

Resolving Pattern Ambiguity for English to Hindi Machine Translation using WordNet

N. Chatterjee, Shailly Goyal
&
A. Naithani



DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY DELHI
NEW DELHI – 110016
INDIA

Introduction



One major difficulty with any Translation System is : *Ambiguity*.

Semantic Ambiguity is talked about more often.

But a more interesting and potentially more difficult one is: *Pattern Ambiguity*.

What is Pattern Ambiguity?



A common notion about translation is:

Structurally similar sentences of the source language give rise to sentences that are structurally similar in the target language too.

- But this does not hold good all the time.
- This leads to the ambiguity regarding:

What is the likely pattern of the translation of a given input?

Translation Patterns of Different English Verbs to Hindi



- In English, a single verb is used to convey different senses.
- However, almost for each of these senses, a specific verb exists in Hindi.

Sentences	Translation of Verb
They run an N.G.O.	<i>chalaanaa</i>
We ran the ad three times	<i>prakaashit karnaa</i>
The river runs into the sea	<i>milnaa</i>
Wax runs in sun	<i>galnaa</i>

- Similarly for *be, go, take, let, give* etc.

Translation Patterns of Different English Verbs to Hindi



- The use of the appropriate Hindi verb can be determined by identifying the sense in which the English verb is used.
- Resolving pattern ambiguity for these verbs is relatively simple.

Case Study: “Have/Had” Sentences



- Most interesting observation in this regard can be made with respect to the “have/had” sentences, i.e., sentences where the main verb is a declension of “have” verb.
- Although the number of possible senses for “have” is relatively less (only 19, as per WordNet 2.0), we have obtained as many as **eleven** translation patterns .

Case Study: “Have/Had” Sentences



- Hindi does not have any equivalent verb.
Typically, translation of “have/had” sentences involves:
 - Verb “*honaa*”, i.e., “to be”.
 - Subject is transformed to a possessive case (with the help of appropriate postpositions).

Example



- Mohan has a good brain
~ Mohan kaa dimaag achchhaa hai
(Mohan's) (brain) (good) (is)

- Mohan has a good pen
~ Mohan ke paas ek achchhii kalam hai
(Mohan) (near to) (a) (good) (pen) (is)

However many different patterns are found!!!

Translation Pattern: P1



Genitive case ending (viz. *kaa*, *ke* or *kii*) is used to convey the sense of the "have" verb.

- The school has good reputation ~

vidyaalay kii achchhii saakh hai
(School) (of) (good) (reputation) (is)

- Mohan had an assistant ~

Mohan kaa ek sahaayak thaa
(Mohan) (of) (a) (assistant) (was)

Translation Pattern: P2



- A genitive case ending is used with the following variations:

English Sentence → Hindi Translation
Object → Subject
Pre-modifying adjective → Subjective complement
Subject → Possessive case of the subject

- Gita has beautiful hair ~

Gita ke baal sundar hain
(Gita) (of) (hair) (beautiful) (are)

- The museum has a magnificent building ~

sangrahaalay kii imaat bhavya hai
(museum) (of) (building) (magnificent) (is)

Translation Pattern: P3



Locative case ending "*ke paas*" is used instead of genitive postposition.

- Rama has a book ~

Rama ke paas ek kitaab hai

(Rama) (with) (a) (book) (is)

- Sona has a B.A. degree ~

Sona ke paas B.A. kii degree hai

(Sona) (with) (B.A.) (of) (degree) (is)

Translation Pattern: P4



English subject is realized as *dative subject* in Hindi.
Postposition "ko" is used with the subject.

- Meera had fever ~

Meera ko bukhaar thaa

(Meera) (to) (fever) (was)

- My uncle has asthma ~

mere chaachaa ko damaa hai

(my) (uncle) (to) (asthma) (is)

Translation Pattern: P5



Postposition "*mein*" is introduced in the translated sentence. Also,

Object → Subject
Subject → Predicative Adjunct

- This city has a museum ~
iss shahar mein ek sangrahaalay hai
(this) (city) (in) (a) (museum) (is)
- This forest has many animals ~
iss jangal mein bahut janwar hain
(this) (forest) (in) (many) (animals) (is)

Translation Pattern: P6



This pattern is characterized by the usage of postposition "par".

- The tiger has stripes ~

baagh par dhaariyan hotii hain

(tiger) (on) (stripes) (are)

- This horse has a saddle ~

iss ghode par ek kaathii hai

(this) (horse) (on) (a) (saddle) (is)

Translation Pattern: P7



The object of the English sentence is realized as Subjective Complement (that is an adjective).

- He has intelligence ~

wah buddhimaan hai

(he) (intelligent) (is)

- This donkey has strength ~

yah gadhaa takatwar hai

(this) (donkey) (strong) (is)

Translation Pattern: P8



- The main verb of the Hindi translation is not "*honaa*".
- The main verb is obtained from the *object* of the English sentence.

- Geeta has regards for old men ~

Geeta buzurgon kii izzat kartii hai
(Geeta) (old men) (of) (respect) (does)

- They had a narrow escape ~

we baalbaal bache
(they) (narrow) (escaped)

Translation Pattern: P9



- The main verb of the Hindi translation is not "honaa".
- The verb is **NOT** obtained from any *functional tag* (FT) of the sentence.
- It is obtained from the "sense" in which the verb "have" is used in the English sentence.
 - a) I had tea ~ *maine chai pee*
(I) (tea) (drank)
 - b) She had a difficult time ~
usne mushkil samay bitaayaa
(she) (difficult) (time) (spent)



Translation Pattern: P10

This pattern is observed if along with the <S V O> the sentence has an adjunct.

- Ram has two rupees ~

Ram ke pass do rupay hain

(Ram) (near to) (two) (rupees) (are)

- Ram has two rupees in his pocket ~

Ram ki zeb mein do rupay hain

(Ram's) (pocket) (in) (two) (rupees) (are)

Object → Subject
Adjunct → Predicative adjunct
Subject → Possessive case to predicative adjunct

Translation Pattern: P11



This pattern is observed if along with the <S V O> the sentence has an infinitive verb phrase.

- My children had me buy the car ~

mere bachchon ne mujhse gaadi kharidvaayai

(my) (children) (by me) (car) (forced to buy)

- The teacher had students submit the project ~

shikshak ne chhaatron se project jamaa karvaye

(teacher) (by students) (project) (forced to submit)

Combination of Patterns



Instances where the Hindi translation follows pattern pertaining to two or more classes are also found.

Mohan had a fracture. ~

Mohan kii haddii tootii
(Mohan) (of) (bone) (broke)

→ Features of P1 and P8

Deepa had an argument with Meera ~

Deepa kii Meera se bahas hui
(Deepa) (of) (Meera) (with) (argument) (happened)

→ Features of both P1 and P9

The Problem



- Such a multitude of patterns comes as a hindrance in effective implementation of any MT system.
- Similar problem is observed in many Indian (e.g., Bengali, Oriya) and some International languages (e.g., Hausa) too.

Rule for P10



- Based on *Sentence Structure*.
- IF

The given sentence structure is of the type
<Subject Verb Object Adjunct(PP)>,
&&

The PP satisfies the following two conditions

- a) The head noun of PP is inanimate.
- b) Head of the PP has a genitive pre-modifier that refers to the subject of the sentence.

THEN Translation Pattern P10

Rule for P10



□ Examples:

- The table has dust on its surface ~

mej ki satah par dhool hai

(table) (of) (surface) (on) (dust) (is)

- Sita has vermilion on her forehead ~

Sita ke maathe par sindoor hai

(Sita) (of) (forehead) (on) (vermillion) (is)

Rule for P10



- The pattern is not appropriate if the two conditions are not satisfied.
- Violation of condition (a)
 - Mohan has regards for his uncle. ~
Mohan apne chaachaa ki izzat kartaa hai ← P8
- Violation of condition (b)
 - Sheela has respect in the society. ~
samaaj mein sheela kii izzat hai ← P1

What Happens with other Patterns?



- Inadequacy of syntax:
 - Except for P10 & P11, English sentence structure is same for all patterns, i.e., $\langle S V O \rangle$.

- Inadequacy of constituent words:
 - Verb is same.
 - Subject or object alone does not help in identification of pattern.

Inadequacy of Subject



- Mohan has a good brain
 - ~ Mohan *kaa* dimaag achchhaa hai
- Mohan has a good pen
 - ~ Mohan *ke paas* ek achchhii kalam hai
- Mohan has high fever
 - ~ Mohan *ko* tej bukhaar hai
- Mohan had a sweet apple
 - ~ Mohan ne meethaa seb *khaayaa*

Inadequacy of Object



- Sita has flowers
~ *Sita ke paas phool hain*
- The tree has flowers
~ *ped par phool hain*
- The vase has flowers
~ *phooldaan mein phool hain*
- Meera has flowers in her home
~ *Meera ke ghar mein phool hain*

The Solution



- A rule based scheme is proposed to make decisions regarding the possible structure of the Hindi translation for a given input.

- Following features have been considered:
 - Sense of the verb "have"
 - Semantic feature(s) of the constituent words
 - Lexicographer file information of the subject/object

Senses of “have”



- Focus on the particular sense being conveyed by the verb of the sentence.
- WordNet 2.0 provides 19 different senses of the verb “have”.
- Example base sentences are examined and rules are formed for different senses.

Senses of “have”



- Usage of 5th, 10th and 11th sense of the verb “have” is rare (in our Example base).
- Sense number 6, 8, 9, 12, 13, 14, 15, 17, 18 and 19 are associated with some specific translation pattern.
- For sense numbers 1, 2, 3, 4, 7 and 16 more than one translation pattern is observed.
 - Finer rules are required to determine the possible translation pattern of the given sentence.

Translation Pattern for Different Senses of “have”



Sense Number	Translation Pattern	Sense Number	Translation Pattern
1	P1 or P2	12	P4
2	P1 or P3 or P5 or P6 or (P1 and P8)	13	P11
3	P1 or P3 or P2 or P8 or P9	14	P9
4	P1 or P2	15	P9
6	P9	16	(P1 and P8) or (P4 and P8)
7	P1 or P2	17	P9
8	P1	18	P9
9	P2	19	P9

Semantic Feature of Constituent Words



- The senses of the nouns used as subject/object as given in WordNet 2.0 are studied to design rules.



Object: Body Part

□ IF

- Object is a body part,

&&

- Object has a pre-modifier adjective that is not quantifier

THEN translation pattern P3.

E.g.

- Meera has swollen fingers ~

Meera kii anguliyaan soozii hui hain.

- Radha has good hair ~

Radha ke baal achchhe hain.

Cont...



- In above rule, IF
 - Pre-modifier of object is absent, OR
 - Pre-modifier is a quantifier

THEN translation pattern P1

E.g.

- The elephant has a trunk ~
haathi kii ek soond hai
- Hrithik has eleven fingers ~
Hrithik kii gyaarah anguliyaan hain

Identification of Body Part



- This rule applies if the abstraction level of hypernym of the object contains levels as given below:

i) organ#1

=> body part#1

=> part#7, piece#3

=> thing#12

=> entity#1

OR

ii) body covering#1

=> covering#1, natural covering#1, cover#5

=> natural object#1

=> object#1, physical object#1

=> entity#1

Remarks



- These rules are neither *necessary* nor *sufficient* for a particular pattern.
- *Size* of the example base is not the key for framing the rules.
- Rather the *variety* of examples plays the pivotal role.

Lexicographer Files



- Lexicographer files in WordNet 2.0 are the files containing all the synonyms logically grouped on the basis of syntactic category.
- For example, *noun.act*, *noun.animal*, *noun.person*.
- According to WordNet, noun has 26 different logical groupings, and thus 26 lexicographer files.

Lexicographer Files



- Also, there are "have/had" sentences where subjects are pronouns.
- These can be taken care of under any of the 26 categories (primarily as *noun.person*).
- There can be imperative sentences where the subject is silent (e.g. Have this book). This is considered as a separate case.
- Thus we have 27 possibilities for subjects, and 26 possibilities for objects to deal with.

Lexicographer Files



- On studying subject and object of our database sentences, a **27*26** matrix has been constructed.
- The matrix suggests the translation patterns obtained in different combination of subject and object.
- In our example base there are no sentences with subjects as *noun.motive*, *noun.phenomenon*, *noun.process*, *noun.feeling*, *noun.possession* or *noun.relation*.
- Similarly, no sentences with objects as *noun.motive* or *noun.relation*.
- These columns and rows are discarded from the matrix.
- Therefore, the final matrix has **21*24 = 504** cells.

Case 1: Empty Cells



- Out of the 504 cells, 297 cells are empty.
- For example, when the subject is *noun.attribute* and object is *noun.animal*.
- For these 297 situations no translation rules need to be formed.

Case 2: Single Entry in a Cell



- There are 85 (out of 504) cells which have only one entry. For these 85 combinations of subject and object, pattern ambiguity can be resolved directly.
- For example:

Subject Sense	Object Sense	Pattern
Noun.act	Noun.state	P1
Noun.animal	Noun.substance	P5
noun.group	noun.quantity	P1
Noun.plant	Noun.state	P5

Case 3: Two or Three Patterns



- For some columns and rows only two or three patterns are observed.
- For example
 - Subject is *noun.act*, then the patterns observed are P1 or P5
 - Object is *noun.shape*, then possible patterns are P5 or P6.
- The system need not explore all the 11 possibilities.
- Rather, it may furnish two or three translations of the sentence and obtain user feedback.

Case 4: Subject-Object Combinations



- There exist some subject-object combinations with only two or three observed patterns.
- For example:
 - If the subject is *noun.artifact*, and object is *noun.communication*, then the patterns observed are P5 or P6.
 - If the subject is *noun.act*, and object is *noun.cognition*, then possible translation patterns are P1 or P5.
- As in Case 3, here too the pattern ambiguity can be resolved through user feedback.

Case 5: Densely Occupied Cells



- There are 15 cells that are very dense.
- Pattern ambiguity could not be resolved for these sentences so far.

Subject	Object	Pattern Observed
noun.artifact	noun.artifact	P1 - 67, P2 - 35, P5 - 36, P6 - 45
noun.person	noun.artifact	P1 - 25, P2 - 10, P3 - 35, P5 - 10, P10 - 24
noun.location	noun.group	P1 - 7, P2 - 5, P5 - 24, P6 - 7

Summarization of the Work



- This paper first defines the term "pattern ambiguity" that is observed in translation from English to Hindi.
- In this work we have used verb senses and subject-object senses separately.
- So far, we could resolve pattern ambiguity for about 75% of cases, out of about 4000 sentences.

Points to Ponder



- Whether pattern ambiguity in translating English sentences with "have" as its main verb is completely resolvable???
- Two possible approaches can be
 - *By analyzing the context.* But creating a large database containing appropriate context information as well as having "have" sentences is not an easy task.
 - *By considering verb and subject/object senses together* for a given input sentence.
- Presently we are focusing our investigations to these directions.

References



- James Allen. *Natural Language Understanding*. Benjamin Cummings, California, 2nd edition, 1995.
- Bonnie J. Dorr. *Machine Translation: A View from the Lexicon*. MIT Press, Cambridge, MA, 1993.
- Bonnie J. Dorr, Pamela A Jordan, and John W. Benoit. A survey of current research in machine translation. In M. Zelkowitz, editor, *Advances in Computers*, volume 49, pages 1--68. Academic Press, London, 1999.
- WordNet 2.0: <http://www.cogsci.princeton.edu/cgi-bin/webwn>
- 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.
- Deepa Gupta and Niladri Chatterjee. Identification of divergence for English to Hindi EBMT. In *MT Summit IX*, pages 141 -- 148, New Orleans, USA, 2003.
- Nancy Ide and Jean Veronis. Word sense disambiguation: The state of the art. *Computational Linguistics*, 24(1):1 -- 40, 1998.



Thank You

September 24, 2005

Shailly Goyal, IIT Delhi, India
shailly_goyal@yahoo.com



September 24, 2005

Shailly Goyal, IIT Delhi, India
shailly_goyal@yahoo.com

Object: Abstract Noun



□ IF

- Object is "some abstract noun",
&&

- Object has a pre-modifier that is not a quantifier
THEN translation pattern P1.

E.g.

- This food has good smell. ~
iss khaane kii khushbuu achchhii hai.
- She has sweet voice. ~
uskii aawaaz miithii hai.

Which Abstract Nouns?



- Based on partial hypernyms.
 - i) creativity#1, creativeness#1, creative thinking#1
 - => ability#2, power#3
 - => cognition#1, knowledge#1, noesis#1
 - => psychological feature#1
 - OR ii) sound#1
 - => sound property#1
 - => property#3
 - => attribute#2
 - => abstraction#6
 - OR iii) perception#3
 - => basic cognitive process#1
 - => process#3, cognitive process#1, mental process#1
 - => cognition#1, knowledge#1, noesis#1
 - => psychological feature#1