, Vol. 0 (2012), No. 0, pp. 00–00

Improving Subject-Verb Agreement in SMT

Eva Vanmassenhove, Jinhua Du, Andy Way

ADAPT Centre/School of Computing, Dublin City University, Ireland,

eva.vanmassenhove2@mail.dcu.ie, jinhua.du@adaptcentre.ie, away@computing.dcu.ie

Abstract. Ensuring agreement between the subject and the main verb is crucial for the correctness of the information that a sentence conveys. While generating correct subject-verb agreement is relatively straightforward in rule-based approaches to Machine Translation (RBMT), today's leading statistical Machine Translation (SMT) systems often fail to generate correct subject-verb agreements, especially when the target language is morphologically richer than the source language. The main problem is that one surface verb form in the source language corresponds to many surface verb forms in the target language. To deal with subject-verb agreement we built a hybrid SMT system that augments source verbs with extra linguistic information drawn from their source-language context. This information, in the form of labels attached to verbs that indicate person and number, creates a closer association between a verb from the source and a verb in the target language. We used our preprocessing approach on English as source language and built an SMT system for translation to French. In a range of experiments, the results show improvements in translation quality for our augmented SMT system over a Moses baseline engine, on both automatic and manual evaluations, for the majority of cases where the subject-verb agreement was previously incorrectly translated.

Keywords: Subject-Verb Agreement, Statistical Machine Translation, Hybrid MT, Source-Language Preprocessing

1 Introduction

Ensuring agreement between the subject of a sentence and the main verb is crucial for the correctness of the information that a sentence conveys. Any disagreement may lead to ambiguity and therefore affects the adequacy and fluency of a sentence considerably. On the one hand, RBMT produces translations that are syntactically better than SMT, where very obvious errors such as lack of number and gender agreement occur. On the other hand, RBMT systems tend to have problems with lexical selection and fluency in general. Furthermore, they often rely on linguistic resources such as parsers that may fail (España Bonet, C. et al., 2011). While generating correct agreements in translations is relatively straightforward in rule-based approaches to Machine Translation, this is much harder to achieve using state-of-the-art statistical approaches. Indeed, recent research on subject-verb agreement of Persian sentences translated from English revealed that, even for Google Translate – the world's most widely used SMT system – subject-verb agreement remains an issue (Bozorgian and Azadmanesh, 2015).

This can distract human post-editors from the benefits of using SMT as a tool to increase their productivity; as subject-verb agreement is deemed to be 'easy' for both L1 and L2 speakers, translators rightly expect MT systems to get this right, and when they do not, whatever benefits do accrue from using MT as a productivity enhancer are masked by such obvious, 'simple' errors.

The difficulties with agreement arise due to the fact that agreement rules are language-specific and therefore also dependent on the specific morphological structure of a language (Avramidis and Koehn, 2008). The problem of agreement becomes increasingly difficult when dealing with translations from a morphologically poor[1] rich (or richer) language (e.g. French), where one surface verb form in the source language corresponds to several surface verb forms on the target side. This implies that choosing the correctly inflected form of the target word requires additional information that cannot be inferred by merely relying on the source word.

Initially, research on integrating morphological information in SMT aimed to improve translation quality from a morphologically rich language, such as Greek or French, into English – a morphologically poor language (Corston-Oliver and Gamon, 2004; Nießen and Ney, 2004; Birch et al., 2007; Carpuat, 2009; Wang et al. 2012). The difficulty when translating from a morphologically rich language into a morphologically poor one is a many-to-one problem that can be solved by converting the actual word form into lemmas or stems in a pre-processing step. More recently, several strategies have been proposed to translate from a morphologically poor language into a morphologically richer language, i.e. a one-to-many problem. This is, as stated in (Koehn, 2005), a more complex task since grammatical features such as number or gender might need to be inferred during the decoding process. Solutions that have been proposed to handle morphology-related difficulties include: (i) preprocessing of the source data, on the assumption that the necessary information to translate an ambiguous word can be found in its source context (Ueffing and Ney, 2003; Avramidis and Koehn, 2008; Haque et al., 2010), or (ii) a combination of both pre- and post-processing in a two-step translation pipeline. The two-step translation method usually implies first building a translation model with stems, lemmas or morphemes, and then inflecting them correctly (El-Kahlout and Oflazer, 2006; Virpioja et al., 2007; El Kholy and Habash, 2012; Fraser et al., 2012). Both pre- and post-processing of source or target language relies on linguistic resources such as Part-of-Speech (POS)-taggers, chunkers, parsers, and manually constructed dictionaries, all of which – assuming them to be available at all – work to different levels of performance. Koehn and Hoang (2007) proposed, instead of having a pre- and/or post-processing step, to have a tighter integration of linguistic information by introducing factored models. In factored translation models, words are represented as vectors that can contain (apart from the word form) lemmas, POS-tags.

[1] The term morphologically poor versus morphologically rich might be a bit problematic. However, since this is the terminology used in other papers on a similar topic (Ueffing and Ney, 2003; Avramidis and Koehn, 2008 and El Kholy and Habash, 2012) to refer to more analytic vs synthetic languages, we adopted the same terminology.

However, merely adding lemma and POS information will not provide the translation model with the information necessary to select the correctly conjugated verb form in the target language.

In this paper, we propose a hybrid approach to tackle the subject-verb agreement in MT from English – a morphologically poor language – into French, a morphologically richer language. Our approach applies a set of rules to the 'morphologically poor' source-language data in order to render it more 'morphologically rich'. Based on source-side information (including POS-tags and the distance between identified subjects and possible main verbs), we modify the identified verb forms in such a way that instead of a one-to-many relationship between source and target verb forms, we create a one-to-one relationship[2] between them, by mapping the verb form in the source language to a single correct verb form in the target language. Our method thus makes minimal changes to the source language and so avoids creating unnecessary extra sparsity. Note that while Koehn (2010, p.313) observes that in general, agreement errors occur "between multiple words, so simple word features such as part-of-speech tags do not give us sufficient information to detect [them]", our results demonstrate that at least as far as subject-verb agreement errors are concerned, POS information can be very useful indeed when combined with some simple rules.

The verb forms that appear to be specifically difficult to tackle for MT are the 1^{st} and 2^{nd} singular and plural. This is due to the fact that: (i) they are not as common as the 3^{rd} person in written texts, so data-wise are under-represented, and (ii) in English, they share the same verb form with more frequently appearing verb forms, such as the 3^{rd} person plural. However, the context in which the 1^{st} and 2^{nd} person appear is limited since those verb forms can only appear in combination with their specific pronouns (I, you, we). This contrasts with 3^{rd} person singular verb forms which can take any NP, VP, PP or even a whole clause as subject, as we demonstrate in (1):

(1)

VP as SUBJ: [$_{VP}$Being a Man Utd fan] makes no sense!
PP as SUBJ: [$_{PP}$In the army] is not a safe place to be.
S as SUBJ: [$_{S}$That the world is round] is no longer in doubt.

Based on the appearance of a specific pronoun in a sentence, we enriched the closest verb form (within a window size of 4, established empirically) in order to create a one-to-one relationship with the source-language verb forms.[3] We did this for all pronouns except for the third person verb forms since (i) creating a different verb form for the 3^{rd} person based on the appearance of a specific pronoun would only create additional unnecessary sparsity to our data, and (ii) due to the fact that we changed the verb forms of the 1^{st} and 2^{nd} person and the 3^{rd} person singular already has a different form (s-ending), the 3^{rd} person plural will be the only one left with the base form in the present tense. Our method only requires the use of a POS-tagger on the English side in order to retrieve the conjugated verbs and label them according to the closest pronoun.

[2] For most of the verb-forms.

[3] We are aware of the work of Cai et al.(2009) on subject-verb agreement for English using a dependency grammar approach. While this may be of interest for our work, it is not conducted within the specific remit of MT. Nonetheless, we plan to compare our ability to find the subject-verb pairs with theirs in future work.

We believe that our approach can be used to help generate more correct subject-verb agreements in terms of number and person when translating from a morphologically richer language into a more analytic one. However, for some language pairs, more specific information e.g. 'gender' information, might be required in order to select the correct verb form.

The remainder of the paper is structured as followed: The related work is briefly discussed in Section 2. In Section 3, a description of our model enriched with context information from the source-side is presented. In Section 4, we provide our experimental results along with some discussion. In Section 5, we conclude along with avenues for future work.

2 Related Work

In the past ten years, plenty of work has been done on error detection and correction in Machine Translation leading to improvements in translation performance (Vilar et al., 2006; Zhou et al., 2008; Popovic and Ney, 2011; Zeman et al., 2011; Yuan and Felice, 2013; Daems et al., 2014; Wang et al., 2014; Wisniewski et al., 2014). However, in this paper, we focus on modeling the source language in pre-processing step to prevent subject-verb agreement errors and not on the detection and correction of errors in already translated segments.

Agreement rules are language-dependent and become increasingly difficult to 'learn' for an SMT system when source and target languages have significantly different morphological structures. Languages that have a morphologically poor structure such as English where, in the present tense, for example, only the third person singular (infinitive +s) can be distinguished from the others by looking at its surface form, are particularly hard since one verb form in English can be matched with several verb forms in (say) French, as Table 1 illustrates.

see	vois, voyons, voyez, voient, voir
sees	voit

Table 1. Single English surface verb forms mapping to multiple French verb forms

Ueffing & Ney (Ueffing and Ney, 2003) were one of the first to enrich the English source language to improve the correct selection of a target form when still working with word-based SMT. By using POS-tags, they spliced sequences of words together (e.g., 'you go' $\rightarrow$ 'yougo') to provide the source form with sufficient information to translate it into the correct target form. By introducing phrase-based models for SMT, this particular problem of word-based SMT seemed largely solved. However, the language model statistics are sparse and due to an increase in morphological variations they become even sparser which can cause an SMT system to output sentences with incorrect subject-verb agreement even when the subject and verb are adjacent to one another.

Although syntax-based models tend to produce translations that are linguistically correct, the syntactic annotations added increase the complexity which leads to slower training and decoding. Furthermore, in general, phrase-based systems still outperform syntax-based ones. Therefore, we decided to add linguistic knowledge in a phrase-based machine translation system.

Within the field of phrase-based SMT, several works have focused on dealing with problems specific to translations into morphologically richer languages. Generally, those works focus on improving phrase-based SMT by: (i) source-language pre-processing of the source data (Avramidis and Koehn, 2008; Haque et al., 2010), and (ii) a combination of both pre-processing of the source language and post-processing of the target language (Virpioja et al., 2007; Mareček et al., 2011; Fraser et al., 2012; El Kholy and Habash, 2012). Avramidis & Koehn (2008) added per-word linguistic information to the English source language in order to improve case agreement as well as subject-verb agreement when translating to Greek and Czech. To improve subject-verb agreement they identified the person of a verb by using POS-tags and a parser. The information of the person was added to the verb as a tag containing linguistical information. Their initial system suffered from sparsity problems which led to the creation of an alternative path for the decoder with fewer (or no) factors. Although there were no significant improvements in terms of BLEU scores (Papineni et al., 2002), manual evaluation revealed a reduction in errors of verb inflection. Haque et al. (2010) presents two kinds of supertags to model source-language context in hierarchical phrase-based SMT: those from lexicalized tree-adjoining grammar and combinatory categorial grammar. With English as a source language and Dutch as the target language, they reported significant improvements in BLEU scores.

Other research has focused on both pre- and post-processing the data in a two-step translation system. This implies, in a first step, simplifying the source data and creating a translation model with stems (Toutanova et al., 2008), lemmas (Mareček et al. 2011; Fraser et al., 2012) or morphemes (Virpioja et al., 2007). In a second step, an inflection model tries to re-inflect the output data. In Toutanova et al. (2008), stems are enriched with annotations that capture morphological constraints applicable on the target side to train an English–Russian translation model, with target forms inflected in a *post hoc* operation. Two-step translation systems working with lemmas instead of stems were presented in both Mareček et al. (2011) and Fraser et al. (2012). While Mareček et al. (2011) perform rule-based corrections on sentences that have been parsed to dependency trees for English-to-Czech, Fraser et al. (2012) use linear-chain Conditional Random Fields to predict correct German word forms from the English stems. Opting for a pre- and post-processing step is necessary when language-specific morphological properties that indicate various agreements are missing in the source language (Mareček et al., 2011). Note that all the methods described above require (a combination of) linguistic resources such as POS-taggers, parsers, morphological analyzers etc.

In contrast with the research mentioned above, our work focuses only on subject-verb agreement and not on other problems related to translations into morphologically rich languages (e.g. case or other types of agreement). We will show that improving subject-verb agreement when translating from English to French does not require a two-step translation pipeline where both source and target language are remodeled since the morphological structure of French is not as complex as respectively Russian (Toutanova et al., 2008), Czech (Mareček et al., 2011) or German (Fraser et al., 2012). Therefore, as in Avramidis & Koehn (2008) and Haque et al. (2010), we aim to improve subject-verb agreement by building a system that augments the source-language data with extra information drawn from the source-side context. However, unlike that work, we do this by using only a POS-tagger on the English side.

3 Modeling of the Source Language

In this section, we describe in more detail how we enriched the morphologically poor source-side of our translation model in order to correct agreement errors in the target.

We use the Penn TreeBank tagset (Marcus et al., 1994) (the default tagger used in the nltk package)[4] to tag the source sentences. Once the source sentences contain the information from the POS-tagger we can, in the next step, use this information to look for verb forms that agree in person with a subject. These are the non-3^{rd} person singular present ('VBP'), the 3^{rd} person singular present ('VBZ'), the past verb tense ('VBD') and modal verbs ('MD').

Within the already tagged sentences, we search for 1^{st} and 2^{nd} person pronouns ('I', 'you' and 'we'). Once a pronoun is found, we identify the closest verb form (a verb tagged 'VBP', 'VBZ', 'VBD' or 'MD') following the pronoun, within a window of size 4. The verbs found are enriched with information of the pronoun as in Table 2 [5]

I work	I work1sg
You work	you work2
we work	we work1pl

Table 2. Enriching English surface verb forms with POS Information

We distinguish between declarative and interrogative sentences by looking at the last token of the sentence. In case it is a question mark, we identify this sentence as an interrogative sentence. For interrogative sentences, the verb typically (but not always) precedes the pronoun, so in these types of sentences, we first look for a verb appearing before the pronoun (within a window size of 2, established empirically) before looking at the words following it (within a window of size 4, established empirically).

Verbs that are 3^{rd} person singular already distinguish themselves from the others by having the s-suffix. Since we artificially enriched all other verb forms with information from the pronoun they agree with, the 3^{rd} person plural is the only one remaining in its original surface form.

Although our method does not resolve the ambiguity between the third person singular and third person plural in the past tense, it does reduce the complexity of disambiguation problem by converting a one-to-many problem into a one-to-two problem. Furthermore, since 3^{rd} person plural and singular are both very common verb forms, their disambiguation is less problematic for the language model to resolve.[6].

[4] http://www.nltk.org/

[5] In a future stage, we would like to add a few rules in order to detect compound subjects and label the verbs accordingly.

[6] For the 2^{nd} person the ambiguity remains given that the pronoun *you* is identical in both singular and plural. However, in French, the 2^{nd} person plural is also ambiguous since it can be both plural or singular (polite form).

4 Empirical evaluation

4.1 Experimental Setup

To evaluate our approach, we build two types of SMT systems with the Moses toolkit (Koehn et al., 2007): (i) from the original data, that we refer to as *baseline*, and (ii) from our morphologically enriched data, which we refer to in the rest of this paper as *morphologically-enriched* systems. We then score these SMT systems using automatic evaluation metrics as well as manual error analysis and compare them.

For training we use subsets of increasing sizes (respectively 200K, 400K and 600K sentences) of the Europarl parallel corpus (Koehn, 2005) for the English–French language pair. Both the baseline data as well as the morphologically-enriched data are tokenized and lowercased using the Moses tokenizer. Sentences longer than 60 tokens are filtered and not used in our model. We use the default Moses settings to train our systems.

Since we are specifically interested in subject-verb agreement, we want to have as much variety in verb forms as possible for the development set and the test set. Accordingly, we created our development and test sets from the WMT development sets from 2008 until 2013.[7] We select from these data the sentences that contain 1^{st} person and 2^{nd} person pronouns and 3^{rd} person verb forms. The 2098 sentences retrieved, we split into a development set of 1000 sentences and a test set of 1098 sentences. To manually evaluate the performance of the morphologically-enriched SMT against the baseline SMT systems, we randomly extract 60 out of the 1098 input sentences containing at least 10 occurrences of each verb form. Table 3 gives an overview of the number of pronouns appearing in the development set, test set and manual test set.

# of Pronouns in each set	DEV SET	TEST SET	MANUAL TEST SET
I	458	542	28
YOU	317	370	32
HE, SHE, IT	382	339	28
WE	519	417	18
THEY	82	54	11

Table 3. Number of different pronouns in the development set, test set and manual test set.

For our morphologically-enriched system, we use exactly the same training, development and test sets but pre-processed as described in Section 3.

4.2 Results

To score each of the SMT systems we built, we use the automatic evaluation metrics BLEU (Papineni et al., 2002) and TER (Snover et al., 2006). We also perform a manual analysis of the correctness of pronoun-verb agreement on the 60 sentences extracted from the test set used for the automatic evaluation. The results of the automatic and human evaluation are presented in Table 4.

[7] http://www.statmt.org/wmt13/

# of training sentences	BLEU		TER		Manual (in %)	
	Baseline	Person-verb	Baseline	Person-verb	Baseline	Person-verb
200 000	19.7	**19.8**	62.9	**62.7**	77.5	**87.6**
400 000	**20.6**	20.4	62.2	**62.0**	78.4	**84.5**
600 000	**21.6**	21.5	61.4	**61.1**	77.9	**88.2**

Table 4. Evaluation metrics comparing the baseline and the *pronoun-verb* approach.

In terms of BLEU scores, while a small improvement is seen for the first data set (+0.1), there is small decrease(-0.2) for the two larger data sets. As far as TER is concerned, a small improvement is seen for all data sets[8]

However, there is an intrinsic problem in using document-level (or even sentence-level) metrics to try to demonstrate improvements in translation quality when one is focused on a single linguistic phenomenon. As in other works on modeling morphology in MT (Avramidis and Koehn,2008; Mareček et al., 2011), when computing (say) the BLEU score, all *n*-grams are weighted equally. However, this does not take into consideration that not every part of a document (or sentence) contributes equally to the overall adequacy and fluency of the translation, which may lead to an incorrect understanding of the system's actual quality. More precisely for our purposes, a subject-verb agreement error that may considerably influence both the grammaticality (fluency) and the semantics (adequacy) of a translation is treated in equal measure to any other error (Callison-Burch et al., 2006).

Accordingly, it is noteworthy that correcting subject-verb agreement errors leads to translations that are considered better by humans (Mareček et al., 2011). Table 4 demonstrates a similar trend, where we see that for each data set, our system considerably outperforms compared to the equivalent baseline. For the largest data set, our model improves by 10.3% absolute (or 13.2% relative) compared to the equivalent *baseline*.

Note too that as is well-known, SMT systems can generate perfectly good translations which do not result in an improved BLEU score, simply because the output translation differs significantly from the reference. One such example appears in (2):

(2)

Reference: "Nous analysons cela car ces idées ..."
Baseline: "Nous examiné pourquoi ces idées ..."
Our system: "Nous examinons pourquoi ces idées ..."

Here, the *baseline* system inserts a past participle in the position where the main verb should be. Our *morphologically-enriched* system correctly inserts a 1^{st}-person plural verb *examinons* ('examine'), which while being semantically correct, differs from the reference *analysons* ('analyse'). Another example is (3):

(3)

Reference: "Je sais qui tu es", a-t-il dit, gentiment.
Baseline: "Je sais qui vous sont", il a dit aimablement.
Our system: "Je sais qui vous êtes,", il a dit aimablement.

In this example, the *baseline* system inserts a 3^{rd}-person plural form *sont* ('are') after the 2^{nd}-person plural pronoun *vous* ('you'). While our system produces the correct form *êtes*, as the reference contains the 2^{nd}-person singular phrase *tu es*, there is

[8] Both for the BLEU-scores as for the TER-scores the differences are insignificant.

a significant difference between this and the output translation, so no additional benefit in terms of BLEU score accrues; indeed, in this example, the incorrect *baseline* translation obtains *exactly the same* BLEU score as our (arguably) correct morphologically enhanced system.

4.3 Manual Error Evaluation

In order to discover in what circumstances our *morphologically-enriched* system improves over the *baseline*, in this section we describe a detailed manual error analysis we conducted which focuses on pronoun-verb agreement cases[9] We evaluate the outputs of the test sets for the (*baseline* and *morphologically-enriched*) SMT systems.

For each test set s we compute the correctness of the pronoun-verb translation ratio by dividing the correctly translated pronoun-verb pairs ($t^{correct}$) by all pronoun-verb pairs that should have agreement (t^{total}): $C_s = \frac{t_s^{correct}}{t_s^{total}}$. The higher the correctness of pronoun-verb translation ratio, the better.

We compared both the *baseline* and the *morphologically-enriched* systems and identify that our approach leads to SMT systems that produce a more correct translation with respect to subject-verb agreement. In the right-most column of Table 4 we already observed that all three *morphologically-enriched* systems have higher correctness-score for pronoun-verb translation. To get a better view on how the *morphologically-enriched* systems (referred to in Table 5 as *ME1*, *ME2* and *ME3*) perform compared to the *baselines* (*BS1*, *BS2* and *BS3*), we count for every pronoun the pronoun-verb pairs that are translated correctly. The results are presented in Table 5.

% correct pronoun-verb agreement	I	YOU	HE, SHE, IT	WE	THEY	TOTAL
BS1 (200k)	78.57	60.71	93.33	61.11	69.23	72.59
ME1 (200k)	92.86	82.76	93.10	88.24	81.82	87.75
BS2 (400k)	85.71	57.14	93.55	70.59	76.92	76.78
ME2 (400k)	85.71	79.31	90.00	83.33	81.82	84.04
BS3 (600k)	78.57	57.14	93.33	70.59	76.92	75.31
ME3 (600k)	92.86	82.76	93.10	83.33	81.82	86.77
BS (average over all)	80.95	58.33	93.41	67.43	74.36	74.90
ME (average over all)	90.48	81.61	92.07	84.97	81.82	86.19

Table 5. % correctly translated pronoun-verb pairs in *baseline* and *pronoun-verb* approach per pronoun.

From Table 5 results that our approach outperforms the *baseline* systems in terms of agreement for all pronouns except for the 3^{rd} person singular, where the *baseline* and *morphologically-enriched* systems score similarly (respectively 93.41% and 92.07%). The other subject-verb agreements result more difficult for all *baseline* systems. The biggest improvements of the *morphologically-enriched* system can be noted for the 2^{nd} person and the 1^{st} person plural, where over all datasets, we can see an absolute

[9] All the verbs were retrieved and were given a mark if they agreed with their subject. There was only one annotator since the correctness of agreement between subject and verb should not be ambiguous.

increase of 23.28% and 17.54% respectively. When averaging over the three datasets for all pronouns we observe an overall improvement of 11.29% absolute (15.07% relative).

In (4) we show an example of the translations generated by the two different systems. The *baseline* system translates the verb, that agrees with the 1^{st}-person plural pronoun *nous* ('we'), incorrectly as an infinitive *aider* ('to help'). However, our system translates it to the correct verb form *aidons*.

(4)

Source: "**We help** relatives as much as patients" says Nathalie Savard, Director of Care.
Baseline: "**Nous aider** proches autant que les patients" affirme Nathalie Savard, directeur de soins.
Our system: "**Nous aidons** proches autant que les patients" affirme Nathalie Savard, directeur de soins.

Another example is (5):

(5)

Source: 'Then **you can** start breaking them,' Jakub told us.
Baseline: **Vous** ' then **puissent** commencer les enfreindre,"Jakub nous a dit.
Our system: **Vous pouvez** ' then commencer enfreindre," Jakub nous a dit.

In example (5), our system correctly translates the pronoun-verb pair *you can* as *vous pouvez*. However, the baseline translates the verb incorrectly as a subjunctive 3^{rd} person plural verb *puissent*.

In the last example (6) the verb *understood* that agrees with the 1^{st} person singular pronoun *I* is translated by the *baseline* system as a past participle while the *morphologically-enriched* system translates it correctly. This example also illustrates how our system deals with unseen verb-forms. While both systems do not translate the verb *drill* since it is unseen in the training set, our system adds the information of the subject to the verb form *drill1pl*. In our error-analysis, we counted these missing verb forms as errors for both systems.

(6)

Source: Here **I** finally **understood** why **we drill** all the...
Baseline: Ici, **je** enfin **compris** pourquoi **nous drill** tous les...
Our system: Finalement, **j'ai compris** pourquoi **nous ici drill1pl** tous les...

5 Conclusions and Future Work

SMT systems typically have problems in ensuring correct subject-verb agreement when producing translations. This is especially problematic when translating from a morphologically impoverished language like English into a morphologically rich language; given that a huge proportion of the world's translation requirement is from English into some other language, this problem affects many SMT system built to date.

Using a simple POS-based model, we annotate source-language verbs with morphological information to turn the problem from a one-to-many mapping between English surface forms and their multiple target-language equivalents into a series of one-to-one associations.

Testing this on English-to-French, we see improvements (averaged over the three different data set) in subject-verb agreement of 11.29% absolute (or 15.07% relative) compared to the equivalent Moses baseline for our morphologically-enriched system, as measured by a human evaluation. We note the problem in relying on automatic metrics when honing in on specific translational phenomena, as well as the well-known problem of improvements in translation quality not being reflected by increased automatic evaluation scores.

Given this promising result, in future work we would like to apply this technique to other language pairs and data types. In addition, we intend to compare the ability of our simple POS-based verb identification model against more sophisticated approaches such as that of Cai et al. (2009).

6 Acknowledgements

We would like to thank the anonymous reviewers for their insights and feedback. The work has been supported by the Dublin City University Faculty of Engineering & Computing under the Daniel O'Hare Research Scholarship scheme and by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund, by the European Commission as part of the FALCON project (contract number 610879).

References

AVRAMIDIS, E., AND KOEHN, P. (2008). Enriching Morphologically Poor Languages for Statistical Machine Translation. In *Proceedings of ACL-08: HLT*, 15-20 June 2008, Columbus, Ohio, USA, ACL, pp. 763–770.

BIRCH, A., OSBORNE, M., AND KOEHN, P. (2007). CCG Supertags in Factored Statistical Machine Translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, 23 June 2007, Prague, Czech Republic, ACL, pp. 9–16.

BOZORGIAN, M., AND AZADMANESH, N. (2015). A Survey on the Subject-Verb Agreement in Google Machine Translation. *International Journal of Research Studies in Educational Technology, 4*(1), April 2015, pp. 51–62.

CAI, D., HU, Y., MIAO, X., AND SONG, Y. (2009). Dependency Grammar Based English Subject-Verb Agreement Evaluation. In *Proceedings of th 23rd Pacific Asia Conference on Language, Information and Computation*, 3-5 December 2009, City University of Hong Kong, pp. 6371.

CALLISON-BURCH, C., OSBORNE, M., AND KOEHN, P. (2006). Re-evaluating the Role of BLEU in Machine Translation Research. In *EACL-2006: 11th Conference of the European Chapter of the Association for Computational Linguistics*, 3-7 April 2006, Trento, Italy), ACL, pp. 249–256.

CARPUAT, M. (2009). Toward Using Morphology in French-English Phrase-Based SMT. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, 30-31 March 2009, Athens, Greece, ACL, pp. 150–154.

CORSTON-OLIVER, S., AND GAMON, M. (2004). Normalizing German and English Inflectional Morphology to Improve Statistical Word Alignment. In *Machine Translation: From Real Users to Research*, October 2004, Springer, pp. 48–57.

DAEMS, J., MACKEN, L., AND VANDEPITTE, S. (2014). On the Origin of Errors: a Fine-Grained Analysis of MT and PE Errors and their Relationship. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 26-31 May 2014, Reykjavik, Iceland, ELRA, pp. 62–66.

EL-KAHLOUT, I. D., AND OFLAZER, K. (2006). Initial Explorations in English to Turkish Statistical Machine Translation. In *Proceedings of the Workshop on Statistical Machine Translation*, 8-9 June 2006, New York City, USA, ACL, pp. 7–14.

EL KHOLY, A., AND HABASH, N. (2012). Translate, Predict or Generate: Modeling Rich Morphology in Statistical Machine Translation. In *Proceedings of the 16th EAMT conference, 12*, 28-30 May 2012, Trento, Italy, EAMT, pp. 27–34.

ESPAÑA BONET, C., MÀRQUEZ VILLODRE, L., LABAKA, G., DÍAZ DE ILARRAZA SÁNCHEZ, A. AND SARASOLA GABIOLA, K. (2012). Hybrid machine translation guided by a rule-based system. 2011.

FRASER, A., WELLER, M., CAHILL, A., AND CAP, F. (2012). Modeling Inflection and Word-Formation in SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 23-27 April 2012, Avignon, France, ACL, pp. 664–674.

HAQUE, R., KUMAR NASKAR, S., VAN DEN BOSCH, A., AND WAY, A. (2010). Supertags as Source Language Context in Hierarchical Phrase-Based SMT. In *Proceedings of AMTA 2010: The Ninth Conference of the Association for Machine Translation in the Americas*, 27-30 October 2010, Denver, Colorado, USA, AMTA, pp. 210–219.

KOEHN, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *MT Summit, 5*, 12 September 2005, Phuket, Thailand, pp. 79–86.

KOEHN, P. (2010). Statistical Machine Translation. January 2010,Cambridge University Press, Cambridge.

KOEHN, P., HOANG, H., BIRCH, A., CALLISON-BURCH, C., FEDERICO, M., BERTOLDI, N., COWAN, B., SHEN, W., MORAN, C., ZENS, R., DYER, C., BOJAR, O., CONSTANTIN, A., AND HERBST, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *The 45th Annual Meeting of the Association for Computational Linguistics: the Demo and Poster Sessions*, 23-30 June 2007, Prague, Czech Republic, ACL, pp. 177–180.

KOEHNM P, AND HOANG, H. (2007). Factored Translation Models In *EMNLP-CoNLL*, 28-30 June, Prague, Czech Republic, pp. 868–876.

MARCUS, M., KIM, G., MARCINKIEWICZ, M., MACINTYRE, R., BIES, A., FERGUSON, M., KATZ, K., AND AND SCHASBERGER, B. (1994). The Penn Treebank: Annotating Predicate Argument Structure. *Proceedings of the Workshop on Human Language Technology (HLT '94)*, 8-11 March 1994, Plainsboro, New Jersey, USA, ACL, pp. 114–119.

MAREČEK, D., ROSA, R., GALUŠČÁKOVÁ, P., AND BOJAR, O. (2011). Two-step Translation with Grammatical Post-Processing. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, 30-31 July 2011, Edinburgh, Scotland, UK, ACL, pp. 426–432.

NIESSEN, S., AND NEY, H. (2004). Statistical Machine Translation with Scarce Resources Using Morpho-Syntactic Information. *Computational Linguistics, 30* (2), June 2004, 181–204.

PAPINENI, K., ROUKOS, S., WARD, T., AND ZHU, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 6-12 July 2002, Philadelphia, PA, USA, ACL, pp. 311–318.

POPOVIC, M., AND NEY, H. (2011). Towards Automatic Error Analysis of Machine Translation Output. In *Computational Linguistics 37: 4*, December 2011, MIT Press, pp. 657–688.

SNOVER, M., DORR, B., SCHWARTZ, R., MICCIULLA, L., AND MAKHOUL, J. (2016). A Study of Translation Edit Rate With Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Visions for the*

Future of Machine Translation, 8-12 August 2006, Cambridge, Massachusetts, USA, AMTA, pp. 223–231.

TOUTANOVA, K., SUZUKI, H., AND RUOPP, A. (2008). Applying Morphology Generation Models to Machine Translation. In *Proceedings of ACL-08: HLT*, 15-20 June 2008, Columbus, Ohio, ACL, pp. 514–522.

UEFFING, N., AND NEY, H. (2003). Using POS Information for Statistical Machine Translation into Morphologically Rich Languages. In *EACL 2003: Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, 12-17 April 2003, Budapest, Hungary, ACL, pp. 347–354.

VILAR, D., XU, J., DHARO, L.F., AND NEY, H. (2006). Error Analysis of Statistical Machine Translation Output. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, 24-26 May 2006, Genoa, Italy, pp. 697–702.

VIRPIOJA, S., VYRYNEN, J. J., CREUTZ, M., AND SADENIEMI, M. (2007). Morphology-Aware Statistical Machine Translation Based on Morphs Induced in an Unsupervised Manner. In *Proceedings of MT Summit XI*, 11 September 2007, Copenhagen, Denmark, pp. 491–498.

WANG, R., OSENOVA, P., AND SIMOV, K. (2012). Linguistically-Enriched Models for Bulgarian-to-English Machine Translation. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 12 July 2012, Jeju, Republic of Korea), ACL, pp. 10–19.

WANG, Y., WANG, L., ZENG, X., WONG, D.F., CHAO, L.S., AND LU, Y. (2014). Factored Statistical Machine Translation for Grammatical Error Correction. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, 26-26 June 2014, Baltimore, Maryland, pp. 83–90.

WISNIEWSKI, G., KUBLER, N., AND YVON, F. (2014). A Corpus of Machine Translation Errors Extracted from Translation Students Exercises. In *Proceedings of the International Conference on Language Resources and Evaluation*, 26-31 May 2014, Reykjavik, Iceland, ELRA, pp. 3585–3588.

YUAN, Z., AND FELICE, M. (2013). Constrained grammatical error correction using Statistical Machine Translation. *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, 8-9 Aug. 2013, Sofia, Bulgaria, pp. 52–61.

ZEMAN, D., FISHEL, M., BERKA, J., AND BOJAR, O. (2011). Addicter: What is wrong with my translations? *The Prague Bulletin of Mathematical Linguistics, 96*, September 2011, pp. 79–88.

ZHOU, M., WANG, B., LIU, S., LI, M., ZHANG, D., AND ZHAO, T. (2008). Diagnostic evaluation of machine translation systems using automatically constructed linguistic check-points. *Proceedings of the 22nd International Conference on Computational Linguistics*, 18-22 Aug. 2008, Manchester, UK, ACL, pp. 1121–1128.