

HIGHLIGHTS

- study translation hypotheses that can be obtained by iterative, greedy oracle improvement from the one-best hypothesis of a state-of-the-art phrase-based SMT system (*Moses*)
- greedy search guided by an automatic evaluation metric : sentence-level BLEU, using one or several reference translations
- improvements are obtained by a set of operations directly applied on biphrases : move, merge, replace, split, remove and rewrite

OPERATIONS IN ACTION

Sequence of operations improving a translation with respect to some reference

Source	Reference
une majorité ₁ du groupe ppe ₂ soutiendra ₄ donc ₃ la ligne ₅ du ₆ rapport kindermann ₇	the majority of the ppe group will be supporting the line of the kindermann report
seed	a majority ₁ of the ppe group ₂ therefore ₃ support ₄ the line ₅ the ₆ kindermann report ₇
replace	a majority ₁ of the ppe group ₂ therefore ₃ will be supporting ₄ the line ₅ the ₆ kindermann report ₇
split	a majority ₁ of the ppe group ₂ therefore ₃ will be supporting ₄ the ₅ line of ₆ the ₇ kindermann reports ₈
remove	a majority ₁ of the ppe group ₂ 3 will be supporting ₄ the ₅ line of ₆ the ₇ kindermann reports ₈
replace	the majority ₁ of the ppe group ₂ 3 will be supporting ₄ the ₅ line of ₆ the ₇ kindermann reports ₈

REWRITE THE SOURCE

A new operation to access new translations

- replaces the source phrase of a biphrase with some other source phrase, and replaces its translation with the newly attainable translations of this new source phrase

previous	il est évident que parler d'intermodalité présuppose un profond changement de la culture d'entreprise . it is clear that speak intermodality presupposes a profound change in the business culture .
rewrite	il est évident que débat d'intermodalité présuppose un profond changement de la culture d'entreprise . it is clear that discussion on intermodality presupposes a profound change in the business culture .

Examples of English source rewritings and new reachable French reference translation fragment

source	reference	rewrite phrases
abused	dénaturé	different
buying	rachat	purchase
complex	multitude	number series wealth
damaging	désastreuse	disastrous
drivers	des personnels	people
excuse	argument	argument grounds reason

CONCLUSION

- important potential for improvement of current phrase-based SMT systems, demonstrated with one and several reference translations
- current SMT systems have the potential to already produce very good translation hypotheses even for difficult language pairs
- contribution of the *rewrite* operation makes explicit that the specificities of a unique source text sometimes are responsible for the difficulty to translate with respect to some reference translation(s)
- pruning a phrase table using statistical significance reduces oracle performance significantly
- future work : train discriminative rerankers using hypotheses generated by our greedy oracle search to automatically improve translations

TRAINING DATA

- Europarl corpus : intersection for 11 languages (310K bi-sentences)
- we also used the BTEC corpus of basic traveling expressions in order to have the possibility to have several reference translations (7 for the language pair fr-en) (20K bi-sentences)

DATA SIZE AND PHRASE TABLE FILTERING

	BLEU		TER		#. biphrases in phrase table
	baseline	oracle	baseline	oracle	
full	29.1	65.9	54.0	23.5	735,273
/2	28.6	60.8	54.4	27.5	419,716
/4	27.6	55.6	55.4	31.1	239,647
/8	26.1	51.1	56.8	35.1	137,719
/16	25.2	46.0	58.4	39.0	79,837
sigtest	29.1	54.7	54.1	32.3	203,672

- reducing by half the quantity of training data roughly leads to the loss of 1 BLEU or less, but a lot of examples were apparently more useful (loss of 5 BLEU for the oracle)
- results for *sigtest* (significance pruning of Johnson et al. (2007)), with a high oracle loss of 11.2 BLEU, indicates that a significant part of the filtered entries, although apparently poorly scored by the translation system, would have in fact largely benefited the system

OPERATIONS AND BEAM SIZE EFFECTS

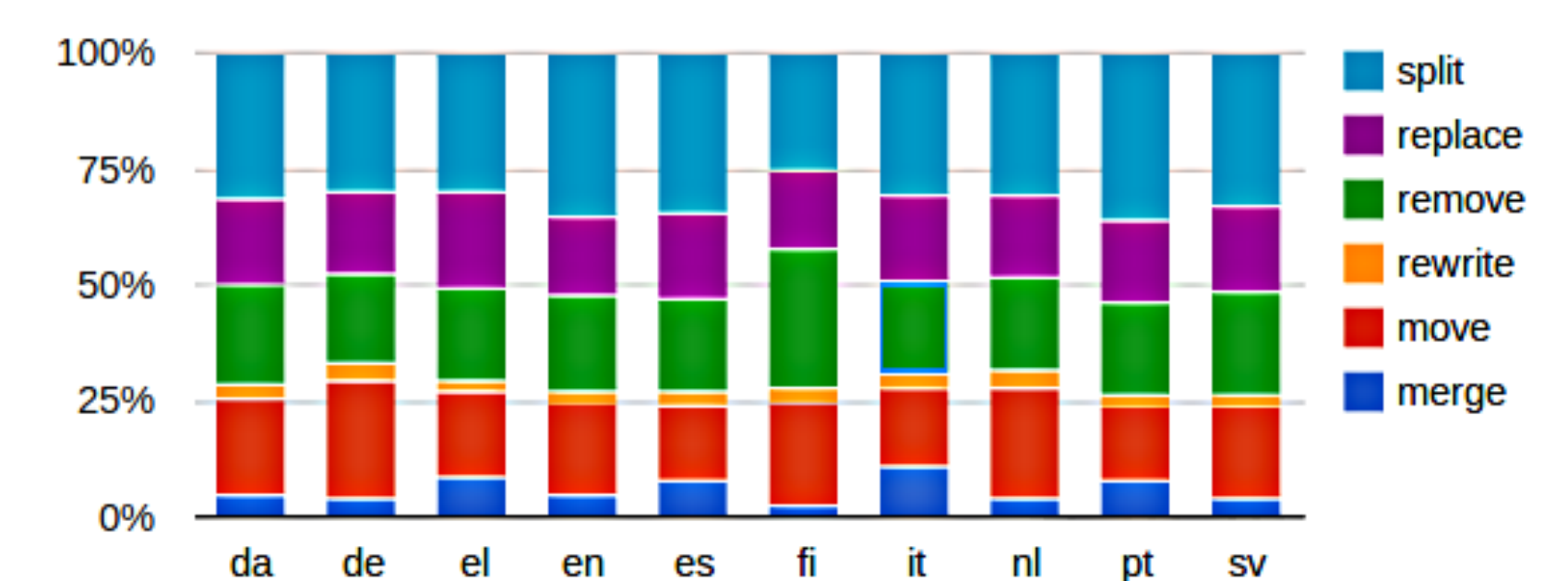
	Europarl fr→en (1 ref.)			BTEC fr→en (7 refs.)		
	BLEU score	TER score	avg # iterations	BLEU score	TER score	avg # iterations
baseline	29.0	54.0	-	59.62	24.60	-
beam size = 1						
merge	31.8	51.7	0.75	60.43	24.32	0.07
move	32.0	53.3	1.01	61.70	24.60	0.16
remove	29.7	50.0	1.03	59.62	24.60	0.00
replace	42.1	42.5	4.40	66.50	23.67	0.91
rewrite	29.8	53.5	0.38	59.69	24.68	0.04
split	45.7	41.3	4.46	69.34	27.07	1.24
all	66.5	23.1	11.04	77.30	23.47	1.92
beam size = 2						
all	66.6	23.0	11.19	77.88	23.16	2.28
beam size = 5						
all	67.8	22.3	11.26	79.06	22.94	2.12

- main improvements reside in hypotheses using generally smaller phrases (*split*) and choosing more appropriate translations for phrases (*replace*)
- increasing beam size helps us to slightly reduce search errors

TARGET LANGUAGES & REWRITE EFFECTS

	BLEU		TER	
	score	+rew	score	+rew
da	baseline 23.2	oracle 58.4 +0.9	baseline 61.3	oracle 29.5 -0.8
de	baseline 17.0	oracle 55.1 +1.4	baseline 68.0	oracle 32.0 -1.2
el	baseline 23.5	oracle 62.8 +1.0	baseline 62.2	oracle 26.5 -0.6
en	baseline 29.0	oracle 66.5 +0.6	baseline 54.0	oracle 23.1 -0.4
es	baseline 35.9	oracle 74.0 +0.5	baseline 49.7	oracle 18.2 -0.5
fi	baseline 11.2	oracle 46.1 +1.2	baseline 79.7	oracle 38.1 -1.2
it	baseline 31.6	oracle 71.2 +1.1	baseline 55.2	oracle 20.4 -1.7
nl	baseline 21.2	oracle 56.3 +1.6	baseline 64.6	oracle 32.4 -0.7
pt	baseline 33.4	oracle 69.8 +0.7	baseline 52.8	oracle 21.5 -0.5
sv	baseline 21.0	oracle 59.9 +1.0	baseline 62.7	oracle 27.8 -1.1

Distribution of operations per target language



- *replace* operations appear uniformly useful for all languages, illustrating the relative inadequacy of the translation models across languages
- *split* operations, more numerous for target languages with good baseline performance, can be attributed to some over-confidence in long bi-phrases

- more *move* operations for translating into German, Dutch and Scandinavian languages
- languages that benefit the most from *rewrite* correspond to languages with lower baseline scores

POS PATTERNS OF REWRITE

- large typology of configurations not limited to strict paraphrase phenomena
- main patterns :
 - Verb → Verb
 - Verb → Noun
 - Noun → Verb

