Inflection rules for Marathi to English in rule based machine translation

Namrata G Kharate¹, Varsha H Patil²

¹Department of Computer Engineering, Vishwakarma Institute of Information and Technology, India
²Department of Computer Engineering, Matoshri College of Engineering and Research Center, India

Article Info

ABSTRACT

Machine translation is an important application in natural language processing. Machine translation means translation from source language to target language to save the meaning of the sentence. A large amount of research is going on in the area of machine translation. However, research with machine translation remains highly localized to the particular source and target languages as they differ syntactically and morphologically. Appropriate inflections result correct translation. This paper elaborates the rules for inflecting the parts-of-speech and implements the inflection for Marathi to English translation. The inflection of nouns, pronouns, verbs, adjectives are carried out on the basis of semantics of the sentence. The results are discussed with examples.

This is an open access article under the CC BY-SA license.

Keywords:
Adpositions
Inflection
Machine translation
Parts-of-speech
Vibhakti

Corresponding Author:
Namrata G Kharate
Department of Computer Engineering
Vishwakarma Institute of Information and Technology
India
Email: namrata.kharate@viit.ac.in

1. INTRODUCTION

Machine translation is one of the emphasis applications in natural language processing (NLP). Institutions and organizations in India have started working on machine translation systems for Indian languages and have gained satisfactory results [1], [2]. Communication plays important role in life of people. There are many languages used for communication all around the world and good literary works are available in every language. It is not possible to learn all the languages and so there is a need to develop effective machine translation means for targeting multiple languages. English is the language used my majority of the world population for official work, literary work, and all sorts of communication. Marathi is primary language and mostly used in Indian state Maharashtra. It is found that about 71 million people speak Marathi and variety of literature and novels are available in Marathi and hence there is a need for Marathi to English translation [3]. Researches have published the work mostly related to pair of languages and some standard tools are also available for translation [4]-[6]. But it is found that more contribution is needed for Marathi to English translation. As the structure and the grammar vary for the source and target languages, the restructuring and grammatical rules need to be observed correctly. This paper mainly discusses the inflectional rules related to the Marathi-English language pair. Rules are discussed with examples. Rules plays important role in rule based machine translation. This paper includes the literature review related to inflections, importance of adpositions in linguistics, proposed work, inflectional rules, research method, results, and discussion.

Tidke and Sugandhi [7] presented the implementation of the inflection for English to Marathi translation for parts of speech like nouns, pronouns, verbs, and adjectives. Wren and Martin [8] written book on English grammar in which various rules are given for English word inflection. Conway [9] has discussed the problem of English plurals and claimed that even at the lexical level; it can be a complex matter to
correctly inflect the individual words of a sentence to reflect their number, person, mood, and case. As per related work there is no sufficient work is done infection on Indian languages. In this paper we are working on Marathi language. As Marathi language is very rich in morphology so it’s little bit difficult to study. Adpositions are words which can occur before or after a word, phrase or clause that is necessary to complete the meaning of a given sentence. Adpositions are mainly categorized as: Prepositions, postpositions and circumpositions.

- Prepositions: Prepositions mean the words which occur before the complement [9]. A preposition occurs in English language. These prepositions are usually converted into postpositions in Marathi language.
- Postpositions: Postpositions mean the words which occur after the complement [10]. Postpositions occur in Marathi language. Example: पक्षी झाडावर बसलेले आहेत. Here ‘वर’ is the postposition.
- Circumpositions: Circumpositions are words appear on before and after the complement. Circumpositions used in English language. Example: I will play regularly from now on. Here ‘from ... on’ are the circumpositions. English language has SVO structure and Marathi language has SOV structure. The languages which follow SVO structure use prepositions. Hence, during translation of a Marathi sentence to an English sentence, there is a necessary to change the postpositions of language Marathi which is the source language of the research to prepositions of language English which is the target language of the research. So, translation postpositions to prepositions are a main problem which needs to be solved by inflecting the nouns, verbs, and cases (Vibhakti). Depending upon suffix attached to Marathi word there is a change in words position of English words while translating.

2. PROPOSED METHOD

In this paper we are designing inflection rules to POS tags (noun, pronoun, adjective and verb) of Marathi language. As shown in Figure 1, output of tokenization [11] and stemming is provided to morphology analysis. We are taking help of shallow parser to retrieve part of speech tags and its morphology analysis. Morphology analysis describes multiplicity, gender, person, and tense of verb. Before implementing inflection module, we have to define rules for inflection of each POS tag. Generating the appropriate inflection of a word is needed to keep the correct inflection of the word in English [12], [13]. Words can be classified in two types based on the inflection [14], [15]: inflectional words and non-inflectional words. The inflectional words are noun, pronoun, adjective and verb. The non-inflectional words are adverb, preposition, interjection, and conjunction. The words are inflected on the basis of changing gender (masculine, feminine, neuter), multiplicity (singular, plural), tense (present, past, future), person (first, second, third) and case (genitive/possessive case).

Figure 1. Proposed work system architecture

2.1. Inflection module

2.1.1. Noun inflection rules

Declension means inflection of nouns in English language. Noun paradigms inflect for number (singular or plural) but not for gender or case except genitive/possessive case. Generally, nouns are made...
plural by appending -s but this approach fails miserably on many special cases such as: class → classes, story → stories and box → boxes. So, there are some pure suffix-based approaches as given in Table 1.

The suffixes which mostly added in noun plural inflections in English language are: -s, -es, -ves, -ies, -en, -ee, -e, and -ices. Conway [9] has discussed the problem of English plurals and claimed that even at the lexical level; it can be a complex matter to correctly inflect the individual words of a sentence to reflect their number, person, mood, and case. Out of the three noun cases, inflection occurs in only possessive case. Possessive case is used to denote authorship, origin, and ownership. Inflection of nouns in the possessive case is carried out by adding of -'s or -s’ to the end of a noun. Table 2 includes the noun case inflection.

| Terminating Strings of the root word | Plural Inflection | Examples |
|-------------------------------------|-------------------|----------|
| Noun ending with -s/-sh/-ch/-x/o/-us | Adding -es        | Class-Classses |
| Nouns ending with -y & preceded by a consonant | -y replace by -i and add -es | City-Cities |
| Nouns ending with -f/-fe | -f/-fe replace by v and add -es | Wife-Wives |
| Nouns ending with -oo | Replace by -ee | Foot-Feet |
| Nouns ending with -an | Replace by -en | Man-Men |
| Nouns ending with -ix | Replace by -ices | Matrix-Matrices |
| General noun | Adding -s | Book-Books |
| | | Desk-Desks |

| Original word type | Inflection Rule | Examples |
|-------------------|----------------|----------|
| Noun-singular     | Add "'s"       | The boy’s school |
| Noun-plural and ends with 's' | Add "'s" | Boys’ school |
| Noun-plural but does not ends with 's' | Add "'s" | Men’s club |
| Two nouns are closely connected | Add "'s" to second noun | Karim and Salim’s Bakery |
| Nouns telling distance/space/ weight | Add "'s" | I want a day’s leave. |
| | | Shila will be back in a month’s time. |

Postpositions in Marathi occur as prepositions in English [16]. Translating Marathi sentence to English sentence requires conversion of postposition to preposition [17]. For example:

एक मार्ग पुणेरून गोवा जातो → One road goes from Pune to Goa.

In above example the suffix ला comes as a postposition in Marathi whereas the word to come as a preposition in English. Thus, postposition processing involves attachment of preposition before prepositional object. Preposition also undergoes inflections according to the suffix attached to postpositional object. In Marathi there are seven cases, each having its own functional meaning and suffixes. There are different prepositions are used according to suffix attached [18] as given in Table 3.

| Case (Vibhakti) | Marathi suffix | English Suffix |
|----------------|----------------|---------------|
| Nominative     | ना, ना         | to/for        |
| Accusative     | ना, ना         | to/for        |
| Instrumental/Agent | नी     | By/with      |
| Dative         | ना, ना         | for/to        |
| Ablative       | ना, ना         | from          |
| Genitive/Possessive | शा/ शी/ शे   | of/’s         |
| Locative       | ना, ना         | in            |
2.2. Verb inflection

Inflection of verbs in English is called conjugation. The conjugation of a verb gives the different verb forms either by inflection or by combination with parts of other verbs (auxiliary verb) which shows mood, tense, number, and person. English verbs are inflected for tense. A verb lexeme has at most five forms i.e., third person singular form, past tense, progressive particle, perfect or passive particle form. In fact, most verbs have only four forms, because the past tense and the perfect (or passive) particle forms are the same. This is true for all regular verbs. In third person singular there are few variations. In present third person singular, suffix -s is added to both regular and irregular verb. If verbs are ends with a sibilant consonant, then suffix -es is added and if verbs end with -y preceded by a consonant have then -y changed to -i- and then the suffix -es is added. Table 4 includes verb-third person singular form inflection.

There are some variations for the progressive participle. The suffix -ing is added to all verbs to get progressive participle form. Most of the verbs add “ing” to the end without changing the spelling, but for some verb’s spelling in present participle form little bit different according to the specific environment. There are different rules according to verbs ends with as indicated in Table 5.

| Original Word Type | Inflection Rule | Examples |
|--------------------|----------------|----------|
| Verbs ending in -ch/-sh/-x/x/2 | Add -es | Watch → Watches |
| Verbs ending in a consonant y | Changing the y to i and add -es | Try → Tries |

Table 4. Verb-third person singular form inflection

| Original Word Type | Inflection Rule | Examples |
|--------------------|----------------|----------|
| verbs ending in silent “e” | Delete -e & add -ing | Bake → Baking |
| verbs with a short, stressed vowel sound | double the final consonant and add -ing | Swim → Swimming |
| verbs ending in -ie | change -ie to -y and add -ing | Lie → Lying |
| verbs ending in -c | Add -k and -ing | Frolic → Frolicking |

| Original Word Type | Inflection Rule | Examples |
|--------------------|----------------|----------|
| verbs ending in silent “e” | Delete -e & add -ing | Bake → Baking |
| verbs with a short, stressed vowel sound | double the final consonant and add -ing | Swim → Swimming |
| verbs ending in -ie | change -ie to -y and add -ing | Lie → Lying |
| verbs ending in -c | Add -k and -ing | Frolic → Frolicking |

Past tense and past participle form are generated by adding -ed to regular verbs, for example walk-walked. Past tense and past participle form are generated by adding -ed to irregular verbs. There are mainly three types of irregular verbs. First type of Verbs in which all the three forms i.e., base form, past tense and past participle form are the same e.g., put – put – put. Next type of verbs in which second and third forms are the same e.g., sit – sat – sat and third type of verbs in which all three forms are different e.g., drink – drank – drunk. All this indicates that inflection for verbs in English requires more consideration than simply adding the affixes -s, -ing, and -ed. Conjugation of verb by combination with parts of other verbs e.g., auxiliary verb, plays vital role in translation of Marathi to English sentence [17]. Verb tense is decided according to action in a sentence is happening e.g., in the present, future, or past. There are four forms in each tense type. Regular verbs follow a standard rule when conjugated according to tense. Conjugation of the regular verb is indicated in Table 6. $V_1$ stands for base form of verb, $V_2$ for past tense of verb, $V_3$ for progressive participle form of verb and $V_{ing}$ for perfect or passive participle form of verb. For Marathi language type of tense is identified from suffix attached to verb and auxiliary verb used as indicated in Table 7. Table 6 shows rules for verb conjugation in tenses according to suffix attached to Marathi verb.

2.3. Adjective inflection

There are three forms of adjective in English grammar. They are called the degrees of comparisons i.e., positive degree, comparative degree, and superlative degree. Positive degree of an adjective is the adjective in its simple form. Adjectives are inflected to get comparative and superlative forms.

Generally, for superlative and comparative forms, adjectives are generated by adding the suffixes -er and -est to the positive form, respectively. There are some exceptional rules as shown in Table 8. Few adjectives in which comparative and superlative are not formed from positive, for example: Good–Better–Best. It can be concluded that adjective inflection in English is also more complicated than following simple rules of grammar.
2.4. Pronoun inflection

A pronoun is a word that can be substituted for a noun or a noun phrase. Pronoun inflection is similar to noun inflection. The words are inflected on the basis of changing gender i.e., masculine, feminine and neuter; multiplicity i.e., singular, plural; and case i.e., nominative, accusative, and possessive. Pronoun inflection rules are given in Table 9.
### Table 9. Pronoun inflection

| Person | Gender | Nominative | Possessive | Accusative |
|--------|--------|------------|------------|------------|
| First Person | M/F | I | My, mine, me | Our, ours, us |
| Second Person | M/F | You | Your, yours | You |
| Third Person | M | He | His | Their, theirs |
| Third Person | F | She | Her, hers | Her |

3. **RESEARCH METHOD**

While implementation of the inflection, there is a necessity of the information of each word i.e., POS tags, gender tags, tense, multiplicity, and degree, which are identified from Shallow parser developed by IIIT, Hyderabad, India. It provides the system with the morphological analysis of a Marathi sentence. The parser provides output in Shakti standard format [19], [20]. It provides the root word, POS tag, tense, gender, multiplicity, direct or oblique case, suffix, Vibhakti and other details important to identify the role of the word in the sentence. The output is represented as a sequence of abbreviated features, with each attribute is having a fixed position and meaning in sequence. Following eight cases are occurs in morph output: <fsaf = 'root, lcat, gend, num, pers, case, vibh, suff'>.

- Root indicates the root word of the word morphed.
- Lcat gives the lexical category of the word. The values it can take are: Noun (n), pronoun (pn), verb (v), adjective (adj), adverb (adv), and number (num).
- Gend gives the gender of the word in context. The values it can take are male (m), female (f), neutral (n).
- Num gives the impression of the word being singular(sg) or plural(pl) in nature.
- Pers gives whether the speech of the word is in the first person (1), second (2) or the third person (3).
- Case gives whether the noun has a direct or an oblique case depending on the sentence and usage.
- Vibh is the Vibhakti of the word.
- Suff identifies the suffix of the word if it contains any.

For example: पर्यटकांनी NN <fsaf=पर्यटक, n, m, pl, o, ना, ना' name="पर्यटकांना">.

Cases and tenses are identified from word endings as per defined in database, for example as shown in Table 3. Using the above retrieved information, we can apply various inflection rules as discussed in inflection module to get the correct inflection. The inflected words then mapped to the SVO structure of English to generate the correct translation [21]. We have 25,000 Marathi-English sentences from tourism domain from TDIL.

4. **RESULTS AND DISCUSSION**

In testing we considered 7000 sentences of tourism domain and we tested for it. Output of our system i.e., inflected words compared with our reference sentences from data set and it is observed that we got 88-90% of accuracy. While testing word sense disambiguation is also considered [5], [6], [22]-[25]. Four test cases are discussed.

- **Example 1**

  Marathi sentence: शहराचे वातावरण पर्यटकांना आनांद देते. The result is in Table 10.

| Marathi Word | Marathi Word After Suffix Separation | English Lemma | Inflected Word | Rules to Get Inflected |
|--------------|-------------------------------------|---------------|----------------|------------------------|
| शहराचे    | शहर    | city           | City’s         | Noun table, case inflection table |
| वातावरण    | वातावरण | atmosphere     | atmosphere     | -                      |
| पर्यटकांना   | पर्यटक    | tourist        | to tourists    | Case table, Accusative Plural |
| आनांद      | आनांद    | pleasure       | pleasure       | -                      |
| देते        | दे        | give           | gives          | Verb table, 3rd Person singular |
Example 2
Marathi sentence: अमृतसर सुवणग मंदिरचे शहर आहे. The result is in Table 11.

| Marathi Word | Marathi Word After Suffix Separation | English Lemma | Inflected Word | Rules to Get Inflected |
|--------------|-------------------------------------|---------------|----------------|------------------------|
| अमृतसर      | अमृतसर                            | Amritsar      | Amritsar       | -                      |
| सुवणग मंदिर  | सुवणग मंदिर                          | Golden Temple | Gold Temple    | Noun case inflection table |
| शहर         | शहर                                 | city          | city           | -                      |
| आहे          | आहे                                  | is            | is             | Verb table, simple present third person |

Example 3
Marathi sentence: गल महालाची रचना इ.स 1534 बहादुरशाह गुजरातीने केली होती. The result is in Table 12.

| Marathi Word | Marathi Word After Suffix Separation | English Lemma | Inflected Word | Rules to Get Inflected |
|--------------|-------------------------------------|---------------|----------------|------------------------|
| जल महाल     | जल महाल                           | Jal           | of Jal Mahal    | Noun case inflection table |
| रचना         | रचना                               | Construction  | Construction    |                        |
| इ. स बहादुरशाह | इ. स बहादुरशाह                 | AD 1534       | 1534           |                        |
| गुजरातीने    | गुजरातीने                           | Guajarati     | by Guajarati    |                        |
| केली होती   | केली होती                           | done          | done            | Verb table, 3rd person past perfect tense |

Example 4
Marathi sentence: केंद्रपाडा जिल्हा लोहमागाला जोडलेला नाही. The result is in Table 13. In the above all cases, all example gives the inflection of pronouns, nouns, verbs according to the inflection rules discussed and defined in tables from inflection module.

| Marathi Word | Marathi Word After Suffix Separation | English Lemma | Inflected Word | Rules to Get Inflected |
|--------------|-------------------------------------|---------------|----------------|------------------------|
| केंद्रपाडा      | केंद्रपाडा                            | Kendrapara    | Kendrapara     |                        |
| जिल्हा         | जिल्हा                               | district      | district       |                        |
| लोहमागाला     | लोहमागाला                           | Rail route    | to Rail route  | Noun case inflection table, Dative case |
| जोडलेला       | जोडलेला                             | join          | joined         |                        |
| नाही          | नाही + नाही                          | is not        | is not         | Verb inflection, past tense form |

5. CONCLUSION
In the field of machine translation for Indian languages, a great amount of work has been done but for Marathi the research is limited. There is no work done on rule based Marathi to English machine
translation. This paper focuses on the issue of Marathi to English translation with proper inflection with 88-90% accuracy. This paper attempts to provide the detailed description of the rules required for inflecting the words for machine translation from Marathi to English. Ultimately it helps in appropriate translation which was confirmed by the results.

REFERENCES

[1] A. Godase and S. Govilkar, “Machine Translation Development for Indian Languages and its Approaches,” International Journal on Natural Language Computing, vol. 4, no. 2, pp. 55–74, 2015, doi: 10.5121/ijnle.2015.4205.

[2] M. A. S. Khan, S. Amada, and T. Nishino, “Sublexical Translations for low-resource language,” Proceedings of Workshop on Machine Translation and Parsing in Indian Languages (MTPIL-2012), 24th International Conference on Computer Linguistics (Coling12), 2012, pp. 39–52.

[3] Census of India, “Abstract of speakers’ strength of languages and mother tongues–2001”. [Online]. Available: https://censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.htm.

[4] K. T. Blerao, V. S. Wadne, and S. V. Phulari, “A novel approach for Interlingual example-based translation of English to Marathi,” International Journal of Emerging Technology and Advanced Engineering, vol. 4, no 3, pp. 414–417, 2014.

[5] S. Dave, J. Parikh, and P. Bhattacharyya, “Interlingual-based English–Hindi machine translation and language divergence,” Machine Translation, vol. 16, no. 4, pp. 251–304, 2001, doi: 10.1023/A:10219022704523.

[6] N. Patel, B. Patel, R. Parikh, and B. Bhatt, “A Survey: Word Sense Disambiguation,” International Journal of Advances in Research and Foundation in Computer (IJAFC), vol. 2, 2015.

[7] C. Tidke, S. Binayakya, S. Patil, and R. Sugandhi “Inflection Rules for English to Marathi Translation” International Journal of Computer Science and Mobile Computing, IJCSMC, vol. 2, no. 4, pp. 7-18, 2013.

[8] P. Wren and H. Martin, High School English Grammar and Composition. S Chand Publication, 2016.

[9] D. Conway, “An algorithmic approach to English pluralization,” in Proceedings of the Second Annual Perl Conference, SanJose, CA, USA, 1998.

[10] G. R. Navalkar, The Student’s Marathi Grammar. Asian Educational Services, 2001

[11] R. Sugandhi, D. Pisharoty, P. Sidhaye, H. Uipat, S. Wandkar, and R. Khope, “Managing Tokens for Machine Translation from English to Marathi,” International Journal of Engineering, Science and Research (IJESR), vol. 1, no. 3, 2011.

[12] A. R. Joshi and N. Sasikumar, “Constructive approach to teach inflections in Marathi language,” Proceedings of National Conference on Advances in Technology and Recent Developments, Mumbai, India, 2008, pp. 10-16.

[13] S. B. CharugatraTidke and P. Shivani, “Inflection Rules for English to Marathi Machine Translation,” IJCSMC, vol. 2, no. 4, pp. 7–18, 2013.

[14] A. R. Joshi, M. Sasikumar, “Constructive approach to teach inflections in Marathi language,” www.cdacmumbai.in/design/corporate_sitel...pdf...CATIML1.pdf

[15] D. Walker and R. Amsler, “The Use of Machine Readable Dictionaries in Sublanguage Analysis,” in Analyzing Language in Restricted Domains, Grishman and Kittredge (eds), LEA Press, pp. 69-83, 1986.

[16] G. V. Garje, G. K. Kharate, and H. Kulkarni, “Transmuter: An Approach to Rule-based English to Marathi Machine Translation,” International Journal of Computer Applications, vol. 98, no. 21, pp. 33–37, 2014.

[17] N. G. Kharate and V. H. Patil, “Challenges in Rule Based Machine Translation from Marathi to English,” 5th International Conference on Advances in Computer Science and Information Technology (ACSTY-2019), 2019, pp. 45-54, doi: 10.5121/csit.2019.91005.

[18] N. G. Kharate and D. H. Patil, “Handling Challenges in Rule Based Machine Translation from Marathi to English,” International Journal on Natural Language Computing, vol. 8, no. 4, pp. 39–49, 2019, doi: 10.5121/ijnlc.2019.8404 39.

[19] B. Kulkarni, P. D. Deshmukh, and K. V. Kale, “Syntactic and Structural Divergence in English-to-Marathi Machine Translation,” IEEE 2013 International Symposium on Computational and Business Intelligence, New Delhi, 2013, pp. 191-194, doi: 10.1109/ISCBI.2013.46.

[20] A. Bharati, R. Sangal, and D. M Sharma, “SSF: Shakti Standard Format Guide,” 2007.

[21] A. Adapanawar, A. Garje, P. Thakare, P. Gundawar, and P. Kulkarni, “Rule based English to Marathi translation of Assertive sentence,” International Journal of Scientific & Engineering Research, vol. 4, no. 5, 2013.

[22] W. Black et al., “Introducing the Arabic WordNet Project,” Proceedings of the 3rd Global Wordnet Conference, 2006, pp. 295-300.

[23] M. Gupta, S. Yadav, S. Sharma, and S. Yadav, “Word Sense Disambiguation Using HindiWordNet and Lesk Approach,” IPSAI International Journal of Computer Science (IJCS), vol. 1, no. 6, pp. 12-17, 2013.

[24] R. Sinha and R. Mihalcea, “Unsupervised Graph-based Word Sense Disambiguation Using Measures of Word Semantic Similarity,” International Conference on Semantic Computing (ISC 2007), 2007, pp. 363-369, doi: 10.1109/ISCSC.2007.87.

[25] K. Neeraja and B. P. Rani, “Approaches for Word Sense Disambiguation: Current State of The Art,” International Journal of Electronics Communication and Computer Engineering, vol. 6, no. 5, pp. 197-201, 2015.
BIOGRAPHIES OF AUTHORS

Ms. Namrata G Kharate is pursuing Ph.D. in Computer Department with specialization in natural language processing from Savitribai Phule, Pune University. She received her BE from Savitribai Phule Pune University & M.E from Savitribai Phule Pune University. Working as an Assistant Professor in Vishwakarma institute of information and technology, Pune.

Dr. Varsha H. Patil is Professor, University of Pune, and Head, Computer Science Department, at Matoshri College of Engineering & Research Centre, Nashik. She is also serving as the Vice Principal of the same institute. She has received her Ph.D. in Computer Engineering from B.V.CO.E, Bharati Vidyapeeth Pune. She has received her M.E in Computer Engineering from COEP, Pune. She has more than 20 years of teaching experience and several research papers, published in national and international journals of repute, to her credit. She is a member of the Board of Studies of Computer Engineering at the University of Pune.