Hypothesis Refinement Using Agreement Constraints in Machine Translation

Ankur Gandhe
Carnegie Mellon University,
USA
ankurgan@andrew.cmu.edu

Rashmi Gangadharaiah
IBM Research,
India
rashgang@in.ibm.com

Abstract

Phrase-based machine translation like other data driven approaches, are often plagued by irregularities in the translations of words in morphologically rich languages. The phrase-pairs and the language models are unable to capture the long range dependencies which decide the inflection. This paper makes the first attempt at learning constraints between the language-pair where, the target language lacks rich linguistic resources, by automatically learning classifiers that prevent implausible phrases from being part of decoding and at the same time adds consistent phrases. The paper also shows that this approach improves translation quality on the English-Hindi language pair.

1 Introduction

Data driven Machine Translation approaches have gained significant attention as they do not require rich linguistic resources such as, parsers or manually built dictionaries. However, their performance largely depends on the amount of training data available (Koehn, 2005).

When the source language is morphologically rich and when the amount of data available is limited, the number out-of-vocabulary (OOV) increases thereby reducing the translation quality. Popovic and Ney (2004) applied transformations to OOV verbs. Yang and Kirchoff (2006) used a back-off model to transform unknown words, where, the phrase-table entries were modified such that words sharing the same root were replaced by their stems. Others (Freeman et al., 2006; Habash, 2008) found in-vocabulary words that could be treated as morphological variants.

Translating into a language that is rich in morphology from a source language that is not morphologically rich also has limitations. The main reason for this is that the source language does not usually contain all the information for inflecting the words in the target half. For language-pairs that have limited amounts of training data, it is unlikely that the Translation model comes across all forms of inflections on the target phrases. Hence, some mechanism is required in order to generate these target phrases with all possible inflections and at the same time be able to filter out the implausible hypotheses.

Certain approaches (Toutanova et al., 2008; Minkov et al., 2007; Green et al., 2012) predict inflections using syntactic and rich morphological sources for the target language. This approach cannot be applied on resource poor languages such as, Hindi or other Indian languages, which lack such rich knowledge sources. Ramanathan et al. (2009) use factored models to incorporate semantic relations and suffixes to generate inflections and case markers while translating from English to Hindi but do not consider the problem of agreement between phrases in the target sentence. William and Koehn (2011) suggested an approach to eliminate inconsistent hypotheses in a string-to-tree model by adding unification-based constraints to only the target-side of the synchronous grammar. Although tranfer-based MT (Lavie, 2008) uses rich feature structures, grammar rules and constraints are manually developed. In addition, rules formed for one language-pair cannot be applied to another language pair. However, it is possible to model these rules as a classification problem: Given the set of source language features that influence the inflection of the target word, we try to predict the best possible target class. The target class could be the either spontaneous words or inflections of words.

This paper, specifically looks at translating from English to Hindi to predict a) Subject case markers, b) Object case markers and c) Verb phrase inflections. In many PBSMT systems, once the

International Joint Conference on Natural Language Processing, pages 429–437,
Nagoya, Japan, 14-18 October 2013.

phrase-pairs have been extracted, it is no longer required to store the training corpus from which the phrase-pairs were extracted. However, while dealing with many morphologically rich languages, the morphological variants of the target phrase not only depend on their source phrase but also on the context in which the source phrase appeared. Hence, it is beneficial to incorporate source-side features while decoding and most PBSMT systems do not use any other information from the input sentence other than the source phrase itself.

This paper presents an approach to improve the translation quality while translating from a morphologically poor language (such as, English) to a target language that is morphologically rich without using any rich resources such as, parsers or morphological analyzers. The contributions of the paper are summarized as follows:

- The approach detects inconsistent hypotheses generated by the translation model by treating the task as a classification problem.
- The approach also predicts plausible target phrases that agree with the features extracted from the input sentence.
- The paper also shows how the incorporation of source-specific features during decoding results in better translations.

Section 2 provides motivating examples to understand the importance of the task at hand.

2 Motivation

We demonstrate the usefulness of our approach on Indian languages as they are rich in morphology. They are also considered as resource-poor and low-density languages due to the lack of data availability and the absence of rich knowledge sources like morphological analyzers or syntactic parsers. Hindi has a free word-order where the constituents are identified through case markers.

A few approaches generate the right inflection by a) capturing all possible variations within the target phrase (Gandhe et al., 2011) and b) use the language model to select the most fluent phrases. However, the following problems still remain:

1) Many language models typically use 4-gram or 5-gram models (even lower when the data available is scarce). Example 1a has a subject (Ram) that is masculine (masc)-3rd person (3)-singular(sg)-present progressive(pp) and example 1b, has a subject (Sita) that is feminine (fem)-3rd person(3)-singular(sg)-present progressive(pp). This difference in gender, changes the inflection on the auxiliary Hindi verb *raha*, from *'a'* (in 1a) to *'i'* (in 1b). It should be noted that lower order n-gram language models fail to obtain the right translation due to the long distance dependency between the subject (*Ram / Sita*) and the verb phrase (*khel raha hai / khel rahi hai* corresponding to *is playing* in English) in the target language.

Example 1a:
S: Ram is playing with the grand master .
T: *Ram grand master ke saath khel* ***raha*** *hai .*
(Ram grand master with play+*3*+*sg*+*masc*+*pp*)

Example 1b:
S: Sita is playing with the grand master .
T: *Sita grand master ke saath khel* ***rahi*** *hai .*
(Sita grand master with play+*3*+*sg*+*fem*+*pp*)

2) Language models are insufficient to produce the right inflections. Consider the case shown in example 2, where the translation of the English pronouns (*he/she*) is same in Hindi (both translate to *Woh*). The inflection on the axillary verb phrase (*raha hai / rahi hai*) is still being decided by the gender of the subject (*he/she*). Even if a higher order language model is employed, the language model gives equal preference to both the translations as the information about the gender of the subject is completely absent in the Hindi translation. Hence, the information that *Woh* corresponds to masculine in example 2a and feminine in example 2b has to come from the source sentence (*He/She*).

Example 2a:
S: He is playing chess .
T:*Woh chess khel* ***raha*** *hai .*
(he chess play+*3*+*sg*+*masc*+*pp*)

Example 2b:
S: She is playing chess .
T:*Woh chess khel* ***rahi*** *hai .*
(she chess play+*3*+*sg*+*fem*+*pp*)

3) Most often in PBSMT systems, the subject and verb phrases are far apart and hence are extracted independently, as in the case of example 1. Since there are no constraints during decoding on which phrases to choose, mis-matched phrases

may get picked. Apart from verb inflections, the presence of the case-marker '*ne*' (shown in example 3) on the subject blocks the transfer of the subject's gender onto the verb phrase and the verb phrase instead gets inflected with the gender of the object*(apple)*. This blocking/presence of case markers is also not captured by traditional PBSMT systems.

Example 3a:
S: He ate an apple .
T: *us ne* *seb* ***khaya*** .
(he apple ate+*3*+*sg*+*masc*+*past*)

Example 3b:
S: She ate an apple .
T: *us ne* *seb* ***khaya*** .
(she apple ate+*3*+*sg*+*masc*+*past*)

3 Model

The agreement constraints can be applied to either the translation model or the language model, such that implausible combination of phrases are not picked for the best hypothesis. In our approach, we apply the agreement constraints on the translation model by filtering phrase-pairs which have an incorrect inflection on the target phrase. Since the problem of inconsistent output is mainly due to the subject, object and verb phrases, we determine agreement constraints only for these target words. For instance, suppose a 'female' gender inflection is expected on the target verb. Then, any phrase that contains 'male' gender inflection on the verb will produce an inconsistent translation and hence should be penalized. We can also add phrase-pairs when the correct inflection is not present in the phrase table.

The easiest way to filter the inconsistent phrase-pairs is to create manual rules to look at the English source side that specify the possible set of target translations and discard the rest. For instance, using example 3 in Section 2, we could create a manual rule, "When the English verb tense is '*past*', Hindi subject takes the case marker '*ne*' and the verb phrase takes the gender and number of the 'subject' ". However, this is time consuming and it is difficult to create an exhaustive list of such rules. Hence, it is imperative that we learn these rules from data. In this paper, we use multi-class support vector machine (Crammer and Singer, 2001) classifiers that use features only from the input source sentence to predict possible target case marker/inflections for the subject, object and verb phrases in the target sentence. We treat these as the allowed inflections on the target phrases and penalize phrase-pairs that do not contain the predicted target inflections. This methodology is expected to prevent implausible sentences being translated and improve the overall fluency of the translated sentence.

4 Classification

We model the prediction of the possible target inflections for a given input sentence as a classification problem. We build different classifiers[1] to predict the target inflections of parts of the input sentence for which the translations are dependent on long range morphological rules. The features that we use for the different classifiers are listed in Section 5. The classifiers built are as follows:

Subject Classifier (SubCM) and Object Classifier (ObjCM): predicts the case marker on the subject and the object.

Verb Phrase Classifier (Vp): is used to predict the inflections on the verbs.

4.1 Subject and Object Classifier

Subject and Object phrases, when translated from English into a morphological rich language, often contain inflections of gender and number. Some languages also generate a case marker to denote the subject or the object. If such a case marker is not present, the target sentence often may not make sense. For our experiments from English to Hindi, we looked at predicting the correct case marker. To obtain the possible case markers that can come after a subject or the object in target language (in our case Hindi), we look at all the case markers following a subject and those that follow the object. If a language has linguistic resources such as parsers, this can be done easily. Since Hindi, and many other languages do not have a good parser, we make use of automatic word alignments obtained from bilingual data to project the subject information from English to Hindi, and determine the case markers following the subject and the object on the target side. Using this technique, we found 4 classes for the subject classifier and 3

[1] we use the libsvm library: http://www.csie.ntu.edu.tw/ cjlin/libsvm/

classes for the object classifier. For the prediction of the classes, we use all the noun phrase features (in Section 5.1), tense feature of the verb phrase (in Section 5.2) and tense conjugate features (in Section 5.3).

SubCM	**ObjCM**	**Vp**
NULL	NULL	*X raha tha* (was *X*+ing)
ne	*ko* (of)	*X+nA chahiye* (should *X*)
ke (of)	*mein* (in)	*X+nI chahiye* (should *X*)
ki (of)		*X+A gayA* (was *X*+ed) ...

Table 1: Classes defined for different classifiers.

4.2 Verb Phrase Classifier 1

Verb phrases contain morphological information about the gender, number, person, tense and aspect of the sentence. It is hence important to produce the right inflections and auxiliary verbs. Since it is impractical to have a class for each verb, we convert the verb phrases to an abstract form and also predict the target verb phrase in its abstract form. For instance, the verb phrase 'was playing' will be generalized to 'was X+ing' form and the corresponding predicted class would be *'X raha tha'*.

A simple approach to find the possible output forms of the classifier is to mine the target language data for all the verb phrases, rank them by frequency and filter them based on a threshold to yield the different forms that the verbs can take in the language. The aggregated verb phrases can be normalized by replacing the root verb in these phrases by an 'X' tag to obtain the possible abstract forms for the target verb phrases. For Hindi, verb phrases were identified by using a simple part-of-speech (POS) tagger to tag the monolingual data and to capture continuous sequences of 'V' tags. We found 120 Hindi verb classes in all. Some of these classes are listed in Table 1. We use all the features listed in Section 5.

4.3 Verb Phrase Classifier 2

Having too many classes for verb phrases causes the following problems: a) During our initial experiments we found that out of the 120 verb classes specified by us, only 60 were present in the bilingual training data. This reduces the chances of predicting a correct class since the classifier does not see all classes during training. b) The classifier sees only a few instances of each class. To simplify the verb phrase prediction, we split the prediction such that instead of predicting each verb form, we predict each 'kind' of inflection that modifies the verb phrase. Since each verb phrase in our training data contains information about the gender, number and person, each class now has ample amount of training examples.

Gender Classifier (VpG): This classifier predicts the gender inflections on the target verb phrases using features from the source sentence.

Number Classifier (VpN): This classifier predicts the number inflections on the target verb phrases using features from the source sentence.

Person Classifier (VpP): This classifier predicts the Person information of the target verb phrases given the source sentence features.

The three classifiers have two, two and three classes, respectively. The predicted gender, number and person is then used to select the target verb form:

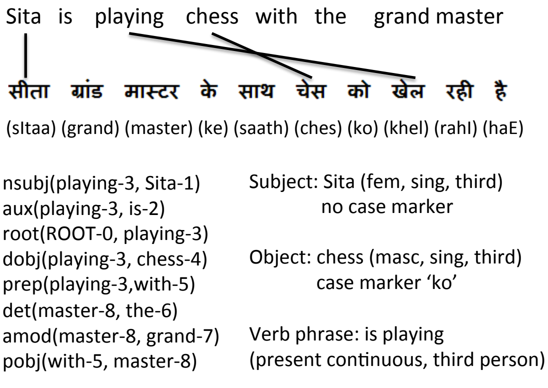
Base Verb form Function: Given the input English verb phrase, this function outputs all possible translations (that is, with all possible inflections and auxiliary verbs) of the given verb form. For example, for the verb phrase 'is playing' in the example in Section 1, this function will produce 12 target verb forms, one each for possible combinations of elements from the sets (masculine and feminine), (singular and plural) and (first, second and third person). The function for producing the list of verb forms given the English verb form is implemented using machine alignments and monolingual data as done in Gandhe et al. (2011). It uses parallel data to extract all the source-target verb phrase-pairs from the word-aligned data. These source-target verb phrase-pairs are converted into an abstract form by replacing the root verb with an 'X' (as done in Section 4.2). Aggregating this over a large amount of parallel data and filtering out the low frequency phrase-pairs gives us translations of a source verb form into its corresponding target forms. The gender, number and person for each of the target verb forms can be found out by looking at the inflections, suffixes and auxiliary verbs.

4.4 Training

We use an English parser to parse the source sentence and obtain the different features. Using the alignments of the subject, object and verb phrase, we project them onto the target language and extract the expected output case-marker/inflections for each of the three cases (SubCM, ObjCM, Vp) and assign it the corresponding class. Our approach is not limited to hand-alignments. Alignments obtained from automatic aligners can also be used. Since hand-alignments were available beforehand, we made use of these alignments in this work. We will explore the usability of automatic aligners as future work. We now briefly describe the features that we used for the above classifiers.

5 Features

Given the parse tree of an English sentence, we determine the subject noun phrases and the object noun phrases for each of the verb phrases present in the input sentence giving *(subject,object,verb)* triples. We also determine the morphological information about the subject, object and verb phrases in sentence (in Sections 5.1 and 5.2). Most of the features described are boolean, unless specified otherwise. Figure 1 shows an example of an English-Hindi word-aligned sentence-pair. The dependency parse of the English sentence is used to determine the source subject (sita), object(chess) and the verb phrase (is playing). Features are calculated over these phrases and the target words aligned to them in the word alignments are used to create the training examples for the three classifiers.


Sita is playing chess with the grand master
सीता ग्रांड मास्टर के साथ चेस को खेल रही है
(sItaa) (grand) (master) (ke) (saath) (ches) (ko) (khel) (rahI) (haE)

```
nsubj(playing-3, Sita-1)        Subject: Sita (fem, sing, third)
aux(playing-3, is-2)                     no case marker
root(ROOT-0, playing-3)
dobj(playing-3, chess-4)        Object: chess (masc, sing, third)
prep(playing-3,with-5)                   case marker 'ko'
det(master-8, the-6)
amod(master-8, grand-7)         Verb phrase: is playing
pobj(with-5, master-8)          (present continuous, third person)
```

Figure 1: An English parse with features.

5.1 Noun Phrase Features

The inflection on the verb phrase is influenced by 3 attributes of a noun phrase:

Gender: Unlike English, most Indian languages have a gender *(male/female)* for every subject and object. To determine the gender of an English word, we take its most common Hindi translation and assign the gender of this translation to the English word. Gender of Hindi words can be determined by mining the Hindi monolingual data for *(noun phrase,verb phrase)* pairs using a simple POS tagger on Hindi data. POS taggers are now easily available for most Indian languages. However, no other rich sources such as, parsers or morphological analyzers are used on the target language. We then assign the gender of the verb phrase suffix ('*a*' for masculine and '*I*' for feminine) to the words in the noun phrase. Doing this over a large amount of data gives us the list of nouns with their gender. For example, the Hindi word 'kItAb' is seen with verb phrases such as, 'pad*I*','d*I*', etc. in the monolingual data. Since 'kItAb' occurs most with verb phrases ending in suffix 'I', its gender is 'female'. The English word 'book' translates most often to *'kItAb'* and is hence assigned the gender 'female' and the corresponding feature value of 1. For words like, 'house', which are determined to be 'male', the value is 0.

Number: Similar to the gender, the singularity or plurality of the noun phrase influences the inflection on the verb phrase. The plurality of the English noun can be determined by using a POS tagger and looking for a '*NNS*' tag or in case of pronouns, a finite list of pronouns. Hence, nouns in plural form and the pronouns, 'they','us','them', were given the feature value as 1. For all other singular words, the value is 0.

Presence of case marker: Perhaps the most important feature, the presence or absence of a case marker on the target subject and object phrase decides the transfer of inflections from the noun phrases to the verb phrase (examples of Section 2). This is not a source side feature, since case markers are present on the noun phrases in the target language. We cannot use the case marker information directly as we do not have the target side information. Hence they are used in two steps: a) Subject and Object classifiers (Section 4) are used to predict the noun phrase *(subject,object)* case markers and b) The predicted case markers are used as an input to the verb phrase classifier. This feature is not used as an input to the subject and object classifiers. If a subject/object case marker is present, the features are valued 1, else 0.

5.2 Verb Phrase Features

The verb phrase features influence the tense, aspect and person of the target verb phrase as well as the case marker presence on the noun phrases. The verb phrase extracted from the dependency parse of the input sentence are morphologically segmented (Minnen et al., 2001) and the different aspects of the verb phrase are obtained from it.

Tense Features: The tense features tell the presence or absence of *Present, Past and Future* tense. For instance, for the verb phrase 'was explained', the present and future features take the value 0 and the past feature takes the value 1.

Aspect Features: The aspect features are important in deciding the final form and and the auxiliaries in the target sentence. We label the features as *simple, progressive and perfect*. In this case, a verb phrase with a 'ing' suffix is said to be progressive, whereas a verb phrase with 'have' and its inflections is said to be perfect. For example, the phrase 'has been explaining' will have both progressive and perfect features with value 1.

Mood Features: The mood features capture the *obligation, conditional and probability* mood in the input English sentence by looking at the modal verbs which are required to produce the corresponding auxiliary verbs in Hindi.

Number: English verb forms with plurality inflection translates into plurality of the Hindi verbs.

Person: English auxiliary verb '*am*' denotes the presence of first person. By looking at the subject of the verb in the dependency parser, *(first, second or third)* the person information can be assigned to the verb phrase.

5.3 Conjugate Features

These features capture the more language-specific nuances that together decide the transfer of inflections from nouns to verbs. These features try to emulate the behavior of grammar rules.
Case marker-Gender: When a case marker is not present on the noun phrase, the inflection from them is likely to be transfered to the verb phrase. For this case, we assign this feature the same value as the gender of the noun phrase. When a case marker is present, information is blocked and hence we assign a null value to this feature.

Case marker-Number: This feature captures blocking of the number information and takes a value 0 or 1 depending on the presence or absence of case marker.

Tense-Gender: When the tense of the sentence is past, it is likely that the gender information is blocked. Hence, when the tense is past, this feature is assigned a null value. Otherwise, the value is same as the value of the gender feature.

Tense-Number: Similar to the previous one, except that this captures the blocking of number information.

6 Decoding

We used a PBSMT system, similar to Tillman et al. (2006), to decode and this required slight modifications to incorporate our approach. The extracted phrase-pairs have phrase translation probabilities and lexical probabilities estimated (similar to Papineni et al. (2002)). The input sentence is passed through a parser to determine the subject, object and the verb phrases in the sentence. Various features mentioned in the previous section are computed during run time and the classfiers are used to predict the subject case marker, object case marker and the verb phrase inflection. The agreement constraints can be applied as:

Hard Removal: All phrase-pairs that do not agree with the predicted case marker or inflections are removed from the phrase table before the hypothesis search.

Soft Removal: The agreement model outputs the prediction probabilities for different target case markers or inflections. This probability score can be used as a feature in the phrase table and trained on a development data set.

Addition: If the predicted case marker or inflection is not present in the original phrase table, the correct phrase-pair can be added by automatically generating the target phrase.

The input sentence is fed into the agreement model to produce the constraints for the subject, object and verb phrases. We use the hard constraint and addition techniques during decoding. Applying soft constraints will be done in future work. For subject and object phrases, we aggregate the phrase-pairs in the phrase table which contain the English source word. From these, all phrase-pairs that do not agree with the predicted case markers on the target side are filtered. In addition, if the predicted case marker is not present in the phrase table, we add the phrase-pair with the right case marker into the phrase table. This is done by looking for the most common target translations of the source word and appending the predicted case marker to them. For verb phrases, we aggregate the phrase-pairs containing the English verb phrase.

All phrase-pairs which do not have the predicted target verb phrase inflections are filtered. Since we do not know the complete translation of the source verb phrase at this step, we look only for the predicted target verb phrase's inflection and auxiliary verbs. If no correct verb phrase form is found in the phrase table, the target phrase is generated using the most common translation of the English verb and the phrase-pair is added. Inorder to score these new phrase-pairs, we can make use of the automatically generated bilingual dictionaries created during the automatic word-alignment phase. The phrase-pairs and entries in the dictionaries can be stemmed to their base forms (removing inflections) using Ramanathan et al. (2003). In cases where there are multiple instances of the same verb (caused due to stemming) present in the modified dictionary, the average of the probabilities is taken. The lexical probabilities for the phrase-pairs can then be estimated as given in Papineni et al. (2002) from the modified dictionaries. To obtain the phrase translation probabilities, the scores from the classifiers are converted to a score between 0 and 1 using a logistic function ($1/(1+e^{-score})$, where, $score$:classifier's score) and then re-normalized such that the sum of probabilities of all the target phrases for a particular source phrase is one (and vice versa). In the case of 'Verb Phrase Classifier 2' (Section 4.3), the scores from each of the classifiers is first converted to a score between 0 and 1 using a logistic function, summed and then re-normalized.

7 Experiments

We first report the results of prediction of noun phrases and verb phrases and proceed on to report the results of using them in PBSMT.

7.1 Prediction Evaluation

To aggregate the classes required for subject, object and verb phrase classifiers, we used 1.4 million Hindi monolingual sentences crawled from the web. We pos-tagged this data using iit kgp Hindi pos tagger [2]. The monolingual data, along with 280,000 automatic alignments of sentence-pairs, was used to apply the technique suggested in Gandhe et al. (2011) to build the base verb form function described in Section 4.2. The svm classifiers were trained and tested using libsvm [3]. To extract the features from manually aligned sentences, we used the Stanford Parser[4] to obtain the English dependency parse trees. The source English side was morphologically segmented using morpha (Minnen et al., 2001) and the target Hindi side was segmented using an approach described in Ramanathan et al. (2003).

Table 2 gives the accuracies of the classifiers when trained with a particular set of features. The conjugate features make a significant improvement to all the three classifiers. Hindi object case markers are easier to predict than subject case markers since the objects usually do not occur with a case marker. Also, the subject case markers show a high dependency on the verb phrase features, which is explained by grammatical rules, according to which tense and structure of the verb phrase decide the case marker on the subject. It is important to remember here that the verb phrase classifier uses the output of the case-markers predicted by noun classifiers as a feature.

Features	**SubCM**	**ObjCM**	**Vp**
NounFeat	0.63	0.81	-
Noun+VerbFeat	0.72	0.84	0.58
Noun+Verb+ConjFeat	**0.75**	**0.87**	**0.61**

Table 2: Prediction accuracy for the classifiers.

The prediction accuracy is low for the **Vp** classifier even with conjugate features due to the large number of classes. Most classes do not have sufficient training examples and a few classes were

[2] http://nltr.org/snltr-software/

[3] http://www.csie.ntu.edu.tw/ cjlin/libsvm/

[4] http://nlp.stanford.edu/software/lex-parser.shtml

even absent in the training data. When we split this classification into separate tasks as explained in Section 4.3 and later combine the output of individual classifiers to obtain the predicted verb phrase, we obtain a much better accuracy. The results of this configuration are shown in Table 3. Since the verb phrase classifier uses case-markers as a feature, we also analyze the importance of these for verb phrase prediction and study 3 different settings: a) Removing the case marker (CM) feature, b) Using Gold case markers from the reference and c) Using the predicted case markers. Although the prediction accuracies are best for GoldCM, using the predicted case markers results in only a slight drop in accuracy.

	VpN	VpG	VpP	Overall
No CM	0.83	0.62	0.95	0.58
Gold CM	**0.87**	**0.86**	**0.95**	**0.74**
Pred CM	0.85	0.83	0.95	0.70

Table 3: Prediction accuracy for verb phrase inflections.

7.2 Machine Translation Evaluation

The system was trained on 285,000 automatically aligned sentences. The baseline system uses the standard decoding algorithm while our approach prunes the phrase table before decoding. We measure the translation quality using a single reference BLEU (Papineni et al., 2002). The test set contains 715 sentences from the News domain. Table 4 gives the comparison of the baseline with the two systems (Note: In both systems, the case marker features are obtained from the predictions of the subject and object classifiers):

Pred1: Verb phrase prediction as a single task (Table 2)

Pred2: Verb phrase prediction split into individual components (Table 3).

	BLEU	Adequacy	Fluency
Baseline	15.43	3.75	2.23
Pred1	15.45	3.87	2.41
Pred2	**15.58**	**3.93**	**2.79**

Table 4: BLEU score and Human Judgment.

The BLEU score increase is small on Pred1 but was significantly better with Pred2 with $p < 0.0001$ with the Wilcoxon Signed-Rank test (Wilcoxon, 1945) performed by dividing the test file into 10 equal subfiles (as done in Gangadharaiah et al. (2010)). On analysis of the reference, we found the tense of the verb phrases in the Hindi reference to be different from that of English. Also, often the presence of auxiliary verbs *'hona'* in the Hindi reference changed the structure of the verb phrase. The output produced by our system is more literal and in congruence with the grammar of the input sentence. Callison et al. (2006) list the disadvantages of using BLEU. The differences in translations between the proposed approaches and the baseline are most often a correction of inflection, and sometimes this resulted in better selection of neighboring words by the language model. BLEU failed to accommodate these improvements, hence we also performed human evaluation to judge the quality of the translations on adequacy and fluency using a scale of 1-5[5].

We gave 100 randomly picked sentences from the test set to a single human judge. We see that our approach (Table 4) has a greater impact on fluency, suggesting that grammatical agreement is important for fluency. Adequacy improvement can be attributed to the correct translations of the case markers and the tense information.

8 Conclusion and future work

We modeled the task of case marker and inflection prediction as a classification task.The prediction accuracies show that the inflections on the verbs are highly influenced by the case markers on the subjects and objects. Similarly, the case markers on subjects are affected by the tense of the verb phrases. Since all the features are extracted from the source side, this approach can be easily applied for improving translation quality from English to any morphologically rich foreign language. More work can be done on creating features that encode the grammatical rules we might have missed.

Even though the gain in translation quality with the BLEU score was small, human evaluation showed that this approach helps in improving the fluency and adequacy of the sentence and hence makes it more readable. Future work can be on using more than one possible case marker-verb phrase constraints (i.e., as a soft constraint) for a given input and applying this approach for other language-pairs where the target language is morphologically rich.

[5] We used the scale defined in http://projects.ldc.upenn.edu/TIDES/Translation/TransAssess04.pdf

References

C. Callison-Burch, M. Osborne, and P. Koehn. 2006. Re-evaluating the role of bleu in machine translation research. In *EACL*, pages 249–256.

K. Crammer and Y. Singer. 2001. On the algorithmic implementation of multiclass kernel-based vector machines. In *Journal of Machine Learning Research*, pages 265–292.

A. T. Freeman, S. L. Condon, and C. M. Ackerman. 2006. Cross linguistic name matching in english and arabic: a "one to many mapping" extension of the levenshtein edit distance algorithm. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, HLT-NAACL '06, pages 471–478.

A. Gandhe, R. Gangadharaiah, K. Visweswariah, and A. Ramakrishnan. 2011. Handling verb phrase morphology for indian languages in machine translation. In *Proceedings of the International Joint Conference on Natural Langauge Processing*. Asian federation for NLP.

R. Gangadharaiah, R. D. Brown, J. Carbonell. 2010. Monolingual distributional profiles for word substitution in machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 320–328.

S. Green and J. DeNero. 2012. A class-based agreement model for generating accurately inflected translations. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, ACL, pages 146–155.

N. Habash. 2008. Four techniques for online handling of out-of-vocabulary words in arabic-english statistical machine translation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, HLT-Short '08, pages 57–60.

P. Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT Summit*.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, F. Marcello, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, E. Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. *Annual Meeting of the Association for Computational Linguistics (ACL)*, demonstration session.

A. Lavie. 2008. Stat-xfer: a general search-based syntax-driven framework for machine translation. In *Proceedings of the 9th international conference on Computational linguistics and intelligent text processing*, CICLing, pages 362–375.

E. Minkov, K. Toutanova, and H. Suzuki. 2007. Generating complex morphology for machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, ACL, pages 128–135.

G. Minnen, J. Carroll, and D. Pearce. 2001. Applied morphological processing of english. In *Natural Language Engineering*.

K. Papineni, S. Roukos, T. Ward and W.J. Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL-2002: 40th Annual meeting of the Association for Computational Linguistics*, pages 311–318.

M. Popovic and H. Ney. 2004. Towards the use of word stems and suffixes for statistical machine translation. In *Proceedings of The International Conference on Language Resources and Evaluation*.

A. Ramanathan and D. Rao. 2003. A Lightweight Stemmer for Hindi. *Workshop on Computational Linguistics for South-Asian Languages*, EACL.

A. Ramanathan, H. Choudhary, A. Ghosh and P. Bhattacharyya. 2009. Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pages 800–808.

C. Tillman. 2006. Efficient Dynamic Programming Search Algorithms for Phrase-based SMT. In *Proceedings of the Workshop CHPSLP at HLT'06*.

K. Toutanova, H. Suzuki, and A. Ruopp. 2008. Applying morphology generation models to machine translation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* pages 514–522.

F. Wilcoxon. 1945. *Individual comparisons by ranking methods.* Biometrics, 1, 80-83, *http://faculty.vassar.edu/lowry/wilcoxon.html.*

P. Williams and P. Koehn. 2011. Agreement constraints for statistical machine translation into german. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, WMT, pages 217–226.

M. Yang and K. Kirchhoff. 2006. Phrase-based backoff models for machine translation of highly inflected languages. In *Proceedings of the 21st International Conference on Computational Linguistics*, pages 1017–1020.