



A Framework for Managing Verb Phrase Effective and Easy English-Hindi Machine Translation

VIVEK DUBEY and H.R. SHARMA

Computer Science & Engineering , SSCET, Bhilai, CG, (India)
Computer Science & Engineering, RCET, Raipur, CG, (India)

(Received: February 12, 2012; Accepted: June 04, 2012)

ABSTRACT

Automatic Machine Translations from one to another language have been the subject of great attention of computational linguistics for many years. In English- Hindi Machine Translation, verb tuning is a vital operation. Present paper is an approach to describe easy English-Hindi verb phrase mapping. This work results satisfactory in Machine Translation over type of English sentences. It is observed that carefully analysis of tense, mood and aspect of English sentences leads to proper translation in Hindi through the efficient use of verb phrase mapping.

Keywords: Machine Translation, Syntax, Semantic, Ambiguity, Free-Order, Mapping, English-Hindi Grammar (noun, pronoun, etc), English-Hindi Sentence Structure, Tagging.

INTRODUCTION

English and Hindi languages are of two different grammar structure suffered with ambiguities of various types such as lexical ambiguity, semantic ambiguity, syntactic ambiguity¹ during English-Hindi Machine Translation (EHMT) due to verb morphology. Hindi verb phrase depends on tense & form of the sentence and gender (masculine, feminine, or neuter), person (First, Second or Third), number of subject (Singular or Plural) and case (possessive, subject, or object). Consider a simple sentence, for example, "He reads a book" in English, but it turn into in Hindi as, "वह एक किताब पढ़ता है". Now again consider similar structure English sentence as, "She reads a book", and in Hindi, it is as "वह एक किताब पढ़ती है". In both examples,

verb "reads" is same in English, but in Hindi, verbs are different as "पढ़ता है" and "पढ़ती है". For Machine Translation (MT), English grammar is categorized for verb phrase structure², which involves different combinations of tenses (present, past, future), aspects and moods (declarative/assertive, yes/no question, wh-question, imperative/instructive). However with respect to English-Hindi translation, there is also ambiguity, noticed³ in understanding the sense of the sentence; rather, the difficulty is in deciding the correct structure of the Hindi translation.

Consider two sentences, in English, "Vedang has a car" and "Vedang has excitement", but in Hindi, both has a different sense as "वेदांग(Vedang) के पास कार है (has) एक(a) गाड़ी(car)" and

“वेदांग(Vedang) कौत्रहे (has) आवेश(excitement)”. In such English sentence, verb may be modified as “के~पास~हे” or “को~हे”. In English-Hindi Machine Translation, such consideration reading verb phrases are an essential operation.

Hindi is free order language but the orders of words are almost fixed in English language. Therefore, grammatically pattern can easily be extracted from tagged English sentence using rule based, example based or statistics based approach. A shallow analysis for English⁴ has been described for identifying the verb phrases in Machine Translation system from English to Bengali.

Words are set of characters or simply a string. All words are categorized using Parts-Of-Speech's keywords, like DT, PRP, RB, VB, NN etc, known as Parts of Speech tag or POS Tag. These tags are total 45 according to Penn Treebank tagset⁵, 61-tag C5 used by Lancaster UCREL project's CLAWS tagger to tag the British National Corpus (BNC)⁶ and 146-tag C7 tagset according to⁷. However, in this paper, we have been using small 45-tag Penn Treebank tagset. We have developed verb phrase analyzer module of tagged English sentence using rule based and tested in our developed Rule-Based Machine Translation (RBMT) system. We used Genia Tagger for tagging English Sentences. The rest of the paper is organized as follows. Section 2-5 discusses verb “to be”, present, past, and future sentences structure respectively. Section 6 introduces syntax of verb phrase. Section 7 presents system framework and lastly, section 8 gives conclusion and further work.

Verb ‘To Be’ Structure

Verb ‘To be’ (Is, Are, Am, Was, Were), Has/Have/Had and Do/Does/Did has been used as finite verb in sentence. Such sentences are structured as affirmative, negative, and interrogative.

Verb ‘is/VBZ’ in affirmative, as the VP, “[He/She is]”, is modified as ‘है’ and in negative, as the VP [He/She/NN is RB] is modified as ‘नहीं है’ and in interrogative as ‘क्या > है’, for grammar pattern ‘Is’. Similarly, rules are for ‘are/VBP’ and ‘am/VBP’

Verb ‘was/VBD’ in affirmative and negative sentence for VP [I/He/She/NN was] is modified as ‘+था / थी’ and in interrogative as ‘क्या > था / थी’ for VP [Was] and similarly for ‘were/VBD’.

With verb ‘have’ and ‘has’, there are always ambiguity⁸. However, Verb ‘has’ in affirmative and negative sentence is modified as ‘+पास~रखता / ती है’ and in interrogative as ‘क्या> पास~रखता / ती है’. Verb ‘have’ in affirmative and negative sentence is modified as ‘+पास~रखते हो / हैं’ and in interrogative as ‘क्या > पास~रखते हो / हैं’. Verb ‘had’ in affirmative and negative sentence is modified as ‘+पास~रखता था / थी / थे’ and in interrogative as ‘क्या > पास~रखता था / थी / थे’.

Verb ‘do/does/did’ usually are used as only affirmative sentence and word ‘no’ sometime be used as negative sentence. Verb ‘does/VBZ’ is modified as ‘करता है’ for VB [He does] and verb ‘do/VBP’ as [You do] = करते हो and verb ‘did/VBD’ as [They did] = करते थे.

Consider English affirmative sentence “He is a student”. Verb phrase analyzer initially

Table 1. Verb "To be" (is/ are/ am)

Affirmative Sentence (Subject+ is/are/am+ Complement)			
English Sentence	Steps	Verb Phrase Analyzer	Verb Phrase in Hindi Sentence
He is a student.	First	1. [He/PRP] ← NP 2. [is/VBZ] ← VP 3. [a/DT student/NN] ← NP	---
	Second	1. [He/PRP] 2. [He/PRP is/VBZ] 3. [a/DT student/NN]	1. वह 2. [He is] = है 3. एक विद्यार्थी

partitions tagged English sentence into sequence of semantically related words as Noun Phrase (NP), Verb Phrase (VP), etc and then modifies verb phrase as “subject+verb”, illustrated in Table 1. Similarly, other sentences structured as negative and interrogative are approached.

Present Tense Structure

There are four categories of present tense structure as simple/indefinite, continuous, perfect, and perfect continuous and English verbs are modified as following:

- A. Simple present Structure = ता है / ती है / ता हूँ / ते हैं / ते हो
For example, [I/PRP read/VBP] \equiv [I VBP] = पढ़ ता / ती हूँ
- B. Present Continuous Structure = रहा है / रही है / रहे हैं / रहे हो
For example, [I/PRP am/VBP going/VBG] \equiv [I am VBG] = जा रहा / ही हूँ

- C. Present Perfect Structure = चुका है / चुकी है / चुके हैं / चुका हो / चुका हूँ
For Example, [I/PRP have/VBP read/VBN] \equiv [I have VBN] = पढ़ चुका / की हूँ
- D. Present Perfect Continuous Structure = से रहा है / से रही है / से रहे हो / से रहा हूँ Or ते रहा है / ते रही है / ते रहे हो / ते रहा हूँ
For example, [I/PRP have/VBP been/VBN working/VBG] \equiv [I have been VBG] = काय- करता रहा / ही हूँ

Consider simple present structure of negative sentence “He does not read a book”. Verb phrase analyzer initially partitions tagged English sentence into sequence of semantically related words and then modifies verb phrase as “subject+verb”, illustrated in Table 2. Similarly, other sentences structured of present tense as continuous, perfect, and perfect continuous are followed.

Table 2: Verb as Simple Present Tense

Negative Sentence [Subject(S/P)+ VB/VBP/VE Z+ not/RB+ Object]			
English Sentence	Steps	Verb Phrase Analyzer	Verb Phrase in Hindi Sentence
He does not read a book.	First	1. [He/PRP] \leftarrow NP 2. [does/VBP not/RB read/VB] \leftarrow VB 3. [a/DT book/NN] \leftarrow NP	--
	Second	1. [He/PRP] 2. [He/PRP does/VBP not/RB read/VB] 3. [a/DT book/NN]	1. वह 2. [He does RB VB]— नहीं पढ़ता है 3. एक किताब

Past Tense Structure

There are four categories of past tense structure as simple/indefinite, continuous, perfect, and perfect continuous and English verbs are shown in Table 3 and described below.

- A. Simple Past Structure = गया था / गयी थी / गये थे
For example, [I/PRP help/VBD] \equiv [I VBD] = सहायता करा था / थी
- B. Past Continuous Structure = रहा था / रही थी / रहे थे
For example, [I/PRP was/VBD read/VBG] \equiv [I was VBG]= पढ़, रहा था / थी
- C. Past Perfect Structure = चुका था / चुकी थी / चुके थे
For example, [He/PRP had/VBD left/VBN]

- \equiv [He had VBN] = छोड़, चुका था
- D. Past Perfect Continuous Structure = ता रहा था / ता रही थी / ता रहे थे
For example, [I/PRP had/VBD been/VBN working/VBG] \equiv [I had been VBG] = काय- करता रहा था / थी

Consider simple past continuous structure of affirmative sentence “I was reading my book”. Verb phrase analyzer initially partitions tagged English sentence into sequence of semantically related words and then modifies verb phrase as “subject+verb”, illustrated in Table 3.

Similarly, other sentences structured of past tense as present, perfect, and perfect continuous are ruled.

Future Tense Structure

Table 4: Verb as Past Continuous Tense

Affirmative Sentence [Subject(S/P)+was/were+VB G+Object]			
English Sentence	Steps	Verb Phrase Analyzer	Verb Phrase in Hindi Sentence
I was reading my book	First	1. [I/PRP] 2. [was/VBD reading/VB G] 3. [my/PRP\$ book/NN]	--
	Second	1. [I/PRP] 2. [I/PRP was/VBD reading/VB G] 3. [my/PRP\$ book/NN]	1. मैं 2. [I was VB G]- पर उस का / की 3. मेरी किताब

There are four categories of future tense structure as simple/indefinite, continuous, perfect, and perfect continuous and English verbs are modified as illustrated in Table 4 and described below.

- A. Simple Future Structure = gaa /gal /gao
For example, [I/PRP shall/MD help/VB] ≡ [MD VB] = सहायता करेगा/गी
- B. Future Continuous Structure = ता रहूँगा/ ती रहूँगा/ ते रहेंगे/ ते रहोगे
For example, [I/PRP shall/MD be/VB helping/VBG] ≡ [I shall be VBG] = सहायता करता /ती रहूँगा /गी
- C. Future Perfect Structure = चुका हूँगा / चुका होगा / चुकी

होगी / चुके होंगे

For example, [He/PRP shall/MD not/RB have/VB reached/VBN] ≡ [He MD RB have VBN] = नहीं पहुँच चुका होगा

- D. Future Perfect Continuous Structure = ता रहा होऊँगा/ता रहा होगा/ती रहा होगी / ते रहे होंगे
For example, [I/PRP shall/MD have/VB been/VBN reading/VBG] ≡ [IMD have been VBN]= पढ़ता/ती रहा/ही हूँगा/गी

Consider Future Perfect Tense structure of affirmative sentence “I was reading my book”. Verb phrase analyzer initially partitions tagged English sentence into sequence of semantically

Table 4: Verb as Future Perfect Tense

Interrogative Sentence [Shall/Will+ Subject(S/P)+have+VBD/VBN+Object]			
English Sentence	Steps	Verb Phrase Analyzer	Verb Phrase in Hindi Sentence
Shall they have finished their work?	First	1. [Shall/MD] 2. [they/PRP] 3. [have/VBP finished/VBN] 4. [their/PRP\$ work/NN]	--
	Second	1. [Shall/MD] 2. [they/PRP] 3. [have/VBP finished/VBN] 4. [their/PRP\$ work/NN] [their/PRP\$ work/NN]	1. [Shall]- क्या 2. वे 3. [Shall they have VBN]- समाप्त का चुकी है 4. उनका कार्य

related words and then modifies verb phrase as "subject+verb", illustrated in Table 4. Similarly, other sentences structured of past tense as present, perfect, and perfect continuous are ruled.

Syntax of Verb Phrase

Now on the basis of above explanation, Verb Phrase (VP), in general, can be concluded by following syntax.

S → NP VP

VP → Verb NP

VP → Verb NP PP

VP → Verb PP

Verb → (not|no) (is) (are) (am) (was) (were)

(has) (have) (had)

(do) (does) (did) (been)

(Main Verb | s | es | ed | ing)

System Framework

A prototype is developed for English-

Hindi Machine Translation using rule-based approach. Verb phrases of English-Hindi language pattern are organized as shown in Table 5 and parallel English-Hindi word example are mapped as shown in Table 6. We download and update dictionary syntactically and semantically way available at IIIT Hyderabad web site⁹ for English-Hindi words. In functioning, initially system obtain tagged sentence of input English sentence using Geniatagger¹⁰ and then retrieve matched words syntactically and semantically. Next, based on retrieved word, system adapt input sentence using adaptation operations like constituent Word Copy (WC) and Suffix Addition (SA) and lastly generate target sentence. The entire procedure is illustrated, in Fig. 1, using Input English Sentence "He punished his brother". Number of unknown sentences are translated by said system and evaluated with translated by Translation Support System¹¹ and Google Indic system¹².

Table 5: Illustration of English-Hindi Verb Suffix

English Sentence Structure	Hindi Suffix
I VB	ना/नी हूँ
He VBZ	ना है
She is VBG	रही है
He VBD	था
He has VBN	चुका है

Table 6: Illustration of English-Hindi Words

English Words	Hindi Word	Mapping
.....
abbreviate/VBP	संक्षिप्त~कर+	21 1:1
abbreviate/VB	संक्षिप्त~कर+	22 1:1
abbreviated/VBD	संक्षिप्त~कर+	23 1:1
abbreviated/VBN	संक्षिप्त~कर+	24 1:1
abbreviates/VBZ	संक्षिप्त~कर+	24 1:1
abbreviating/VBG	संक्षिप्त~कर+	25 1:1
abbreviation/NN	संक्षिप्ति	26 1:1
.....

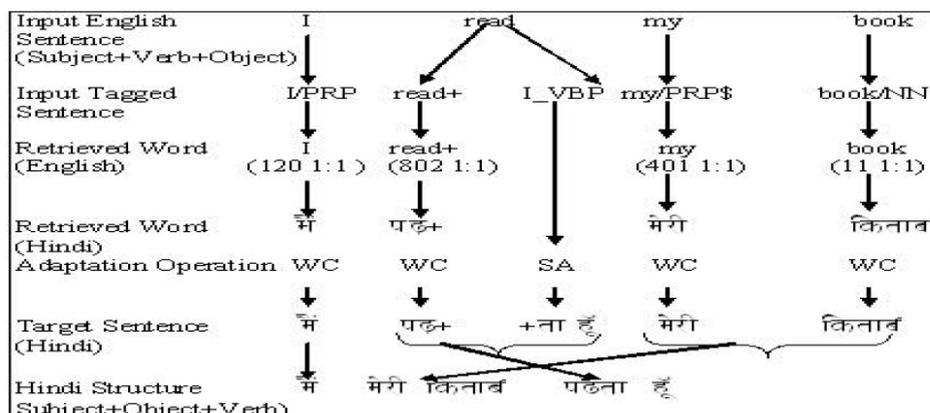


Fig. 1: Illustrate Rule Based Machine Translation

CONCLUSION

Present works expose the implementation of verb phrases in English-Hindi Machine Translation using Rule-Based approach. The process involves segmentation of tagged English sentence, identification and modification of Verb Phrase. It has been observed that such approach

helps a lot smooth English-Hindi Machine Translation. Other phrases like noun phrases, adjective phrases, adverbial phrases, and prepositional phrases can be easily developed and implemented in this way and over all efficiency of the machine translation system may be improved towards target sentence.

REFERENCES

- Bonnie J. Dorr, Pamela A Jordan, and JohnWBenoit. "A survey of current research in machine translation", In M. Zerkowitz, editor, *Advances in Computers*, volume 49, pages 1- 68. Academic Press, London, 1999.
- Quick, R. and Greenbaum S.: 1976, "A University Grammar of English", English Language Book Socitey, Longman.
- Shailly Goyal, Deepa Gupta, and Niladri Chatterjee. "A study of Hindi translation pattern for English sentences with \have as the main verb", In *ISTRANS-2004*, pages 46-51, New Delhi, India, 2004.
- Sudip Kumar Naskar and Sivaji Bandyopadhyay, "A Phrasal EBMT System for Translating English to Bengali", In the *Proceedings of MT SUMMIT X*; September 13-15, 2005, Phuket, Thailand.
- Marcuss, M.,B.Santorini, and M.Marcinkiewicz,1993, "Building a large annotated corpus of English:The Penn Treebank",*Computational Linguistics* 19, 2.
- Garside, R., Leech, G., and McEnergy, "Corpus Annotation", Longman London, and New York, 1997.
- Leech. G., Garside, R. and Bryant, M., "Claws4: The tagging of the British National Corpus", In *Coling-94*, Kyoto, pp. 622-628, 1994
- Niladri Chatterjee,Shailly Goyal and Anjali Naithani, *Workshop Modern Approaches in Translation Technologies 2005-Borovets*, Bulgaria, pp.19-25, 2005.
- WWW1 (http://ltrc.iit.ac.in/onlineServices/Dictionaries/DicDict_Frame.html)
- WWW2(http://www-tsujii.is.s.u-tokyo.ac.jp/GE_NIA/tagger/)
- WWW3 (<http://tdil.mit.gov.in/download/mat.htm>)
- WWW4 (<http://translate.google.com/#>)