

Semantic Constraints in Pre-/Post-Position Disambiguation in Reverse MT

Anil Thakur

Introduction: The paper presents a mechanism, based on the semantic constraints, for the disambiguation of pre-/post-positions for English-Hindi-English reverse machine translation. To obtain machine translation from one natural language to another, we need to examine different syntactic and semantic constraints of both the source and the target languages. However, if the aim is to get a reverse translation of the translated text back into the source text, then both the languages are at the same time the source and the target language. We need to examine the structures of both the languages from the perspective of the source and the target language. In such a case, we need to investigate to what extent the semantic analysis for the source language at the first place can be extended for translating the target texts back into the source text in reverse translation. We may notice that in some cases, the semantic constraints are the same for both the languages (as the source language). However, in many cases, there are gaps in the semantic constraints between the two languages. In this paper, I take some representative adpositions (pre-/postpositions) from English and Hindi and examine the similarities and gaps in the semantic constraints for English-Hindi-English machine translation.

Outlines:

Most of the prepositions and postpositions in English and Hindi respectively are highly polysemous. For instance, *to* in English has (at least) 14 different mapping patterns in Hindi (Sinha & Thakur 2007, 2005). *ko* in Hindi has (at least) 6 mapping patterns in English (Sinha & Thakur 2007, 2005, Singh 2003). In this context the following points are to be taken into account.

- i. Their disambiguation is necessary to obtain exact mappings for these pre-/post-positions in the target language for machine translation
- ii. The disambiguation of these pre-/post-positions is heavily dependent on the semantic constraints: the semantic category of the verb and/or the head noun.
- iii. The semantic constrains in the source language and the target-cum-source language (for reverse MT) is not necessarily similar.
- iv. Separate semantic analysis is needed in each case (in reverse MT).

Semantic Constrains: Disambiguation of Prepositions/Postpositions:

2.1 Semantic Constrains: Similarities:

In this section, some of the representative examples where similar semantic constraints can be useful to formulate rule for disambiguation are presented.

2.1.1. to

i. to => se => to

Example1:

We have talked **to** the Principal. => *hamane prinsipal se baat kii hE* {we principal to talk did be.PR} => We have talked **to** the Principal.

Disambiguation of to=>se:

Rule1: <noun(human/human_org)><verb(communicative)>[to]<noun(human/human_org)>
=> [se]

Disambiguation of se=>to:

Rule2: <noun(human/human_org)><noun(human/human_org)>[se]<verb(commumacative)>
=> [to]

ii. to => ko => to

Example2:

Sita gave the book **to** Ram. => *siitaa-ne raam-ko kitaab dii.* {Sita-ERG Ram-DAT book gave} => Sita gave the book **to** Ram.

Disambiguation of to=>ko

Rule3: <noun><verb(transfer)>[to]<noun(human/human_org)> => [ko]

Disambiguation of ko=>to:

Rule4: <noun><noun(human/human_org)>[ko]<verb(ditransitive)> => [to]

iii. to => ke paas => to

Example3:

Sita went **to** Ram. => *siitaa raam ke paas gaii*. {Sita Ram to went} => Sita went **to** Ram.

Disambiguation of to=>ke paas:

Rule5: <noun><verb(motion)>[to]<noun_x(human)> => [ke paas]

Disambiguation of ke paas=>to:

Rule6: <noun><noun(human)>[ke paas]<verb(motion)> => [to]

iv. to => φ => to

Example4:

Sita went **to** Delhi. => *siitaa dilli gaii*. {Sita Delhi went} => Sita went **to** Delhi.

Disambiguation of to=>φ:

Rule7: <noun><verb(motion)>[to]<noun_x(place)> => [φ]

Disambiguation of φ=>to:

Rule8: <noun><noun(place)>[φ]<verb(motion)> => [to]

v. to => par => to

Example5:

Sita went to the river/mountain. => *siitaa nadii/pahaar par gaii*. {Sita river/mountain to went} => Sita went to the river/mountain.

Disambiguation of to=>par:

Rule9: <noun><verb(motion)>[to]<noun(natural_object)> => [par]

Disambiguation of par=>to:

Rule10: <noun><noun(natural_object)>[par]<verb(motion)> => [to]

2.1.2. with**i. with => ke saath => with**

Example6:

Sita lives with her children. => *siitaa apne baccon ke saath rahatii hE*. [Sita self children with lives] => Sita lives with her children.

Disambiguation of with=>ke saath:

Rule11: <noun><verb>[with]<noun(animate)> => [*ke saath*]

Disambiguation of *ke saath*=>with:

Rule12: <noun><noun(animate)>[*ke saath*]<verb> => [with]

ii. with => se => with

Example7:

He kneeled with respect. => *vah samman se jhukaa*. {he respect with kneeled} => He kneeled with respect.

Disambiguation of with=>se:

Rule13: <noun><verb>[with]<noun(concept)> => [*ke saath*]

Disambiguation of se=>with:

Rule14: <noun><noun(concept)>[*ke saath*]<verb> => [with]

iii. with

Example8:

He ran with a gun. => *vah banduuk lekar dORaa*. => {he gun with ran} => He ran with a gun.

Disambiguation of with=>lekar:

Rule15: <noun><verb(motion)>[with]<noun(instrument)> => [*ke saath*]

Disambiguation of lekar=>with:

Rule16: <noun><noun(instrument)>[*ke saath*]<verb(motion)> => [with]

2.2. Semantic Constrains: Gaps :

In this section are presented examples where different sets of semantic-syntactic constraints are needed in the formulation of disambiguation rules for the adpositions.

i. to => Or => to/from

Example9:

a. Sita went to the left. => *siitaa baayiiN Or gaa*. {Sita left side to went} => Sita went to the left.

b. *siitaa baayiiN Or se aai*. {Sita left to from came} => Sita came from the left.

Disambiguation of to=>Or:

Rule17: <noun><verb(motion)>[to]<noun(place_direction)> => [Or]

Disambiguation of *Or*=>to:

Rule18: <noun><noun(place_direction)>[*Or*]<verb(motion)> => [to]

Gaps in Semantic Constrains: The example in (9b) shows that even if *Or* occurs with a place_direction noun the disambiguation is not necessarily to *to* in English. More detailed analysis is needed to obtain exact mapping of *Or* in reverse Hindi-English MT.

ii. at => *meN* => at/in

Example10:

a. He came at night. => *vah raat meN/ko aayaa*. {he night at came} => He came at night.

b. *vah subah meN aayaa*. {he morning in came} => He came in the morning.

Disambiguation of at=>*meN*:

Rule19: <noun><verb>[at]<noun(time_non-point)> => [*meN*]

Disambiguation of *meN*=>at:

Rule20: <noun><noun(time_non-point)>[*meN*]<verb> => [at]

Gaps in Semantic Constrains:The example in (10b) shows that the semantic category time_non-point of the head noun in Hindi does not necessarily lead the postposition *meN* to be mapped by preposition at in English. Further analysis is needed to obtain exact mapping of *meN* to at in reverse Hindi-English MT.

iii. with => *vaalii* => with/...

Example11:

a. We saw a girl with blue eyes. => *hamane ek niilii aaNkhoN vaalii laRakii dekhii*. {we-
ERG a blue eyes with girl saw} => We saw a girl with blue eyes.

b. *dOkTar-ne usakii kamarvaalii dard Thiik kar dii* {doctor-ERG her waist-of pain cure did gave} => The doctor cured the pain in/of her waist.

Disambiguation of with=>*vaalii*:

Rule21: <noun><verb><noun>[with]<noun(body_parts)> => [*vaalii*]

Disambiguation of *vaalii*=>with:

Rule22: <noun><noun(body_parts)>[*vaalii*]<noun><verb> => [with]

Gaps in Semantic Constrains:The example in (11b) shows that the Rule22 cannot be globally applied to disambiguate the mapping pattern of the Hindi postposition *vaalii*

to English. Further analysis of the semantic constrains in Hindi is required to obtain exact mapping of *vaalii* onto English for reverse Hindi-English MT.

Conclusion:In this short paper, a relatively new area of machine translation has been explored. Generally, linguistic research for machine translation concerns a unidirectional pair of language which requires a different set of linguistic knowledgebase for the purpose of formulation rules. However, it is both linguistically and from the point of view of machine translation research interesting to explore how the linguistic constrains may be manipulated when the linguistic constraints in a pair of MT languages is explored by taking both the language source language as well as target language at the same time. In this paper, we can see that both the translation languages need separate semantic analysis (as the source language) to obtain disambiguation rules. In the present paper, some representative examples of prepositions from English and postpositions from Hindi have been taken for study and it has been shown that in certain case, a similar set of semantic constrains can be used to disambiguate the adposition in the question whereas in many other case, a different sets of semantic constrains are required. It is interesting to note the nature of the output when the target language becomes the source language for the reverse MT.

References:

- Dorr, Bonnie. 1994. Classification of Machine Translation Divergences and a Proposed Solution. *Computational Linguistics*, 4(20). 597-633.
- Halliday, M.A.K. 1960. Linguistics and Machine Translation. In McIntosh, A. and Halliday, M.A.K. (eds.) *Patterns of Language: Papers in General, Descriptive and Applied Linguistics*. Indiana University Press. London.
- Sinha, R.M.K. 2005. Translation Divergence in English-Hindi MT, EAMT. Budapest: Hungary.
- Sinha, R.M.K., etal. 2009. Divergence Patterns in English-Hindi MT. *Computer Society of India Journal*, Vol. 33.
- Singh, Surajabhan.2003. *English-Hindi Translation Grammar*. Delhi: Prabhat Prakashan.
- Srivastava, Murlidhara. 1969.. *The Elements of Hindi Grammar*. Delhi: Moltlal Banarsi Dass.

