

POS-based Word Reorderings for Statistical Machine Translation

Maja Popović, Hermann Ney

Lehrstuhl für Informatik VI - Computer Science Department
RWTH Aachen University
Ahornstrasse 55, 52056 Aachen, Germany
{popovic, ney}@informatik.rwth-aachen.de

Abstract

In this work we investigate new possibilities for improving the quality of statistical machine translation (SMT) by applying word reorderings of the source language sentences based on Part-of-Speech tags. Results are presented on the European Parliament corpus containing about 700k sentences and 15M running words. In order to investigate sparse training data scenarios, we also report results obtained on about 1% of the original corpus. The source languages are Spanish and English and target languages are Spanish, English and German. We propose two types of reorderings depending on the language pair and the translation direction: local reorderings of nouns and adjectives for translation from and into Spanish and long-range reorderings of verbs for translation into German. For our best translation system, we achieve up to 2% relative reduction of WER and up to 7% relative increase of BLEU score. Improvements can be seen both on the reordered sentences as well as on the rest of the test corpus. Local reorderings are especially important for the translation systems trained on the small corpus whereas long-range reorderings are more effective for the larger corpus.

1. Introduction

The goal of statistical machine translation is to translate an input word sequence in the source language into a target language word sequence. Given the source language sequence, we have to choose the target language sequence which maximises the posterior probability. Our translation system models this posterior probability directly as a log-linear combination of seven different models. The most important ones are phrase-based models in both directions. Additionally, we use IBM1 models at phrase level also in both directions, language model of the target language, as well as phrase and word penalties. For detailed description of the system see for example (Vilar et al., 2005; Zens et al., 2005).

In order to improve the translation process, it is possible to perform preprocessing steps before training and translation in both source and target language sequence. If necessary, after the translation the inverse transformations are applied to the generated target sequence. In this work, we apply transformations only on the source language sequence. We investigate possibilities for improving the translation quality by reordering of the source sentence using Part-of-Speech (POS) tags. Source languages in our experiments are Spanish and English whereas target languages are English, Spanish and German.

2. Related Work

Many publications are dealing with the word reordering problem, e.g. (Kanthak et al., 2005), but only few of them are making use of linguistic knowledge about the sentence structure.

The use of morpho-syntactic information for reordering in SMT has been proposed in (Nießen and Ney, 2001). Harmonising word order between German and English has been done by prepending German verb prefixes and by treating interrogative sentences using syntactic information. Other method for harmonising word order between the same language pair has been proposed in (Collins et al.,

2005). They use a German parse tree for moving German verbs towards the beginning of the clause.

In this work, we investigate three languages - Spanish, English and German. The only additional language resource is POS information. Parsing or some other type of detailed information about syntax is not necessary.

3. Novel Methods for Reorderings in the Source Language

Word reordering is an important issue in SMT since differences in word order between the source and the target language are one of the main causes of translation errors. We propose two types of reorderings based on the POS tags:

- local reorderings (convenient for translation from and into Spanish)
- long-range reorderings (convenient for translation from and into German)

3.1. Local reorderings

Adjectives in the Spanish language are usually placed after the corresponding noun, whereas in English and most other non-Romanic languages is the other way round. Therefore local reorderings of nouns and adjective groups might be helpful for translation systems where the Spanish language is involved.

If Spanish is the source language, each Spanish noun is moved behind the corresponding adjective group. If Spanish is the target language, each adjective group in the source language is moved behind the corresponding noun. Examples of some local reorderings are shown in Table 1.

3.2. Long-range reorderings

The verbs in the German language, unlike many other languages, can often be placed at the end of the clause. This is mostly the case with infinitives and past participles. Therefore long-range reorderings of verb groups might be useful

Spanish	motivos económicos y políticos ↓ económicos y políticos motivos
English	economic and political reasons ↓ reasons economic and political

Table 1: Examples of local reordering

for translation systems where the German language is involved.

Reordering of verbs in German as a source language has already been investigated in (Collins et al., 2005). In this work we concentrate on the translation into German. Each verb group in the source language containing infinitive and/or past participle form is moved towards the end of the clause. If an auxiliary verb is present, it stays at its original position. Examples of long-range reordering can be seen in Table 2.

English	we have to offer them our hand in a very symbolic way ↓ we have them our hand in a very symbolic way to offer
German	wir müssen dem Volk ganz symbolisch die Hand reichen
English	I have already answered this question. ↓ I have already this question answered .
German	ich habe diese Frage bereits beantwortet .

Table 2: Examples of long-range reordering

4. Experimental Settings

The experiments have been performed on the European Parliament corpus described in (Koehn and Montz, 2005). It contains Spanish, German and English full form text. The corpus statistics can be seen in Table 3. For each language, POS tags have been generated using a corresponding morpho-syntactic analyser (Freeling for Spanish, EN-GCG for English, GERCG for German). The original corpus consists of about 700k sentences and 15M running words. Since it is adjusted only for translation into English and therefore is not parallel for all language pairs, we also extracted a common trilingual set containing about 670k sentences and 14M running words in order to investigate the German-Spanish language pair. In order to investigate effects of sparse resources, we have randomly extracted a small subset containing about 7k sentences and 144k running words (about 1% of the original corpus).

As already pointed out, reorderings were applied in the source language, then training and search were performed using the transformed source language data. The translation system we used is the phrase-based system described in (Zens et al., 2005). Modifications of the training and search procedure were not necessary.

The evaluation metrics used in our experiments are WER (Word Error Rate), PER (Position-independent word Error Rate) and BLEU (BiLingual Evaluation Understudy) (Papineni et al., 2002).

5. Translation Results

5.1. Local reorderings

Spanish→English: Table 4 presents the results for the translation from Spanish to English. In Table 4(a) it can be seen that local reorderings have reduced the errors and increased the BLEU score. The improvements are even more significant for the small corpus.

Some more details can be seen in Table 4(b). Development and test corpus have been divided into two parts: one containing sentences that have been actually reordered (which is about 60%) and other which remains the same. Then these two sets have been evaluated separately for each translation system. Table 4(b) shows that the reorderings are improving translation quality for both sets, slightly more for the reordered set. This difference is more significant for the small training corpus. This means that the new system allows better learning of models so that the translation quality has been improved both directly as well as indirectly.

From the translation example in Table 4(c) it can be seen that the system trained on the reordered corpus is better capable of producing the correct output.

English→Spanish: The results for this translation direction can be seen in Table 5. The effects presented in Table 5(a) are similar to those for Spanish→English, although the improvements are smaller. The reason for that is probably due to the following characteristics of the Spanish language: although the adjective group is *usually* situated behind the corresponding noun, this is not *always* the case. Therefore some reorderings in English are not really helpful. The results of the separated evaluation shown in Table 5(b) are also similar as those for Spanish→English, again with smaller improvements.

Translation example in Table 5(c) shows the advantage of the new system. Without reordering, “democratic wish” has been translated as “democracy wish”. The output of the new system still does not exactly match the reference, but it is syntactically and semantically correct.

Spanish↔German: The translation results for this language pair show similar tendencies in both translation directions as the results for Spanish and English, but the improvements are smaller. Some differences in the POS labelling of two languages are probably the reason for that (e.g. distinction between adjectives and past participles).

5.2. Long-range reorderings

English↔German: The translation results for English→German are presented in Table 6. Table 6(a) shows that the long-range reorderings are also improving translation quality. It also shows that, unlike for the local reorderings, the improvements are less significant for the small corpus. The most probable reason for this is exactly the long range. Long distances between words within a phrase generally pose problems for translation systems even if the word order in the source and target language has

	Spanish	English	German	English
Train: Sentences	730740		751088	
Running Words+Punctuation	15724914	15222146	15257678	16052330
Vocabulary	113882	72739	205374	74708
Singletons [%]	39.2	38.3	49.8	38.3
Dev: Sentences	2000		2000	
Running Words+Punctuation	60628	58655	55147	58655
Distinct Words	8182	6547	9213	6547
OOVs [%]	0.4	0.2	0.8	0.2
Test: Sentences	2000		2000	
Running Words+Punctuation	60332	57951	54260	57951
Distinct Words	8279	6496	9048	6496
OOVs [%]	0.4	0.2	0.7	0.2

Table 3: Corpus statistics

been harmonised. These problems are especially difficult to overcome if only a small training corpus is available. More detailed results are presented in Table 6(b). It can be noted that for the reordered set (about 60% sentences) the error rates are significantly higher than for the other set. The improvements for the reordered set are larger and for the full corpus there are also consistent improvements of the rest of sentences. For the small corpus, there are small improvements of the BLEU score of the unchanged sentences and a slight deterioration of the error rates.

The translation example in Table 6(c) shows an improvement of the verb group translation. Without reordering, the auxiliary verb “werden” which expresses the future tense has been lost, whereas the new system translated the whole verb group correctly.

Spanish→German: Long-range reorderings also have similar effects for the translation from Spanish into German as for the translation from English into German, but again the improvements are smaller.

For this translation direction we have also investigated the combination of the two reordering types since for this language pair both types of word order differences are present. As expected, for the full corpus the two methods are additive, i.e. the best results have been obtained by applying both local and long-range reorderings. For the small corpus, the best results have been obtained by applying only local reorderings. The results can be seen in Table 7.

6. Conclusion

In this work we presented methods for harmonising word order between source and target language using POS tags of the source language. Two types of reorderings have been investigated depending on the language pair. Experiments showed that both types of POS-based word reorderings improve the translation quality for various language pairs and translation directions. Local noun-adjective reorderings for translation from and into Spanish language are especially helpful when a small amount of bilinugal training data is available. Long-range verb reorderings for translation into German are more effective for the large training corpus.

In future work, we plan to apply the methods on other tasks and other language pairs with different word order (for ex-

ample Catalan↔English). We also plan to investigate possibilities for using reordering probabilities.

7. Acknowledgement

This work was partly supported by the TC-STAR project by the European Community (FP6-506738).

8. References

- M. Collins, P. Koehn, and I. Kučerová. 2005. Clause restructuring for statistical machine translation. In *Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Ann Arbor, Michigan, June.
- S. Kanthak, D. Vilar, E. Matusov, R. Zens and H. Ney. 2005. Novel reordering approaches in phrase-based statistical machine translation. In *Proc. of the ACL Workshop on Building and Using Parallel Texts*, pages 167–174, Ann Arbor, Michigan, June.
- P. Koehn and C. Montz. 2005. Shared task: statistical machine translation between European languages. In *Proc. of the ACL Workshop on Building and Using Parallel Texts*, pages 119–124, Ann Arbor, Michigan, June.
- S. Nießen and H. Ney. 2001a. Morpho-syntactic analysis for reordering in statistical machine translation. In *Proc. MT Summit VIII*, pages 247–252, Santiago de Compostela, Galicia, Spain, September.
- K. Papineni, S. Roukos, T. Ward, and W.J. Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proc. 40th Annual Meeting of the Assoc. for Computational Linguistics*, pages 311–318, Philadelphia, PA, July.
- D. Vilar, E. Matusov, S. Hasan, R. Zens, and H. Ney. 2005. Statistical Machine Translation of European Parliamentary Speeches. In *Proceedings of MT Summit X*, pages 259–266, Phuket, Thailand, September.
- R. Zens, O. Bender, S. Hasan, S. Khadivi, E. Matusov, J. Xu, Y. Zhang, and H. Ney. 2005. The RWTH Phrase-based Statistical Machine Translation System. In *Proc. of the International Workshop on Spoken Language Translation (IWSLT)*, pages 155–162, Pittsburgh, PA, October.

Spanish→English		dev			test		
		WER	PER	BLEU	WER	PER	BLEU
700k	baseline	56.8	43.5	27.4	57.3	44.1	27.4
	reorder adjective	56.0	42.8	28.4	56.7	43.4	28.1
7k	baseline	65.5	50.0	19.6	66.0	50.1	20.0
	reorder adjective	64.0	48.9	21.8	64.9	49.4	22.1

(a) Translation results (%)

Spanish→English			dev			test		
			WER	PER	BLEU	WER	PER	BLEU
700k	reordered	baseline	56.7	42.8	27.9	57.4	43.5	27.6
		reorder adjective	56.3	42.2	28.7	57.1	43.1	28.3
	rest	baseline	57.1	45.3	26.3	57.0	45.6	26.7
		reorder adjective	56.6	44.9	26.7	57.1	45.5	27.0
7k	reordered	baseline	65.9	49.4	19.6	66.5	49.5	19.7
		reorder adjective	64.2	48.3	22.0	65.0	48.6	22.0
	rest	baseline	64.6	51.6	19.5	64.9	51.6	20.9
		reorder adjective	64.4	51.0	20.3	64.9	51.2	21.6

(b) Detailed translation results (%) for reordered and other sentences

original Spanish sentence:	...este sistema no sea susceptible de ser usado como arma política .
reordered Spanish sentence:	...este sistema no sea susceptible de ser usado como política arma .
generated English sentence:	
without reordering:	...this system is not likely to be used as a weapon policy .
with reordering:	...this system is not likely to be used as a political weapon .
reference English sentence:	...the system cannot be used as a political weapon .

(c) Translation examples with and without reordering

Table 4: Translation results for Spanish→English: local noun-adjective reorderings

English→Spanish		dev			test		
		WER	PER	BLEU	WER	PER	BLEU
700k	baseline	58.6	45.6	32.4	58.6	45.6	32.1
	reorder adjectives	58.2	45.5	32.8	58.1	45.1	32.5
7k	baseline	67.1	52.5	24.0	67.0	52.3	23.8
	reorder adjectives	66.4	51.9	25.3	66.4	51.9	25.0

(a) Translation results (%)

English→Spanish			dev			test		
			WER	PER	BLEU	WER	PER	BLEU
700k	reordered	baseline	59.0	45.4	32.4	58.8	45.0	32.0
		reorder adjectives	58.6	45.2	32.8	58.3	44.4	32.5
	rest	baseline	57.0	46.4	32.4	57.9	47.9	32.2
		reorder adjectives	56.8	46.4	32.8	57.6	47.5	32.6
7k	reordered	baseline	67.6	52.4	23.6	67.3	51.8	23.5
		reorder adjectives	66.7	51.6	25.2	66.6	51.3	24.8
	rest	baseline	65.3	52.8	25.3	65.9	54.2	25.0
		reorder adjectives	65.2	52.8	25.7	65.9	54.1	25.7

(b) Detailed translation results (%) for reordered and other sentences

original English sentence:	...the democratic wish of the countries and people in question...
reordered English sentence:	...the wish democratic of the countries and people in question...
generated Spanish sentence:	
without reordering:	...la democracia deseo de los países y personas en cuestión...
with reordering:	...la voluntad democrática de los países y personas en cuestión...
reference Spanish sentence:	...el deseo democrático de los correspondientes estados y pueblos...

(c) Translation examples with and without reordering

Table 5: Translation results for English→Spanish: local noun-adjective reorderings

English→German		dev			test		
		WER	PER	BLEU	WER	PER	BLEU
700k	baseline	68.6	56.6	19.6	68.5	56.4	19.7
	reorder verbs	68.1	55.9	20.4	68.1	55.4	20.8
7k	baseline	77.6	61.9	15.0	77.6	61.7	15.4
	reorder verbs	77.4	61.9	15.5	77.2	61.6	16.0

(a) Translation results (%)

English→German			dev			test		
			WER	PER	BLEU	WER	PER	BLEU
700k	reordered	baseline	70.3	57.8	17.8	70.4	57.4	17.8
		reorder verbs	69.7	56.9	18.8	69.8	56.2	19.0
	rest	baseline	64.9	54.0	23.7	64.0	53.8	24.3
		reorder verbs	64.6	53.8	24.1	64.0	53.5	25.0
7k	reordered	baseline	79.3	62.5	13.5	79.9	62.6	13.6
		reorder verbs	78.9	62.4	14.0	79.1	62.4	14.4
	rest	baseline	73.8	60.6	18.4	72.4	59.6	19.5
		reorder verbs	74.0	60.6	18.6	72.6	59.8	19.6

(b) Detailed translation results (%) for reordered and other sentences

original English sentence:	In spite of these remarks , we shall vote for this report .
reordered English sentence:	In spite of these remarks , we shall for this report vote .
generated German sentence:	
without reordering:	Trotz dieser Ausführungen stimmen wir für diesen Bericht .
with reordering:	Trotz dieser Bemerkungen , werden wir für diesen Bericht stimmen .
reference German sentence:	Trotz dieser Bemerkungen werden wir für den vorliegenden Bericht stimmen .

(c) Translation examples with and without reordering

Table 6: Translation results for English→German: long-range verb reorderings

Spanish→German		dev			test		
		WER	PER	BLEU	WER	PER	BLEU
600k	baseline	68.4	55.0	21.2	68.7	55.2	21.2
	reorder adjectives	68.2	55.0	21.3	68.6	55.2	21.3
	reorder verbs	68.0	55.0	21.3	68.2	55.2	21.3
	reorder adjectives + verbs	67.9	54.7	21.6	68.1	55.1	21.5
6k	baseline	78.9	63.3	14.1	79.1	63.2	14.3
	reorder adjectives	78.3	62.9	14.8	78.6	62.9	14.9
	reorder verbs	78.8	63.3	14.2	78.9	63.0	14.4
	reorder adjectives + verbs	78.5	63.0	14.6	78.8	62.9	14.7

Table 7: Translation results (%) for Spanish→German: local and long-range reorderings