

TALP Phrase-Based System and TALP System Combination for the IWSLT 2006 *IWSLT 2006, Kyoto*

*Marta R. Costa-jussà, Josep M. Crego, Adrià de Gispert,
Patrik Lambert, Maxim Khalilov,
José A.R. Fonollosa, José B. Mariño and Rafael Banchs*

Universitat Politècnica de Catalunya (UPC),
Barcelona, Spain

November 2006



- 1 Introduction
- 2 TALP-phrase system
 - Baseline system
 - Statistical Machine Reordering (SMR)
 - SMR Translation Model
 - SMR Tuple Extraction
- 3 Phrase-Based and N -gram-Based Combination
- 4 IWSLT 2006 Evaluation
- 5 Conclusions

Participation in the IWSLT 2006 Evaluation

- Tasks

- 1 Arabic to English
- 2 Chinese to English
- 3 Italian to English
- 4 Japanese to English

- Systems

- 1 TALP-phrase
- 2 TALP-comb



Alignment and Phrase Extraction

- Alignment:
 - IBM MODEL 4 (GIZA++ (OCH, 2000))
 - 50 CLASSES USING 'MKCLS'
 - LOWER-CASED TEXT
 - UNION
- Phrase Extraction:
 - PHRASE-EXTRACTION (OCH AND NEY, 2004)
 - NO LENGTH LIMIT

Log-linear combination of feature functions

$$\tilde{e} = \underset{e}{\operatorname{argmax}} \left\{ \sum_{m=1}^M \lambda_m h_m(e, f) \right\} \quad (1)$$

- Conditional and posterior probability
- Target Language Model (4-grams, SRILM (Stolcke, 2002))
- Forward and backward lexicon models
- Word bonus
- Phrase bonus

Log-linear combination of feature functions

$$\tilde{e} = \underset{e}{\operatorname{argmax}} \left\{ \sum_{m=1}^M \lambda_m h_m(e, f) \right\} \quad (1)$$

- Conditional and posterior probability
- Target Language Model (4-grams, SRILM (Stolcke, 2002))
- Forward and backward lexicon models
- Word bonus
- Phrase bonus

→ Additionally, the decoder allows for non-monotonic search, taking into account two limits:

- distortion limit m (maximum distance measured in words that a phrase is allowed to be reordered,)
- reordering limit j (maximum number of reordering jumps)

Statistical Machine Reordering (SMR)

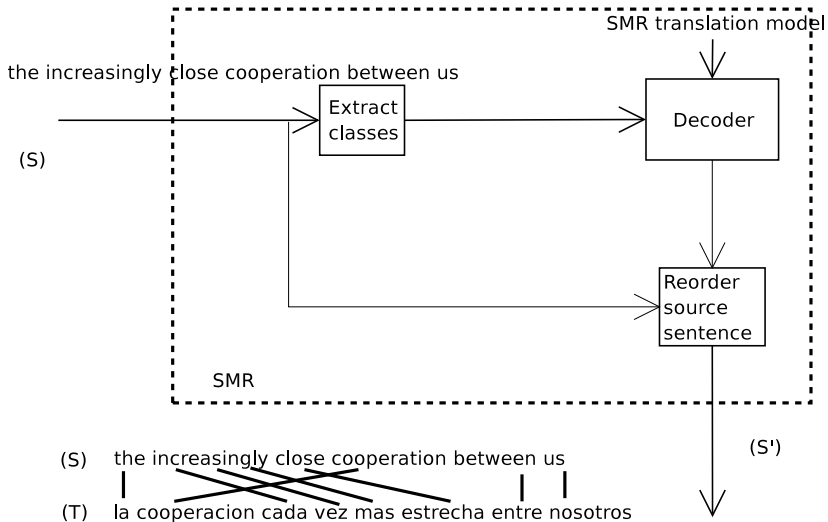
- SMR is applied before SMT.
- SMR uses an SMT system to deal with reordering.
- SMR translates from an original source language (S) to a reordered source language (S'), given a target language (T).

(S) the increasingly close cooperation between us
 (T) la cooperacion cada vez mas estrecha entre nosotros

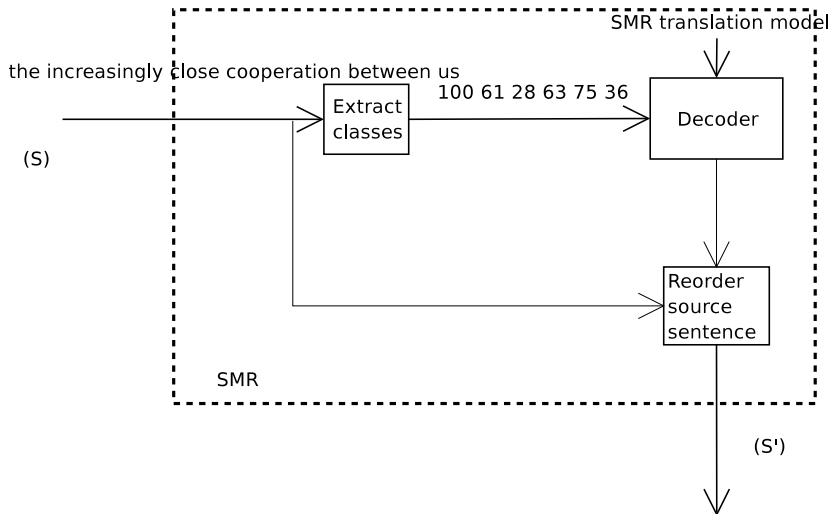
(S') the cooperation increasingly close between us
 (T) la cooperacion cada vez mas estrecha entre nosotros

- In monotonized task $S' \rightarrow T$, training (word alignment) is easier, and also translation (less reordering problems)

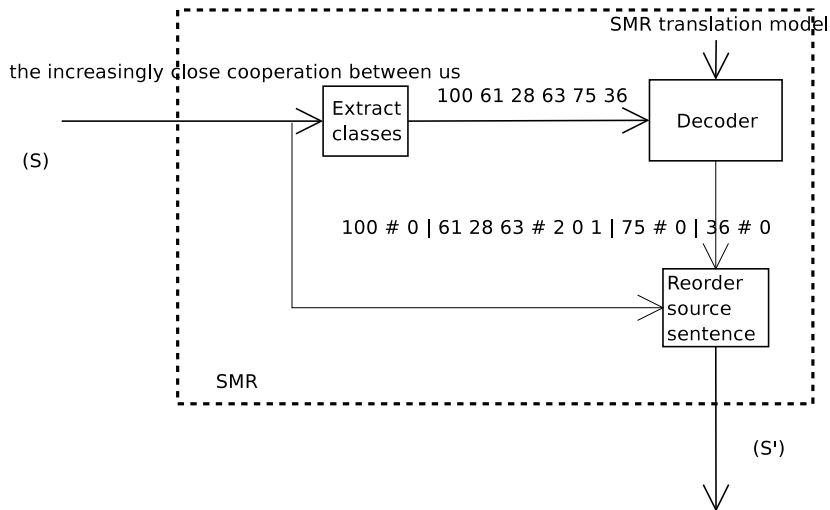
SMR Block Diagram



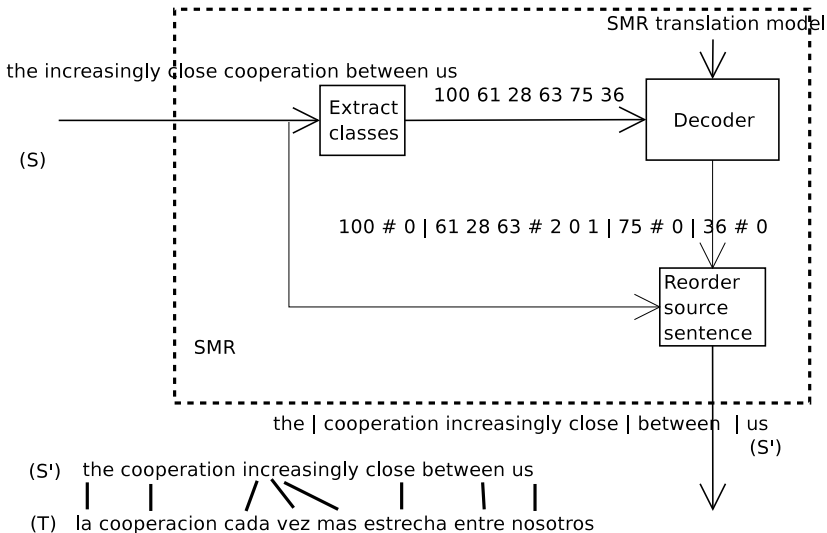
SMR Block Diagram



SMR Block Diagram



SMR Block Diagram

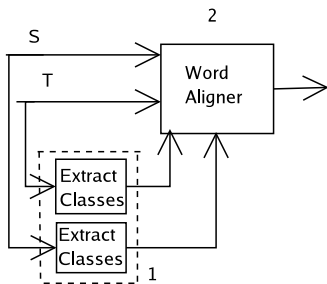


SMR Translation Model

→ Building the SMR translation model...

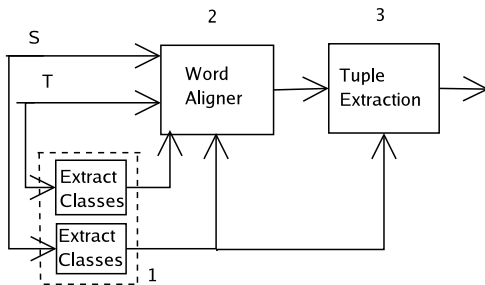
SMR Translation Model

→ Building the SMR translation model...



SMR Translation Model

→ Building the SMR translation model...



SMR Tuple Extraction

(a) bilingual S2T tuple

increasingly close cooperation #	cooperacion cada vez mas estrecha #	0-1 0-2 0-3 1-4 2-0
(source)	(target)	(word alignment)
		(wrд_src-wrd_trg)

(b) many-to-many word alignment \longrightarrow many-to-one word alignment

P_ibm (increasingly, mas) > P_ibm (increasingly, cada)

P_ibm (increasingly, mas) > P_ibm (increasingly, vez)

increasingly close cooperation #	cooperacion cada vez mas estrecha #	0-3 1-4 2-0
----------------------------------	-------------------------------------	-------------

(c) bilingual S2S' tuple

increasingly close cooperation #	2 1 0
(source)	(new order)

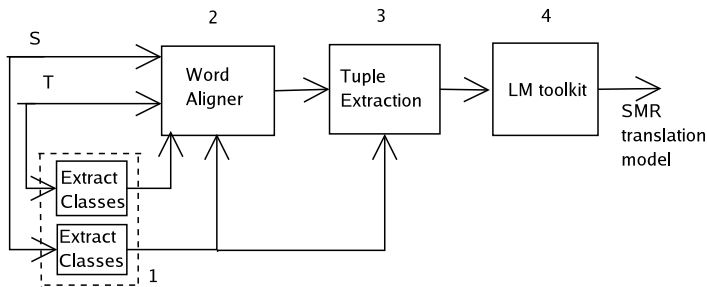
(d) delete tuples whose source fragment is NULL

(e) class substitution

C61 C28 C63 #	2 1 0
---------------	-------

SMR Translation Model

→ Building the SMR translation model



TALP-comb

Given the 1best outputs of each system, we perform a log-linear combination, which is computed by using the following models:

- IBM Model 1 for the sentence in the source to target direction.
- IBM Model 1 for the sentence in the target to source direction.
- Target language models: 2gram, 3gram and 5gram.
- Word bonus.

And we optimize the weights of each model with the simplex algorithm.

Results

Results for the TALP-phrase in all the tasks of the IWSLT 2006.

- Results for case-insensitive evaluations and without punctuation marks
- Two TALP-phrase systems: using SMR technique or not.
- In all tasks, except for the Italian to English, we used a non-monotonic search (with $m = 5$ and $j = 3$).

Results

Results for the TALP-phrase in all the tasks of the IWSLT 2006.

- Results for case-insensitive evaluations and without punctuation marks
- Two TALP-phrase systems: using SMR technique or not.
- In all tasks, except for the Italian to English, we used a non-monotonic search (with $m = 5$ and $j = 3$).

Task	System	Dev		Test		Eval	
		BLEU	NIST	BLEU	NIST	BLEU	NIST
zh2en	TALP-phrase p (SMR)	19.29	6.57	46.33	8.95	20.08	6.42
	TALP-phrase c1	20.36	6.75	44.87	8.56	20.06	6.26

Results

Results for the TALP-phrase in all the tasks of the IWSLT 2006.

- Results for case-insensitive evaluations and without punctuation marks
- Two TALP-phrase systems: using SMR technique or not.
- In all tasks, except for the Italian to English, we used a non-monotonic search (with $m = 5$ and $j = 3$).

Task	System	Dev		Test		Eval	
		BLEU	NIST	BLEU	NIST	BLEU	NIST
zh2en	TALP-phrase p (SMR)	19.29	6.57	46.33	8.95	20.08	6.42
	TALP-phrase c1	20.36	6.75	44.87	8.56	20.06	6.26
ar2en	TALP-phrase p (SMR)	27.07	7.15	55.34	10.28	22.20	6.54
	TALP-phrase c1	25.95	7.07	54.06	10.24	23.66	6.70

Results

Results for the TALP-phrase in all the tasks of the IWSLT 2006.

- Results for case-insensitive evaluations and without punctuation marks
- Two TALP-phrase systems: using SMR technique or not.
- In all tasks, except for the Italian to English, we used a non-monotonic search (with $m = 5$ and $j = 3$).

Task	System	Dev		Test		Eval	
		BLEU	NIST	BLEU	NIST	BLEU	NIST
zh2en	TALP-phrase p (SMR)	19.29	6.57	46.33	8.95	20.08	6.42
	TALP-phrase c1	20.36	6.75	44.87	8.56	20.06	6.26
ar2en	TALP-phrase p (SMR)	27.07	7.15	55.34	10.28	22.20	6.54
	TALP-phrase c1	25.95	7.07	54.06	10.24	23.66	6.70
it2en	TALP-phrase p	41.66	9.08	62.68	10.69	35.55	8.32
	TALP-phrase c1 (SMR)	41.65	8.92	61.45	10.46	35.55	8.32

Results

Results for the TALP-phrase in all the tasks of the IWSLT 2006.

- Results for case-insensitive evaluations and without punctuation marks
- Two TALP-phrase systems: using SMR technique or not.
- In all tasks, except for the Italian to English, we used a non-monotonic search (with $m = 5$ and $j = 3$).

Task	System	Dev		Test		Eval	
		BLEU	NIST	BLEU	NIST	BLEU	NIST
zh2en	TALP-phrase p (SMR)	19.29	6.57	46.33	8.95	20.08	6.42
	TALP-phrase c1	20.36	6.75	44.87	8.56	20.06	6.26
ar2en	TALP-phrase p (SMR)	27.07	7.15	55.34	10.28	22.20	6.54
	TALP-phrase c1	25.95	7.07	54.06	10.24	23.66	6.70
it2en	TALP-phrase p	41.66	9.08	62.68	10.69	35.55	8.32
	TALP-phrase c1 (SMR)	41.65	8.92	61.45	10.46	35.55	8.32
jp2en	TALP-phrase p (SMR)	15.37	6.01	48.93	9.54	14.51	5.58
	TALP-phrase c1	17.04	6.40	47.52	9.82	15.09	5.82

TALP Phrase-based System

- The SMR improved the results in the internal test.
- In the final evaluation, these conclusions can not be generalized, which might be explained by the high number of unknown words in the eval set

Set	Chinese	Arabic	Italian	Japanese
development	71	165	138	66
test	50	55	79	25
evaluation	106	220	186	202

Table: *Number of unknown words in development, test and evaluation sets.*

Results

Results for the TALP-phrase, TALP-tuple and the combination of both in all the tasks of the IWSLT 2006.

(case-insensitive and without punctuations marks)

Task	System	Dev		Eval	
		BLEU	NIST	BLEU	NIST
zh2en	TALP-phrase p (SMR)	19.29	6.57	20.08	6.42
	TALP-phrase c1	20.36	6.75	20.06	6.26
	TALP-tuple p	19.75	6.64	20.34	6.22
	TALP-tuple c1	19.69	6.59	19.80	6.39
	TALP-comb	21.19	6.69	20.21	5.97

Results

Results for the TALP-phrase, TALP-tuple and the combination of both in all the tasks of the IWSLT 2006.

(case-insensitive and without punctuations marks)

Task	System	Dev		Eval	
		BLEU	NIST	BLEU	NIST
zh2en	TALP-phrase p (SMR)	19.29	6.57	20.08	6.42
	TALP-phrase c1	20.36	6.75	20.06	6.26
	TALP-tuple p	19.75	6.64	20.34	6.22
	TALP-tuple c1	19.69	6.59	19.80	6.39
	TALP-comb	21.19	6.69	20.21	5.97
ar2en	TALP-phrase p (SMR)	27.07	7.15	22.20	6.54
	TALP-phrase c1	25.95	7.07	23.66	6.70
	TALP-tuple p	29.27	7.52	23.83	6.80
	TALP-tuple c1	29.48	7.46	23.60	6.72
	TALP-tuple c2	28.75	7.40	23.40	6.65
	TALP-tuple c3	29.09	7.41	23.10	6.67
	TALP-comb	30.29	7.41	23.95	6.60

Results

Results for the TALP-phrase, TALP-tuple and the combination of both in all the tasks of the IWSLT 2006.

(case-insensitive and without punctuations marks)

Task	System	Dev		Eval	
it2en	TALP-phrase p	41.66	9.08	35.55	8.32
	TALP-phrase c1 (SMR)	41.65	8.92	35.55	8.32
	TALP-tuple p	43.05	9.21	37.38	8.59
	TALP-tuple c1	43.05	9.20	31.13	8.46
	TALP-tuple c2	43.63	9.24	37.55	8.49
	TALP-tuple c3	41.60	9.15	36.21	8.35
	TALP-comb		44.13	9.04	37.74

Results

Results for the TALP-phrase, TALP-tuple and the combination of both in all the tasks of the IWSLT 2006.

(case-insensitive and without punctuations marks)

Task	System	Dev		Eval	
it2en	TALP-phrase p	41.66	9.08	35.55	8.32
	TALP-phrase c1 (SMR)	41.65	8.92	35.55	8.32
	TALP-tuple p	43.05	9.21	37.38	8.59
	TALP-tuple c1	43.05	9.20	31.13	8.46
	TALP-tuple c2	43.63	9.24	37.55	8.49
	TALP-tuple c3	41.60	9.15	36.21	8.35
	TALP-comb	44.13	9.04	37.74	8.41
jp2en	TALP-phrase p (SMR)	15.37	6.01	14.51	5.58
	TALP-phrase c1	17.04	6.40	15.09	5.82
	TALP-tuple p	16.59	6.34	14.61	5.27
	TALP-tuple c1	18.20	6.37	15.17	5.18
	TALP-comb	19.36	6.42	15.66	5.51

TALP System Combination

- In the combination, we observe a general improvement of the BLEU score, whereas the NIST score seems to decrease.
- It can be explained by the particular features used: shorter outputs benefit both from IBM model 1 and language model, while the word bonus is not enough to compensate this.

Conclusions

- We have presented the TALP-phrase and TALP-comb systems for IWSLT 2006 evaluation.
- The SMR reordering technique was expected to improve translation quality in the evaluation set as it did in the internal test set. The failure of doing so may be explained by the increased number of unknown words in the evaluation set compared to the test set.
- We are currently studying ways to cope with unknown words in the SMR technique.
- The very basic combination of TALP-phrase and TALP-tuple systems succeeded in improving BLEU score, but not NIST score. Nevertheless, it is an encouragement to investigate more elaborate ways to perform the combination.