Problems and Challenges in Hindi to Bangla Translation: Some Empirical Observation and Workable Solutions

NILADRI SEKHAR DASH

This paper presents in brief the methods and strategies that are adapted for translating Hindi texts into Bangla in the project titled 'Indian Languages Corpora Initiative' (ILCI), funded by the DeitY, MeitY, Govt. of India. The basic task of translation is done manually by a team of translators (including the present author) who have exhibited good linguistic skill both in Hindi and Bangla language with a clear purpose that the eventual output can be utilized as benchmarked translated texts for machine learning works as well as for teaching translation methodology to new generation of translators. With application of some translation support tools and structured knowledge resources available, the team has translated more than 80000 sentences from Hindi to Bangla. This paper presents some of the problems and challenges that the translators have faced as well as the strategies they have applied to overcome the challenges. Due to brevity of space, I have discussed here some of the representative problems and their possible solutions with an expectation that this may be useful for future tasks of manual and machine translation between the two languages.

Keywords: translation, Hindi, Bangla, lexical replacement, pronoun, equivalence, divergence, copula

Introduction

The act of translating Hindi texts into Bangla is not an age-old profession. This has started, at a larger scale, quite recently although some time-tested imaginative and informative texts are translated in a bi-directional way with a high degree of accuracy, which has been possible due to sheer translation skill and excellent linguistic knowledge of the scholars who are engaged in translation. In most cases, translation is done manually and the success of a translation is directly proportional to the knowledge, expertise, and skill of a translator (Dave, Parikh & Bhattachrayya 2001). Surprisingly, not much effort is made to understand the issues and problems involved in translation between the two languages in a scientific manner, although the actual work of translation is, at present, a vibrant practice among the members of both the speech communities.

The present paper is, perhaps, the first of this kind where real corpus-based examples are taken into analysis to identify the problems as well as to suggest

solutions for translation between Hindi and Bangla in both directions. The relevance of this study may be attested in manual and machine translation and in designing courses for teaching translation methods to students and professionals.

In Section 2, I highlight the factors that are vital in translating Hindi texts into Bangla; in Section 3, I report how the process of lexical replacement is invoked to generate acceptable translation in Bangla; in Section 4, I show how proper and common nouns used in Hindi text are transferred into Bangla through transliteration; in Section 5, I discuss how Hindi pronominal forms are translated into Bangla keeping their connotative significance intact; in Section 6, I show how we solve the problems of lexical mismatch; in Section 7, I record some syntactic proximities; in Section 8, I describe the degree of syntactic divergence noted in translation; in Section 9, I highlight the problem of 'conflict of copula' which has been a challenge to the translators; in Section 10 (conclusion), I look into the impact and importance of the present study in the process of development of Hindi-Bangla machine translation system which is an elusive dream for the present Indian MT scientists.

Translating Hindi Texts into Bangla

Translation is a text processing and text production task involving two different languages: the source language and the target language. It involves several conscious and careful means and methods for retaining the original sense of the source language, preserving all semantic and stylistic properties of the source language and representing these in an appropriate way in the target language. This marks translation as a complex cognitive process which demands proper comprehension of texts in the source language and incorporation of a production process in the target language (Dorr, Ayan & Habash 2004). The task of translation, on the part of translators, requires a tremendous amount of linguistic competence in both the source and the target language along with acknowledged competence in the domains from where texts are selected for translation. Keeping these arguments as maxims we focus here on the strategies that we adopt for translating health and tourism texts from Hindi to Bangla (Fig. 1).
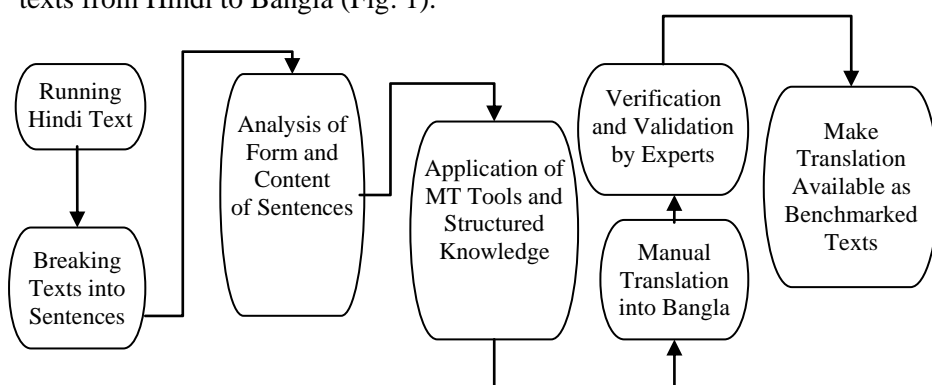
FIGURE 1. Process and strategies used in Hindi to Bangla Translation

It is implied in the selection of domains of the source texts in Hindi that these texts, by their nature, are less ambiguous, because their broad text types are predefined, which incidentally belong to non-literary and non-imaginative text type. That means these are informative text, the translation of which is more challenging because we do not have liberty to deviate from the level of accuracy of information presented in the source texts (Gupta & Chatterjee 2001; Gupta & Chatterjee 2003). Moreover, we have to maintain consistency and balance in the transfer of data and information to Bangla in such a manner that the source information is neither curtailed nor fabricated. Besides, we have to evoke similar kinds of sense and impact in Bangla, as texts of both the domains are directed for specific target audience (i.e., common people). Since such hurdles are abundant at every stage, we have no other alternative but to face the challenges and resolve these problems in a very pragmatic manner taking texts, contexts, and purposes into consideration (Dash 2015).

In essence, while translating Hindi texts into Bangla, we have to deal with several linguistic and extralinguistic issues linked with lexical difference, semantic gap, syntactic mismatch, grammatical difference, conceptual gaps, paradigmatic gap, lexical divergence, lexical convergence, etc. (Dorr 1992; Dorr 1994; Toury 1995). Most of these issues evolve from the differences in typological structures, inherent sociocultural divergence, and from different ecolinguistic settings induced into the language and their users (Janssen 2004). Since all these issues are not within our control, it is our responsibility to devise strategies that can reduce the grey shades in translation. The problems that we have faced in translating Hindi texts into Bangla are classified into 5 types, namely the followings (Toury 1995; Orozco & Amparo 2002).

    (a) Lexical issues (relating to usage and choice of words),
    (b) Syntactic issues (relating to the structure of sentences),
    (c) Semantic issues (relating to the meaning of words and sentences),
    (d) Sociocultural issues (relating to discourse and cultural aspects), and
    (e) Humanistic issues (relating to skill and expertise of translators).

Since it is not possible to describe all these issues in this paper, we shall try to illustrate only a few lexical and syntactic issues which we have encountered while translating Hindi texts into Bangla. Other issues will be discussed in details in subsequent papers under construction.

Lexical Replacement

At the time of translating Hindi text into Bangla, we note that the number of problems at the lexical level, in contrast to sentence, is quite less. Therefore, we translate by following a few simple processes of replacement of lexical items that are conceptually equivalent in the two languages (Sinha & Thakur 2004). For instance, Hindi words denoting universal objects, natural elements,

and global entities are translated into Bangla by use of exact or near-exact equivalent terms available in Bangla (Table 1).

Hindi	Bangla	Gloss
ghar	ghar	'house'
pāhāṛ	pāhāṛ	'mountain'
phul	phul	'flower'
phal	phal	'fruit'
bṛkṣa	bṛikṣa	'tree'
nadī	nadī	'river'
sāgar	sāgar	'sea'
ākāś	ākāś	'sky'
jamin	jamin	'land'
sūrya	sūrya	'sun'
chandra	chandra	'moon'
tārā	tārā	'star'
hawā	hāoyā	'air'

TABLE 1. Translational equivalence in universal concepts

Since these are mostly universal lexical items, both the languages do possess a specific set of terms to refer these concepts (Santos 1990). During translation, we do not face much problem to deal with such terms as simple replacement of source language terms by conceptually equivalent terms from the target language can solve the problem.

In a similar fashion, we effectively replace many Hindi adjectival forms by Bangla equivalent adjectival forms in translation, since these adjectival forms are normally similar in sense denotation in both the languages, e.g., *khūbsurat : sundar, bhālo* 'beautiful'; *choṭi : choṭo* 'small', etc. The same strategy is followed with adjectives of comparison also.

In the same manner, we translate the events designating action, processes, happenings, etc. from Hindi to Bangla by using exact or near-exact equivalents, e.g., *dauṛ : dauṛ* 'run', *dauṛnā : dauṛāno* 'running', *silwāi : selāi* 'stitching', *likhnā : lekhā* 'to write', etc. Also, abstract entities designating qualities, quantities, degrees of objects, events and action are translated in a similar manner, as in, *dayā : dayā* 'sympathy', *māyā : māyā* 'affection', *mamtā : mamatā* 'sympathy', *pyār : prem* 'love', *bhakti : bhakti* 'devotion' , *saphed : sādā* 'white', *kālā : kālo* 'black', *pilā : halud* 'yellow', etc.

The personal and kinship terms that refer to meaningful connections between the people living in a society or family are also translated from Hindi to Bangla following the process of simple lexical replacement, such as, *dādā : dādā* 'elder brother', *mā : mā* 'mother', *pāpā : bābā* 'father', *bhāiyā : bhāi*

'younger brother', *bahan : bon* 'sister', *mausi : māsi* 'aunt', *vyakti : byakti* 'individual', etc.

Finally, most of the Hindi finite and non-finite verbs are translated in Bangla with the Bangla equivalent finite and non-finite verb forms, e.g., *jāyegā : yābe, khāyegā : khābe, rote huye : kāndte thākā,* etc. The honorific expressions and titles used in Hindi are kept intact in Bangla translation, e.g., *śrī : śrī, śrīmatī : śrīmatī, śrīmān : śrīmān, kumārī : kumārī,* etc. Similarly, the Hindi numerals like *ek* 'one', *do* 'two', *tīn* 'three', *chār* 'four', *anek* 'many' are also transliterated in Bangla as *ek, dui, tin, char, anek,* etc. without translation since these forms do not require to be translated.

Translation of Proper and Common Nouns

Most of the Hindi proper nouns (e.g., place names, person names, item names, etc.) are retained unchanged in Bangla translation. That means these are only transliterated from the Devnāgarī script to the Bangla script (and not translated) simply because these are 'non-translatable' units. This is a common practice in translation across all language pairs that the proper names used in source language text should be transliterated into target language. Otherwise, it will generate many unwanted problems which may eventually kill the content and quality of a translation. This is a deliberate act of preserving proper nouns, since it does not disturb the act of translation as well as does not hamper the act of information transfer and preservation from source language to target language (Table 2).

Hindi	Bangla	English
Dilli	Dilli	'Delhi'
Vārānasī	Benāras	'Varanasi'
Mumbāi	Mumbāi	'Mumbai'
Uttarāṅchal	Uttarāṅchal	'Uttaranchal'
Hindusthān	Hindusthān	'Hindustan'

TABLE 2. Translation of proper names through transliteration

On the contrary, we translate most of the Hindi generic nouns (e.g., common nouns, material nouns, collective nouns, and abstract nouns, etc.) into Bangla, e.g., *ādmi : byekti, pāni : jal, ladki : meye, ladkā : chele, aurat : mahilā, dost : bandhu, peṛ : gāch,* etc. This is because most of these nouns are different in surface form but similar in meaning in both the languages considered in our translation. We, however, note that many English names are used in Hindi texts in Devnāgari script. At the time of translation, we reproduce these names in Bangla through transliteration (Table 3).

English	Hindi	Bangla
tomato	ṭamāṭar	ṭamyāṭo
green tea	grīn ṭī	grin ṭi
broccoli	brokalī	brākoli
vitamin	viṭāmin	bhiṭāmin
accident	aksiḍenṭ	ayaksiḍenṭ
disco	ḍisko	ḍisko
fast food	phāsṭ phuḍ	phāsṭ phuḍ
Bible	bāibel	bāibel
supreme court	suprim korṭ	suprim korṭ

TABLE 3: Transliteration of terms into Bangla directly from English

The notable thing is that at the time of translation our focus seems to be on the original English term than on the Hindi term, which is actually a transliterated form from English. Therefore, we by-passed the Hindi term to adopt the English term for better representation in Bangla.

Many common English nouns that are used in our daily life (e.g. *plate, glass, scale, rubber, company, police*, *plane*, *station, message, train, note,* etc.) are mostly kept intact. We transliterate these forms as they have already entered into our common vocabulary through the process of naturalization (Dash, Chowdhury & Sarkar 2009). Moreover, these forms do not possess any serious hindrance in comprehension. However, there are some cases, where the translation process is different. For instance, English nouns like *chair, cup,* etc. are translated in Hindi as *kursi, peyālā,* etc. but we retain them as transliterated forms in Bangla (e.g., *cheyār, kap*, etc.). This means that we use the naturalized English forms rather than using the Hindi words like *kursi* 'chair', *kedārā* 'chair', *peyālā* 'cup'. The basic strategy that we adopt in this case may be understood from the diagram given below (Fig. 2).
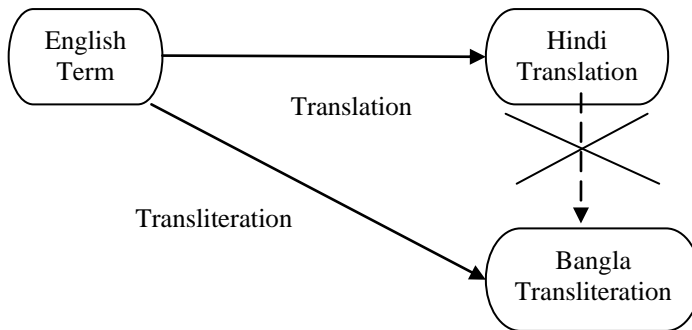

FIGURE 2. Transliterating from English, not translating from Hindi

Many scientific and technical terms of English are used in Hindi texts. Most of these terms are retained unchanged in Bangla translations. For

example, we use Bangla terms like *ḍāyābeṭik* and *kyānsār* for the English terms *diabetic* and *cancer* rather than opting for Hindi terms like *madhumeha* and *karkaṭ rog*. Similarly, we use Bangla transliterated forms like *ṭuriṣṭ* and *phaṭogrāphār*, respectively, for English *tourist* and *photographer*, rather than using the Hindi words *bhramaṇārthī* and *chitragrāhak*.

Translation of Pronouns

There is a level of semantic consistency in use of pronominal forms both in Hindi and Bangla due to which it is assumed that translation of pronominal forms between the two languages will not create many problems. In reality, however, this does not appear to be true at the time of actual translation. Although less number of problems is noted in case of translating personal pronouns between the languages, deceptive meanders are noted in case of pronouns of other classes, particularly for demonstrative, relative, and interrogative pronouns.

We consistently replace all Hindi personal pronominal forms by equivalent Bangla personal pronominal forms in translation with a lexical database of pronominal equivalent forms as the following examples show:

Dataset 1:

meine	: āmi	'I'
tu	: tui	'you'
hāmārā	: āmāder	'ours'
merā	: āmār	'mine'
meri	: āmār	'mine'
terā	: tomār	'yours'
teri	: tomār	'yours'
uskā	: or	'his'
uski	: or	'her'
unkā	: tār	'his'
unki	: tār	'her', etc.

Dataset 2:

āp	: āpni	'you' (singular)
āp	: āpnārā	'you' (plural)
tum	: tumi	'you' (singular)
tum	: tomrā	'you' (plural)
ham	: āmi	'I' (singular)
ham	: āmrā	'we' (plural)

Dataset 1 shows that simple replacement of Hindi pronominal forms with similar Bangla pronominal forms can solve the problem for a human or a machine translator. The problem arises in case of Dataset 2, as we have to decide which of the two alternative forms of Bangla should be the most appropriate translation for the Hindi form, since Bangla has two different

forms based on number, while Hindi has only one form irrespective of number. Normally, in case of manual translation, this problem is solved by a human translator just by looking at the full sentence in the source Hindi text and taking into consideration the syntactic and topical contexts surrounding the pronominal forms used in the text (Dash 2008). In case of machine translation, however, it is a tough task, since a machine cannot decide which one of the two alternatives should be accepted as an appropriate one (Bentivogli & Pianta 2000). How the lexical selection will be made is an open question for an automatic translation system.

The Hindi honorific and non-honorific pronominal forms are translated in a uniform manner into Bangla. However, an interesting thing is noted here. These pronominal forms can also act as function words. That means, both in Hindi and Bangla, a demonstrative pronoun may also be identified as a relative or indefinite pronoun in specific syntactic functions, such as the followings (1a-2c):

(1) a. Hindi : woh ladkā *jā rahā hæy*.
(1) b. Bangla : oi cheleṭi yāchhe.
(1) c. English : 'That boy is going'

(2) a. Hindi : woh *jā rahā hæy*.
(2) b. Bangla : se yāchhe.
(2) c. English : 'He is going'

Clearly, the Hindi form *woh* can be interpreted in two different ways, based on the context of its use in text, which may influence us in selection of appropriate one between two available forms in Bangla (*oi* and *se*) (one is a demonstrative and the other is a pronoun). In Hindi, the form *woh* is a single lexical item that can act as a demonstrative and as a pronoun based on the context of its use. In Bangla, there are two separate forms to mark respective functions, as the above examples show. Again, this cognitive equivalence selection problem is a challenge to a machine translation system. We are yet to come out with a solution.

The Hindi pronominal phrase *koi nahi* 'no one' can be translated into Bangla in two different ways: *keu nay* 'no one' and *keu nei* 'no one is there'. At the time of translation, we have to adopt one of the two alternatives based on the structure and meaning of the Hindi sentences (3a-5c).

(3) a. Hindi : woh aur koi nahi, Rām hæy.
(3) b. Bangla : se ar keu nay, Rām.
(3) c. English : 'He is none but Ram.'

(4) a. Hindi : who aur koi nahi, Rām thā.
(4) b. Bangla : se ar keu nay, Rām chila.
(4) c. English : 'He was no one else, but Ram.'

(5) a. Hindi   : wahā pe aur koi nahi hæy.
(5) b. Bangla : sekhāne ar keu nei.
(5) c. English : 'There was none.' Or 'None was there.'

The examples given above (3a-5c) show that translation of Hindi pronominal forms into Bangla is not an easy problem as we assumed. Rather, it asks for application of supporting linguistic knowledge and information to generate the most accurate translation.

Lexical Mismatch or False Friend

Usually, orthography does not have any role in translation (Goyal, Gupta, & Chatterjee 2004). Since a translation is not a transliteration, the role of orthography used in two different languages should not be an issue in translation. This argument is realized at the time Hindi to Bangla translation and it should be noted that during translation, we do not encounter any issue relating to the orthographic difference, except in the case of transliteration of proper names, which is a different topic beyond the scope of the present paper.

Hindi			Bangla	
pataṅg	'kite'	::	pataṅga	'insect'
sandeś	'information'	::	sandes	'sweets'
bacan	'promise'	::	bacan	'speech/number'
cāul	'boiled rice'	::	cāul	'non-boiled rice'
sucnā	'information'	::	sucanā	'introduction'
mārg	'road'	::	mārg	'means of salvation'
bāli	'tender age'	::	bāli	'sand'
ghaṛi	'hour'	::	ghaṛi	'watch/clock'
sāmānya	'general'	::	sāmānya	'meagre'
nāriyel	'green coconut'	::	nārikel	'ripe coconut'
nibedan	'appeal/request'	::	nibedan	'presentation'
āvedan	'impact'	::	ābedan	'appeal/request'
prabāl	'new leaf'	::	prabāl	'coral'

TABLE 4. Some examples of Hindi-Bangla lexical mismatch

On the other hand, we find that orthographic similarity is not a guarantee for semantic similarity. That means lexical match at the level of orthography between the two languages (i.e., Hindi and Bangla) does not guarantee that the words are similar in meaning. We observe that there are large numbers of words in Hindi and Bangla, which are orthographically similar but different in meaning. Many of them differ in sense even though they are orthographically

identical in Devanāgarī and Bangla script. For instance, examples (Table 4) show that these words are orthographically identical but semantically different in two languages. These are the instances of 'lexical mismatch' or 'false friend'.

This refers to the concept of 'semantic marking' conditioned by adjoining terms as opposed to 'syntactic markings' where particular meaning of a word that is intended clearly specifies grammatical constructs in which it occurs. Such forms create problems in bidirectional translation. It highlights a unique linguistic phenomenon where a particular word refers to a particular concept in Hindi but refers to a different concept in Bangla. That means, although a lexical item in both the languages are orthographically same (perhaps same in pronunciation and etymology also), it actually refers to two different concepts. This is a frequent translation problem which requires special attention at the time of Hindi to Bangla translation.

Syntactic Proximity

Since Hindi and Bangla are sister languages, they are supposed to have same word order in sentence formation. This can reduce much of our work load as we can adopt the process of linear arrangement of lexical items in Bangla following the word order in Hindi. However, at certain situations, we make some alternations in syntactic patterns to account for O-S-V, V-O-S, and S-O-V formations to produce acceptable constructions in Bangla to make room for stylistic variations and to account for the predicate structures of respective languages. Some of the syntactic changes that we invoked in Bangla translation are summarised below.

We translate simple declarative Hindi sentences into Bangla with minimum change keeping the structure of phrases and sequence of word order unchanged in the target language (6a-6c).

(6) a. Hindi   : shiva ballabpur nāmak sthān me Shiva nivās karte hæin.
(6) b. Bangla : Śib ballabhpur nāmak sthāne śib basabās karen.
(6) c. English : 'Lord Shiva lives at a place called Shiva Ballabhpur.'

In general, we translate most of the simple Hindi interrogative sentences into Bangla without changing their word order. The only change that we do in most of the translations is the displacement of the interrogative pronoun from the sentence-medial position in Hindi to the sentence-final position in Bangla translation, as the following example show (7a-7c).

(7) a. Hindi   : bon ṭibi kī kæyā pehchān hai tathā iskā kæyā ilāz hæy?
(7) b. Bangla : bon ṭibir lakṣaṇ kī, ār er chikitsā kī ?
(7) c. English : 'What are the symptoms and treatment of bone T.B.?'

Hindi has a useful way of forming tag questions. One can simply use the verb form *hæy* with a negative particle *nā* at the end of the declarative sentence to evoke the desired impact. In case of translation of such constructions into Bangla, we follow almost the same strategy by introducing

the negative particle *nā* at the end of the sentence preceded by a demonstrative pronoun (8a-8c).

(8) a. Hindi   : rām æisā kar sekhtā hæy, hæy nā ?
(8) b. Bangla : rām eman karte pāre, tāi nā ?
(8) c. English : 'Ram can do such things, can't he?'

However, in place of using the Hindi term (i.e., *hæy*), we use a demonstrative pronoun (i.e., *tāi*) to generate the desired impact. This serves the purpose in two ways: first, it helps us to maintain structural equivalence between source and target sentence for system training; second, it becomes a natural translation. This kind of construction is acceptable in standard Bangla.

Similar to the translation of tag questions, translation of imperative sentences from Hindi to Bangla does not ask for any extra effort from us. In a simple manner, we are able to translate these sentences into Bangla with no (or marginal) change in word order (9a-10c).

(9) a. Hindi   : is anuṣṭhān ki ānand lijiye.
(9) b. Bangla : ei anuṣṭhāner ānanda nin.
(9) c. English : "Enjoy this programme"

(10) a. Hindi   : āp kā chasme yā kanṭækṭ lens kā rakh-rakhāb karen.
(10) b. Bangla : āpnār chasmā bā kanṭyākṭ lenser rakṣaṇābekṣaṇ karun.
(10) c. English : "Take care of your spectacle or contact lens."

Similarly, translation of exclamatory sentences from Hindi to Bangla hardly requires extra labour except on the choice of appropriate lexical items from the target language. Since this construction follows the strategies already applied to other types of sentences stated above, we do not discuss them here.

Syntactic Divergence

Translation from Hindi to Bangla reveals some interesting information which challenges age-old belief. We find that 'syntactic convergence' (i.e., structural proximity at sentence level) is less than 'syntactic divergence' between the languages. That means conceptual equivalence among sentences of the two languages does not confirm their structural equivalence (Hurtado de Mendoza 2008). This leads us to adopt various linguistic strategies to produce structurally and semantically equivalent sentences in Bangla for the input Hindi sentences, some of which are discussed below.

In case of multiword Hindi verb forms, we use single-word verb for Bangla, because by using a single-word verb, we can easily capture the sense and translate the idea expressed by the Hindi multiword-verb form (11a-11c).

(11) a. Hindi   : rām ne ye khānā khā rahā thā.
(11) b. Bangla : rām ei khābār khācchila
(11) c. English : 'Ram was eating food'

The example shows that the multiword verb form *khā rahā thā* can be rightly translated into Bangla by a single inflected form *khācchila*, which is fit to capture the entire load of TAM of the Hindi verb form. This gives us a strategic advantage in translation of sentences of this type, as replacement of multiword verb forms by a single-word inflected verb form becomes a common method of translating Hindi sentences into Bangla (MWVF = SWVF).

The Hindi case marker *-ne* is normally used in the ergative sense in present and past perfect forms of the main verb. We translate this into Bangla with a zero case marker (12a-13c). It helps us to formulate a translation rule: whenever there is an ergative construction in Hindi with a case marker *-ne*, the marker can be dropped in Bangla and the subject can be used in its non-inflected form. This happens because Bangla drops the ergative case marker.

(11) a. Hindi   : Ram ne bolā.
(12) b. Bangla   : Ram balla.
(12) c. English   : 'Ram said.'

(13) a. Hindi   : main ne śunā hæy.
(13) b. Bangla   : āmi śunechi.
(13) c. English   : 'I have heard'

It is linguistically ascertained that because of the feature of grammatical gender, the gender markers in Hindi agree with final forms of the main verbs (Sinha & Thakur 2005b). This brings in a kind of grammatical constraint in formation of Hindi sentences where gender is an integral part of the construction. Since this phenomenon is not observed in Bangla syntax, we have liberty to ignore the use of gender marker with words while translating Hindi sentences (14a-15c). The most important part of this is that it generates a well-formed rule in Hindi to Bangla translation in the manner that Hindi gender markers can totally be ignored in Bangla translation because while Hindi is a gender-sensitive language, Bangla is gender neutral, even though both are so-called 'sister languages'. Parental property is a costly heritage which is difficult to carry through generations with equal manifestation in every successor.

(14) a. Hindi   : Oh *ladki jāti* hæy.
(14) b. Bangla   : oi meyeṭi *yācche*.
(14) c. English   : 'That girl is going'

(15) a. Hindi   : Oh *ladka jātā* hæy.
(15) b. Bangla   : oi cheleṭi *yācche*.
(15) c. English   : 'That boy is going'

Another regular feature of Hindi is that in case of a negative sentence, the negative particle usually comes immediately before the main verb. This is different from Bangla, because the negative particle comes immediately after the main verb in Bangla (16a-17c). Therefore, we adopt it as a rule and apply

it in a consistent manner to translate negative Hindi sentences with accuracy and naturalness.

(16) a. Hindi	: main ghar nahi jaungā
(16) b. Bangla	: ami bāri yābo nā
(16) c. English	: 'I shall not go home'

(17) a. Hindi	: āpne eisa chitra kabhi nahi dekhā
(17) b. Bangla	: āpni eman chabi kakhono dekhen ni
(17) c. English	: 'You have never seen such a picture'

Hindi has a set of plural suffixes to be used with countable nouns, namely, *-iya* (e.g., *ladkiyā* "the girls", etc.), *-on* (e.g., *ladkon* "the boys", etc.). Similarly, there are specific number- denoting affixes to be tagged with countable nouns in Bangla: (a) singular makers: *-tā, -ti, -khānā, -khāni, -tuku,* etc. and (b) plural markers: *-guli, -gulo, -gulā, -diga, -der,* etc. At the time of translating these words from Hindi to Bangla, the omission of singularity marker in Bangla may result in two different readings of the source Hindi sentence causing ambiguity (18a-19c).

(18) a. Hindi	: ladke ne ladki ko phul diyā.
(18) b. Bangla	: chelerā meyederke phul dila.
(18) c. English	: 'Boys gave flowers to girls'

(19) a. Hindi	: ladke ne ladki ko phul diyā.
(19) b. Bangla	: cheleṭi meyeṭike phul dila.
(19) c. English	: 'The boy gave flower to the girl'

The examples show that the same Hindi sentence can have two different possible translations in Bangla. Therefore to avoid ambiguity, we opt to tag *-tā* or *-ti* with Bangla nouns (agent and recipient) to dissolve ambiguity and to generate acceptable translations.

Bangla nouns and pronouns, based on their 'case identity' in sentences can take specific case markers (i.e., *-ke, -e, -te, -ye, -āy, -er, -r,* etc.). They can also take emphatic particles (i.e., *-i, -o,* or *-to,* etc.) immediately after the case markers. The use of such forms shows notable variations in respect to the two languages considered in translation. This phenomenon needs to be accounted for at the time of translation. Since this is a wide topic of discussion, we opt this out from the present paper.

The Copula Conflict

The copula is a deceptive lexical entity. It is very prominent and vibrant in Hindi but mostly silent and inactive in Bangla. Thus it creates strong challenges in translation between the two languages. In Hindi, sentences in simple present tense require the presence of copula *hæy* "is" preceded by a verb, a sequence quite different from Bangla. In Bangla, sentences in simple present tense do not need a copula (most often it does not require for

declarative or existential sentences). The copula *hæy* in Hindi has, therefore, different types of manifestation in Bangla translation.

Type (a): Redundancy of *hæy*

The Hindi copula *hæy* may be an optional element in Bangla translation because it is not necessary to be represented in a declarative Bangla sentence (20a-21c).

(20) a. Hindi	:	manobhāv par niyantraṇ rākhnā kaṭhin hotā *hæy*.
(20) b. Bangla	:	manobhāber upar niyantraṇ rākhā kaṭhin.
(20) c. English	:	'It is difficult to control mentality'
(21) a. Hindi	:	sundarban pakṣī premio ke liye swarg ki tarah *hæy*.
(21) b. Bangla	:	sundarban pakhipremikder kāche svarger mato.
(21) c. English	:	'Sundarban is a heaven for the bird lovers'

In these sentences, it is clear that we can translate a Hindi sentence into Bangla quite rightly without translating the Hindi copula *hæy*. This does not minimize the naturalness of a translation.

Type (b): Presence of *hæy*.

For some sentences, the presence of *hæy* in Hindi has to be retained and rightly represented in Bangla translation. Otherwise, the original sense of the Hindi sentence cannot be properly captured in translation. If we look at the following examples (22a-22c), we can see that Hindi *hæy* is represented as *hala* in Bangla. The most interesting thing is that the Hindi sentence is written in the present tense, while Bangla translation is produced in the past tense. Even then, the past form of the copula (i.e., *hala*) is actually denoting a present sense. This is a unique feature of the Bangla language, where in some situations, the past form of a verb can denote present sense, and in a reverse manner, the present form of a verb can denote past sense.

(22) a. Hindi :	sundarban kā sabse romāñchak hissā nadī kī yātrā *hæy*.
(22) b. Bangla :	sundarbaner sab theke romāñchakar byāpār *hala* nadīte bhramaṇ.
(22) c. English :	'The most exciting event of Sundarban is a journey by the river'

Type (c): Hindi *hæy* as Bangla *pāre*

In some sentences, we translated the Hindi *hæy* as a modal verb (i.e., *pāre*) in Bangla. Then we couple it with the preceding main verb in Bangla translation (23a-23c).

(23) a. Hindi :	is rāste se vi Banāras siṭi hote huye Chābnī jā sakte *hæy*.
(23) b. Bangla :	ei rāstā dhareo Benāras śahar haye Chābnī *yete pāren*.
(23) c. English :	'By this road (you) can go Chabni through Varanasi city'

Type (d): Hindi *hæy* as Bangla *āche*

In some sentences we translate the Hindi copula *hæy* with the existential verb *āche* "has" in Bangla as the following examples show (24a-24c).

(24) a. Hindi : Mugalsarāi sṭeśan se 128 māil kī durī par Ayodhyā lāin *hæy*.

(24) b. Bangla : Mogolsarāi sṭeśan theke 128 māil duratve Ayodhyā lāin *āche*.

(24) c. English : 'The Ayodhya line is located 128 miles away from Mughalsarai station'

Type (e): Hindi *hæy* as Bangla *rayeche*

In a similar manner, we translate the Hindi copula *hæy* as *rayeche* "is present" in case of some sentences in Bangla (25a-25c).

(25) a. Hindi: chandra pravu digambar jain kṣetra Rājasthān kī Alwār jile me sthita *hæy*.

(25) b. Bangla : Chandra prabhu digambar jaina kṣetra Rājasthāner āloyār jelāy abasthita *rayeche*.

(25) c. English : 'Chandraprabhu Digambar Jaina place is located in the Alwar district of Rajasthan.'

Type (f): Hindi *hæy* as Bangla *geche*

Finally, in some sentences, we render the Hindi copula *hæy* as Bangla verb *geche* "has gone" in Bangla translation (26a-26c).

(26) a. Hindi   : Bhopāl sṭeśan se ek lāin Ujjain jāti *hæy*.

(26) b. Bangla : Bhupāl sṭeśan theke ekṭi lāin ujjayinī *geche*.

(26) c. English : 'A track runs towards Ujjain from Bhopal station'

The types of translation of Hindi copula *hæy* into Bangla are neither exhaustive nor final. It is far more diversified and complex than we encounter and present here. In fact, we can find out many more types and subtypes of the phenomenon if we analyse and translate a large number of Hindi *hæy* constructions into Bangla and other Indian languages (including English). This is just a tip of an iceberg, which is hinted here with an expectation that this particular feature will be taken into further exploration in translation between Hindi and Bangla.

The sub-classifications of *hæy* presented above and their corroborative examples indicate that the translation of the Hindi copula (i.e., *hæy*) into Bangla is a challenging task as this does not follow any predefined patterns or rules (Chesterman 1993). While in some cases (type a), it is worth removing, in some cases its presence is mandatory, and in some other cases, it can be rendered into various other lexical realizations in Bangla to make the translations acceptable. This creates a tough challenge to an MT system as it fails to formulate any rule-based pattern for translating the Hindi copula into

Bangla. We require more rigorous analysis of the types with examples obtained from different Hindi-Bangla parallel translation corpus to understand its functional diversities, which, however, it is not attempted in this paper.

Conclusion

Translation is a challenging task because we have to retain semantic, syntactic, and stylistic equivalence of the source text into the target text (Sinha & Thakur 2005a). In general, the most problematic area we note in translation is the single words and their roles in creating deviations based on the context of their use in texts. Similar observation also stands valid in case of compounds and multiword units which also exhibit semantic deviations in accordance with their usage in different kinds of text (Cao 1996). Therefore, we argue that at the time of translation we must take into account all kinds of lexical whimsicalities before we elicit appropriate equivalence in the target language.

We apply some methods and strategies to deal with specific lexical and syntactic problems in translation of Hindi texts into Bangla. We do this keeping manual and machine translation in mind, since machine translation between Hindi to Bangla (bidirectional) is going to grow as a new area of technical and commercial exploration. This inspires us to look into the issues and challenges involved in translation and present these in this paper with some examples taken from Hindi-Bangla parallel translation corpus. From this study, future translators can know what kinds of problem they are going to face and what methods they should apply to solve the problems. Moreover, people engaged in developing Hindi to Bangla machine translation technology can also use this information and insight for successful training of their systems.

References

BENTIVOGLI, L., AND PIANTA, E. 2000. Looking for Lexical Gaps. *Proceedings of the Ninth Euralex International Congress.* 663-660. Stuttgart: Germany.

CAO, DEBORAH. 1996. Towards a Model of Translation Proficiency. *Target* 8(2). 325-340.

CHESTERMAN, ANDREW. 1993. From 'is' to 'ought: Translation Laws, Norms, and Strategies. *Target* 5(1). 1-20.

DASH, NILADRI SEKHAR. 2008. Context and Contextual Word Meaning. *Skase Journal of Theoretical Linguistics* 5(2). 21-31.

DASH, NILADRI SEKHAR, DUTTA CHOWDHURY, PAYEL AND SARKAR, ABHISEK. 2009. Naturalization of English Words in Modern Bangla: A Corpus-based Empirical Study. *Language Forum* 35(2). 127-142.

DASH, NILADRI SEKHAR. 2015. Syntactic, Lexical, and Semantic-cum-Cultural Barriers in English to Bengali Translation. *International Journal of Translation* 27(1&2). 49-73.

DAVE, S., PARIKH, J., and BHATTACHRAYYA, P. 2001. Interlingua Based English-Hindi Machine Translation. *Journal of Machine Translation* 16(4). 251-304.

DORR, BONNIE, AYAN, N. and HABASH, N. 2004. Divergence Unraveling for Word Alignment of Parallel Corpora. *Natural Language Engineering.* 1(1): 1-17.

DORR, BONNIE. 1992. The Use of Lexical Semantics in Interlingual Machine Translation. *Machine Translation* 7(1). 135-193.

DORR, BONNIE. 1994. Machine Translation Divergences: A Formal Description and Proposed Solution. *Computational Linguistics* 20(4). 597-633.

GOYAL, S., GUPTA, D. and CHATTERJEE, N. 2004. A Study of Hindi Translation Patterns for English Sentences with 'have' as the Main Verb. *Proceeding of iSTRANS,* 1. 46-51.

GUPTA, D. and CHATTERJEE, N. 2001. Study of Divergence for Example-Based English-Hindi machine translation. *Proceedings of the STRANS 2001,* IIT-Kanpur.132-139.

GUPTA, D. and CHATTERJEE, N. 2003. Identification of Divergence for English to Hindi Ebmt. *Proceeding of MT Summit-IX.* 41-48.

HURTADO DE MENDOZA, A. 2008. The Problem of Translation in Cross-Cultural Research on Emotion Concepts. *International Journal of Dialogical Science* 3. 241-248.

JANSSEN, MAARTEN. 2004. Multilingual Lexical Database, Lexical Gaps, and Simullda. *International Journal of Lexicography* 17(1). 136-154.

OROZCO, MARIANA, and AMPARO, HURTADO ALBIR. 2002. Measuring Translation Competence Acquisition. *Meta* 47(3). 375-402.

SANTOS, DIANA. 1990. Lexical Gaps and Idioms in Machine Translation. *Proceedings of Coling* 2. 330-335. Helsinki: Finland.

SINHA, R.M.K. and THAKUR, A. 2004. Synthesizing Verb Form in English to Hindi Translation: The Case of Mapping Infinitive and Gerund in English to Hindi. *Proceeding of iSTRANS* 1. 52-55.

SINHA, R.M.K. and THAKUR, A. 2005a. Translation Divergence in English-Hindi Machine Translation. *Proceeding of EAMT-10th Annual Conference.* 245-254. Budapest: Hungary.

SINHA, R.M.K. and THAKUR, A. 2005b. Divergence Patterns in Machine Translation between Hindi and English. *Proceeding of MT Summit-X.* 346-353. Phuket: Thailand.

TOURY, GIDEON. 1995. *Descriptive Translation Studies and Beyond.* Amsterdam and Philadelphia: Benjamin.

***