. *Asamiya Jatiya Abhidhan*, Asom Jatiya Prakash

Shikhar Kr Sarma,Utpal Saikia, Mayashree Mahanta, Himadri Bharali. 2012, *Assamese Vocabulary and Assamese Wordnet Building: An Analysis*. Global Wordnet Conference, Matsue, Japan

Shikhar Kr. Sarma, Moromi Gogoi, Rakesh Medhi and Utpal Saikia. 2010. *Foundation and Structure of Developing an AssameseWordnet*, Department of Computer Science Gauhati University, Proceed-

ings of the 5th Global Wordnet Conference,Narosa Publishing House.

Maja Stanojević. *Cognitive Synonomy: A General Overview*, Linguistics and Literature Vol. 7, No 2, 2009, pp. 193 – 200, Facta Universitatis