Title: English-Hindi MT Divergence

Ms. Neelakshi Goel

gneelakshi123@gmail.com

Abstract: Divergence identified with mapping examples between two or more natural languages is a typical marvel. The examples of divergence between two languages should be distinguished and procedures conceived to handle them to get right interpretation from one language to another. In the writing on MT, a few endeavors have been made to order the sorts of interpretation uniqueness between a couple of common languages. In any case, the issue of phonetic disparity is such a mind boggling marvel, to the point that a considerable measure more should be done in this zone to recognize further classes of dissimilarity, their suggestions and between relatedness as well as the ways to deal with handle them. In this paper, we take Dorr's (1994) order of interpretation dissimilarity as base and analyze the interpretation examples in the middle of Hindi and English to find further subtle elements and ramifications of these divergences. The essential objective of the paper is to call attention to distinctive sorts of interpretation divergences in Hindi and English MT that have not been examined in the current writing.

Keywords: Divergence, Interlingua

Introduction: Hindi-English dialect pair for (machine) interpretation introduces a rich instance of uniqueness at distinctive linguistic and also additional syntactic levels. It is essential to recognize the distinctive sorts of divergences to acquire right interpretation for Hindi sentences to English and the other way around. The interpretation divergences has been inspected in the writing on MT from distinctive hypothetical points of view with the end goal of their legitimate characterization and taking care of (Dorr 1990a, 1990b, 1993, 1994, Barnett et al 1991a, Barnett et al 1991b, see Dorr 1994 for a brief audit of them). Dave et al (2001) talk about a note worthy's percentage classes of interpretation uniqueness as proposed in Dorr (1993) and oultine an UNL based interlingua approach for the treatment of an interpretation's percentage divergences in the middle of English and Hindi. Gupta et al (2003) talk about an interpretation's percentage divergences for English-Hindi MT, in light of the arrangement proposed in Dorr (1994) and recommend a bound together approach for their distinguishing proof and determination. On the other hand, the issue of interpretation dissimilarity is intricate one and various critical uniqueness issues have stayed out of the extension in the current chips away at the theme. In this paper, we look at the changed regions of interpretation divergences both from Hindi to English and English to Hindi machine interpretation viewpoints. We take Dorr's arrangement of interpretation uniqueness as the purpose of takeoff to look at the point of dissimilarity in Hindi and English dialect pair. In Section 2, we talk about arrangement of interpretation dissimilarity as proposed in Dorr (1994) and present applicable illustrations from Hindi-English interpretation pair to look at in the matter of what degree the present characterization can be received for these cases and to what degree we need further classes/classifications of difference to represent the cases of interpretation divergences we experience in Hindi-English and English-Hindi MT dialect sets. In segment 3, we look at further points of uniqueness in the middle of Hindi and English MT. We talk about the interpretation divergences under distinctive points of syntax and present our perceptions on their order.

2. Dorr's Classification and Divergence

2.1 Dorr's Classification

Dorr (1994) has recognized seven classes of interpretation divergences. These classes are: (i) Thematic Divergence, (ii) Promotional Divergence, (iii) Demotional Divergence, (iv) Structural Divergence, (v) Conflational Difference, (vi) Categorial Divergence, and (vii) Lexical Divergence. The classes of interpretation difference have been characterized to

represent diverse sorts of interpretation divergences found in a couple of interpretation dialects. She calls attention to that the interpretation divergences emerging out of idiomatic usage aspectual knowledge, discourse knowlwdge, domain knowledge, or world knowledge remain out of the extent of her paper (Dorr 1994).

2.2. Hindi-English and English-Hindi MT Divergence

2.2.1. Thematic Divergence

Thematic divergence refers to those divergences that arise from differences in the realization of the argument structure of a verb. The Hindi counterpart of an English example in which the subject NP occurs in the dative case whereas the subject NP in English is in the nominative case can be cited as a type of thematic divergence

John likes Mary.

=> i. *jOn mErii-ko pasand karataa hE.*

{John Mary-ACC like do be.PR}

ii. *jOn-ko merii pasand aaii.*

{John-DAT Mary like came}

iii. *jOn-ko mErii pasand hE.*

{John-DAT Mary like be.PR}

2.2.2. Promotional and Devotional Divergence

Promotional and demotional divergences or Head-swapping divergences emerge where the status (lower or higher) of a syntactic constituent in one dialect is influenced in another dialect. Case in point, when a word intensifying component in one dialect is acknowledged by a verbal component, it constitutes an instance of special dissimilarity and an inverse case will bring about demotional difference.

2.2.3. Structural Divergence

Structural divergences are illustrations where a NP contention in one dialect is acknowledged by a PP assistant/sideways NP in another dialect. The verb "enter" in an English sentence, for example, "he entered the room" => vah kamare meN paravesh kiyaa {he room in enter did} takes a NP contention 'the room' though its Hindi partner pravesh karanaa takes a PP assistant kamare meN {room in} (Dave et al, 2001).

2.2.4. Conflational Divergence

The sense conveyed by a single word in one language requires at least two words of the other language.For example, "*He stabbed me*" will be translated as "u*sne mujhe chaaku se maaraa*".The English word "*stab*" has no one-word equivalent in Hindi, and therefore the introduction of the word "chaaku" was necessitated.

2.2.5. Categorial Divergence

It defines the changes in category. For example, the predicate is adjectival in one language but nominal in other language.The English sentence "*I am feeling hungry.*" will be translated into Hindi as " *mujhe bhukh lag rahii hai.*"

In English "*hungry*" is adjective and but in Hindi "*bhukh*" (*hunger*) becomes the noun.

2.2.6. Lexical Divergence

The event is lexically realized as the main verb in one language but as a different verb in other language.

Consider the sentence "*They run into the room.*" Its Hindi translation is "w*oh daurte huye kamre mein ghus gaye*"

The event is lexically realized as the main verb "*run*" in English but as a different verb "*ghus*" (literally (*to enter*)) in Hindi, and "*run*" is used as participle.

3. Divergence in English-Hindi MT

Still there are many types of divergence which cannot be clearly accounted for within the existing classification.

3.1 Word order and its implication

Some of the word order related divergence can be handled within the defined classification. But still needs more exploration.

Example: Interpretation of question particle *kya in Hindi is dependent of word order facts of English and Hindi.*

Eg:

> *aap kya kha rahe hai?*
> ➢ *What are you eating?*
> *kya aap kha rahe hai?*
> *aap kha rahe hai kya?*
> ➢ *Are you eating?*

3.2 Replicative words

Hindi has certain replicative words for which there is no exact English word.

Almost all kinds of words can be replicated to denote a number of different functions in Hindi.

Eg. *bachchaa bachchaa*
dekhate-dekhate

3.3 Stative words

A class of verb such as '*sit*', '*stand*', '*sleep*', and '*wake up*' exhibit divergence with respect to the realization of their aspectual and participial forms.

For these verbs in English, there is no distinction between the progressive aspectual form and the participial form.

For instance, '*sitting*' can mean either *betha hua* or *beth raha* in Hindi. However, in this case, the reverse translation also causes divergence.

3.4 Expressive and Echo words

There is no exact parallel word available for these lexical items in English.

This may be related to the socio-cultural and even anthropological aspects of a natural language which may use in one language but not in other.

Eg: patte khar khara rahe hai
The leaves were making a *khar khara* sound.

3.5 Honorific

Honorific features are expressed by several linguistic markers including the use of plural pronoun and plural verbal inflections.

Eg:

> *unake pitaa aaye hEN.*
> His father has come.

> *unakaa nOkar aayaa hE.*
> His servant has come.

3.6 Determiner System

English has indefinite articles that mark the indefiniteness of the noun phrase overtly.

Hindi lacks an overt article system and different devices are used to realize the indefiniteness of a noun phrase in Hindi.

These type of divergence is related to more than one aspect of grammar.

For instance, mapping of a bare NP in Hindi onto an NP with an article '*a-an*/*the*' in English is dependent on a detailed syntactic and semantic analysis of the noun phrases in both the languages

Eg: *laRakaa aayaa*
> ➢   The/*a boy came
>   *jangal meN sher hE.*
> ➢   There is a lion in the forest.
>   *sher jangal meN hE.*
> ➢   The lion is in the forest.

3.7 Morphological gaps

In Hindi, there are certain type of passive construction that when used marks a certain kind of modality function but in English counterparts of such Hindi sentences can only partially express the exact meaning

Eg:

> *raam se shiishaa TuuT gayaa.*
> i. The glass got broken by Ram.
> ii. Ram broke the glass (unintentionally).

3.8 Conjuctions, Particles and Passive words

Using different conjunctions, punctuation marks, and particles in Hindi gives rise to another source of divergence.

Some of these particles such as *ki*, *na*, *yaa* and *vaalaa* have functional roles in Hindi that are mapped in English by different means than can be identified on the basis of the syntactic structure.

Conclusion

On the basis of the discussion presented in this paper, we have shown that the translation divergence between Hindi and English machine translation is more varied and complex than the works in the existing literature can accommodate and account for. To obtain correct translation, we need to examine the different grammatical as well as some of the extra-grammatical characteristics of both Hindi and English to exhaustively identify the types of translation divergence in this pair of translation languages. Some of the topics, particularly those related to socio-cultural aspects of language need further exploration in light of the complexity in their formalization.

References

1. ABBI, Anbita (1992). Reduplication in South Asian Languages: An Areal, Typological and Historical Study. New Delhi: Allied Publishers.

2. BARNETT, Jim, Inderjeet Mani, Paul Martin, Elaine Rich (1991a). 'Reversible Machine Translation: What to do When Languages Don't Line Up'. In Proceedings of the Workshop on Reversible Grammar in Natural Language Processing, ACL-91, 61-70, University of California, Berkley, CA.

3. BARNETT, Jim, Inderjeet Mani, Elaine Rich, Chinatsu Aone, Kevin Knight, Juan C. Martinez (1991b). 'Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT'. In Proceedings of Machine Translation Summit, 25-32, Washington, DC.

4. DAVE, Sachi, Jignashu Parikh, and Pushpak Bhattacharyaa (2001). 'Interlingua-based EnglishHindi Machine Translation and Language Divergence'. Machine Translation 16 (4), pp. 251304.

5. DORR, Bonnie (1990a). 'A Cross-Linguistic Approach to Machine Translation'. In Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 13-32, Linguistic Research Center, The University of Texas, Austin, TX.

6. DORR, Bonnie (1990b). 'Solving Thematic Divergence in Machine Translation'. In Proceedings of the 28th Anual Conference of the Association for Computational Linguistics, 127-134, University of Pittsburg, Pittsburg, PA.

7. DORR, Bonnie (1993). Machine Translation: A View from the Lexicon. Cambridge, Mass: The MIT Press.

8. DORR, Bonnie (1994). 'Classification of Machine Translation Divergences and a Proposed Solution'. Computational Linguistics, Vol. 20 (4), pp. 597-633.

9. DORR, Bonnie, Necip Fazil Ayan, and Nizar Habash (2004). 'Divergence Unraveling for Word Alignment of Parallel Corpora'. Natural Language Engineering, 1 (1), pp. 1-17.

10. GOYAL, Shaily, Deepa Gupta and Niladri Chatterjee (2004). 'A Study of Hindi Translation Patterns for English Sentences with "Have" as the Main Verb'. In Proceeding of iSTRANS, Vol. 1, pp. 46-51.

11. GUPTA, Deepa, and Niladri Chatterjee (2003). Identification of Divergence for English to Hindi EBMT. In Proceeding of MT Summit-IX, pp. 141148.

12. SINHA, R.M.K. and Anil Thakur (2004). 'Synthesizing Verb Form in English to Hindi Translation: Case of Mapping Infinitive and Gerund in English to Hindi'. In Proceeding of iSTRANS, Vol. 1, pp. 52-55.