THE CRITERION
AN INTERNATIONAL JOURNAL IN ENGLISH

Bi-Monthly Refereed and Indexed Open Access eJournal

8th Year of Open Access

The Criterion

Editor-In-Chief: Dr. Vishwanath Bite

www.the-criterion.com

About Us: http://www.the-criterion.com/about/

Archive: http://www.the-criterion.com/archive/

Contact Us: http://www.the-criterion.com/contact/

Editorial Board: http://www.the-criterion.com/editorial-board/

Submission: http://www.the-criterion.com/submission/

FAQ: http://www.the-criterion.com/fa/

ISSN 2278-9529

Galaxy: International Multidisciplinary Research Journal

www.galaxyimrj.com

Computational Translation of English Text into Marathi: Noun Phrase Issues

Gadekar M. B.
Shriman Bhausaheb Zadbuke Mahavidyalaya,
Barshi, Dist- Solapur
State – Maharashtra
&
Dr. Ronge P.T.
Principal,
Arts and Commerce College, Madha
Dist- Solapur
State – Maharashtra

Abstract:

The paper briefs the experimental activity of computational translation process of English Text into Marathi which encounters numerous problems. It deals in detail with translation of English Noun Phrase (ENP) into Marathi Noun Phrase (MNP) engrossed with several major and minor issues at not only lexical level, but wide range aspects that turn this overtly simple matchmaking task into highly complex multi-layered and multi-componential analysis. The constitution of ENP has inherently three properties viz. Pre-modifier, Head and Post-modifier. Every element in the phrase has its own unique and peculiar membership from word classes which differs from one another. None of them is indiscriminably similar. With these strictly rigid rule governance to the phrase elements and their syntactic indifference they need to be exchanged in terms of the Target Language (TL) i.e. Marathi with tremendous differences at every level. But the exchange is not only the verbal, syntactic or grammatical, it is semantic. It makes this task more intellectually laborious and challenging. The paper is an attempt to introduce this phrase level processing problem.

Keywords: Computational Linguistics, Computational Translation, Noun Phrase, English-Marathi Translation, Phrase Processing, Pre-modifier, Post-modifier, Head.

1. Introduction

The highly intricate nature of human language along with multi-layered complexity in terms of other languages is an issue in itself. When it is considered at the translational level, there are still more. As it turns out to be machine issue, there problem becomes more serious. These are concerns due to the multidisciplinary nature of computational translation and its interrelated disciplines with wide range interrelated branches with blurring borders of distinction. Computational linguistics (CL), Machine Translation (MT), Artificial Intelligence (AI) and Natural Language Processing (NLP) have superficially identical, but deeply diverse of their objectives. In the present context, instead of becoming more prosaic in discussion on

interrelations, it is the Computational Linguistics, in its strict sense, is practically considered. The branch of knowledge, unlike Natural Language Processing, is more linguistic than computational. The paper runs over the practical issues of processing a natural language phenomenon in terms computational constraints. Among other linguistic issues, the semantic divergence poses numerous challenges, initially at lexico-Semantic level with furtherance at discursive practices with its gradual linguistic hierarchical development

As mentioned above, the boundaries are blurred. Here is the clarification to the term Computation Linguistics and demarcation of the boundary. As well this is an attempt to justify the position of machine translation in Computational Linguistics. The following definitions usefully help the way out to reach the conceptual framework.

2. Computational Linguistics (CL)

The computational Linguistics is more computational than linguistics. Thus branch of knowledge attempts to resolve the linguistic issues with the application of computer. The study of the branch is application specifically subcategorized further. The following definition is useful to catch the meaning.

"Computational Linguistics might be considered as a synonym of automatic processing of natural language since the main task of computational linguistics is just the construction of computer programmes to process words and texts in natural language" (Patrice John,11)

The definition highlights the focal areas of the study of computational linguistics. It has its focus on Hyphenation, Spell-checkers, Machine Translation, Data Extraction, Summarization etc. Machine Translation is area which inititated the investigation and experimentation. The diversified sub-divisioning of the branch has widened the width and depth of the study. The following details are helpful to understand the concept of MT.

2.1 Machine Translation (MT)

However, the paper is related to Machine Translation, it needs to be clarified what is MT. The following definition is useful.

"Machine translation means automatic translation of a text of a source language to the target language by machine, the computer." (S. Kanakraj, 69)

The definition itself introduces the issue that the translation phenomenon is related to source and target language. It is at least bilingual issue. The following details are associated with English as a source language and Marathi as target language. English and Marathi are linguistically diverse languages. Nothing is common in them. As a result, there are graphological, lexical, syntactical, grammatical and many other such issues. The development upto dictionary level is ever so far successful. However, the text translation is attempted at C-DAC, IIT Kanpur, Mumbai, Hyderabad, yet it is at lab-level. The reason behind them is the failure to match the linguistic and computational threads. Among the linguistic issues given above, the syntactic and grammatical issues pose greater challenges. And they begin with

Phrases. The Noun Phrase (NP) is the most prominent among them. The ENP is syntactically similar to the larger extent with MNP. However, the difference causes the problem. Furthermore, the complications arise on the compatibility on computational grounds. While translation English phrases into Marathi the Noun Phrases encounter various level problems. These problems are discussed with details.

3. Noun Phrase (NP)

The English Noun Phrase can systematically be considered on three different grounds. i.e. formal, functional and semantic properties of the phrase ingredients. A translational platform, as it necessitates, consists of formal logical units processable in terms of mathematical requirements. And to meet the needs, the formal and functional framework indirectly embodies technical aspects to be processed by machine. The linguistic analysis of English phrases hereunder initiates the issue with appropriate introductory background for English-Marathi Machine Translation (MT) problems. Structurally, an English Noun Phrase (ENP) is peculiar in terms of larger commonalities with Marathi from linguistic perspectives. The following details about the NP are at conceptual level and its. Linguistic elements are useful to comprehend the nature of the issue.

3.1 Nature of NP

For the general purpose formal discussion, the definition of Noun Phrase is helpful to understand the underlying concept.

A noun is an idea, object, event, incident, element, or any abstract or concrete entity. The word used to name them is a noun. To identify a noun, apart from this working definition, there are several other elements which consist of structure, function and meaning. The following details are useful to know them in particular.

Structurally, an NP includes Pre-modifiers, Head and Post-modifiers. However, these elements are diverse and unique for their morphological properties, functional features and semantic layers. The following formulaic representation of an NP is in most part modeled after some of structural Linguists like Leonard Bloomfield, Noam Chomsky and Geoffrey Leech with certain modification required for computational processing. The computer languages with rigid structures are needed to be processing Natural Languages (NL) with disparate issues at linguistic, cultural, social and contextual levels. The structure of ENP is organized below.

3.2 Structural Elements of NP

As discussed earlier an NP constitutes Pre-modifier, Head and Post-modifier. There is tremendous variety in the Noun Phrase. The Noun Phrase consists of premodifiers and postmodifiers. The head of the Noun Phrase has wide range of word classes consisting of nouns, pronouns, enumerators, adjectives etc. The premodifiers also consists of determiners, enumerators, adjectives, nouns and genitive phrases. And among the postmodifiers there can be prepositional phrases, relative clauses, adverbs, adjectives, and Noun Phrases.

Noun Phrase is the most important phrase and it is next to verb phrase. It performs maximum functions in the clause. The structure and function of Noun Phrase is given below.

Noun Phrase = (Mn + H+Mn)

This formulaic presentation vividly shows that the NP is constituted with premodifier, Head and Postmodifier. Head is the most important element. The number, gender and case of the noun controls the meaning-shade of premodifiers and postmodefiers. The following word classes occur as Head of NP.

3.2.1 Head-
1) Noun – (John)
2) Pronoun – (it)
3) Adjective – (the poor)
4) Enumerator – (three)
5) Genitive Phrase (God's)

The heads of Noun Phrases in general have six different possibilities of semantic variations based on number and gender. These varieties are tagged separately. They are given below as a representative example.

Head Database			
Sr.	**Word**	**Tag**	**Meaning**
1	boy	NMS	मुलगा
2	boys	NMP	मुलगी
3	girl	NFS	मुलगी
4	girls	NFP	मुली
5	table	NNS	मेज
6	tables	NNP	मेज

Fig.1 Head Database

The Head Database in Fig. 1 above is the representation of database for Noun as a head. It shows that semantically a Noun can have at least six different possibilities from the perspectives of gender and number. The premodifiers in NP come unmarkedly but the MNP occur with graphological modification in their translated version. It creates problems for the machine to process. Apart from these elements, there are still more that affect the choice of modifiers meaning-shade, and they are adjectives. The Premodifiers representative Database is given below to acquaint with the problem. It is not possible and feasible to encompass such a vast issue in a paper because it is multifaceted and multi-dimensional.

3.2.2 Modifiers

The Noun Phrases take both kinds of modifiers; premodifiers and postmodifiers. The premodifiers and postmodifiers are listed below.

A) PreModifiers

There are number of modifiers in a NP.

1) Determiners (the man)
2) Enumerators (seven people)
3) Adjective (sweet faces)
4) Noun (garden fence)
5) Genitive Phrase (Fred's Uncle)
6) Adverbs (quite a noise)

The premodifiers in Noun Phrase consist of different kinds of word classes as referred above. The most interesting and conflicting category for the translation purpose is adjective because it does not have any special marker for number and gender, but while translating, there are vowel changes in Marathi. It creates confusion and difficulties in processing. The generation of following database interface is helpful to resolve the issues.

PrMDB					
AdjMS	AdjMP	AdjFS	AdjFP	AdjNS	AdjNP
black	black	black	black	black	black
big	big	big	big	big	big
heavy	heavy	heavy	heavy	heavy	heavy
kind	kind	kind	kind	kind	kind

Fig. 2 Premodifiers' Database

The above given Premodifiers Database shows that the appropriate choice of the meaning-shade from the table is chosen with the properties of noun or head in correspondence with their semantic properties viz. number, gender, case etc. Here Adj. stands for adjective. The number include singular(S) and Plural(P), whereas, the gender, include Masculine (M), Feminine (F) and Neuter (N). All the tags have these forms.

B) Postmodifiers

There are following postmodifiers in a Noun Phrase. All the above given are premodifiers. Post modifiers that come after head in a Noun Phrase.

1) Prepositional Phrases (PP)
   (The aim (of my life))
2) Relative Clause (RCL)
   (The train [passing a platform])
3) Adjectives(something strange)
4) Adverbs (the girl inside)

The post modifiers in a Noun Phrase can be prepositional phrases, relative clauses, adverbs, adjectives and Noun Phrases. As this research work does not consider postmodification, the database interfaces are not prepared and discussed. It is left for future research considerations.

3.3 Noun Phrase- Computational Issue

The NP in itself has numerous other issues related to different dimensions. They are discussed below:

3.3.1 Phrase elements

Among the phrase elements, there are the issues of heads and modifiers. As discussed earlier, the head of Noun Phrase can be a noun, pronoun, adjective, enumerator and Genitive Phrase. The most common of the phrase issues in the elements of a phrase. Mostly, they are related to nouns and adjectives. The noun can consist of number, gender and case. These elements in English come without special marking and in Marathi when they are considered for translation, they are inflected by certain markers to Gender and case elements. These elements while translating into target language, the machine cannot choose between the alternatives correctly. This is the basic problem that it needs the separator database and separate programming to make the appropriate choice. The issue of case element can be resolved partially with the order or the sequence of a sentence in basic structure order in a declarative sentence. However, if the structures are derived, they influence the choice. And automatically a sentence is mistranslated because of the subject along with the noun As far as the head is concerned, the major issue of the premodifier is associated with adjectives, and adjective can modify a noun in a Noun Phrase where the choice of the correct meaning-shade according to the gender and number leads to wrong translation, because the number and gender control the adjective meaning shade as well as the adjectives of size, shape, volume and colour can affect the appropriate choice according to the head of the Noun Phrase. To resolve this problem, it is necessary to consider the separate database creation for the sake of independent semantic layers for every category of words.

3.3.2 Phrase Case

One of the important issues of machine translation related to Noun Phrase is the issue of case of the Noun Phrase. The Noun Phrase which is coming at the beginning of the clause, usually, stands as subject, and they are usually in nominative or subjective case. However, if the same Noun Phrase is followed by a verb can stand as an object or complement maybe in objective case. The basic problem lies here with the fact that the computer cannot process them with the existing database. It needs separate provision and separate programming and to process them separately. When there are complex structures with intersected clause relations, it is difficult for a computer program to judge the given Noun Phrase as an object of the first clause or subject of the following. In such cases the processing becomes difficult and challenging. It also needs the separate treatment to be given to a Noun Phrase to translate it from English to Marathi, because in English there are no separate markers for a Noun Phrase to stand as a subject and object or complement. However, in Marathi the same phrase

undergoes the change from the case perspective this is major problem because the Noun Phrases are the major phrase types which perform maximum functions in a given clause. It can stand as a subject, object, complement or adverbial.

3.3.3 Case of Pronouns

The next issue related to the Noun Phrase is the issue centering around pronouns. The pronouns in English are unmarked in the sense that the same pronoun can be in Marathi in two different cases, and this case conflict occurs usually when they stand as subject and affects the meaning. For example *He is a boy*.is different from *He has a son.* In these two examples, the same pronoun is controlled by two different cases in Marathi. In the first case, it is unmarked with any case related suffixation. However, in the second example the issue becomes critical, when it becomes the issue of objective case that differs from this one. The pronoun *her* is both possessive pronoun and objective pronoun. In such cases, it is difficult to choose the correct one. It needs separate programming to process the pronouns. Such problems affect the machine translation process and resolution of such issue in given plain programming is impossible.

4 Conclusion

It is vivid with this discussion that NP is the most important phrase as it performs the functions of subject, object, complement, adverbial etc in a clause. When it is an issue with mechanical translation, it has several machine level complications which are discussed above. The discussion leads to the conclusion that the gravity of the problem is unresolvable in current system with traditional methods. It needs the creation of separate databases with their classification into semantic properties which are well defined with their registral or domain-centric identifiers. Simply the same NP can stand as a subject or object, but its semantic divergence, the change in its case creates problems. The internal elements or the phrase constituents in a give language are organized hierarchically cannot be represented equally in the alternate language because of multi-level diversity. The paper is useful for the furtherance of semantic database creation.

Works Cited:

Hocket, Charles. *A Course in Modern Linguistics.* OUP. 1970

John, Patric and David Christopher. *Computational Linguistics.* Commonwealth Publishers. 2011

Kanakaraj, S and N Kalaithasan. *Translatology*. Prem Publishers. 2003