

Impact of Systematic and Efficient Adaptation Methodology in English-Hindi EBMT

Vivek Dubey

Associate Professor

Computer Sc. & Engg. Dept. SSCET, Bhilai, CG, India

H.R.Sharma

Professor

Computer Sc. & Engg. Dept. RCET, Bhilai, CG, India

ABSTRACT

It is no of doubt; the output of Example Based Machine Translation is always consequence of adaptation procedure over retrieved examples. There are always scopes for refining the resources structure of retriever examples. In this paper, we will describe an efficient approach to organize English-Hindi parallel examples taken from online available machine translation system instead of any linguistic database and its implementation towards English-Hindi Example-Based Machine Translation.

Keywords

Morpheme, Word-Processing, Tagging, English-Hindi Grammar (noun, pronoun, etc), Syntax, Sentence Structure

1. INTRODUCTION

For generating the quality output of Example Based Machine Translation (EBMT), a systematic organization and an efficient implementation of English-Hindi parallel examples are always needed for adaptation retrieved examples. Various schemes have been directed with adaptation aspect likes based on phrase and word modification [1], difference in segments of input sentence and example sentence [2], full and partial case [3], case-based reasoning [4], share lexical items with the input [5], the structural similarity between the input and example sentence [6], and many other are found in literature, e.g. GEBMT [7-10], EDGAR [11] and TTL [12].

However, ten adaptation operations [13] are identified for modifying retrieved examples. We select only five adaptation operations as constituent Word Replacement (WR), constituent Word Addition (WA), constituent Word Deletion (WD), Suffix Addition (SA), and Copy (CP) and implement adaptation procedure on ranked retrieved phrase-word examples first matched first tune-up basis.

We also developed a prototype English-Hindi EBMT and English-Hindi parallel examples considering tense and form of the verb, Subject/Object functional slot, wh-family interrogative sentences and kind of sentences like affirmative, negative, interrogative and negative interrogative sentences and tested over unknown sentences. The result of proposed system is very much promising. In this paper, the second section we will describe organization of English-Hindi parallel examples as corpus, third section will demonstrate the implementation of adaptation in English-Hindi machine translation and forth section will concludes.

2. ORGANIZATION OF ENGLISH-HINDI EXAMPLES

Initially, we select sentences from open dictionary available at IIT Hyderabad web site [14] and generate tagging sentences using geniatagger [15] and use as English examples. Later we translate these English sentences in Hindi using translation support system [16] and google indic online translation software [17] and use as parallel Hindi examples.

In following section, we will explore organization of verb morphology, adjective morphology and possessive morphology used in parallel English-Hindi example sentences.

2.1 Verb Morphology

An English-Hindi verb morphological variations depend on pronoun (1st person (I/PRP), 2nd person (You/PRP) and 3rd person (He/PRP, She/PRP)), tense(Present, Past, Future) and form of verb (base (VB), 3sg pres (VBZ), gerund (VBG), past (VBD), past participle (VBN)).

Consider a case for Same Tense (ST), Same Verb (SV) and Different Person Pronoun (DPP). Let the input sentence is “She goes” and retrieved example sentence is “He goes”. After Part-Of-Speech (POS) tagging, Input Sentence becomes as “She/PRP go+/She_VBZ and retrieved example sentence becomes as “He/PRP go+/He_VBZ. So, for adaptation process, operations WR (for She as in Hindi vah), CP (for go as in Hindi jaa) and SA (for She_VBZ as in Hindi tI hO) are applied as illustrated in Figure 1.

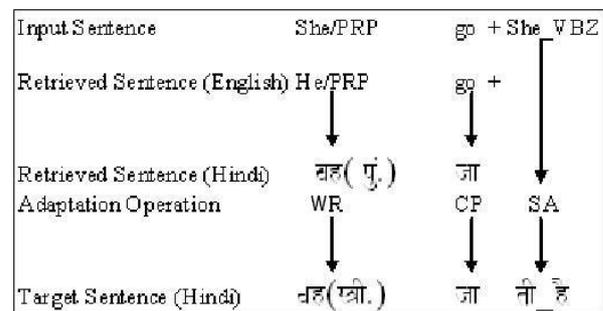


Figure 1. Illustration of Verb Morphology in English Sentence

Since we are not using operation SR, so by default operation SA will be invoked in every suffix case. Similarly for other combinations of tense, its form and person pronoun, different operation combinations are summarized as illustrated in Table-1.

Table 1. Illustration of Other Verb Morphology in English Sentences

Input Sentence (English)	Sentence Type	Retrieved Sentence (English)	Operations
She goes	ST+SV+SPP	She goes	CP CP SA
	ST+SV+DPP*	He goes	WR CP SA
She goes	ST+DV+SPP	She eats	CP WR SA
	ST+DV+DPP	He eats	WR WR SA
She is going	DT*+SV+SPP	She goes	CP WA WR SA
	DT+SV+DPP	He went	WR WA WR SA
She went	DT+DV+SPP	She is eating	CP WD WR SA
	DT+DV+DPP	He has taken	WR WD WR SA

*DT (Different Tense), DV (Different Verb), DPP (Different Personal Pronoun)

Verb morphology for such cases are organized in three columns, 1st column is used for English Examples, 2nd column is used for Hindi Examples and 3rd column is used for mapping English-Hindi Examples as illustrated in Table-2 and Hindi Suffix are organized in two columns, 1st column is used for English Sentence Structure and 2nd column is used as equivalent in Hindi Suffix as illustrated in Table-3.

Table 2. Illustration of English-Hindi Examples

English Example	Hindi Example	English-Hindi Mapping
I/PRP go/VB	मैं जा+	1:1 2:2
He/PRP goes/VBP	वह(पुं.) जा+	1:1 2:2
She/PRP is/VBZ going/VBG	वह(स्त्री.) + जा+	1:1 2:2 3:3
He/PRP went/VBD	वह(पुं.) जा+	1:1 2:2
He/PRP has/VBZ goen/VBN	वह (पुं.) + जा+	1:1 2:2 3:3

Table 3. Illustration of English-Hindi Verb Suffix

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

2.2 Adjective Morphology

English-Hindi adjective morphological variations depend on its next word's POS as number (singular (S) and plural (P)) of noun's gender (masculine singular/plural noun (NNM/NNSM) and feminine singular/plural noun (NNF/NNSF)). Let us consider segment of English input sentence as “good boy” and retrieved chunk as “good girl”. After Part-Of-Speech (POS) tagging, Input Sentence becomes as “good/JJ_NNM boy/NNM and retrieved example sentence becomes as “good/JJ_NNF girl/NNF. So, for adaptation process, operations CP (for good as in Hindi AcC), SA (for JJ_NNSM as in Hindi a) and WR (for boy as in Hindi laDka) are applied as illustrated in Figure 2.

Input Sentence	good/JJ + NNM	boy/NNM
Retrieved Sentence (English)	good/JJ	girl/NNF
Retrieved Sentence (Hindi)	अच्छ	लडकी
Adaptation Operation	CP SA	wR
Target Sentence (Hindi)	अच्छ	लडकी

Figure 2. Illustration of Adjective English Word

The English-Hindi examples with adjective word are organized in three columns table as illustrated in Table 4 and organization of its suffixes is illustrated in Table 5.

Table 4. Illustration of English-Hindi Adjective Examples

English Example	Hindi Example	English-Hindi Mapping
a/DT good/JJ girl/NNF	एक अच्छा लडकी	1:1 2:2
good/JJ sister/NNF	अच्छा बहन	1:1 2:2
boy/NNM	लडका	1:1

Table 5. Illustration of English-Hindi Adjective Suffix

English Sentence Structure	Hindi Suffix
DT JJ NNM	ए
DT JJ NNF	ई
DT JJ NNSM	ए

2.3 Possessive Case Morphology

The possessive case morphological variations depend on pronoun (pronoun (1st person (my/PRP\$), 2nd person (your/PRP\$) and 3rd person (his/PRP\$, her/PRP\$)) and its next adjective (JJ) or/and noun as my brother or my good brother. Consider, for input English sentence's segment as “my sister”, the output in Hindi is as “maor_+_ I bahna”. Table 6 illustrates required operations for retrieved English sentence as “my sister”, “my brother”, “his sister” and Table 7 and Table 8 illustrates English-Hindi examples and English-Hindi possessive case suffix respectively.

Table 6. Illustration of Possessive Case Morphology

Input Sentence (English)	Sentence Type	Retrieved Sentence (English)	Operations
my sister	PRP\$+NNF	my sister	CP SA CP
my sister	PRP\$+NN	my book	CP SA WR
my sister	PRP\$+NNSF	his sisters	WR SA WR

Table 7. Illustration of English-Hindi Possessive Case

English Example	Hindi Example	English-Hindi Mapping
my sister	मेरा बहन	1:1 2:2
my book	मेरा किताब	1:1 2:2
his sisters	बहका	1:1

Table 8. Illustration of English-Hindi Possessive Case Suffix

English Sentence Structure	Hindi Suffix
my NNF	†
my NNM	†
my NNSF	†
my NNSM	†

3. IMPLEMENTATION

In example based English-Hindi machine translation, above described methodology of adaptation is effective and reliable. Consider a English input sentence as “She punished his brother” and its tagging sentence as “She/PRP punished/VBD his/PRP\$ brother/NNM” and retrieved English sentences as shown in table 9. Adaptation process for target sentence in Hindi is illustrated in Figure 2.

Table 9. Retrieved English-Hindi examples

English Example	Hindi Example	English-Hindi Mapping
He/PRP eats/VBZ food/NN	वह खा+ना है भोजन	1* 1:1* 2:2 3:3
His/PRP\$ brother/NN is/VBZ a/DT good/JJ boy/NN	उसका भाई + एक अच्छा लड़का	2 1:1 2:2 3:3 4:4 5:5 6:6
She/PRP is/VBZ punishing/VBG students/NNS	वह + दण्ड दे+रही है विद्यार्थी	3 1:1 2:2 3:3 4:4
There/EX are/VBP punished/VBN students/NNS	वहाँ + कण्डन विद्यार्थी	4 1:1 2:2 3:3 4:4
Teacher/NN punished/VBD them/PRP	गुरु दण्ड दि+या उन्हें	5 1:1 2:2 3:3

1* Example Number, 1:1* EngWordNo:HinWordNo

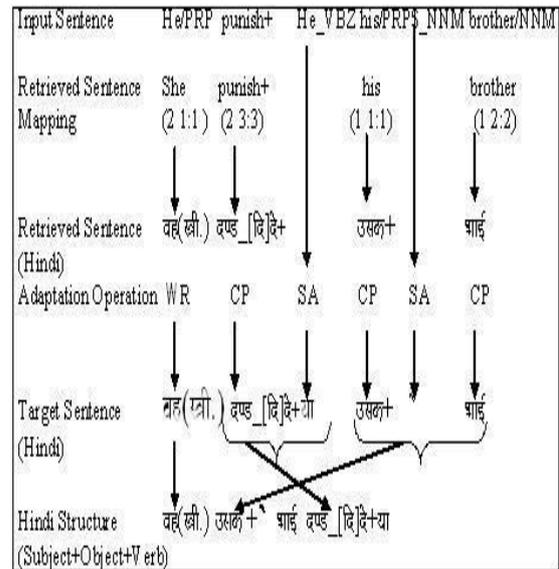


Figure 2. Illustration of Adjective English Word

4. CONCLUSIONS

A novel adaptation scheme EBMT has been developed and proposed a systematic organization along with syntactically and semantically implementation model of the said system. Parallel examples has been also designed using online English-Hindi translated TSS and Google Indic system instead of any linguistic knowledge and the comparative evaluation performance has been carried out. Proposed system is resulting satisfactory.

5. REFERENCES

- [1] Veale, T. and Way, A. - “Gaijin: A template-driven bootstrapping approach to Example-Based Machine Translation, International Conference, Recent Advances in Natural Language Processing”, Tzigov Chark, Bulgaria., pp. 239-244, 1997
- [2] Shiri, S., Bond, F. and Takhshi, Y. - “A Hybrid Rule and Example-Based Method for Machine Translation”, Proceedings of the 4th Natural Language Processing Pacific Rim Symposium: NLP97, Phuket, Thailand, pp. 49-54, 1997.
- [3] Collins, B. - “Example-Based Machine Translation: an Adaptation-Guided Retrieval Approach”, PhD thesis, University of Dublin, Trinity College, 1998.
- [4] Somers, H. - “EBMT seen as case-based reasoning”, MT Summit VIII Workshop on Example-Based Machine Translation, Santiago de Compostela, Spain, pp. 56-65, 1997.
- [5] McTait, K. - “Translation Pattern Extraction and Recombination for Example-Based Machine Translation”, PhD thesis, Centre for Computational Linguistics Department of Language Engineering, UMIST, 2001.
- [6] Jain, R. - “HEBMT: A Hybrid Example-Based Approach for Machine Translation (Design and Implementation for Hindi to English)”, PhD thesis, I.I.T. Kanpur, 1995.
- [7] Brown, R. D. - “Example-Based Machine Translation in the pangloss system”, Proceedings of the 16th International Conference on Computational Linguistics

- (COLING-96), Copenhagen, Denmark, pp. 169-174, 1996.
- [8] Brown, R. D. - "Adding linguistic knowledge to a lexical Example-Based Translation System", Proceedings of the Eighth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99), Chester, UK, pp. 22-32, 1999.
- [9] Brown, R. D. - Automated generalization of translation examples, Proceedings of the Eighteenth International Conference on Computational Linguistics, pp. 125-31, 2000
- [10] Brown, R. D., "Transfer-rule induction for Example-Based Translation", Proceedings of the MT Summit VIII Workshop on Example-Based Machine Translation, Santiago de Compostela, Spain, pp. 1-11, 2001.
- [11] Carl, M. and Hansen, S., "Linking translation memories with Example-Based Machine Translation", Proceedings of Machine Translation Summit VII, Singapore, pp. 617-624, 1999.
- [12] Guvenir, H. A. and Cicekli, I., "Learning translation templates from examples", Information System 23, 353-363, 1998.
- [13] Gupta Deepa and Chatterjee N, "A Systematic Adaptation Scheme for English-Hindi Example-Based Machine Translation", In Proceeding of the STRANS-2002, Indian Institute of Technology Kanpur, 2002.
- [14] WWW1 (http://ltrc.iiit.ac.in/onlineServices/Dictionaries/DicDict_Frame.html)
- [15] WWW2 (<http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/tagger/>)
- [16] WWW3 (<http://tdil.mit.gov.in/download/mat.htm>)
- [17] WWW4 (<http://translate.google.com/#>)