Dealing with Replicative Words in Hindi for Machine Translation to English

R.M.K Sinha
Indian Institute of Technology Kanpur
rmk@iitk.ac.in

Anil Thakur
Indian Institute of Technology Kanpur
anilt@iitk.ac.in

Abstract

The South Asian languages are well-known for their replicative words. In these languages, words of almost all the grammatical categories can occur in their reduplicative form. Hindi is one such language which is quite rich in having various types of replicative words in its lexicon. The traditional grammars and some of the research works have discussed the topic to some extent, particularly from the point of view of their descriptions and classifications. However, a detailed study of the topic becomes significant in view of the complexity involved in handling of such replicative words in the area of natural language processing, particularly for machine translation. In this paper, we discuss different types of replicative words in Hindi and their syntactic and semantic characteristics to formulate rules and strategies to identify their multiple functions and mapping patterns in English for machine translation from Hindi to English.

1 Introduction

The phenomenon, in a natural language, where a word repeats itself in a sequence without any morphological variation has been referred to as "complete word reduplication". In this paper, we call this process as replication, retaining the definition[1] given in (Abbi, 1992): "those paired constructions when a single word or a clause is repeated once in the same sentence without any phonological or morphological variation". Replicative words are the typical characteristics of the South Asian languages. All types of words can occur in replicated form to denote various kinds of functions. The examples in (1) illustrate the process of replication of the different types of words in Hindi.

(1) a. Noun Noun
ghar ghar {house house}
'every house'
b. Verb Verb
ruk ruk {stop stop}
'after stopping'
c. Pronoun Pronoun
apanaa apanaa {self self}
'(one's) own'
d. Adjective Adjective
baRii baRii {big big}
'quite big'
e. Numeral Numeral
ek ek {one one}
i. 'every one'
ii. 'one by one'
f. Adverb Adverb
dhiire dhiire {slow slow}
i. '(quite) slowly'
ii. 'gradually'

In (1), we notice that the replication of the different categories of words have different semantic consequences. For instance, when a noun occurs in the replicated form, it has a quantificational function that gives a sense of each of the Noun, as in (1a). In the case of a verbal replication, the resultant meaning is generally a kind of repetitive completive aspect marker, as in (1b). The replication of adjective generally functions as a kind of intensifier (1d) and same appears to hold in case of adverbial replication (1f). The process of replication is more productive in case of words of some categories than the others. For instance, prenominal replication is restricted only to a certain kind of reflexive pronoun in Hindi (*apanaa apanaa* {self}) and other pronoun resist replication (e.g. **tum tum* {you you}). Besides, there are also further variations in the kind of functions and the resultant meanings of a replicated unit depending on

[1] Abbi (1992) defines reduplication as "repetition of all or a part of a lexical item carrying a semantic modification". In the sense of this definition, reduplication can be either partial or complete. She also includes lexical-couplets such as *khaanaa-piinaa* {eating-drinking} and echo words such as *khaanaa-vaanaa* {eating, etc} in the category of reduplication. In this paper, however, we have taken only the case of complete reduplication. Hence the term "replication" used here can help make differentiation between these two types of reduplication.

different factors such as the nature of the relevant words, nature of the predicate, etc.

The question as to whether the replicative words are arbitrary or amenable to generalization or rules have been discussed in literature (Abbi, 1992; Friedrich, 1979) and the general view is that they are not completely arbitrary and can be examined with a view to generalizing their syntactic and semantic constraints. For the purpose of machine translation, the replicated words of the fixed/idiomatic semantic nature have to be handled differently (may be at the level of the lexicon itself) than those that follow a systematic pattern. For this purpose, we need to examine the syntax and semantics of the different types of the replicated words in Hindi. We also need to examine the mapping patterns of the different types of these words to obtain correct translation of the constructions involving replicated words. Although some of the existing works (Abbi, 1992, 1977, 1975) go into some aspects of categorization and generalization of the replicative words in Hindi, they do not examine the issue of their mapping patterns in English. Besides, the classifications and generalizations already made need further examination with respect to many new patterns that have not been explored in the existing literature. In (Abbi, 1992, 1975) the issue has been discussed mainly from typological and historical perspectives. It discusses verbal replications in detail with relation to their aspectual functions but no detail discussion is attempted about the nominal replication. In the existing works, the issue has not been discussed in any form from the point of view of machine translation from Hindi to English involving replicative words and their mapping patterns. Another important issue pertaining to replicative words is the fact that the occurrence of a replicated words sequence does not always constitute replication. In a number of cases, they can also represent two separate words with independent structural status.

In this paper, we discuss these issues in detail by examining the different types of replicated words, their syntactic-semantic constraints and their mapping patterns in English. The issue has been examined from the point of view of machine translation from Hindi to English. Thus the constructions involving replicative words of ambiguous nature have been specially examined and rules have been formulated for their disambiguation. In section 2, we examine the different types of replicated words in Hindi and their mapping patterns in English. In doing so, we take into consideration both syntactic and semantic constraints of the replicated words. In section 3, we formulate rules and strategies to identify the different functions/meanings of the replicated words. We take into account the nature of the verb and other relevant elements to disambiguate between multiple meanings of the replicated words.

2 Types of Replication in Hindi and their Mapping Patterns in English

2.1 Noun-Noun Replication

The replication of the nominal elements is associated with a number of different functions.

I. Quantification: A nominal element is replicated to denote a quantified noun phrase which is mapped in English by {every Noun}. Only a certain kinds of nouns are productively replicated for this type of function. Nouns indicating place are commonly used in the process of replication of this type.

(2) a. *vah galii galii ghuumataa rahaa.*
{he street street wander PROG2}
'He kept on wandering in every street.'
b. *baccaa baccaa gandhiijii ko jaanataa hE.*
{child child Gandhiji ACC know be.PR}
'Every child knows Gandhiji.'

Place denoting nouns such as *galii* {street}, *shahar* {city}, *gaaNv* {village}, *ghar* {house}, *darawaajaa* {door}, *dukaan* {shop}, *konaa* {corner}, *desh* {country}, etc. have very common occurrence in replicative forms. We also notice that some common nouns are also used in replicative form with identical function, as in (2b) above. However, this is restricted to only a handful of common nouns as is shown by the unacceptability of a construction where *baccaa* {child} has been replaced by another similar type of common noun *aadamii* {man}, as in (3a). In this case, the use of the relevant quantifier element is obligatory to get the intended meaning, as in (3b).

(3) a. **aadamii aadamii gandhiijii ko jaanataa hE.*
{man man Gandhiji ACC know be.PR}
b. *har aadamii gandhii jii ko jaanataa hE.*
{every man Gandhiji ACC know be.PR}
'Every man knows Gandhiji.'

[2] Abbreviations: CMT: Completive Aspect, CPP: Conjunctive Participle Particle, FU: Future Tense, IMP: Imperfective Aspect, PROG: Progressive Aspect Marker, PR: Present Tense, PST: Past Tense, *: Not Possible; Hindi words are shown in *italics* (letter '*c*' is used to denotes the sound of '*ch*' as in China).

II. Coordinate Conjunction: Almost all kinds of nouns can occur in their replicated form to denote coordinate conjunction. In this case, the typical coordinate conjunction particle *Or* {and} is (optionally) dropped. In the English counterpart of such examples, the relevant word {and} is obligatory. Some examples are listed in (4).

(4) a. *aadamii aadamii meM phark hE.*
{man man in difference be.PR}
'There is difference between man and man.'
b. *aaNkhoN aaNkhoN meN ishaaraa ho gayaa.*
{eyes eyes in hint be went}
'There was exchange of hints between eyes and eyes.'

In (4), the occurrence of the postposition *meN* {in} seems to be the deciding factor for differentiating the noun-noun sequence in (4) from those in (2). However, let us see the examples in (5a), where the noun-noun replication is followed by the postposition *meN* {in} but this does not make it identical to the one in (4). In (5a), the noun-noun sequence that appears like a replication is actually two separate words which occur in two independent grammatical positions. The first noun *paanii* {water} is the subject NP whereas the second noun *paanii* {water} is the locative adverbial NP/PP. In (5b), we notice that a true replicated unit (see (2a) above) can also occur immediately followed by the postposition *meN* {in}. Thus it appears that the occurrence of a particular postposition cannot always be a reliable factor and we need to examine other factors, too, such as the nature of the nominal or the adjectival elements in the predicate and the type of the verb.

(5) a. *paanii paanii meN jaa milaa.*
{water water in go meet}
'The water went in water.'
b. *galii galii meN shOr hE.*
{street street in commotion be.PR}
'There is commotion in every street.'

Thus to identify the exact function of a replicative word, we need to examine not only the following postposition but also the type of the predicate. However, there still remain certain cases where multiple interpretations cannot be avoided. Further, to differentiate between the types of replicative words and identify whether a replicative sequence of words constitute a replicated word or not, we need to examine different factors in a sentence such as the nature of the neighboring words, the nature of the verb, etc.

III. Vocative: Sometimes a noun can be replicated to denote vocative function. For instance, in (6), *cor cor* {thief thief} is used to make an "addressive" sound. This type of replication is common in all types of words including verbs, as in (6c).

(6) a. *us-ne cor cor (kii) aawaaj lagaaii.*
{he-ERG thief thief of sound give}
'He shouted, thief thief.'
b. *sab-ne use cor cor kar ke maarane ko dORaa.*
{all-ERG him thief thief do CPP beating for ran
i. 'All ran to beat him shouting thief'
ii. 'All ran to beat him considering him a thief.'
c. *koii bacaao bacaao cilaa rahaa thaa.*
{someone save save shout PROG be.PR}
'Someone was shouting "save, save".'

IV. Pseudo-replication: Certain noun-noun sequences look like replicated words but they actually occupy the different grammatical positions in a sentence. In (7a-b), the first noun constitutes the subject NP whereas the second noun is the predicative noun. In (7c), the first noun is the subject NP whereas the second is the object NP. The typical objective Case marker *se* {with} or *ke saath* {with} that can optionally occur with the second noun is not present in this case but the whole sentence retains that meaning.

(7) a. *maaN maaN hotii hE.*
{mother mother be be.PR}
'Mother is mother.'
b. *dost dost naa rahaa.*
{friend friend not remained}
'The friend did not remain friend.'
c. *dost dost mil gayaa.*
{friend friend meet went}
'The friends got together.'

(8) a. *hiiraa hiiraa ko kaaTataa hE.*
{diamond diamond ACC cut be.PR}
'The diamond cut the diamond.'
b. *aadamii aadamii se laRataa hE.*
{man man with fight be.PR}
'The man fights with a man.'

In the examples in (8), the sequence of identical nouns does not form replicative words. The second noun in the identical noun-noun sequence is followed by a Case marker (postposition) and the nature of the verb demands that the second noun be taken as the object NP in the sentence. Thus both the factors; nature of postposition and the nature of

the verb can help in identifying such noun-noun sequence as non-replicative words.

2.2 Numeral-Numeral Replication

The replication of numeral results in denoting unit / group meaning. Sometimes the words that follow these numeral replicative words decide the actual meaning of the replicative words. For instance, in (9c & 10a) the replicative word is followed by *kar* {do} and in such cases, the replicative words always means group of X(numeral) each. The behavior of *ek* {one} is different from other numerals in the sense that *ek* {one} does not make a group and it also has some context-based idiomatic meanings (9).

(9) a. *vah ek ek ko dekh le-gaa.*
{he one one ACC see take-FU}
'He will take on each of them/everyone.'
b. *yah baat is shahar meN ek ek jaantaa hE.*
{this matter this city in every one know be.PR}
'Everyone in this city knows this matter.'
c. *us-ne ek ek kar sabhii kitaabeN dekh liiN.*
{he-ERG one one do all books see took}
'He saw/checked all the books one by one.'

(10) a. *do do kar aao.*
{two two CPP come}
'Come in a group of two each.'
b. *ek (ek) rupaye ke do do kele milate hEN.*
{one one rupees of two two banana get be.PR}
'The banana is sold at rupee one for two.'
c. *ek ek ruum meN caar caar bEThate hEN.*
{one one room in four four sitting be.PR}
'A group of four sits in each room.'

2.3 Adjective-Adjective Replication

Adjectives are one of the most common grammatical categories to occur in replicative forms. In replicative forms, they denote different types of functions. Their interpretation is based on both the type of the adjective and the type of the noun they qualify. The adjectives denoting color terms generally give intensifier meaning and if the modified noun is in plural number it also gives the sense of distributiveness (Abbi, 1992), as in (11a). The ordinal (numeral) adjective *pahalaa* {first} has idiomatic use. It is interpreted as a kind of focus and can be mapped in English by adjoining {very} to the modifier (11d). Other replicated ordinal numerals behave similar to cardinal numerals (as mentioned above). The word *dusaraa* is homophonous with both ordinal and adjective meanings. For instance, *dusare-dusare din* has ordinal interpretation {every second day} whereas *dusare-dusare log* has adjective interpretation {other people}.

(11) a. *vahaaN par piile piile phuul paRe the.*
{there on yellow yellow flowers lying be.PR}
'There were (lots of) yellow flowers lying there.'
b. *vahaaN mEN-ne gulaab kii baRii baRii phuuleN dekhii.*
{there I-ERG rose of big big flower saw}
'There I saw very big rose flowers.'
c. *unakii do coTii coTii beTiyaaN hEN.*
{his two small small daughter be.PR}
'He has two (very) young daughters.'
d. *yah unakaa pahalaa pahalaa pyaar hE.*
{this their first first love be.PR}
'This is their very first love.'
e. *ham kisii kisii baccoN se mile.*
{we some some children with meet}
'We met some children.'
f. *ve thoRe thoRe kaale the.*
{they little little black be.PST}
'They were (somewhat) little black.'
g. *hamaarii nayii nayii dostii huii hE.*
{our new new friendship be be.PR}
'Our friendship is very new.'

Some of the adjectival elements have fixed meaning in replicative forms. For instance, *kisii kisii* in (11e) is always mapped by {some} in English. *thoRaa thoRaa* (and its variant forms) is always mapped by {little} (11f) in English.

2.4 Adverb-Adverb Replication

Adverbs occur in replicative forms to denote a number of functions. In most of the manner adverbials, adverb replication functions as an intensifier. However, it can be mapped in English either by the use of {very} or null.

(12) a. *usa-ne jor jor se aawaaj lagaaii.*
{he-ERG loud loud with sound gave}
'He shouted (very) loudly.'
b. *vah dhiire dhiire baRaa hone lagaa.*
{he slowly slowly grow start}
'He started growing steadly.'
c[3]. *vah dhiire dhiire calane lagaa.*
{he slowly slowly move start}

[3] A difference in meaning results if this replicative word *dhiire dhiire* is shifted to the sentence-initial position, as below: *dhiire dhiire vah calane lagaa.* {gradually he walk start} => 'Gradually, he started walking.'

'He started walking slowly.'

d. *baccaa dhiire dhiire baRaa hone lagaa.*
{child gradually gradually grow be start}
'Gradually, the child grew.'

e. *vah dhiire dhiire bolataa hai.*
{he slowly slowly speak be.PR}
i. 'He speaks in a low voice.'
ii. 'He speaks slowly.'

(13) a. *kahiiN kahiiN dhuup hE.*
{somewhere somewhere sunshine be.PR}
'There is sunshine somewhere.'

b. *ve kabhii kabhii aate hEN.*
{they sometime sometimes come be.PR}
'They come sometimes.'

2.5 Possessive *apanaa*

The possessive adjective *apanaa* {own} occurs in replicative form. It occurs only with plural subject and its meaning depends on the type of the subject. The extra *apanaa* gives an interpretation of respective or own to the original meaning.

(14) a. *sabhii apanaa apanaa kaam karate hEN.*
{all self self work do be.PR}
'Everyone does his respective/own work.'

b. *ve apane apane ghar gaye.*
{they self self house went}
'Everyone went to his respective/own house.'

c. *sab apane apane ke liye maratаа hE.*
{all own own for die be.PR}
'All are interested in their own kith and kin.'

2.6 Focus Function

Almost all the interrogative, relative and adverbial particles are replicated to denote focus meaning or some other meaning. The meanings are largely fixed and can be handled in the lexicon. Some of these particles/pronouns are *jab jab* {whenever}, *jahaaN jahaaN* {wherever}, *jidhar jidhar* {wherever}, *jis jis* {whoever}, *jEse jEse* {as}, However, some of these particles show context-ambiguity and need to be disambiguated.

(15) a. *aap-ne jab jab bulaayaa, vah aayaa.*
{you-ERG whenever called he came}
'Whenever you called, he came.'

b. *ve jahaaN jahaaN gaye, saphal rahe.*
{they wherever went success live}
'Wherever he went, he was successful.'

c. *unhoN-ne jis jis ko bulaayaa, ve aa gaye.*
{they-ERG whoever ACC called they come went}
'Whoever they called, they came.'

(16) a. *aap kahaaN kahaaN gaye?*
{you where where went}
i. 'Where did you go.'
ii. 'Which places did you visit?'

b. *aap-ne kyaa kyaa dekhaa?*
{you-ERG what what saw}
'What did you see?'

c. *ham-ne kyaa kyaa nahiiN dekhaa*!
{we-ERG what what not saw}
'We saw everything!'

d. *aap kab kab ghar par rahate hEN?*
{you when when house at live be.PR}
'When do you stay at home?'

e. *aap kis kis se mile?*
{you who who with met}
i. 'Who did you see?'
ii. 'How many people did you meet?'

f. *aap kin kin logoN se mile?*
{you who who people with meet}
'Which are the people that you saw?'

2.7 Expressives

These are onomatopoetic expressions and have fixed meanings. They have been categorized into different categories on the basis of their associations with natural sounds, animal sounds, human sounds, and some expressive words related to different senses (see, taste, smell, etc) (Abbi, 1992). Some examples are presented in (17) below for illustration.

(17) a. Animal noises:
kaav-kaav 'cawing' (of crows)
mayaauN-mayaauN 'mewing' (of cats)
hii-hii 'clattering' (of monkey), etc.

b. Noises of natural phenomenon:
Tap-Tap 'dropping' of rains'
san-san / sar-sar 'sound of air'
jhar-jhar 'sound of water'

c. Human noises:
khii-khii 'sound of laughing clumsily'
khoN-khoN 'coughing'
thuu-thuu 'spitting'

2.8 Verb-Verb Replication

Verbs mainly form two types of replicative words: (a) verb verb *kar* and (b) verb-*te* verb-*te*. Both these forms denote different types of adverbial functions. The interpretation depends both upon the nature of replicating verb and the main verb in the sentence (Abbi, 1992; Kachru, 1980). In a simple verb predicate, the main verb is replicated rather than the auxiliary elements (18). In conjunct verb constructions (N/Adj Verb), only the verb is replicated not the nominal or the adjectival constituent (19). In compound verb

constructions (V1 V2), the V1 gets replicated and the auxiliary V2 remains unchanged (20).

(18) a. *vah ro ro kar bola.*
{he cry cry CPP spoke}
'He spoke crying.'
b. *vah padhate padhate thak gaya.*
{he study study tired went}
'He got tired of studying.'

(19) *vah use salaah de de kar thak gayaa.*
{he him advice give give CPP tired went}
'He got tired of advising him.'

(20) *vah yahaaN kaii baar bhaag bhaag kar aayaa hE.*
{he here several times run-away un-away CPP come be.PR}
'He has come here after running away several times.'

(21) a. *ham haarate haarate jiit gaye.*
{we lose lose win went}
i. 'We finally won while losing till the end.'
ii. 'We were losing but finally won.'
b. *vah rote rote haNsane lagaa.*
{he cry cry laugh start}
i. 'He finally started laughing while crying till the end.'
ii. 'He was crying but finally started laughing'

In (21), we notice that the replicating verb and main verb are of antonym nature with respect to each other. In this case the mapping pattern in English is obtained by adding an addition word {finally} as prefixed with main verb and {till the end} is appended to replicating word with {while} as co-joining word.

(22) a. *vah duubate duubate bachaa.*
{he drown drown survive}
i. 'He just escaped from drowning.'
ii. 'He was almost drowned (but just escaped).'
b. *vah piite piite ruka gayaa.*
{he drink drink stop went}
i. 'He just stopped from drinking.'
ii. He stopped just short of drinking.

In (22), we notice that if the main verb is of the type *bachanaa* or *rukanaa* then {from} is used as a co-joining word to the replicated verb and {just} to the main verb.

In (Abbi, 1992), five different functions of these replicated verbal adverbials have been noted. They are: i. Simultaneity, ii. Non-precipitation, iii. Continuation-Duration, iv. Iteration, and v. Sequentiality. This categorization has been made on the basis of the semantics of the different functions. This does not help in capturing the different mapping patterns these adverbial functions of the replicated words exhibit in English. Thus there is a need to have a categorization of these replicated words taking into account their mapping patterns in English. In this paper, we have done this by taking into consideration both the types of the functions and the mapping patterns of these replicated words. Thus we categorize the different functions of these verbal adverbial into five types: i. Time adverbial, ii. Manner adverbial, iii. Reason adverbial, iv. Contrastive, v. Means. We take into account the different categories [4] of verbs to correlate the different functions and mapping patterns of these replicative words. For instance, we notice that three categories of verbs can be exploited to determine/identify the various functions of these replicated verbs. The verbs in both the main clause and the subordinate clause jointly determine the exact function of the replicated words (verbs). The examples in (23-26) illustrate the main functions of the replicated verbs.

(23) Time
a. *das bajate bajate gaRii pahuuNc gaii.*
{ten strike strike train reach went}
'As soon as it struck/was ten, the train arrived.'
b. *haNsate haNsate usakii aaNkhoN meN aaNsu aa gaye.*
{laughing laughing his eyes in tears come went}
'While laughing, tears came into his eyes.'
c. *vah dOrate dOrate gir gayaa.*
{he run run fall went}
'While running he fell down.'

(24) Manner
a. *vah ruk ruk kar kaam karataa hE.*
{he stop stop work do be.PR}
'He works with repeatedly/often halting.'
b. *vah haNsate haNsate bola.*
{he laughing laughing spoke}
'He spoke smilingly.'

(25) Reason

[4] The relevant categories are: com_expression (*ronaa* 'cry'), communicative (*bolanaa* 'speak'), motion (*janaa* 'go'), resultive (*honaa* 'become'), voluntary (*paDhanaa* 'read'), resultive:psych (*thakanaa* 'get tired').

a. *lohaa piiTate piiTate usake haath meN dard ho gaga.*
{iron beat beat his hand in pain be went}
'He got pain in hands due to repeatedly beating the iron.'
b. *vah khaa khaa kar moTaa ho gayaa.*
{he eat eat CPP fat be went}
'He became fat due to excessive eating.'

(26) Means
a. *vah padh padh kar gyaanii ho gaya.*
{he ready read CPP wise be went}
'He became wise by a lot of reading.'
b. *abhayaas karate karate vah nipuN ho gayaa.*
{practice doing doing he skilled be went}
'He became skilled by practicing.'

3 Translating Sentences with Replicative Words

From the foregoing discussions, it is evident that there are varied interpretations of sentences with replicative words in Hindi. Formulating unambiguous rules is quite complex. Further, an elaborate, exhaustive and reliable corpus is not available for Hindi. The usage of replicative words is more common in oral communication and informal presentations. Therefore, any corpus-based technique such as SMT is not going to be successful unless corpus coverage enteds to such communications. However, it is important to understand the underlying process of usage of replicative words by collecting examples from various sources such as from grammar books and from native speakers. The first aim of this work has been to accomplish this task before attempting to translate them. Besides eliciting data from the native speakers and other sources, we have drawn the relevant examples from a corpus of the size of 913468 words. The replicative words occurs 2145 (0.23%) times which contains 366 different words comprising of adjectives (60), verbs (59), adverbials (40), nouns (30) and other words of miscellaneous (interjections, etc) categories.

Some of the replicative words that occur with high frequency are: *saath saath* {together} (184), *alag alag* {separately} (168), *dhiire dhiire* {slowly/gradually} (133), *kabhii kabhii* {sometimes} (132). It is observed that several words that occur very frequently in replicative form in spoken conversation are not found in the written text corpus. Some examples are: *bolate bolate* {speak speak}, *haNsate haNsate* {laugh laugh}, *sunate sunate* {listen listen}, etc. This explains the limitation of the available corpora.

We have adopted a hybrid strategy for dealing with replicative words in our machine-aided translation systems (Sinha, 2004) for Hindi to English translations. We have incorporated hybridization of rule-based and example-based strategies in our systems. It should be noted that the example-based approach need not necessarily be only corpus driven. The examples are entered in an interactive way during the development phase of the system. During this phase, the test sentences are fed both from the corpus and the collected samples and the process of building the example-base is semi-automated. During the development stage, an attempt is made first to devise rules based on patterns of sentences with replicative words in the rule-base. These rules are used for automated pre-editing and paraphrasing of the input sentence wherever applicable. Replicative numerals, and some of the replicative nouns and adjectives fall in this category. In case no definitive pattern emerges, such sentences are handled through example-base. The example-base is first constructed by storing example sentences containing replicative words in their raw form and their translation in English along with their syntactic, semantic and ontological information. The examples are then generalized based on syntactic and semantic closeness of the constituents. A separate partition of examples of sentences with replicative words is maintained. The distance of the input sentence containing replicative words from the stored example sentence is computed with respect to the tags associated with constituents. Besides these, the lexical data-base has entries for replicated words that have a fixed meaning and are directly mapped for translation.

Given below are some of the rules that we have found to be reliable under different categories:

Noun-Noun:
If replicative noun is followed by *meN* and [main verb is not of the type 'motion'],
then
<noun_x> <noun_x> => between<noun_x> and <noun_x>

Numeral-Numeral:
a. **If** replicative number = *ek* and is followed by *kar*, t**hen**
ek ek => one by one
else *ek ek* => everyone | each

b. **If** replicative number ≠ *ek*,
then
<numeral_x> <numeral_x> => group of <numeral_x> each

Adjective-Adjective:

<adj_x> <adj_x> => very/ many / a number of / a lot of [5] <adj_x>

A specific rule is as follows:
If replicative adjective denotes 'color' and no other intensifier such as very precede the replicative adjective,
then
<adj_x-color><adj_x-color> => very <adj_x>

Verb-Verb:
a. **If** the replicative verb is of category com_expression and is followed by *kar* and the main verb is of communicative type,
then
<verb_x><verb_x> => by <verb_x>+ing

b. **If** the replicative verb has a suffix *–te* and the main verb is of the 'resultive:psych' type
then
<verb_x-te><verb_x-te> => due to|of <verb_x>+ing

c. **If** the replicative verb has a suffix '*-te*' and the activity associated with the replicative verb can co-occur with that of the main verb,
then
<verb_x-*te*><verb_x-*te*> => while <verb_x>+ing

d. **If** the replicating verb and main verb are of antonym nature with respect to each other,
then
[<verb_x><verb_x>=>while <verb_x>+ing 'till the end'
<main_verb> => finally <main_verb>]

e. **If** the replicative verb has a suffix '*-te*' and the main verb is of the type *bachanaa* or *rukanaa*,
then
[<verb_x><verb_x>=>from<verb_x>+ing< main_verb> => just <main_verb>]

4 Conclusions

In this paper, we have discussed the phenomenon of replicative lexical elements in Hindi from the point of view of their processing in machine translation from Hindi to English. We have discussed different types of replicative words and their multiple grammatical functions and their mapping patterns in their translation in English. We have collected more than one thousand examples of sentences with replicative words from grammar and linguistics books and also from a considerable number of native speakers. We have examined their syntactic and semantic properties to identify their different functions and also their mapping patterns in English. We have taken into account the syntactic and semantic nature of the relevant words to disambiguate the multiple functions and the mapping patterns in English. We have formulated a hybrid strategy and some rules for the identification of different types of functions that the replicative words denote. Although there is a scope for these rules and examples being further explored, no such formulation and exploration from the angle of machine translation has been reported in the literature to the best of our knowledge.

The research reported here has been implemented in a Hindi-English machine aided translation system (Sinha, 2004) presently under development. The input sentences are paraphrased using the rules outlined here before being fed to the translation engine. The initial results on examination of the sentences from the corpus exhibit coverage of more than 50% for the replicative verbs and more than 90% for the rest of the categories.

[5] It is based on the semantics of the qualified noun.

References

Anvita Abbi. 1992. *Reduplication in South Asian Languages: An Areal, Typological and Historical Study.* Allied Publishers, New Delhi.

Anvita Abbi. 1977. Reduplicated Adverbs in Hindi Indian Linguistics. *Journal of the Linguistic Society of India 38*: 125-135.

Anvita Abbi. 1975. Reduplication in Hindi: A Generative Semantic Study. *Dissertation Abstracts Internacional, Vol. 36*, University of NY (1975).

Pau Friedrich. 1979. The Symbol and its Relative Non-Arbitrariness. In *Language Context and the Imagination*, selected by A.S. Dil, Stanford University Press, Stanford.

K.P. Guru. 1925. *Hindivyakaran* (in Hindi). Nagari Paracarini Sabha, Varansi.

Yamuna Kachru. 1980. *Aspects of Hindi Syntax.* Manohar, Delhi.

Ayendra Sharma. 1958. *A Basic Grammar of Modern Hindi.* Central Hindi Directorate, Government of India.

R.M.K. Sinha 2004. An Engineering Perspective of Machine Translation: Angla Bharti - II and Anu Bharti - II Architecture. In *Proceedings of International Symposium on Machine Translation NLP and TSS*, The McGraw-Hill Companies, New Delhi: 10-17.

K. D. Vajpai. 1958. *Hindi Shabdanushasan* (in Hindi). Nagari Paracarini Sabha, Varansi.