Chunk-based Verb Reordering in VSO Sentences for Arabic-English Statistical Machine Translation

Arianna Bisazza and Marcello Federico
Fondazione Bruno Kessler
Human Language Technologies
Trento, Italy
`{bisazza,federico}@fbk.eu`

Abstract

In Arabic-to-English phrase-based statistical machine translation, a large number of syntactic disfluencies are due to wrong long-range reordering of the verb in VSO sentences, where the verb is anticipated with respect to the English word order. In this paper, we propose a chunk-based reordering technique to automatically detect and displace clause-initial verbs in the Arabic side of a word-aligned parallel corpus. This method is applied to preprocess the training data, and to collect statistics about verb movements. From this analysis, specific verb reordering lattices are then built on the test sentences before decoding them. The application of our reordering methods on the training and test sets results in consistent BLEU score improvements on the NIST-MT 2009 Arabic-English benchmark.

1 Introduction

Shortcomings of phrase-based statistical machine translation (PSMT) with respect to word reordering have been recently shown on the Arabic-English pair by Birch et al. (2009). An empirical investigation of the output of a strong baseline we developed with the Moses toolkit (Koehn et al., 2007) for the NIST 2009 evaluation, revealed that an evident cause of syntactic disfluency is the anticipation of the verb in Arabic Verb-Subject-Object (VSO) sentences – a class that is highly represented in the news genre[1].

Fig. 1 shows two examples where the Arabic main verb phrase comes before the subject. In such sentences, the subject can be followed by adjectives, adverbs, coordinations, or appositions that further increase the distance between the verb

---

[1]In fact, Arabic syntax admits both SVO and VSO orders.

and its object. When translating into English – a primarily SVO language – the resulting long verb reorderings are often missed by the PSMT decoder either because of pure modeling errors or because of search errors (Germann et al., 2001): i.e. their span is longer than the maximum allowed distortion distance, or the correct reordering hypothesis does not emerge from the explored search space because of a low score. In the two examples, the missed verb reorderings result in different translation errors by the decoder, respectively, the introduction of a subject pronoun before the verb and, even worse, a verbless sentence.

In Arabic-English machine translation, other kinds of reordering are of course very frequent: for instance, adjectival modifiers following their noun and head-initial genitive constructions (*Idafa*). These, however, appear to be mostly local, therefore more likely to be modeled through phrase internal alignments, or to be captured by the reordering capabilities of the decoder. In general there is a quite uneven distribution of word-reordering phenomena in Arabic-English, and long-range movements concentrate on few patterns.

Reordering in PSMT is typically performed by *(i)* constraining the maximum allowed word movement and exponentially penalizing long reorderings (distortion limit and penalty), and *(ii)* through so-called lexicalized orientation models (Och et al., 2004; Koehn et al., 2007; Galley and Manning, 2008). While the former is mainly aimed at reducing the computational complexity of the decoding algorithm, the latter assigns at each decoding step a score to the next source phrase to cover, according to its orientation with respect to the last translated phrase. In fact, neither method discriminates among different reordering distances for a specific word or syntactic class. To our view, this could be a reason for their inadequacy to properly deal with the reordering peculiarities of the Arabic-English language pair. In

src:	w **AstdEt** <u>kl mn AlsEwdyp w lybyA w swryA</u>$_{Subj}$ <u>sfrA' hA</u>$_{Obj}$ fy AldnmArk .
ref:	*Each of Saudi Arabia , Libya and Syria*$_{Subj}$ **recalled** <u>their ambassadors</u>$_{Obj}$ *from Denmark .*
MT:	He recalled all from Saudi Arabia , Libya and Syria ambassadors in Denmark .

src:	**jdd** <u>AlEAhl Almgrby Almlk mHmd AlsAds</u>$_{Subj}$ <u>dEm h</u>$_{Obj}$ l m$rwE Alr}ys Alfrnsy
ref:	*The Moroccan monarch King Mohamed VI*$_{Subj}$ **renewed** <u>his support</u>$_{Obj}$ *to the project of French President*
MT:	The Moroccan monarch King Mohamed VI his support to the French President

Figure 1: Examples of problematic SMT outputs due to verb anticipation in the Arabic source.

this work, we introduce a reordering technique that addresses this limitation.

The remainder of the paper is organized as follows. In Sect. 2 we describe our verb reordering technique and in Sect. 3 we present statistics about verb movement collected through this technique. We then discuss the results of preliminary MT experiments involving verb reordering of the training based on these findings (Sect. 4). Afterwards, we explain our lattice approach to verb reordering in the test and provide evaluation on a well-known MT benchmark (Sect. 5). In the last two sections we review some related work and draw the final conclusions.

2 Chunk-based Verb Reordering

The goal of our work is to displace Arabic verbs from their clause-initial position to a position that minimizes the amount of word reordering needed to produce a correct translation. In order to restrict the set of possible movements of a verb and to abstract from the usual token-based movement length measure, we decided to use shallow syntax chunking of the source language. Full syntactic parsing is another option which we have not tried so far mainly because popular parsers that are available for Arabic do not mark grammatical relations such as the ones we are interested in.

We assume that Arabic verb reordering only occurs between shallow syntax chunks, and not within them. For this purpose we annotated our Arabic data with the AMIRA chunker by Diab et al. (2004)[2]. The resulting chunks are generally short (1.6 words on average). We then consider a specific type of reordering by defining a production rule of the kind: *"move a chunk of type* T *along with its* L *left neighbours and* R *right neighbours by a shift of* S *chunks"*. A basic set of rules

that displaces the verbal chunk to the right by at most 10 positions corresponds to the setting:

$$T=\text{'VP'}, L=0, R=0, S=1..10$$

In order to address cases where the verb is moved along with its adverbial, we also add a set of rules that include a one-chunk right context in the movement:

$$T=\text{'VP'}, L=0, R=1, S=1..10$$

To prevent verb reordering from overlapping with the scope of the following clause, we always limit the maximum movement to the position of the next verb. Thus, for each verb occurrence, the number of allowed movements for our setting is at most $2 \times 10 = 20$.

Assuming that a word-aligned translation of the sentence is available, the best movement, if any, will be the one that reduces the amount of distortion in the alignment, that is: *(i)* it reduces the number of swaps by 1 or more, and *(ii)* it minimizes the sum of distances between source positions aligned to consecutive target positions, i.e. $\sum_i |a_i - (a_{i-1} + 1)|$ where a_i is the index of the foreign word aligned to the ith English word. In case several movements are optimal according to these two criteria, e.g. because of missing word-alignment links, only the shortest good movement is retained.

The proposed reordering method has been applied to various parallel data sets in order to perform a quantitative analysis of verb anticipation, and to train a PSMT system on more monotonic alignments.

3 Analysis of Verb Reordering

We applied the above technique to two parallel corpora[3] provided by the organizers of the NIST-MT09 Evaluation. The first corpus (Gale-NW) contains human-made alignments. As these refer to non-segmented text, they were adjusted to

---

[2]This tool implies morphological segmentation of the Arabic text. All word statistics in this paper refer to AMIRA-segmented text.

[3]Newswire sections of LDC2006E93 and LDC2009E08, respectively 4337 and 777 sentence pairs.
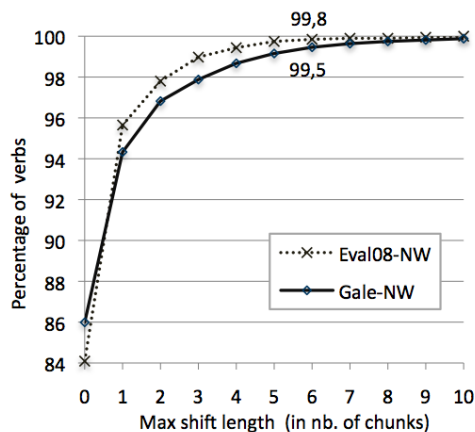
Figure 2: Percentage of verb reorderings by maximum shift (0 stands for no movement).

Figure 3: Distortion reduction in the GALE-NW corpus: jump occurrences grouped by length range (in nb. of words).

agree with AMIRA-style segmentation. For the second corpus (Eval08-NW), we filtered out sentences longer than 80 tokens in order to make word alignment feasible with GIZA++ (Och and Ney, 2003). We then used the *Intersection* of the direct and inverse alignments, as computed by Moses. The choice of such a high-precision, low-recall alignment set is supported by the findings of Habash (2007) on syntactic rule extraction from parallel corpora.

3.1 The *Verb's Dance*

There are 1,955 verb phrases in Gale-NW and 11,833 in Eval08-NW. Respectively 86% and 84% of these do not need to be moved according to the alignments. The remaining 14% and 16% are distributed by movement length as shown in Fig. 2: most verb reorderings consist in a 1-chunk long jump to the right (8.3% in Gale-NW and 11.6% in Eval08-NW). The rest of the distribution is similar in the two corpora, which indicates a good correspondence between verb reordering observed in automatic and manual alignments. By increasing the maximum movement length from 1 to 2, we can cover an additional 3% of verb reorderings, and around 1% when passing from 2 to 3. We recall that the length measured in chunks doesn't necessarily correspond to the number of jumped tokens. These figures are useful to determine an optimal set of reordering rules. From now on we will focus on verb movements of at most 6 chunks, as these account for about 99.5% of the verb occurrences.
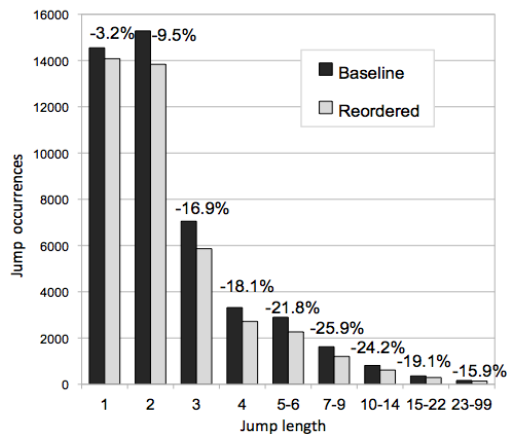
3.2 Impact on Corpus Global Distortion

We tried to measure the impact of chunk-based verb reordering on the total word distortion found in parallel data. For the sake of reliability, this investigation was carried out on the manually aligned corpus (Gale-NW) only. Fig. 3 shows the positive effect of verb reordering on the total distortion, which is measured as the number of *words* that have to be *jumped* on the source side in order to cover the sentence in the target order (that is $|a_i - (a_{i-1} + 1)|$). Jumps have been grouped by length and the relative decrease of jumps per length is shown on top of each double column.

These figures do not prove as we hoped that verb reordering resolves *most* of the long range reorderings. Thus we manually inspected a sample of verb-reordered sentences that still contain long jumps, and found out that many of these were due to what we could call "unnecessary" reordering. In fact, human translations that are free to some extent, often display a global sentence restructuring that makes distortion dramatically increase. We believe this phenomenon introduces noise in our analysis since these are not reorderings that an MT system needs to capture to produce an accurate and fluent translation.

Nevertheless, we can see from the relative decrease percentages shown in the plot, that although short jumps are by far the most frequent, verb reordering affects especially medium and long range distortion. More precisely, our selective reordering technique solves 21.8% of the 5-to-6-words jumps, 25.9% of the 7-to-9-words jumps and 24.2% of the 10-to-14-words jumps, against

only 9.5% of the 2-words jumps, for example. Since our primary goal is to improve the handling of long reorderings, this makes us think that we are advancing in a promising direction.

4 Preliminary Experiments

In this section we investigate how verb reordering on the source language can affect translation quality. We apply verb reordering both on the training and the test data. However, while the parallel corpus used for training can be reordered by exploiting word alignments, for the test corpus we need a verb reordering "prediction model". For these preliminary experiments, we assumed that optimal verb-reordering of the test data is provided by an *oracle* that has access to the word alignments with the reference translations.

4.1 Setup

We trained a Moses-based system on a subset of the NIST-MT09 Evaluation data[4] for a total of 981K sentences, 30M words. We first aligned the data with GIZA++ and use the resulting *Intersection* set to apply the technique explained in Sect. 2. We then retrained the whole system – from word alignment to phrase scoring – on the reordered data and evaluated it on two different versions of Eval08-NW: plain and oracle verb-reordered, obtained by exploiting word alignments with the first of the four available English references. The first experiment is meant to measure the impact of the verb reordering procedure on training only. The latter will provide an estimate of the maximum improvement we can expect from the application to the test of an optimal verb reordering prediction technique. Given our experimental setting, one could argue that our BLEU score is biased because one of the references was also used to generate the verb reordering. However, in a series of experiments not reported here, we evaluated the same systems using only the remaining three references and observed similar trends as when all four references are used.

Feature weights were optimized through MERT (Och, 2003) on the newswire section of the NIST-MT06 evaluation set (Dev06-NW), in the original version for the baseline system, in the verb-reordered version for the reordered system.
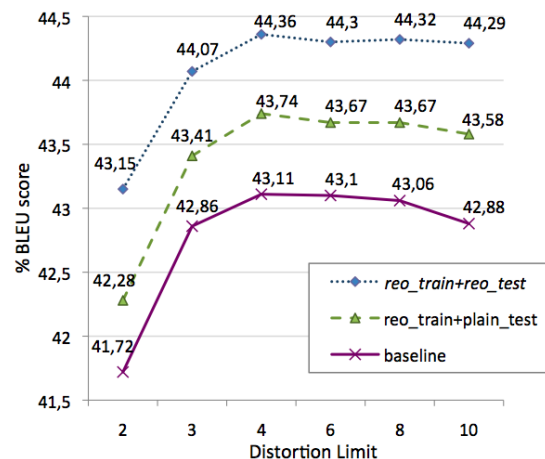

Figure 4: BLEU scores of baseline and reordered system on plain and oracle reordered Eval08-NW.

Fig. 4 shows the results in terms of BLEU score for *(i)* the baseline system, *(ii)* the reordered system on a plain version of Eval08-NW and *(iii)* the reordered system on the reordered test. The scores are plotted against the distortion limit (DL) used in decoding. Because high DL values (8-10) imply a larger search space and because we want to give Moses the best possible conditions to properly handle long reordering, we relaxed for these conditions the default pruning parameter to the point that led the highest BLEU score[5].

4.2 Discussion

The first observation is that the reordered system always performs better (0.5∼0.6 points) than the baseline on the plain test, despite the mismatch between training and test ordering. This may be due to the fact that automatic word alignments are more accurate when less reordering is present in the data, although previous work (Lopez and Resnik, 2006) showed that even large gains in alignment accuracy seldom lead to better translation performances. Moreover phrase extraction may benefit from a distortion reduction, since its heuristics rely on word order in order to expand the context of alignment links.

The results on the oracle reordered test are also interesting: a gain of at least 1.2 point absolute over the baseline is reported in all tested DL conditions. These improvements are remarkable, keeping in mind that only 31% of the train and 33% of the test sentences get modified by verb reordering.

---

[4]LDC2007T08, 2003T07, 2004E72, 2004T17, 2004T18, 2005E46, 2006E25, 2006E44 and LDC2006E39 – the two last with first reference only.

[5]That is, the histogram pruning maximum stack size was set to 1000 instead of the default 200.

AR: w [>kdt]_{VP} [mSAdr rsmyp]_{NP} [wjwd rAbT]_{NP} [byn AlAEtdA'At]_{PP} .
and confirmed sources official being link between the-attacks
CH.1 CH.2 CH.3 CH.4 CH.5

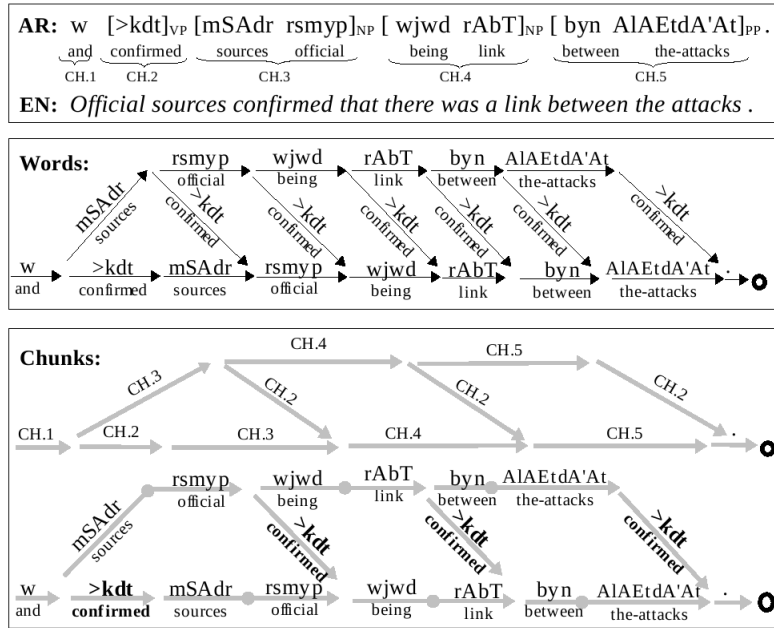**EN:** *Official sources confirmed that there was a link between the attacks .*

Figure 5: Reordering lattices for Arabic VSO sentences: word-based and chunk-based.

Concerning distortion, although long verb movements are often observed in parallel corpora, relaxing the DL to high values does not benefit the translation, even with our 'generous' setting (wider beam search). This is probably due to the fact that, with weakly constrained distortion, the risk of search errors increases as the reordering model fails to properly rank an exponentially growing set of permutations. Therefore many correct reordering hypotheses receive low scores and get lost in pruning or recombination.

5 Verb Reordering Lattices

Having assessed the negative impact of VSO sentences on Arabic-English translation performance, we now propose a method to improve the handling of this phenomenon at decoding time. As in real working conditions word alignments of the input text are not available, we explore a reordering lattice approach.

5.1 Lattice Construction

Firstly conceived to optimally encode multiple transcription hypothesis produced by a speech recognizer, word lattices have later been used to represent various forms of input ambiguity, mainly at the level of token boundaries (e.g. word segmentation, morphological decomposition, word decompounding (Dyer et al., 2008)).

A main problem when dealing with permuta-

tions is that the lattice size can grow very quickly when medium to long reorderings are represented. We are particularly concerned with this issue because our decoding will perform additional reordering on the lattice input. Thanks to the restrictions we set on our verb movement reordering rules described in Sect. 2 – i.e. only reordering between chunks and no overlap between consecutive verb chunks movement – we are able to produce quite compact word lattices.

Fig. 5 illustrates how a chunk-based reordering lattice is generated. Suppose we want to translate the Arabic sentence *"w >kdt mSAdr rsmyp wjwd rAbT byn AlAEtdA'At"*, whose English meaning is *"Official sources confirmed that there was a link between the attacks"*. The Arabic main verb *>kdt (confirmed)* is in pre-subject position. If we applied word-based rather than chunk-based rules to move the verb to the right, we would produce the first lattice of the figure, containing 7 paths (the original plus 6 verb movements). With the chunk-based rules, we treat instead chunks as units and get the second lattice. Then, by expanding each chunk, we obtain the final, less dense lattice, that compared to the first does not contain 3 (unlikely) reordering edges.

To be consistent with the reordering applied to the training data, we use a set of rules that move each verb phrase alone or with its following chunk by 1 to 6 chunks to the right. With this settings,
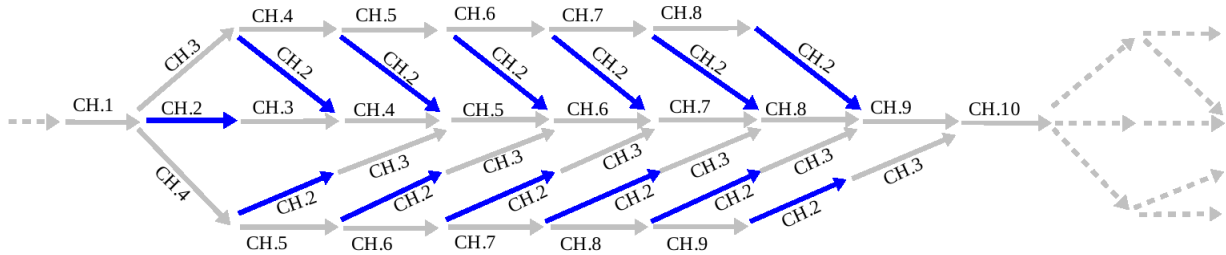
Figure 6: Structure of a chunk-based reordering lattice for verb reordering, before word expansion. Edges in boldface represent the verbal chunk.

our lattice generation algorithm computes a compact lattice (Fig. 6) that introduces at most $5 \times \Delta S$ chunk edges for each verb chunk, where ΔS is the permitted movement range (6 in this case).

Before translation, each edge has to be associated with a weight that the decoder will use as additional feature. To differentiate between the original word order and verb reordering we assign a fixed weight of 1 to the edges of the plain path, and 0.25 to the other edges. As future work we will devise more discriminative weighting schemes.

5.2 Evaluation

For the experiments, we relied on the existing Moses-implementation of non-monotonic decoding for word lattices (Dyer et al., 2008) with some fixes concerning the computation of reordering distance. The translation system is the same as the one presented in Sect. 4, to which we added an additional feature function evaluating the lattice weights (*weight-i*). Instead of rerunning MERT, we directly estimated the additional feature-function weight over a suitable interval (0.002 to 0.5), by running the decoder several times on the development set. The resulting optimal weight was 0.05.

Table 1 presents results on three test sets: **Eval08-NW** which was used to calibrate the reordering rules, **Reo08-NW** a specific test set consisting of the 33% of Eval08-NW sentences that actually require verb reordering, and **Eval09-NW** a yet unseen dataset (newswire section of the NIST-MT09 evaluation set, 586 sentences). Best results with lattice decoding were obtained with a distortion limit (DL) of 4, while best performance of text decoding was obtained with a DL of 6.

As we hoped, translating a verb reordering lattice yields an additional improvement to the reordering of the training corpus: from 43.67% to 44.04% on Eval08-NW and from 48.53% to

48.96% on Eval09-NW. The gap between the baseline and the score obtainable by oracle verb reordering, as estimated in the preliminary experiments on Eval08-NW (44.36%), has been largely filled.

On the specific test set – Reo08-NW – we observe a performance drop when reordered models are applied to non-reordered (plain) input: from 46.90% to 46.64%. Hence it seems that the mismatch between training and test data is significantly impacting on the reordering capabilities of the system with respect to verbs. We speculate that such negative effect is diluted in the full test set (Eval08-NW) and compensated by the positive influence of verb reordering on phrase extraction. Indeed, when the lattice technique is applied we get an improvement of about 0.6 point over the baseline, which is still a fair result, but not as good as the one obtained on the general test sets.

Finally, our approach led to an overall gain of 0.8 point BLEU over the baseline, on Eval09-NW. We believe this is a satisfactory result, given the fairly good starting performance, and given that the BLEU metric is known not to be very sensitive to word order variations (Callison-Burch et al., 2006). For the future, we plan to also use specific evaluation metrics that will allow us to isolate the impact of our approach on reordering, like the ones by Birch et al. (2010).

System	DL	eval08nw	reo08nw	eval09nw
baseline	6	43.10	46.90	48.13
reord. training +				
plain input	6	43.67	46.64	48.53
lattice	4	44.04	47.51	48.96
oracle reord.	*4*	*44.36*	*48.25*	*na*

Table 1: BLEU scores of baseline and reordered system on plain test and on reordering lattices.

6 Related Work

Linguistically motivated word reordering for Arabic-English has been proposed in several recent works. Habash (2007) extracts syntactic reordering rules from a word-aligned parallel corpus whose Arabic side has been fully parsed. The rules involve reordering of syntactic constituents and are applied in a deterministic way (always the most probable) as preprocessing of training and test data. The technique achieves consistent improvements only in very restrictive conditions: maximum phrase size of 1 and monotonic decoding, thus failing to enhance the existing reordering capabilities of PSMT. In (Crego and Habash, 2008; Elming and Habash, 2009) possible input permutations are represented through a word graph, which is then processed by a monotonic phrase- or n-gram-based decoder. Thus, these approaches are conceived as alternatives, rather than integrations, to PSMT reordering. On the contrary, we focused on a single type of significant long reorderings, in order to integrate class-specific reordering methods into a standard PSMT system.

To our knowledge, the work by Niehues and Kolss (2009) on German-English is the only example of a lattice-based reordering approach being coupled with reordering at decoding time. In their paper, discontinuous non-deterministic POS-based rules learned from a word-aligned corpus are applied to German sentences in the form of weighted edges in a word lattice. Their phrase-based decoder admits local reordering within a fixed window of 2 words, while, in our work, we performed experiments up to a distortion limit of 10. Another major difference is that we used shallow syntax annotation to effectively reduce the number of possible permutations. A first attempt to adapt our technique to the German language is described in Hardmeier et al. (2010).

Our work is also tightly related to the problem of noun-phrase subject detection, recently addressed by Green et al. (2009). In fact, detecting the extension of the subject can be a sufficient condition to guess the optimal reordering of the verb. In their paper, a discriminative classifier was trained on a rich variety of linguistic features to detect the full scope of Arabic NP subjects in verb-initial clauses. The authors reported an F-score of 61.3%, showing that the problem is hard to solve even when more linguistic information is available. In order to integrate the output of the classifier into a PSMT decoder, a specific translation feature was designed to reward hypotheses in which the subject is translated as a contiguous block. Unfortunately, no improvement in translation quality was obtained.

7 Conclusions

Word reordering remains one of the hardest problems in statistical machine translation. Based on the intuition that few reordering patterns would suffice to handle the most significant cases of long reorderings in Arabic-English, we decided to focus on the problem of VSO sentences.

Thanks to simple linguistic assumptions on verb movement, we developed an efficient, low-cost technique to reorder the training data, on the one hand, and to better handle verb reordering at decoding time, on the other. In particular, translation is performed on a compact word lattice that represents likely verb movements. The resulting system outperforms a strong baseline in terms of BLEU, and produces globally more readable translations. However, the problem is not totally solved because many verb reorderings are still missed, despite the suggestions provided by the lattice. Different factors can explain these errors: poor interaction between lattice and distortion/orientation models used by the decoder; poor discriminative power of the target language model with respect to different reorderings of the source.

As a first step to improvement, we will devise a discriminative weighting scheme based on the length of the reorderings represented in the lattice. For the longer term we are working towards bringing linguistically informed reordering constraints inside decoding, as an alternative to the lattice solution. In addition, we plan to couple our reordering technique with more informative language models, including for instance syntactic analysis of the hypothesis under construction. Finally we would like to compare the proposed chunk-based technique with one that exploits full syntactic parsing of the Arabic sentence to further decrease the reordering possibilities of the verb.

src:	w A$Ar AlsnAtwr AlY dEm h m$rwEA ErD ElY mjls Al$ywx
ref:	*The Senator referred to his support for a project proposed to the Senate*
base MT:	The Senator to support projects presented to the Senate
new MT:	Senator noted his support projects presented to the Senate

src:	mn jAnb h hdd >bw mSEb EbdAlwdwd Amyr AlqAEdp b blAd Almgrb AlAslAmy fy nfs Al$ryT b AlqyAm b slslp AEtdA'At w >EmAl <rhAbyp Dd AlmSAlH w Alm&ssAt AljzA}ryp fy AlEdyd mn AlmnATq AljzA}ryp
ref:	*For his part , Abu Musab Abd al-Wadud , the commander of al-Qaeda in the Islamic Maghreb Countries , threatened in the same tape to carry out a series of attacks and terrorist actions against Algerian interests and organizations in many parts of Algeria*
base MT:	For his part threatened Abu Musab EbdAlwdwd Amir al-Qaeda Islamic Morocco country in the same tape to carry out a series of attacks and terrorist acts against the interests and the Algerian institutions in many areas of Algiers
new MT:	For his part , Abu Musab EbdAlwdwd Amir al Qaida threatened to Morocco Islamic country in the same tape to carry out a series of attacks and terrorist acts against the interests of the Algerian and institutions in many areas of Algiers

src:	w ymtd Alm$rwE 500 km mtr w yrbT Almdyntyn Almqdstyn b mdynp jdp ElY sAHl AlbHr Al>Hmr .
ref:	*The project is 500 kilometers long and connects the two holy cities with the city of Jeddah on the Red Sea coast.*
base MT:	It extends the project 500 km and linking the two holy cities in the city of Jeddah on the Red Sea coast .
new MT:	The project extends 500 km , linking the two holy cities in the city of Jeddah on the Red Sea coast .

Figure 7: Examples showing MT improvements coming from chunk-based verb-reordering.

References

Alexandra Birch, Phil Blunsom, and Miles Osborne. 2009. A quantitative analysis of reordering phenomena. In *StatMT '09: Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 197–205, Morristown, NJ, USA. Association for Computational Linguistics.

Alexandra Birch, Miles Osborne, and Phil Blunsom. 2010. Metrics for MT evaluation: evaluating reordering. *Machine Translation*, Published online.

Chris Callison-Burch, Miles Osborne, and Philipp Koehn. 2006. Re-evaluation the role of BLEU in machine translation research. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, April.

Josep M. Crego and Nizar Habash. 2008. Using shallow syntax information to improve word alignment and reordering for smt. In *StatMT '08: Proceedings of the Third Workshop on Statistical Machine Translation*, pages 53–61, Morristown, NJ, USA. Association for Computational Linguistics.

Mona Diab, Kadri Hacioglu, and Daniel Jurafsky. 2004. Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks. In Daniel Marcu Susan Dumais and Salim Roukos, editors, *HLT-NAACL 2004: Short Papers*, pages 149–152, Boston, Massachusetts, USA, May 2 - May 7. Association for Computational Linguistics.

Christopher Dyer, Smaranda Muresan, and Philip Resnik. 2008. Generalizing word lattice translation. In *Proceedings of ACL-08: HLT*, pages 1012–1020, Columbus, Ohio, June. Association for Computational Linguistics.

Jakob Elming and Nizar Habash. 2009. Syntactic reordering for English-Arabic phrase-based machine translation. In *Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages*, pages 69–77, Athens, Greece, March. Association for Computational Linguistics.

Michel Galley and Christopher D. Manning. 2008. A simple and effective hierarchical phrase reordering model. In *EMNLP '08: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 848–856, Morristown, NJ, USA. Association for Computational Linguistics.

Ulrich Germann, Michael Jahr, Kevin Knight, Daniel Marcu, and Kenji Yamada. 2001. Fast decoding and optimal decoding for machine translation. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 228–335, Toulouse, France.

Spence Green, Conal Sathi, and Christopher D. Manning. 2009. NP subject detection in verb-initial Arabic clauses. In *Proceedings of the Third Workshop on Computational Approaches to Arabic Script-based Languages (CAASL3)*, Ottawa, Canada.

Nizar Habash. 2007. Syntactic preprocessing for statistical machine translation. In Bente Maegaard, editor, *Proceedings of the Machine Translation Summit XI*, pages 215–222, Copenhagen, Denmark.

Christian Hardmeier, Arianna Bisazza, and Marcello Federico. 2010. FBK at WMT 2010: Word lattices for morphological reduction and chunk-based

reordering. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR*, Uppsala, Sweden, July. Association for Computational Linguistics.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic.

Adam Lopez and Philip Resnik. 2006. Word-based alignment, phrase-based translation: What's the link? In *5th Conference of the Association for Machine Translation in the Americas (AMTA)*, Boston, Massachusetts, August.

Jan Niehues and Muntsin Kolss. 2009. A POS-based model for long-range reorderings in SMT. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 206–214, Athens, Greece, March. Association for Computational Linguistics.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

F. J. Och, D. Gildea, S. Khudanpur, A. Sarkar, K. Yamada, A. Fraser, S. Kumar, L. Shen, D. Smith, K. Eng, V. Jain, Z. Jin, and D. Radev. 2004. A smorgasbord of features for statistical machine translation. In *Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*, Boston, MA.

Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In Erhard Hinrichs and Dan Roth, editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 160–167.