

HIGHLIGHTS

- best system 😊 but degrades the performance of the original MT system 😞
- main difficulty of the task: inconsistencies in the annotation

DATA ANALYSIS

Main Findings

- almost 25% of all edits are unique
- 60% of all edits are substitutions
- 3.42% of all edits are related to punctuation
- 1.16% of all edits are solely a change in case
- most word substitution do not change the PoS of the word
- many substitutions consist in correcting verb endings

edits	occurrences	edits	occurrences
+j	286	+la	108
+,	267	-el	107
+de	247	+el	102
+que	231	-los	101
-,	202	+los	92
-que	164	-se	92
-la	164	+en	88
+a	156	+se	85
-de	146	su → tu	71
+'	117	+las	68

substitution	count
VERB → VERB	2,372
NOUN → NOUN	1,243
ADP → ADP	605
ADJ → ADJ	571
PRON VERB → VERB	225
DET → DET	224
VERB → NOUN	178
NOUN → VERB	169
DET NOUN → DET NOUN	151
total	6,560

POST-EDITING AS MONOLINGUAL MT

- a MT system ‘translate’ the automatic translation (i.e. output of the system) to its post-edited version;
- MT system: NCODE our in-house MT system that implements a n -gram based translation model
- features:
 - 3-gram bilingual LM on words
 - 4-gram bilingual LM on PoS
 - 4-gram monolingual LM estimated on the post-edited sentences
 - conventional features: word and phrase penalties, lexical features, ...
 - no reordering was allowed

	train	development	test
MT output	23.32	23.21	22.91
APE-MT output	21.64	23.95	23.57

hTER score achieved by a MT system train to predict the post-edition from the MT output.

We are over-correcting 😞

RULE-BASED APPROACH

- focus on very precise categories of errors to avoid over-correction and to be able to analyze prediction errors
- three simple rules were considered:
 - **prediction of word case:** translated word must have the same case the source word it is aligned with;
 - **prediction of !/; and ?/¿:** enforce the punctuation to be the same as the source sentence and find the correct position to insert the inverted mark;
 - **prediction of verbal endings:** use WIKTIONARY conjugation table to generate a list of candidates and a LM to select the most probable form.

	hTER
baseline	23.320
+case correction	23.396
+punctuation correction	23.708
+verb correction	24.217

hTER score of our rule-based approach (dev. data).

Still degrading performance 😞

Failure Analysis:

- in the post-edited corpora, interrogative sentences do not always have an inverted interrogation mark → we are over-correcting
- correction of case: very good performance when restricted to the words that have been post-edited

CONCLUSION: OPEN QUESTION OR DIFFICULTIES

- are we correcting MT errors or normalizing MT output?
- we need more consistent annotations!
- many errors of our systems are due to tokenization and alignment errors

ACKNOWLEDGMENTS

This work was partly supported by ANR project Transread (ANR-12-CORD-0015).

<http://transread.limsi.fr/>