21

LEXICAL AMBIGUITY IN HINDI–MARATHI MACHINE TRANSLATION SYSTEM

(IN THE CONTEXT OF HOMONYMY)

Kamble Prakash Abhimannu

In the age of modernization, linguists cannot forget that India is a multilingual country. Problem of multilingualism was solved by translation but in the age of information technology machine is the competitor of language. In the field of machine translation there are many problems. One of the major problems is homonymy disambiguation. This issue is being considered at the world level. This is good opportunity for us to discuss this subject in the context of Indian languages. The problem of Homonymy is not only in machine translation but also in the area of Human Translation, quick translations (interpretation), new language learners (Language Acquisition), and young children of normal intelligence, Artificial Intelligence specialists, because it causes word ambiguity. The homonymy words come from original words (Tatsam, Tadbhav) of native languages or dialects. These words can also be combinations of foreign languages and Hindi form. Homophones are pronounced with the same sound but have different meaning. Because Devanagari script is sound based (Dhwanimulaka), if there is a difference in pronunciation then the difference appears in the spelling also.

Before proceeding to the formula of Homonymy words problem of Disambiguation, a detailed explanation of these terms and introduction is important because it helps in understanding the original depth of the problem.
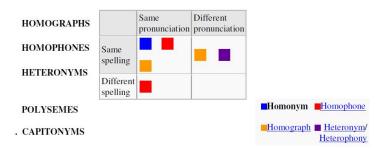
HOMONYMY WORD DEFINITION

"In linguistics, a *Homonym* is one of a group of words that share the same pronunciation but have different meanings, and are usually spelled differently or same."[1]

---

[1] http://en.wikipedia.org/wiki/Homonym

The above definition can be interpreted that" the same pronunciation ambiguous word combinations, that often have the same accent, but their meaning is different and spellings could be same or different." A few words in Hindi have similar pronunciations, the meaning is also same, but have different spellings, but the number of these words is very less. These words are exceptions in homonymy. Homonymy is divided in many types, let's see what they are, so that more and more modified forms of the words can be found which will help us in moving on.

We can categorize words on the combinations of Pronunciation, Homograph and ambiguity so that we can easily understand the difference in words like homophony and Homograph.

KINDS OF HINDI HOMONYMY

No	Spelling	pronunciation	Meaning
1	Same spelling	Same pronunciation	different meaning
2	Different spelling	Same pronunciation	different meaning

In the lexical data the word sense is decided on the basis of word derivation or grammatical category. But the problem arises when the computer reads Lexical Data which has been given in order of lexicon in the storage input. But output is not identical with the input used in the sentence because this formula is not good to identify the sense of the sentence. In the source language Lexical data doses not refers to the group sense because of this computer dictionary are not able to obtain transfer true sense in the target language. It is necessary to increase the percentage of Dictionary refers sense meaning with the mathematical algorithmic programs and linguistically formed algorithmic programs that help in increasing sense meaning. According to the partition of Homonymy word and grammatical category can be useful to get the correct meaning and sense. But translators and the general public frequently ask the same question that "the best existing program through only a

series of option, translator got to do can be presented as or better than the best selection to do a translation."[2] The answer to this question on algorithm can also be given by homonymy disambiguation. Even with the same focus on amendment, big problems can be divided in small parts to make it easy to solve them. When the Small problems will be solved, big problems will be automatically solved with them. Similarly, this Paper is a part of one of the biggest problem in machine translation; it is an effort to resolve ambiguity. That is taken account by the following translation from Hindi to Marathi. In this paper Hindi is the source language. The problem of Homonymy first originates in source language because so more attention is given to Hindi. It is necessary to solve the problem in the same language. The target language can be used only to translate the solution (meaning) of the problematic homonymy words. Some examples are given below.

Homophone (Same pronunciation – same spelling – different meaning)

No	Hindi word	Grammatical category	Hindi meaning	Marathi meaning
1	कुल/kul	Noun, Mas, Singular,	कुनबा/joad	कुळ/kul (naun,mas, singular)
	कुल/kul	Noun, Mas, Singular	जोड/joad	एकूण/ekun (Noun,mas,singular)
2	अचल/achal	Adj	पर्वत/parvat	पर्वत/parvat (Adj)
	अचल/achal	Adj	स्थिर/darja	स्थिर/sthir (Adj)
3	कोटि/koti	Noun, Fem	दर्जा/darja	दर्जा/darja (noun,masc)
	कोटि/koti	Noun, Fem	करोड/karod	करोड/karod (Adj)

Table No- 1

The words are represented by graph and graphs are useful in solving the problem of homonymy in Hindi and devnagari scripts. A problem like homophony depends on the pronunciation and in Hindi; pronunciation depends on the shape of the word. Each word has a different shape and meaning. The first task should be to find the correct meaning through the shape of the word because meaning is the soul of the language. This would help in the case of differently shaped words but not for the similar ones. The problem of similar shaped words can be solved only by using algorithm, corpus and grammar. This would not be very difficult because the number of similar shaped words and similarly pronounced words is very less.

Similar pronunciation-homophonic-different spelling-different meaning (Homophons)

---

[2] कंप्यूटर अनुवाद:प्रयोग और विधि - प्रो.रीतारानी पालिवाल, अनुवाद पत्रिका (कंप्यूटर अनुवाद विशेषांक -२) अप्रैल-जून २००४ पेज - ५९

No	Hindi word	Grammatical Category	Hindi Meaning	Marathi meaning
1	दिन/din	Noun Mas.	दिवस/divas	दिवस(Noun mas.)
	दीन/deen	Adj/N	गरीब/garib	गरीब (Adj.)
2	बलि/bali	Noun Mas.	बलिदान/balidan	बलि (Noun,Mas)
	बली/balee	Adj.	बलवान/balavan	शक्तिशाली (Adj.)
3	बाजि/baji	Noun Mas.	घोडा/ghoda	घोडा (Noun, Mas)
	बाज़ी/bazee	Noun Fem.	दाॅव/danv	बाजी (Noun,Fem)

Table No:- 2

In Hindi the difference in spelling means difference in pronunciation. Word pairs like aviraam-abhiram; sam-sham; asan-asann are similarly pronounced. But in English words can be found that have different spellings but the same pronunciation like – wood-would; know-no; pain-pen; sun-son; etc. These words are called "homophones" in English.[3] Although the samesound –different meaning words are not really homograph, but still they can be equated with homophones. Though homophones are recognized by dwanyatmakata but this problem loses its importance in Hindi because of the different spellings. It can be solve easily by algorithm.

Some nouns are unclear like homonyms, that have same pronunciation but their meaning is different. They have the same spelling and sound.

Homophones (Same pronunciation – Same sound-same spelling- different meaning)

No	Hindi word	Grammatical Category	Hindi Meaning	Marathi meaning
१	देव/dev	Noun Mas.	भगवान	देवता (Noun,Mas)
	देव/dev	Noun	नाम	नाव (Noun)
२	धवल /dhaval	Adj.	सफेद	पांढरा(Adj)
	धवल/dhaval	noun	नाम	नाव (Noun)
३	दिवाकर/divakar	Noun Mas.	सूर्य/sury	सूर्य (Noun,Mas)
	दिवाकर/divakar	noun	नाम/nam	नाव (Noun)

Table no. 3

The problem of ambiguity 'same sound-different meaning' words is equally found in all the languages. Especially in the Nouns that are frequently used in daily life. Young

---

[3] हिंदी में अनेकार्थता का अनुशीलन - डॉ.त्रिभुवन ओझा करीम सिटी कॉलेज, जमशेदपुर, विश्वविद्यालय प्रकाशन, वाराणसी १९९४ पेज - ७७

children and machines have to face a lot of difficulties because of them. But Scholars cannot put these words under homonymy. Therefore they are not being researched.

Same pronunciation–Different sound– same spelling–different meaning (Homograph)

No	Hindi word	Grammatical Category	Hindi Meaning	Marathi meaning
१	टेस्ट/taste	Noun	स्वाद/swad	चव/chav (Noun)
	टेस्ट/test	Verb transitive	इम्तहान/imtiyan	परिक्षा/parikShya(Noun)
२	इंटरेस्ट/interest	Noun	दिलचस्पी/dilchaspi	रुचि/ruchi (Noun)
	इंटरेस्ट/interest	Verb transitive	ब्याज/byaj	व्याज/vyaj (Verb Tran.)
२	कंडिशन/condition	Verb transitive	हालत/halat	परस्थिति/paristhiti(Verb Tran.)
	कंडिशन/condition	Verb transitive	प्रतिबंध/pratibandh	नियम/niyam(Verb Tran.)

Table no. 4

The practice of taking words from other languages is a practice of language development but the other side of this language development is an increase in homonymy, which creates ambiguity in the language. In Hindi there are a lot of words which came from many languages. Despite similarities in the sound and the pronunciation words have different meanings. Dr. Tribhuvan Ojha(1994) has divided these words in three categories that can clarify the meanings a little.

1. Words that have been assimilated in Hindi with their original pronunciation fall in this category. Like - June, foot, boot etc.

2. In this category are words that have been accepted in Hindi with minimal sound change.

Like- chauk (chalk), acaadamee (academy)

3. Words that are tadbhav forms of English words. Like bam (bomb), kaag (cork) etc.

These kinds of words can be clarified by categorizing them grammatically with meaning specification. The simplest language game is the naming game in which all objects are uniquely identifiable, like persons, but unlike chairs or apples. Such objects can be categorized with unireferential categories. Because of this, each object can be uniquely named or labeled with a proper name. If in this setup both the speaker and the hearer know the topic (for example because the speaker points to it

as is commonly done in many experiments and models), and if the probability of re-inventing an existing word is zero, then no homonymy can arise: an agent can always associate the correct meaning with an unknown and unique word. The most important task is to disambiguate the homonyms using language tools more and more. Such as: - corpus, tagger, morphological analyzer and special dictionary in which only homonyms lexical data. So that work can be sped up.

I put here Some Common Methods to resolve Hindi-Marathi machine translation Homonymy ambiguity. First preference is to create a special dictionary for homonymy. In this dictionary only those words will be entered that are a type of homonymy with categories by homonymy type Homograph, Homophone, Polysemy, Heteronyms, Capitonyms. In this dictionary we enter the sentence which is more useful to that special word disambiguation word; this is only a sentence example. By this sentence we take right word sense and meaning. This example sentences are of three or four types which are using any homonymy single word. Four type of Useful lexical data collection for homonymy is

(1) Baseline (Unmodified)
(2) Hindi-Marathi homonymy lexical data with sentence and examples (Additional Homonymy added words)
(3) Homophone (Additional Homophone)
(4) Homograph words (Additional Homograph)
The second method that I prefer to go with is Contextual method of functional homonymy disambiguation, developing for every functional homonymy type group of rules defining the syntactic context of the homonym disambiguation and forming the group control structure which defines the rule application order. Text analysis is one of the good examples to get the exact sense, as rather frequently a syntactic method of building homogeneous groups is used in homonymy disambiguation. In this analysis also we use a pre-syntactic analysis technique of homonymy disambiguation effective at the stage of sentence analysis. Provide Grammatical, world knowledge and linguistic level knowledge to system about the word. Morph syntactic analysis and Lexical-semantic analysis is also useful in this tool. To develop a "One Subsystem provides the realization of homonymy disambiguation method based on collocations and the one based on the ontology linguistic frame. In the development of these methods, the engineering approach is used, which allows selecting the typical frequent language cases, which are actively used in technical language."[4] Some useful language tools 1.Morphological analyzer 2.Tagger 3.Spacial Dictionary 4. Chunker 5. Ontological tools 6. word net.

---

[4] Integral Technology of Homonymy Disambiguation in the Text Mining system " Lota"- Oiga Nevzorova , Julia Zin'kina, Nicoly Pjaktin

Sometimes corpus gives only an accumulation of information but this term is not useful to disambiguate the homonymy word, rather the utility should be focused on the target.

Like:- १.१ आना (Noun)=(हिंदी)   <भिखारी> <का> <आठ> <आना> <खो> <गया>

(मराठी) <भिका-याचे> <चार> <आने> <गमावले> <.>

२. आना (Verb) (हिंदी) - <उसका> <आना> <मेरे> <लिए> <कितना> <सुखद> <था> <|>

(मराठी) - <त्याच> <येणं> <माझ्या> <साठी> <किती> <सुखद> <होतं> <.>

with the help of homonymy word dictionary corpus can be modified clarity must come in the meaning of given sentence in the corpus.

In this paper care has been taken that just simplification of meaning transfer is not done. This effort is to make such a mechanism that can convey the minimum meaning itself. This Paper presented Hindi-Marathi Machine translation in the context of the literal searching of correct meaning, to mark the ambiguity of effort and minimum but successful communication tool.

REFERENCES

1. A cross-situational learning algorithm for damping homonymy in the guessing game - Joachim De Beule1, Bart De Vylder1 and Tony Belpaeme2 Vrije Universiteit Brussel, Belgium, University of Plymouth, United Kingdom

2. Integral Technology of Homonymy Disambiguation in the text mining system "LOTA"

- Olga Nevzorova, Vladimir Nevzorov, Julia Zin'kina, Nicolay Pjatkin

3. Particle Homonymy and Machine Translation

- K6roly F&bricz, JATE University of Szeged, Egyetem u. 2.

Hungary I[- 6722

4. Children's difficulty in learning homonyms*

- MARTIN J. DOHERTY

Department of Psychology, University of Stirling

5. Native and L2 processing of homonyms in sentential context

- Kerrie E. Elston-Gu¨ ttler*, Angela D. Friederici

Max Planck Institute of Human Cognitive and Brain Sciences, Leipzig, Germany

6. Particle Homonymy and Machine Translation

- K6roly F&bricz, JATE University of Szeged,

Egyetem u. 2.

Hungary I[- 6722

7. Learning Form-Meaning Mappings in Presence of Homonymy: a linguistically motivated model

of learning inflection

- Katya Pertsova, University of California Los Angeles

8. प्रामाणिक हिंदी शब्द-रचना एव वतेनी प्रकाश -

9. हिंदी में अनेकार्थता का अनुशीलन - डॉ.त्रिभुवन ओझा करीम सिटी कॉलेज, जमशेदपुर,

   विश्वविद्यालय प्रकाशन, वाराणसी

10. कंप्यूटर अनुवाद:प्रयोग और विधि-प्रो.रीतारानी पालिवाल, अनुवाद पत्रिका (कंप्यूटर अनुवाद विशेषांक-२) अप्रैल-

    जून २००४ पेज-५९

http://en.wikipedia.org/wiki/Homonym

http://assortedmaterial.googlepages.com/EnglishIndex.htm

http://www.tribuneindia.com/2000/20000819/windows/roots.htm