
LANGUAGE IN INDIA

Strength for Today and Bright Hope for Tomorrow

Volume 14:5 May 2014
ISSN 1930-2940

Managing Editor: M. S. Thirumalai, Ph.D.
Editors: B. Mallikarjun, Ph.D.
Sam Mohanlal, Ph.D.
B. A. Sharada, Ph.D.
A. R. Fatihi, Ph.D.
Lakhan Gusain, Ph.D.
Jennifer Marie Bayer, Ph.D.
S. M. Ravichandran, Ph.D.
G. Baskaran, Ph.D.
L. Ramamoorthy, Ph.D.
C. Subburaman, Ph.D. (Economics)
Assistant Managing Editor: Swarna Thirumalai, M.A.

Translation Divergences in Hindi-Nepali Machine Translation

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

=====

Abstract

This paper examines the translation divergence in Hindi-Nepali Machine Translation which occurs due to the divergent patterns of the two languages in the grammatical level. The primary task of this paper is to identify the different types of translation divergences in the context of the Hindi and Nepali MT with a view to classify them according to the well-defined theoretical framework as proposed in the existing literature. In this paper, the different areas of translation divergences both from Hindi to Nepali and Nepali to Hindi machine translation perspectives have been examined. Dorr's classification of translation divergence has been taken as the basis in order to examine the topic of divergence between the Hindi and Nepali language pair. However, the divergences like Promotional and Demotional Divergence have not been

Language in India www.languageinindia.com ISSN 1930-2940 14:5 May 2014

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

Translation Divergences in Hindi-Nepali Machine Translation

examined in this paper as the case of such divergence is very rare or unavailable in this language pair.

1. Background

Generally speaking, machine translation is a translation of text from one language to another with the help of computers. It aims at achieving the translation quality as human translators do but it is impossible or hard to achieve this goal due to the divergent characteristics of each and every language. Every language has its own patterns in its grammatical and extra grammatical levels. Though, Hindi and Nepali belong to the same language family, Indo-Aryan, they have their own distinct linguistic patterns which result in rich divergence in machine translation. This paper aims at identifying such divergences as this is necessary in order to get a good quality translation for the Hindi sentences in Nepali and vice-versa.

The literatures on MT have examined the translation divergences from different theoretical perspectives for the purpose of their proper classification and handling. Dorr 1990a, 1990b, 1993, 1994, Barnett, et al. 1991a, Barnett, et al. 1991b, and Dorr 1994 are the main works in this regard. The issue of translation divergence has not been discussed yet in the context of machine translation between Hindi and Nepali. The primary task of this paper is to identify the different types of translation divergences in the context of Hindi and Nepali MT with a view to classifying them according to the well-defined theoretical framework as proposed in the existing literature. In this paper, the different areas of translation divergences both from Hindi to Nepali and Nepali to Hindi machine translation perspectives have been examined. Dorr's classification of translation divergence has been taken as the basis in order to examine the topic of divergence between the Hindi and Nepali language pair. However, the divergences like Promotional and Demotional Divergence have not been examined in this paper as the case of such divergence is very rare or unavailable in this language pair.

2. Dorr's Classification and Divergences in Hindi-Nepali Machine Translation

2.1. Dorr's Classification

Dorr (1994) has identified six classes of translation divergences. They are: (i) Thematic Divergence, (ii) Promotional and Demotional Divergence, (iii) Structural Divergence, (iv) Conflational and Inflational Divergence, (v) Categorical Divergence, and (vi) Lexical Divergence.

Language in India www.languageinindia.com ISSN 1930-2940 14:5 May 2014

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

Translation Divergences in Hindi-Nepali Machine Translation

In this paper main concentration is on the translation divergences arising out of grammatical aspects of the translation of languages. In the following section, the main classes of translation divergences as proposed in Dorr (1994) with some illustrative examples from Hindi and Nepali are illustrated.

2.2. Divergences in Hindi-Nepali MT

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 the Nepali sentences in which the subject NP occurs in the dative case whereas the subject NP in Nepali is in nominative case can be cited as the type of thematic divergence.

1. (a) **ram ko** puraskar mila (H)

Ram-DAT award get-3.SG.PST

ram-le puraskar pajo (N)

Ram-NOM award get-3SG.NPST

Gloss- *Ram got the award.*

- (b) **apko** dik^{hai} degi (H)

You-DAT see-3.SG.NPST

tapai-le dek^hnu hunec^hə (N)

You-NOM see-3.SG.NPST

Gloss- *You will see.*

In the examples above, the subjects of Hindi sentences with dative case are in the patient role whereas the Nepali counterparts with nominative case have been realized in the agent role. This type of divergence arises due to the attestation of voice differences in the Hindi and Nepali pair of language.

2.2.2. Conflational and Inflational Divergence

A Conflational Divergence results when two or more words in one language are translated by one word in another language. The opposite case is referred to as Inflational Divergence.

Hindi and Nepali language pair shows lots of examples in the case of Conflational Divergence as in examples below:

2 (a) hām **pəidəl cəlkər** zænge (H)

We foot walk go-2.PL.NPST

hami **hīderə** zæne^həũ (N)

We walk go-2.PL.NPST

Gloss- *We will go by foot.*

(b) vo log **vapəs ləut** gæe (H)

They people return back -3.PL.PST

unihəru **p^hərkie** (N)

They return-3.PL.PST

Gloss- *They returned back.*

In the examples above, the Nepali counterparts for Hindi *pəidəl cəlkər* and *vapəs ləut gæe* are *hīderə* and *p^hərkie* respectively. Hence, the case can be regarded as conflational divergence as two lexical items of Hindi are realized by one lexical item in Nepali.

Likewise, the cases of Inflational Divergence are also attested in the Hindi–Nepali language pair as in the following examples:

(2) a. **zane** wo din kəb ægə (H)

Don't Know that day when come-3.SG.NPST

t^haha c^həinə tyo din kəile aulə (N)

Know not that day when come-3.SG.NPST

Gloss- *Don't know when the day will come.*

Language in India www.languageinindia.com ISSN 1930-2940 14:5 May 2014

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

Translation Divergences in Hindi-Nepali Machine Translation

b. **kəbʰi** həm zaenge (H)

Some time we go-2.PL.NPST

kunəi belə hami zane^{həũ} (N)

Some time we go-2.PL.NPST

Gloss- *We will go someday.*

Here, *zane* in 2 (a) and *kəbʰi* in 2 (b) have been realized by two lexical items in Nepali and hence are the cases of inflational divergence.

2.2.3. Categorical Divergence

Categorical divergence arises due to the mismatch between parts of speech in the sentences of the pair of translation languages. It is the most common type of divergence that occurs in any pair of language. In the Hindi and Nepali language pair, most of the cases of categorical divergence arise due to the lack of equivalent of Hindi verbal noun in the Nepali language. The verbal nouns of the Hindi are realized as infinite verb in Nepali as in the sentences below:

3(a) həm hatʰi ki **səvəri** ka məza utʰa rəhē hē (H)

We elephant ride fun have-2.PL.NPST

hami hattima **cəḍne** məzza lirəheka cʰəũ (N)

We elephant ride fun have-3.PL.NPST

Gloss- *We are having fun of elephant riding.*

(b) jəhā **zəl bʰərav** ka karjə hota he (H)

Here water storage task be-3.SG.NPST

jəha pani **bʰərne** kam huncʰə (N)

Here water storage task be-3.SG.NPST

Gloss- *Water storage task is done here.*

In the given examples, the Hindi verbal nouns like *səvəri* and *bʰərav* are realized by infinite verb forms in the Nepali like *cəḍne* and *bʰərne* respectively. Some further examples of

such Hindi verbal nouns which lack equivalence in the Nepali are *c^{hi}dka^u*, *bikri*, *səp^hai*, *k^hudai*, *pəɭtai*, *bizai*, *u^thau*, etc. Beside the cases of mismatch between verbal nouns and infinite verb, we get some cases of mismatch between adverb and verbal elements: A non-finite verb form *tutlakər* in the Hindi is translated as adverbial element *tote* in the Nepali language as in the sentence like the following:

4 (a) wo **tutlakər** bolta he (H)

He stutter speak -3.SG.NPST

u **tote** bolc^hə (N)

He stutter speak-3.SG.NPST

Gloss- *He speaks stutter.*

2.2.4. Lexical Divergence

Lexical divergence arises due to the lack of an exact translation map for the construction of a language into another. It can be seen as the case where the choice taken for the target language word is not the literal translation for a source language word. In this paper, an attempt has been made to examine lexical divergence under different sub-topics like unavailability of equivalence, phrasal verbs, kinship terms, and words with ambiguity as follows:

2.2.4.1. Unavailability of Equivalence

5(a) həme vəhã se **guzərna** hẽ (H)

We there from go-2.PL.NPST

Gloss- *We will go from there.*

(b) usne muz^he kitab **dilaja** (H)

He me book give-CAUS.3.SG.NPST

Gloss- *He manages to give me a book.*

For the above bold words, the Nepali has no exact equivalent and hence, has the tendency to burrow those words into the language.

2.2.4.2. Phrasal Verbs

Another type of Lexical Divergence arises when we get difference in the realization of phrasal verbs in the Hindi and Nepali language pair. For example,

6 (a) uska mehnət **rəŋg laja** (H)

usko pəriʃrəm səpʰəl bʰəjo (N)

Gloss- *His hard labour paid.*

(b) muz^{he} dek^hkər usne **datō t̥əle ungli d̥əbaja** (H)

məlai dek^herə u tin cʰək pərjo (N)

Gloss- *he became surprised to see me.*

2.2.4.3. Kinship Terms

It is obvious that different societies classify kinship relations differently and therefore they use different systems of kinship terminology for different relations and we also know that it is language specific phenomenon. But in the context of Machine Translation, it can be regarded as a hindrance when we come across with the lack of equivalence between the pair language as in the Hindi and Nepali. There are three types of problems regarding the translation of the Hindi kinship terms into the Nepali as follows:

- i. Terms having ambiguous meaning;
- ii. Terms having no equivalence; and,
- iii. Terms which need to be translated by paraphrasing.

Example (first type)

7 (a) **bʰəija** (brother) (H)

dazu or **bʰai** (elder or younger brother) (N)

(b) **behen** (sister) (H)

didi or **bəhini** (elder or younger sister) (N)

Language in India www.languageinindia.com ISSN 1930-2940 14:5 May 2014

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

Translation Divergences in Hindi-Nepali Machine Translation

- (c) **dada** (paternal grandfather) (H) and **nana** (maternal grandfather)

Both terms are translated by one term in Nepali, i.e., **baze**

Example (second type)

- 8(a) **muhbola beta** (adopted son) (H)

- (b) **māngetār** (fiancee) (H)

There is no equivalence for these terms in Nepali.

Example (third type)

- 9 (a) **cāceri behen** (cousin) (H)

kakaki c^hori (uncle's daughter) (N)

2.2.4.4. Words with ambiguity

Words having ambiguous meaning often create problems in Machine translation. Let us illustrate this with the help of some examples from the Hindi and Nepali language pair as follows:

- The Hindi word **log** is realized by two different meanings; sometimes as **manche** (people) and sometimes as **-hāru**, a plural suffix in the Nepali:

- 10 (a) **log** kitna nic^he gir saktā he (H)

manche kati tala z^harnā sākchā (N)

Gloss- *People can stoop to such level*

- (b) vo **log** acc^he hē (H)

uni**hāru** ramra c^han (N)

Gloss- *They are good.*

- pas** sometimes realized as a locative case marker **kāhū** and sometimes it got the meaning **nāzik** (near) in the Nepali:

- 11(a) muz^he uske **pas** zana he(H)

məlai ukəhā zanu cʰə (N)

Gloss- *I have to go there*

(b) pas me hi he mera gʰər (H)

nəzikəi cʰə mero gʰər (N)

Gloss- *My home is near*

- Likewise, the Hindi word **zara** sometimes got null sense and sometimes has meaning **əlikəti** (little bit) in Nepali:

12 (a) **zara** əpna cehra to dekʰo (H)

apʰno ənuhar tə hera (N)

Gloss- *Look at your own face.*

(b) vəkt ka **zara** bʰi əndaʒa nəhi rəha(H)

səməjəko **əlikəti** pəni kʰjal rəhenə (N)

Similarly, the Hindi word **aur** is realized in three different forms in Nepali; sometimes as a conjunct **əni** (and), sometimes as **ərko** (another) and sometimes as **əzʰə** (more):

13 (a) ram **aur** shjam (

ram **əni** shjam (N)

Gloss- *Ram and Shyam.*

(b) vo zəmana **aur** tʰa (H)

tjo zəmana **ərko** tʰijo (N)

Gloss- *That was a different time.*

(c) həmə **aur** mehnət kərni hogi (H)

hamile **əzʰə** pəriʃrəm gərnu pərnechə (N)

Gloss- *We have to do more labor.*

- The Hindi particle **vala** is realized by about seven different forms in Nepali as in the sentences below:

- lədkə **vala** - keta **pəkʰə**
- kismət **vala** - bʰagjə**mani**
- dʰoti **vala** - dʰoti laune
- gaũ**vala** - gaũko **basinda**
- kərne **vala** he- gərne **bʰæko** cʰə
- kər dene **vala** - bənaidine **kʰale**
- sadi **vala** gʰər – bihe gʰər

2.2.5. Structural Divergence

According to Dorr (1994), Structural Divergences are examples where an NP argument in one language is realized by a PP adjunct/oblique NP in another language as in the example given by Sinha, et al (2005) from English and Hindi language pair like “*he entered the room*” => *vəh kəmərə mē prəvef kija* {he room in enter did} takes an NP argument ‘*the room*’ whereas its Hindi counterpart *prəvef kərna* takes a PP adjunct *kəmərə mē* {room in}. The divergence of this type is yet not attested in the case of the Hindi-Nepali language pair. However, the divergence results from the differences in the structure of sentences during coordination process and this may be seen as an example of structural divergence between the two languages as in given sentences:

14 (a) həm itna becēn hē **ki** kəb tum əoge (H)

I so eager be CONJ when you come-1.SG.NPST

mə timi kəile aũcʰəu **bʰəni** atur cʰu (N)

I you when come CONJ eager be-1.SG.NPST

Gloss- *I am so eager that when will you come.*

(b) je **nə** **keval** sundər he **bəlki** accʰa bʰi he (H)

This not only beauty be but good also be-3.SG.NPST

jo sundər **matrə** **nəbʰəerə** ramro pəni cʰə (N)

This beauty only not good also be-3.SG.NPST

Gloss- *This is not only beautiful but is also good.*

In the first example, structurally the sentences of the two languages differ from each other in the sense that the subordinate sentence *kāb tum aoge* of Hindi comes after the conjunct particle *ki* whereas *b^hāni*, the conjunct of Nepali adjoined the subordinated sentence reversely, i.e. subordinated sentence of Nepali *timi kāile aũc^hau* comes before the conjunct. Similarly, in the second example, coordination pattern of the two languages differ in structures as *nā keval* and *balki*, the Hindi conjuncts, are positioned between the two different places in the coordinated sentences whereas in Nepali, *matrā nāb^hāerā* is conjunct which adjoined two sentences being together in the medial position.

3. Conclusion

In this paper, the issue of classification of translation divergence for MT between the Hindi and Nepali has been examined. The classification of translation divergence as proposed in Dorr (1994) and EAMT 2005 has been regarded as the basis in order to capture the various types of translation divergence between Hindi and Nepali. On the basis of the discussion presented in this paper, it is clear that the translation divergence between Hindi and Nepali machine translation is more varied and complex. To obtain correct translation, we need to examine the different grammatical as well as some of the extra-grammatical characteristics of both Hindi and Nepali 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

- BARNETT, Jim, Inderjeet Mani, Paul Martin, Elaine Rich. (1991a). *Reversible Machine Translation: What to do when Languages Don't Line Up*. In the proceedings of the Workshop on Reversible Grammar in Natural Language Processing, ACL-91, 61-70. Berkeley, CA: University of California.
- BARNETT, Jim, Inderjeet Mani, Elaine Rich, Chinatsu Aone, Kevin Knight, Juan C. Martinez (1991b). *Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT*. In the proceedings of Machine Translation Summit, 25-32, Washington, DC.
- DORR, Bonnie (1990a). *A Cross-Linguistic Approach to Machine Translation*. In the proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 13-32. Austin, TX : Linguistic Research Center, The University of Texas,.
- DORR, Bonnie (1990b). *Solving Thematic Divergence in Machine Translation*. In the proceedings of the 28th Annual Conference of the Association for Computational Linguistics, 127-134: Pittsburg, PA University of Pittsburg,.
- DORR, Bonnie (1993). *Machine Translation: A View from the Lexicon*. Cambridge, Mass: The MIT Press.
- DORR, Bonnie. (1994). *Classification of Machine Translation Divergences and a Proposed Solution*. Computational Linguistics, Vol. 20 (4), 20(4):597-633
- DORR, Bonnie, Necip Fazil Ayan, and Nizar Habash . (2004). *Divergence Unraveling for Word Alignment of Parallel Corpora*. Natural Language Engineering, 1 (1), 1-17.
- R.M.K. Sinha and Anil Thakur. (2005a). *Translation Divergence in English-Hindi MT*. In the

Krishna Maya Manger
Research Scholar
Department of Nepali
University of North Bengal
Rajaram Mohanpur
Siliguri,
Dist- Darjeeling 734013
West Bengal
India
krishnamanger@gmail.com