Journal of Multimedia Information System VOL. 8, NO. 4, December 2021 (pp. 221-232): ISSN 2383-7632 (Online)
http://doi.org/10.33851/JMIS.2021.8.4.221

Hindi Correspondence of Bengali Nominal Suffixes

Sanjay Chatterji*

Abstract

One bottleneck of Bengali to Hindi transfer based machine translation system is the translation of suffixes of noun. The appropriate translation of a nominal suffix often depends on the semantic role of the corresponding noun chunk in the sentence. With the availability of a high performance Bengali morphological analyzer and a basic Bengali parser it is possible to identify the role of each noun chunk. This information may be used for building rules for translating the ambiguous nominal suffixes. As there are some similarities between the uses of Bengali and Hindi nominal suffixes we find that the rules may be identified by linguistically analyzing corpus data. In this paper, we identify rules for the ambiguous four Bengali nominal suffixes from corpus data and evaluate their performances. This set of rules is able to resolve a majority of the nominal suffix ambiguities in Bengali to Hindi transfer based machine translation system. Using the rules, we are able to translate 98.17% Bengali nouns correctly which is much better than the baseline ILMT system's accuracy of 62.8%.

Key Words: Nominal Suffix, Chunk, Transfer based Machine Translation, Bengali, Hindi.

I. INTRODUCTION

In Bengali and Hindi, a noun chunk comprises a noun or pronoun root, its inflections and postpositions, adjectives and demonstratives. In a typical Rule Based Machine Translation System (RBMT), the exact sense of the root part of the noun, the pronoun, the adjective and the demonstrative in a sentence are identified by a Word Sense Disambiguation (WSD) module and then they are translated using a bilingual root dictionary. However, the suffixes are generally not disambiguated by the WSD module. In this article, we address the problem of handling the Hindi translation of Bengali nominal suffixes.

Our study reveals that translation of suffixes from Bengali to Hindi is not one to one. The appropriate translation often depends on the semantic role of the noun chunk in the sentence. Automatic identification of semantic role is a non-trivial task and it can be facilitated by finding different features of the noun chunks. For example, consider the Bengali suffix -ra attached with a noun in the Bengali sentence "rAmera khide peYechhe." [Ram is hungry.]. The corresponding noun in this sentence acts as subject and this suffix should be translated to the Hindi marker -ko. But, the Bengali noun with suffix -ra in the Bengali sentence "rAmera chhele Achhe." [Ram has a son.] has possession role and should be translated to the Hindi marker -kA.

We wish to study the morphological, syntactic and semantic features of Bengali language. We also wish to study the application of different suffixes in different values of these features. Then, by observing the Hindi translations of Bengali suffixes, we wish to formulate rules for each translation. In the rule, we use only those features which are required to disambiguate the current translation from other translations.

Bengali suffixes used for indicating cases can be of four types, namely, null-marker (also referred to as 0-marker), ra-marker, ke-marker and te-marker. Each type may have different surface forms for singular and plural numbers.

In Bengali, the nominal inflections for case and number are handled together. The surface forms of the suffix indicating case and number depend on the orthographic forms of the root with which the suffix is attached. In this paper the rules for translating each of the four suffix markers are described with related examples.

In Bengali sentences, a suffix may be used in multiple roles. For example, in the Bengali sentence "madhu parIkShA dite dillI yAchchhe." [Madhu is going to Delhi to sit for examination.] the noun chunks "madhu" [Madhu], "parIkShA" [examination], and "dillI" [Dehli] all have -0 suffix but their respective roles in this sentence are subject (karta), reason (hetu) and locative (adhikaran). The noun with a specific role may contain multiple bibhaktis. For example in the Bengali sentences "bA.Dite yAchchhi." [I am going home.] and "bA.Di yAo." [Go to home.], the roles of "bA.Dite" having -te suffix and "bA.Di" having -0 suffix both are locative (adhikaran).

The remaining of this article is arranged as follows. Some work related to rule based translations are presented in Section 2. The morphological features of Bengali nouns

Manuscript received October 24, 2021; Revised November 18, 2021; Accepted November 20, 2021. (ID No. JMIS-21M-11-036)

Corresponding Author (*): Sanjay Chatterji, Department of Computer Science and Engineering, Indian Institute of Information Technology Kalyani, West Bengal, India, sanjayc@iiitkalyani.ac.in

are gender, number, person, specificity, and emphasizer and the syntactico-semantic features are case and nominal modifier. These features are indicated by different suffixes. We discuss the possible suffixes for each such features and the possible roles in Section 3. The overall technique, corpus and rule formats are described in Section 4.

The rules are described with related examples in Section 5-8. The evaluation is carried out using human annotation guided by a parallel corpus similar to the technique presented by Condon et al. [11].

In Section 9, we conclude the article. In this paper Bengali (B:) input sentences, Hindi (H: and TH:) translations and Output Hindi (OH:) sentences are written in Indian language TRANSliteration (ITRANS) [10] format.

II. RELATED WORKS

A parallel corpus based approach for disambiguating the prepositions and case suffixes for Basque language has been proposed by Agirre et al. [1]. This general approach could also be applied to find translation equivalences for prepositions of any language. A detailed study related to the orthographic changes due to addition of suffixes with Bengali words is discussed by Bhattacharya et al. [5].

In the last three decades, multiple architectures have been proposed for building machine translation systems. The two main paradigms of machine translation approaches are linguistic based paradigm and non-linguistic based paradigms [13]. Rule Based Machine Translation (RBMT) is a linguistic based paradigm which relies on built-in linguistic rules and dictionary entries [18].

Rosetta [2] is an RBMT system which implements 'isomorphic transfer grammar method' which is based on the Montague's compositional grammar [16]. Om Transliteration [17] is a rule based machine translation system for Indian languages. Recently, a common transfer based machine translation framework has been proposed by Sangal [20] and has been implemented for some major Indian languages under Indian Language to Indian Language Machine Translation (ILMT) project funded by Deity, MCIT, Government of India. In the ILMT framework, there are three parts, namely, source language analyzer, source to target language transferor and target language generator.

Statistical tools are widely used in Natural language processing [21-26]. However, Harris [14] has suggested that the grammatical differences of a pair of languages may be identified using a set of rules called transfer grammar rules. Klima [15] has discussed the correspondence between English and some other languages using a set of transfer grammar rules.

Language dependent grammatical rules may be used for transferring the structures of the source language to the structures of the target language in the ILMT framework. Avinesh [3] has developed a transfer grammar component for the Indian languages which is able to process such rules. Some work has also been done on transfer grammar rules for translation between Bengali Hindi language pair. Chatterji et al. [7] have identified the grammatical rules for translating Hindi pronouns to possible Bengali pronouns. Dash [12] and Prasad [19] have investigated the uses of Hindi and Bengali pronouns in corpus data, respectively.

III. IMPLICATIONS OF BENGALI SUFFIXES

In this section, we discuss the uses of Bengali suffixes in different dependency relations. We first study, for each of the dependency relations which are the possible suffixes that may be used. Then, we study how they are used to indicate the dependency relations.

3.1. Suffixes of Bengali Cases and Nominal Modifiers

The markers applicable to each of the Bengali cases and nominal modifiers are discussed below.

When karta acts as the doer of an action then it may take null-marker or te-marker. The karta which experiences something may take ra-marker. Some karta of the passive verb and causative verb may take te-marker. Other kartas, including the complement of the karta, take null-marker.

The karma of the transitive verb takes null-marker or ke-marker. In plural number, such karmas may take null-marker or ke-marker with plural specificity (gulo) or the plural forms of ra-marker (dera). With ditransitive verbs, the direct, indirect, purposive and predicative objects take null-marker, ke-marker, ke-marker and null-marker, respectively. The te-marker is often also used with karan karak.

The adhikaran may indicate place, time, domain and state of the action. The te-marker and null-marker are used in such cases.

The nouns indicating "reason" use te-marker. The nouns indicating destination also take te-marker. The nouns which are used in comparison with karta, take the combinations of suffix and postposition (written with an underscore) "ra_theke", "ra_cheYe" or "ra_tulanAYa". Similarly, the nouns which are used in similarity with karta, take the suffix-postposition combinations "ra_mata" or "ra_samAna". The sambandha modifier takes ra-marker. The sanyogmulak modifier takes null-marker.

3.2. Use of Different Markers in Bengali Cases

A suffix marker may be used with multiple cases and with multiple nominal modifiers. We present the possible cases and modifiers with which the Bengali nominal markers are attached in Table 1.

Table 1. Possible Attachments of Bengali Markers which use Suffixes.

Suffix	Possible case of the noun chunk
null-marker	Karta, Karma, Adhikaran, Sanyogmulak Modifier
ra-marker	Karta, Karma, Sambandha Modifier
ke-marker	Karta, Karma
te-marker	Karta, Karan, Adhikaran, Reason, Destination, Sanyogmulak

IV. TECHNIQUE USED FOR FINDING HINDI TRANSLATION OF BENGALI MARKERS

We wish to find appropriate translations of the Bengali markers to Hindi. The markers which have unique Hindi translation may be translated using a parallel list of Bengali and Hindi markers. Our objective is to formulate rules for finding the appropriate translation of each marker in its context.

4.1. Mapping of Bengali Suffix to Hindi Counterpart

A Bengali suffix may be translated to a Hindi suffix or a combination of Hindi suffix and postposition as shown in Example 1 (a) and (b), respectively.

Example 1
(a) B: tumi AmAke ekaTi ba;i dAo. [You give me a book.]
H: tuma mujhe eka kitAba do.
(b) B: ei mandire ekaTi kAlira mUrti Achhe. [There is an idol of Kali in this temple.]
H: isa mandira me.N eka kAli kI mUrti hai.

Similarly, a combination of Bengali suffix and postposition may be translated to a Hindi suffix or a combination of Hindi suffix and postposition as shown in Example 2 (a) and (b), respectively.

Example 2
(a) B: AmAra dbArA ei saba habe nA. [These things can not be done by me.]
H: mujhase yaha saba nahI hogA.
(b) B: bA.Di theke be.DiYe gelAma. [I went out from home.]
H: mai ghara se nikala gayA.

4.2. Overview of the Bengali Nominal Marker Tran slation Technique

There are four suffix markers (See Table 3), namely, nullmarker, ra-marker, ke-marker and te-marker. We have identified the different contexts of use of the suffixes and formed rules for translating them in each of the contexts.

Among the suffix markers, the null-marker occurs most frequently. It may be used to indicate karta, karma, adhikaran and sanyogmulak relations. Depending on different roles it may be translated to the Hindi null-marker or to one of the Hindi postpositions “ne”, and “ko”.

Bengali ra-marker may be used with karta, and sambandha. Depending on role, the Bengali ra-marker suffixes -ra, -era, -Yera, -dera, -edera, and -Yedera may be translated to Hindi postpositions “kA”, “ki”, and “ko”.

Bengali ke-marker may be used with karta and karma. Depending on their roles, the Bengali ke-marker suffixes -ke, -derake, -ederake, and -Yederake may be translated to Hindi markers -ko, -e, etc.

Bengali te-marker may be used with karta, karan and adhikaran karaks and with nouns indicating reason and destination. It is also attached with sanyogmulak modifers. Depending on their role, the Bengali te-marker suffixes -te, -ete, and -Yete may be translated to Hindi null-marker or postpositions “sAtha”, “kA”, “se”, “me.N”, “pe”, and “lie”.

4.3. Mapping of Bengali Suffix to Hindi Counterpart

We used a corpus to help us formulate the rules for translation of ambiguous nominal suffixes. We refer to this as the *validation corpus*. This validation corpus is POS annotated. We have worked on translation of these ambiguous suffixes. We have been able to come up with rules for appropriate translation of the suffixes.

We use the KGPBenTreebank corpus of [8] for testing the effects of these rules. To identify the roles of the suffixes, the rules use dependency features of this treebank and POS features, named entity features, other morphological features, and chunk features provided by the analyzer modules of the DMT system.

We observed that there may be two reasons of the mistakes. The features returned by the modules may be incorrect or the rule required to translate the marker may be incorrect. The number of mistakes and the reasons are discussed against each marker.

4.4. Format of the Transfer Rules

We formulate some rules for translating Bengali suffixes to Hindi markers. The rules have two parts: LHS (Left Hand Side) and RHS (Right Hand Side). The LHS represents the role of the Bengali suffix and the RHS represents the changes that need to be done in its translation. The LHS and RHS parts are separated by ⇒ symbol. The format of the rule is shown below.

Rule Format SM SC <SW SF1=SV1...> (HW HF1=HV1...) ((HHW HHF1= HHV1 ...)) {DW DF1=DV1...} [FW1 FF11=FV11... FW2 FF21=FV21... ...] ⇒ TM

Each rule has a rule number. The rule set is prepared considering Bengali as the source language and Hindi as the target language. The LHS indicate the features of the

Bengali language that need to be satisfied to fire the rule and RHS indicate the Hindi translation. There are two mandatory fields and four optional fled in LHS. The terms used in each field are defined below.

- In LHS, SM stands for source language suffix. SC stands for the category of the source language word with which the suffix is attached.
- SW, SF and SV stand for source language word with which the marker is attached, the feature of that word and corresponding value.
- HW, HF and HV are the head word to which the source language word is related, feature of the head word and value of the corresponding feature.
- HHW, HHF and HHV are the head word of the word to which the source language word is related, feature of the head of head word and value of the corresponding feature.
- DW, DF and DV are the dependent word which is the dependent of source language word, feature of the dependent word and value of the corresponding feature.
- In RHS, the TM stands for target language marker.

There may be multiple (feature = value) combinations any or all of which need to be satisfed and a feature may have any of the multiple values. The (feature = value) combinations are separated by || (OR) or && (AND) symbols to indicate any and all. 3 dots (...) are used to indicate multiple (feature = value) combinations and multiple felds. The (feature = value) combinations shown for source words (as a condition), is also used for target words (as an effect).

In the following sections, we first discuss the translation rules of Bengali case markers. The rules are fired in the order of specificity. More specific rules are fired before more general rules. The effects of the rules are compared with the baseline ILMT rule based machine translation system.

V. HINDI TRANSLATIONS OF BENGALI NULL-MARKERS

We now study the translation of Bengali null-marker to Hindi. A Bengali null-marker may be translated to the Hindi null-marker or to one of the Hindi postpositions “ne”, and “ko”.

When the Bengali null-marker suffix is attached to a word which is the karta of an intransitive verb then the marker is translated to Hindi suffix -0. It is translated to Hindi suffix -0 when it is attached to a word which is the karta of a transitive verb which is not in past perfect tense. We refer both of these two uses of null-marker to null (1). The Bengali null-marker is translated to Hindi postposition “ne” when the head word, with which the marker is attached, is a karta of the verb which is in the past perfect tense. We refer to such use of null-marker as null (2). The Bengali null-marker attached to a non-specific object, is translated to Hindi suffix -0. We refer to such null marker use as null (3). When the specificity value of the Bengali null-marker attached to the object of transitive verb is true it is translated to the Hindi postposition “ko”. We refer to such null-marker use as null (4). When the null-marker is attached to a word which is the adhikaran, part of complex predicate or nominal modifier then it is translated to Hindi suffix -0. We refer to such uses of null-marker as null (5). We now discuss the rules for translating Bengali nullmarker and show their effects in translating Bengali sentences.

5.1. Rule for translation of null (1) marker

The rules for translating null (1) are presented in Rule 1.
Rule 1. null n|pn <drel=k1> (lcat=v&&type=intrans) ⇒ 0*
null n|pn <drel=k1> (lcat=v&&type=trans &&tense=not past_perfect) ⇒ 0*

In the Bengali clause of Example 3, the karta is attached to the transitive verb which is in simple past tense. So this is null (1) marker and the second rule of Rule 1 is used to translate this marker.
Example 3
B: mohana gItikAke balalo. [Mohan said to Gitika.]
H: mohana gItikA ko kahA.

5.2. Rule for translation of null (2) marker

The rule for translating null (2) is presented in Rule 2.
Rule 2. null n|pn <drel=k1> (lcat=v&&type=trans &&tense=past_perfect) ⇒ 0_ne*

The karta of the Bengali sentence of Example 4, is attached to the verb which is in past perfect tense. Hence, this is null (2) marker and Rule 2 is used to translate this marker.
Example 4
B: rAma ji~NgAsA karechhilo. [Ram asked.]
H: rAma ne puchhA thA.

5.3. Rule for translation of null (3) marker

The rule for translating null (3) is presented in Rule 3.
Rule 3. null n|pn <specificity=false&&drel=k2> ⇒ 0*

In the Bengali sentence of Example 5, the karma is nonspecific. So this is null (3) marker and Rule 3 is used to translate this marker.
Example 5
B: sandIpana mAchha khAchchhe. [Sandipan is eating fish.]
H: sa.NdIpana machhalI khA rahA hai.

5.4. Rule for translation of null (4) marker

The rule for translating null (4) is presented in Rule 4. Here, k2* indicates any subdivision of karma.
Rule 4. null n|pn <specificity=true&&drel=k2> ⇒ 0_ko*

In the Bengali sentence of Example 6, the karma is specific. So this is null (4) marker and Rule 4 is used to translate this marker.
Example 6
B: sandIpana mAchhaTA dekhachhe. [Sandipan is looking at the fsh.]
H: sa.NdIpana machhalI ko dekha rahA hai.

5.5. Rule for translation of null (5) marker

The rule for translating null (5) is presented in Rule 5. Here, k7* indicates any subdivision of adhikaran.
Rule 5. null n|pn <drel=k7|pof|nnmod> ⇒ 0*

In the Bengali sentence of Example 7, one null-marker is attached with an adhikaran. So this is (5) marker and Rule 5 is used to translate this marker.
Example 7
B: Ami bA.Di yAchchhi. [I am going to home.]
H: mai ghara jA rahA hu.N.

5.6. Evaluation and Analysis of Bengali null-marker Translation Rules

Null-marker is highly frequent nominal marker in the corpus. We have analyzed 50 Bengali sentences of the KGPBenTreebank for observing the effects of the null-marker translation rules. The number of occurrences of each of the null-marker are shown in Table 2.

Table 2. Number of Occurrences of Different Uses of null-marker in 50 Bengali Sentences of KGPBenTreebank.

null-marker	# of Occur.	null-marker	# of Occur.
null(1)	35	null(5) with adhikaran	3
null(2)	18	null(5) with pof	8
null(3)	16	null(5) with nmod	6
null(4)	2		

The baseline system translates each null-marker to the most frequent Hindi suffix -0. The number of correct translation of null-marker by the baseline system and by the proposed method are shown in Table 3.

Table 3. Number of Correct Translation of null-marker in 50 Bengali Sentences of KGPBenTreebank by the Baseline and Proposed Rule Based Systems.

System	# of Correct Translations out of 88 null-markers
Baseline	68
Proposed	83

Among all these null-markers, there are 5 mistakes while translating them using the proposed rules. The mistakes are analyzed below.

Sometimes the -TA suffix is added with objects not as specifier. In such cases the Rule 4 is fired wrongly. One incorrect translation due to this mistake is shown in Example 8 (a).

There are some other cases, where -TA is used as specifier, but the null-marker does not follow the Rule 4. One incorrect translation due to this mistake is shown in Example 8 (b). In both the sentences the "ko" postposition should not be used, as shown in correct Hindi translation.
Example 8
(a)B: tumi kichhuTA kheYe nAo. [You eat something.]
OH: tuma kuchha ko khA lo.
TH: tuma kuchha khA lo.
(b)B: chhe.De yAoYATA uchita naYa. [You should not leave.]
OH: chho.Da jAnA ko uchita nahI.N hai.
TH: chho.Da jAnA uchita nahI.N hai.

VI. HINDI TRANSLATIONS OF BENGALI ra-MARKER

We now study the translation of Bengali ra-marker to Hindi. A Bengali ra-marker may be translated to Hindi postpositions "kA", "ki", or "ko".

The Bengali ra-marker in singular form is translated to the combination of Hindi suffix -0 and Hindi postposition "kA" when the corresponding chunk is attached to a masculine noun chunk by genitive (sambandha) dependency relation. This is translated to the combination of Hindi suffx -0 and Hindi postposition "ki" when the corresponding chunk is attached to a feminine noun chunk by genitive (sambandha) dependency relation. In both the cases, the noun with which the ra-marker is attached is nonspecific and in direct (not oblique) form. We refer to both of these two uses of ra-marker as ra(1).

Sometimes the noun to which the ra-marker noun is attached by r6 relation is not present in the sentence. In that case the ra-marker noun does not have r6 relation. It takes the relation taken by the absent noun. We refer to such uses of the ra-marker as (2). The noun with which such ra-marker is attached is in oblique form. Often we do not get the gender feature of such absent noun. In that case, (2) is translated to the general Hindi marker 0_kA.

Bengali uses ra-marker for its dative karta constructions. We refer to such use of ra-marker as (3). Dative relation is not used in the dependency set. It is considered as experiencer karta (3) is translated to the combination of Hindi suffix -0 and Hindi postposition "ko".

Bengali ra-marker is also used with karma. We refer to such ra-marker as ra(4). Similar to ra(3), ra(4) is also translated to the combination of Hindi suffix -0 and Hindi postposition “ko”. We now discuss the rules for translating Bengali ra-marker and show their effect in translating Bengali sentences.

6.1. Rule for translation of ra(1) marker

The rules for translating (1) are presented in Rule 6.
Rule 6. ra n|pn <drel=r6&&form=direct> (lcat=n|pn&& gender=masculine) ⇒ 0_kA
ra n|pn <drel=r6&&form=direct> (lcat=n|pn&& gender=feminine) ⇒ 0_ki

In the Bengali sentence of Example 9, the ra-marker noun is attached to a feminine noun by r6 dependency relation. So this is ra(1) marker and Rule 6 is used to translate this marker.
Example 9
B: jitera didi khuba bhAlo meYe. [Jit’s elder sister is a very good girl.]
H: jIta ki ba.DI bahena bahuta achchhI la.DakI hai.

6.2. Rule for translation of ra(2) marker

The rule for translating ra(2) is presented in Rule 7.
Rule 7. ra n|pn <form=oblique> ⇒ 0_Ka

In the Hindi translation of such Bengali ra-marker, after the kA or ki postposition, the Hindi pronouns “yaha”, “baha”, and “isa” are used. In the Bengali sentence of Example 10, the ra-marker noun is in oblique form. So this is ra(2) marker and Rule 7 is used to translate this marker.
Example 10
B: sAYanadIperaTA AmAke dAo. [Give me the Sayandeep’s one.]
H: sAyanadIpa kA yaha mujhe do.

6.3. Rule for translation of ra(3) marker

The rule for translating ra(3) is presented in Rule 8.
Rule 8. ra n|pn <drel=k1e> ⇒ 0_ko

In the Bengali sentence of Example 11, the ra-marker noun is experiencer karta. So this is ra(3) marker and Rule 8 is used to translate this marker.
Example 11
B: mitAlIra khide peYechhe. [Mitali is hungry.]
H: mitAlI ko bhuka lagI hai.

6.4. Rule for translation of ra(4) marker

The rule for translating ra(4) is presented in Rule 9.
Rule 9. ra n|pn <drel=k2> ⇒ 0_ko*

In the Bengali sentence of Example 12, the ra-marker noun is karma. So this is ra(4) marker and Rule 9 is used to translate this marker.
Example 12
B: Ami bAchchhAdera beshi pachhanda kari. [I like children more.]
H: mai bachcho.N ko jyAdA pasa.Nda karatA hu.N.

6.5. Evaluation and Analysis of Bengali ra-marker Translation Rules

We have analyzed the KGPBenTreebank corpus containing 4167 sentences (56,514 words) for observing the effects of the ra-marker translation rules. The number of occurrences of each of the ra-marker are shown in Table 4.

Table 4. Number of Occurrences of Different Uses of ra-marker in the Bengali Sentences of KGPBenTreebank.

ra-marker	# of Occurrences
Attached with sambandha	3397
Attached as oblique	3
Attached with karta	64
Attached with karma	42

The baseline system translates each ra-marker to the most frequent Hindi suffix -kA. The number of correct translation of ra-marker by the baseline system and by the proposed method are shown in Table 5.

Table 5. Number of Correct Translation of ra-marker in the Bengali Sentences of KGPBenTreebank by the Baseline and Proposed Rule Based Systems.

System	# of Correct Translations out of 3503 ra-markers
Baseline	2722
Proposed	3492

There are 11 errors while translating the ra-markers using the proposed rules. Some of the mistakes are analyzed below.
Sometimes the sambandha (r6) dependency relation links two words of two different clauses. However, as we capture the inter clause dependency relations between the verbs of two clauses the r6 relation is not identified properly. Therefore, the first rule of Rule 6 could not be used for translation. The translation of ra-marker is incorrect in such cases. One such mistake is shown in Example 13 (a). Similar mistake is also observed when r6 links two words of two different sentences.
In some question sentences where the ra-marker noun indicates the experiencer karta, the ra-marker needs to be translated to the Hindi postposition “kA”. Rule 8 is unable to translate such ra-markers. One such mistake is presented in Example 13 (b).
Example 13

(a)B: jiYA mohanera Ara TiYA gaganera strI. [Jiya is Mohan’s wife and Tiya is Gagan’s wife.]
OH: jiyA mohana kA aura TiyA gagana ki patnI hai.
TH: jiyA mohana ki aura TiyA gagana ki patnI hai.
(b)B: rAmera ki abasthA? [How is Ram?]
OH: rAma ko kyA hAla hai?
TH: rAma kA kyA hAla hai?

VII. HINDI TRANSLATIONS OF BENGALI ke-MARKER

We now study the translation of Bengali ke-marker to Hindi. A Bengali ke-marker may be translated to Hindi markers -ko, -e, etc.

The ke-marker which is used with karta is referred to as ke(1). This ke(1) is translated to the combination of Hindi suffix -0 and Hindi postposition “ko”. Sometimes, we use -ko as a suffix instead of as a postposition.

The ke-marker which is used with karma is referred to as ke(2). This ke(2) is also translated to the combination of Hindi suffix 0 and Hindi postposition “ko” or to -ko suffix.

The ke-marker attached with some pronouns (either in karta or karma position) may be translated to either –e or -ko Hindi suffixes. Among them the -e suffix is more frequent in modern Hindi. The alternative pair of Hindi translations of such Bengali pronouns are shown in Table 6. We searched each of the Hindi word forms using Google search engine. The number of documents containing each word is shown (using curly brace) with that word. Here, M stands for million. We observe that the words in second column have more occurrences that the words in third column. We refer to such ke-markers as ke(3).

Table 6. Two Alternative Forms of Hindi Translations of Some Bengali ke-marker Pronouns along with the Number of Documents (in Millions) Containing that Word as Returned by Google Search Engine.

Pronoun	One Word Form (with -e suffix)	Alternative Word Form (with -ko suffix)
isa	ise {1.84M}	isako {1.82M}
usa	use {2.19M}	usako {0.46}
mai	mujhe {2.05M}	mujhako {1.29M}
hama	hame.N {1.57M}	hamako {1.22M}

The Bengali ke-marker is also used in plural number form. We refer to such ke-markers as ke(4). This ke(4) is translated to the combination of Hindi plural suffix -o.N and Hindi postposition “ko” when it is attached with noun. When it is attached with pronoun then instead of the suffix -o.N a Hindi word (logo.N) is added before the postposition “ko”. We now discuss the rules for translating Bengali kemarker and show their effects in translating Bengali sentences.

7.1. Rule for translation of ke(1) marker

The rule for translating ke(1) is presented in Rule 10.
Rule 10. ke n|pn <drel=k1 number=singular> ⇒ 0_ko*
In the Bengali sentence of Example 14, the ke-marker noun is in singular number and is karta of the sentence. Hence, this is ke(1) marker and Rule 10 is used to translate this marker.
Example 14
B: jaYarAmake yete habe. [Jayram has to go.]
H: jayarAma ko jAnA pa.DegA.

7.2. Rule for translation of ke(2) marker

The rule for translating ke(2) is presented in Rule 11. Here, k2* indicates any subdivision of karma.
Rule 11. ke n|pn <drel=k2 number=singular>⇒ 0_ko*

In the Bengali sentence of Example 15, the ke-marker noun is in singular number and is the karma of the sentence. So this is ke(2) marker and Rule 11 is used to translate this marker.
Example 15
B: Ami manIShAke Dekechhi. [I have called Manisha.]
H: mai.Nne manIShA ko bulAyA hai.

7.3. Rule for translation of ke(3) marker

The rule for translating ke(3) is presented in Rule 12.
Rule 12. ke pn <list2.txt number=singular> ⇒ e

In the Bengali sentence of Example 16, the ke-marker pronoun is a member of list2.txt and is in singular number. So this is ke(3) marker and Rule 12 is used to translate this marker.
Example 16
B: AmAke yete habe. [I have to go.]
H: mujhe jAnA pa.DegA.

7.4. Rule for translation of ke(4) marker

The rules for translating ke(4) are presented in Rule 13.
Rule 13. ke n <number=plural> ⇒ o.N_ko
ke pn <number=plural> ⇒ 0_logo.N_ko

Some instances of the translation of Bengali ke-marker attached with plural noun and pronoun, are shown in Example 17 (a)-(b). In the Bengali sentence of Example 17(a), the ke-marker noun is in plural number. So this is ke(4) marker and the first rule of Rule 13 is used to translate this marker. In the Bengali sentence of Example 17 (b), the ke-marker pronoun is in plural number. So this is ke(4) marker and the second rule of Rule 13 is used to translate this marker.

Example 17
(a) B: chheleke meYera theke beshi subidhA deoYA haYa. [Boys are given more facilities than girls.]
H: la.Dako.N ko la.Dakio.N se jyAdA subidhA diyA jAtA hai.
(b) B: tomAderake DAkA haYechhe. [You(plural) have been called.]
H: tuma logo.N ko bulAyA gayA hai.

7.5. Evaluation and Analysis of Bengali ke-marker Translation Rules

We have analyzed the KGPBenTreebank corpus for observing the effects of the ke-marker translation rules. The number of occurrences of each of the ke-marker are shown in Table 7.

Table 7. Number of Occurrences of Different Uses of ke-marker in the Bengali Sentences of KGPBenTreebank.

ke-marker	# of Occurrences
Attached with karta	21
Attached with transitive karma (k2t)	218
Attached with mukhya karma (k2m)	1
Attached with gauna karma (k2g)	29
Attached with uddyeshya karma (k2u)	81
Attached with Bidheya karma (k2s)	2
Attached with nnmod, pnmod, and pronmod	6
Attached with special pronouns of list2.txt	103

The baseline system translates each ke-marker to the most frequent Hindi suffix -ko. The number of correct translation of ke-marker by the baseline system and by the proposed method are shown in Table 8.

Table 8. Number of Correct Translation of ke-marker in the Bengali Sentences of KGPBenTreebank by the Baseline and Proposed Rule Based Systems.

System	# of Correct Translations out of 461 ke-markers
Baseline	409
Proposed	443

There are 18 mistakes while translating the ke-marker using the proposed rules. Some of the mistakes are analyzed below.

Sometimes the plural form of the ke-marker is represented by the Bengali suffix -dera. In that case often the translation rules for ke-marker fails to translate them.

Such suffixes are translated by the translation rules for ra-marker. One such mistake is shown in Example 18 (a). With some Hindi pronouns "logo.N" postposition is not used. Therefore, Rule 13 makes mistake in translating the ke-marker attached with such pronouns. One such mistake is presented in Example 18 (b).

Example 18
(a) B: Ami oi skulera chheledera pachhanda kari nA. [I don't like the boys of that school.]
OH: mai usa skula ki la.Dako.N kA pasanda nahI.N karatA hu.N.
TH: mai usa skula ki la.Dako.N ko pasanda nahI.N karatA hu.N.
(b) B: toderake okhAne yete habe. [You (familiarsingular) have to go there.]
OH: tujha logo.N ko bahA.N jAnA pa.DegA.
TH: tujhako bahA.N jAnA pa.DegA.

VIII. HINDI TRANSLATIONS OF BENGALI te-MARKER

W We now study the translation of Bengali te-marker to Hindi. A Bengali te-marker may be translated to Hindi null-marker or postpositions "sAtha", "kA", "se", "me.N", "pe", and "lie".

The te-marker which is used with karta is referred to as te(1). te(1) is translated to the Hindi suffix -0 when it is attached with the single karta (general karta) of the sentence.

In some constructions, this te(1) is also used with the pair of karta which are referred to as simple and associative karta. The te(1) attached with general karta is translated to Hindi suffix -0 and the te(1) attached with associative karta is translated to the combination of Hindi suffix -0 and postpositions "ke" and "sAtha".

When te-marker is attached with karan we refer to it as te(2). This te(2) is translated to the combination of Hindi suffix -0 and Hindi postposition "se" marker.

When te-marker is attached with adhikaran it is translated to the combination of Hindi suffix -0 and Hindi postposition "me.N". The Hindi postposition "pe" is also used in place of "me.N" though in case of adhikaran "pe" is less frequent than "me.N". The te-marker attached with the noun indicating destination is also translated to these Hindi markers. But, in case of destination "pe" is more frequent than "me.N". We use te(3) to indicate the te-marker of both adhikaran and destination.

The noun with te-marker may be used to indicate reason or purpose. When it is used to indicate reason then it is translated to the Hindi postposition "se". When it is used to indicate purpose then it is translated to the Hindi postposition "lie". We refer to these two uses of te-marker as te(4).

When te-marker is attached with sanyogmulak noun modifiers then generally same noun with te-marker follows this and it has the POS category Reduplication (REDUP). This reduplicated te-marker noun may indicate either the continuation or plurality. We refer to both of these uses of the te-marker as te(5). When it indicates continuation, we use the root form of the Hindi translation of the first occurrence. When it indicates plurality, we delete one occurrence of the te-marker noun in Hindi translation and another occurrence is pluralized.

We now discuss the rules for translating Bengali temarker and show their effects in translating Bengali sentences.

8.1. Rule for translation of te(1) marker

The rules for translating te(1) are presented in Rule 14.

Rule 14. te n|pn <drel=k1m> ⇒ 0
te n|pn <drel=k1a> ⇒ 0_ke_sAtha

Some instances of the translation of Bengali te-marker attached with Karta are shown in Example 19 (a)-(b). In the Bengali sentence of Example 19 (a), the te-marker is attached with a general karta. So this is te(1) marker and Rule 14 is used to translate this marker. In the Bengali sentence of Example 19 (b), the first te-marker is attached with associative karta and the second te-marker is attached with general karta. So both are te(1) marker and Rule 14 is used to translate these markers.

Example 19
(a)B: mAnuShe kathA bale. [Humans talk.]
H: loga bAta karate hai.
(b)B: rAjAYa rAjAYa yuddha haYa. [Kings fght with kings.]
H: rAjA ke sAtha rAjA kI yuddha hotI hai.

8.2. Rule for translation of te(2) marker

The rule for translating te(2) is presented in Rule 15.

Rule 15. te n|pn <drel=k3> ⇒ 0_se

In the Bengali sentence of Example 20, the te-marker noun is attached with a Karan of the sentence. So this is te(2) marker and Rule 15 is used to translate this marker.

Example 20
B: Ami penasile likhate bhAlobAsi. [I like to write with a pencil.]
H: mai pe.Nsila se likhanA pasa.Nda karatA hu.N.

8.3. Rule for translation of te(3) marker

The rules for translating te(3) are presented in Rule 16.

Rule 16. te n|pn <drel=k7> ⇒ 0_me.N*
te n|pn <drel=des> ⇒ 0_pe

One instance for each of the translation of Bengali te-marker attached with Adhikaran and Destination are shown in Example 21 (a) and 21 (b), respectively. In the Bengali sentence of Example 21 (a), the te-marker is attached with a adhikaran. Hence, this is te(3) marker and the first rule of Rule 16 is used to translate this marker. In the Bengali sentence of Example 21 (b), the te-marker is attached with destination. Hence, this is te(3) marker and the second rule of Rule 16 is used to translate this marker.

Example 21
(a)B: Ami dillIte Achhi. [I am in Delhi.]
H: mai dillI me.N hu.N.
(b)B: Ami dillIte inTArabhiu dite yAchchhi. [I am going to Delhi to appear for interview.]
H: mai dillI pe i.NTarabiu dene ke lie jA rahA hu.N.

8.4. Rule for translation of te(4) marker

The rules for translating te(4) are presented in Rule 17.

Rule 17. te n|pn <drel=rh> ⇒ 0_se
te n|pn <drel=ru> ⇒ 0_ke_lie

One instance of the translation of Bengali te-marker attached with Reason and Purpose are shown in Example 22 (a) and (b), respectively. In the Bengali sentence of Example 22 (a), the te-marker is attached with a Reason. Hence, this is te(4) marker and the first rule of Rule 17 is used to translate this marker. In the Bengali sentence of Example 22 (b), the te-marker is attached with Purpose. Hence, this is te(4) marker and the second rule of Rule 17 is used to translate this marker.

Example 22
(a) B: bhaYe bhule yAYa debatAra nAma. [Name of God is forgotten out of fear.]
H: Dara se bhula jAtA hai debatA ki nAma.
(b) B: Ami parIkShAYa basachhi. [I am sitting for examination.]
H: mai parIkShA ke lie baiTha rahA hu.N.

8.5. Rule for translation of te(5) marker

The rules for translating te(5) are presented in Rule 18. The second occurrence of the word in both the translation rules, is translated according to the rules presented in Rule 14 to 17. We consider the second translation rule more frequent and that is considered as the general case. The first translation rule is considered for some specific words listed in list3.txt.

Rule 18. te REDUP <list3.txt> ⇒ 0
te REDUP ⇒ DEL [number=plural]

Some instances of the translation of Bengali te-marker attached with reduplicated nouns are presented in Example 23 (a)-(b). In the Bengali sentence of Example 23 (a), the te-marker reduplicated nouns indicate continuation and are the members of list3.txt. So the first occurrence has te(5) marker and the first rule of Rule 17 is used to translate this marker. Therefore, the first occurrence has -0 suffix. In the Bengali sentence of Example 23 (b), the te-marker reduplicated nouns indicate plurality. So the first occurrence has te(5) marker and the second rule of Rule 17 is used to translate this marker. Therefore, that first occurrence is deleted and the second occurrence is pluralized.

Example 23
(a) B: se dine dine Arao rogA hachchhe. [Day by day he is becoming thiner.]
H: baha dina dina aura bhI patalA ho rahA hai.
(b) B: se rAstAYa rAstAYa ghure be.DAchchhe. [He is roaming in the streets.]
H: baha sarako.N pe ghuma rahA hai.

8.6. Evaluation and Analysis of Bengali te-marker Translation Rules

We have analyzed the KGPBenTreebank corpus for observing the effects of the te-marker translation rules. The number of occurrences of each of the te-marker are shown in Table 9.

Table 9. Number of Occurrences of Different Uses of te-marker in the Bengali Sentences of KGPBenTreebank.

te-marker	# of Occurrences
Attached with general karta (k1m)	14
Attached with associative karta (k1a)	5
Attached with karan	116
Attached with reason	72
Attached with adhikaran	716
Attached with destination	511
Attached with reduplications	47
Attached with nnmod, pnmod, etc.	32

The baseline system translates each te-marker to the most frequent Hindi suffix -se. The number of correct translation of te-marker by the baseline system and by the proposed method are shown in Table 10.

Table 10. Number of Correct Translation of te-marker in the Bengali Sentences of KGPBenTreebank by the Baseline and Proposed Rule Based Systems.

System	# of Correct Translations out of 1513 te-markers
Baseline	296
Proposed	1445

There are 68 mistakes while translating them using the proposed rules. Some of the mistakes are analyzed below. Instead of noun or pronoun, the te-marker is also used with verbs or vmod (adverb). These are nominal forms of the verb. As the rules for translating the te-marker are prepared for noun and pronoun, they fail to translate such derived forms. One such mistake is shown in Example 24(a). When te-marker is attached with an incidence then often the above rules are unable to translate such te-markers. One such mistake is shown in Example 24 (b).

Example 24
(a) B: dUratba kamAte sthitishakti kame yAYa. [By decreasing distance the stability decreases.]
OH: durI ghaTAne ke lie sthitishakti ghaTa jAtI hai.
TH: durI ghaTane se sthitishakti ghaTa jAtI hai.
(b) B: tini 2011 sAle durghaTanAYa mArA gechhena. [He died in an accident in 2011.]
OH: be 2011 sAla me.N durghaTanA se mara gae.N hai.N.
TH: be 2011 me.N durghaTanA me.N mara gae.N hai.N.

The te-markers in some special noun phrases could not be translated correctly by the rules. Two such examples are presented in Example 25 (a)-(c). The te-marker noun in Bengali sentence of Example 25 (a) is the part of (pof) a complex predicate. The te-marker noun in Bengali sentence of Example 25 (b) is the reduplication. The te-marker noun in Bengali sentence of Example 25 (c) is the karta.
Example 25
(a)B: tumi kichhu mane koro nA. [You do not mind.]
OH: tuma kuchha mana me.N mata karo.
TH: tuma burA mata mAno.
(b)B: se mane mane aruke bhAlobAse. [He loves Aru secretly.]
OH: baha mana mana me.N aru ko pyAra karatA hai.
TH: baha a.Ndara hi a.Ndara aru ko pyAra karatA hai.
(c)B: AmarA sakale khushi. [We all are happy.]
OH: hama loga saba me.N khusa hai.
TH: hama saba khusa hai.

IX. CONCLUSION

Each of the Bengali nominal suffixes are ambiguous. In this paper we study a corpus for identifying the possible Hindi translations of each of the Bengali suffixes. Then a set of rules are created for translating these suffixes depending on their contexts. The effects of the rules are observed a test corpus.

The effect of the rules need to be checked in a machine translation system. The rules for translating the Bengali postpositions also need to be developed. Similar rules can be designed for other language pairs using the corresponding language expertise.

Acknowledgement

The author is grateful to Prof. Sudeshna Sarkar and Prof. Anupam Basu for sharing their knowledge to complete this work. A part of the knowledge used in this work is derived from the Indian Language Machine Translation (ILMT) project, sponsored by MeiTy, Govt. of India where the author has worked for 7 years and from his PhD works.

References

[1] Eneko Agirre, Mikel Lersundi, and David Martinez, "A multilingual approach to disambiguate prepositions and case suffixes," in *Proceedings of the ACL-02 workshop on Word sense disambiguation: recent successes and future directions*, Association for Computational Linguistics, vol. 8, pp. 1–8, 2002.

[2] Lisette Appelo, "A compositional approach to the translation of temporal expressions in the Rosetta system," in *Proceedings of the 11th conference on Computational linguistics*, Association for Computational Linguistics, pp. 313–318, 1986.

[3] PVS Avinesh, "Transfer Grammar Engine and Automatic Learning of Reorder Rules in Machine Translation," Master's thesis. LTRC, IIIT Hyderabad, Hyderabad, India, 2010.

[4] Akshar Bharati, Vineet Chaitanya, and Rajeev Sangal, "Natural Language Processesing: A Paninian Perspective," 1999.

[5] Samit Bhattacharya, Monojit Choudhury, Sudeshna Sarkar, and Anupam Basu, "Inflectional Morphology Synthesis for Bengali Noun, Pronoun and Verb Systems," in *Proceedings of the National Conference on Computer Processing of Bangla (NCCPB 05)*, Dhaka, Bangladesh, pp. 34–43, 2005.

[6] Bamandev Chakravarty. February, Uchchatara Bangla Vyakaran, *A complete text book on higher Bengali grammar* (nineteenth ed.). Akshay Malancha, 2010.

[7] Sanjay Chatterji, Sudeshna Sarkar, and Anupam Basu, "Translations of Ambiguous Hindi Pronouns to Possible Bengali Pronouns," in *Proceedings of the 10th Workshop on Asian Language Resources*, COLING 2012, Mumbai, India, pp. 125–134, 2012.

[8] Sanjay Chatterji, Tanaya Mukherjee Sarkar, Pragati Dhang, Samhita Deb, Sudeshna Sarkar, Jayshree Chakraborty, and Anupam Basu. "A dependency annotation scheme for Bangla treebank," *Language Resources and Evaluation*, vol. 48, no. 3, pp. 443–477, 2014.

[9] Suniti Kumar Chatterji, BHASHA-PRAKASH BANGALA VYAKARAN, *A Grammar of the Bangla Language* (third ed.). Roopa and Company, 2003.

[10] Avinash Chopde. February, ITRANS "Indian Language Transliteration Package", A package for printing text in Indian Language Scripts, available from http://www.aczone.com/itrans/. February, 2000.

[11] Sherri L. Condon, Dan Parvaz, John S. Aberdeen, Christy Doran, Andrew Freeman, and Marwan Awad, "Machine Translation Errors: English and Iraqi Arabic," *ACM Trans. Asian Language Inf. Process*., vol. 10, no. 1, pp. 2-8, 2011.

[12] Niladri Sekhar Dash, "Bangla pronouns-a corpus based study," *Literary and linguistic computing,* vol. 15, no. 4, pp. 433-444, 2000.

[13] Bonnie J. Dorr, Pamela W. Jordan, and John W. Benoit, "A survey of current paradigms in machine translation," *Advances in Computers*, vol. 49, pp. 1–68, 1999.

[14] Zellig S. Harris, "Transfer Grammar," International *Journal of American Linguistics*, vol. 20, no. 4, pp. 259–270, 1954.

[15] Edward S. Klima, "Structure at the lexical level and its implication for transfer grammar," in *Proceedings of the International Conference on Machine Translation of Languages and Applied Language Analysis*, vol. 1. 108, 1962.

[16] Jan Landsbergen, "Machine translation based on logically isomorphic Montague grammars," in *Proceedings of the 9th conference on Computational linguistics*, Academia Praha, vol. 1, pp. 175–181, 1982.

[17] Ganapathiraju Madhavi, Balakrishnan Mini, N Balakrishnan, and Reddy Raj, "Om: One tool for many (Indian) languages," *Journal of Zhejiang University-Science A*, vol. 6, no. 11, pp. 1348–1353, 2005.

[18] Steve Lawrence Manion, "Fluency enhancement: applications to machine translation," Master's thesis. Information & Telecommunications Engineering, Massey University, Palmerston North, New Zealand, 2009.

[19] Rashmi Prasad, "A corpus study of zero pronouns in Hindi: An account based on centering transition preferences," in *Proceedings of the DAARC*, pp. 66–71, 2000.

[20] Rajeev Sangal, "Project Proposal to Develop Indian Language to Indian Language Machine Translation System," IIIT Hyderabad, TDIL Group, Dept. of IT, Govt. of India (2006). Manuscript submitted to *ACM*, 2006.

[21] Abhilash Pathak, Sudhanshu Kumar, Partha Pratim Roy and Byung-Gyu Kim, "Aspect-Based Sentiment Analysis in Hindi Language by Ensembling Pre-Trained mBERT Models," *Electronics*, vol. 10, no. 2641, 2021.

[22] Sudhanshu Kumar, Monika Gahalawat, Partha Pratim Roy, Debi Prosad Dogra and Byung-Gyu Kim, "Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning," *Electronics*, vol. 9 no. 2, p. 374, 2020.

[23] Seo-Jeon Park, Byung-Gyu Kim, "A Robust Facial Expression Recognition Algorithm Based on Multi-Rate Feature Fusion Scheme," *Sensors*, vol. 21, no. 6954, pp. 1-26, 2021.

[24] Young-Ju Choi, Young-Woon Lee, Byung-Gyu Kim, "Residual-based Graph Convolutional Network (RGCN) for Emotion Recognition in Conversation (ERC) for Smart IoT," *Big Data (Mary Ann Liebert)*, vol. 9, no. 4, pp. 279-288, 2021.

[25] Ji-Hae Kim, Byung-Gyu Kim, Partha Pratim Roy, Da-Mi Jeong, "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure," *IEEE Access (IEEE)*, vol. 7, pp. 41273-41285, 2019.

[26] Young-Ju Choi, Young-Woon Lee, Byung-Gyu Kim, "Wavelet Attention Embedding Networks for Video Super-Resolution," in *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR2020)*, pp. 7314-7320, Milan, Italy, Jan 10-15, 2021.

Author

Dr. Sanjay Chatterji earned his B. Tech. degree from Haldia Institute of Technology (VU) in 2003, M. E. degrees from BESUS (Now IIEST) in 2005 and Ph. D. from IIT Kharagpur in 2014 both from the Department of Computer Science and Engineering. He worked in Samsung, India for about 4 years. In 2017, he joined the IIIT Kalyani as an assistant professor.

His research interests include NLP, ML, AI, and Computer Vision.