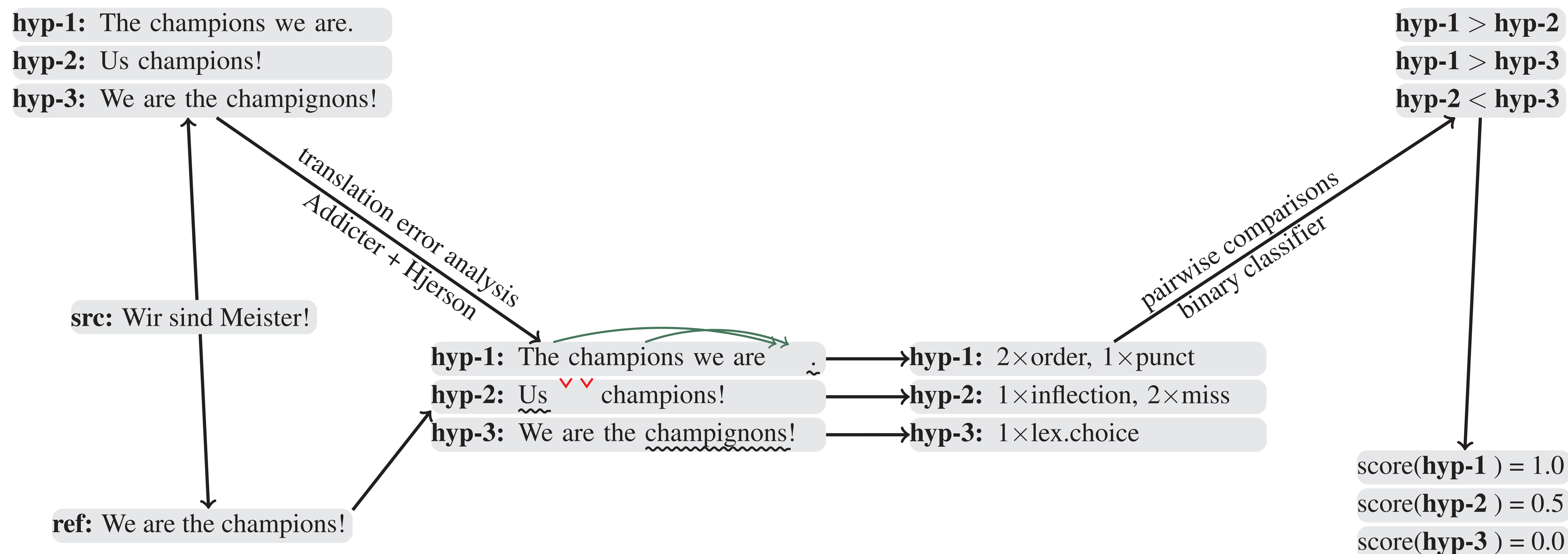


TerrorCat: a Translation Error Categorization-based MT Quality Metric

Mark Fishel@cl.uzh.ch, Rico Senrich@cl.uzh.ch, Maja Popović@dfki.de, Ondřej Bojar@ufal.mff.cuni.cz



Optimization:

- Variable reward to the winner proportional to the confidence estimate of the classifier
- PoS-specific features, e.g.: missing noun/misplaced verb/mis-inflected adj./...
- Generalizing mapping for PoS-tags
- Data from 2007 is an outlier; tried excluding it from the training material
- *Addicter's own alignment vs trained bilingual alignment*
- *Language-pair-specific vs tgt-language-specific vs universal models*

Addicter:

- Ref-hyp word-align (lemma, context)
- Unaligned words: missing/extra
- Aligned words: inflection/lex. choice error (based on the lemmas)
- “Untangle” alignment to find min. amount of reordering errors
- punctuation errors as separate class

Hjerson:

- Computes WER and PER on lemmas
- missing/extra tokens are the ones that contribute to both WER and PER
- inflection errors do not contribute to either, but their lemmas don't match
- order errors contribute only to WER
- The remaining are lex.choice errors

Pairwise Comparison:

- Binary SVM classifier, linear kernel
- WMT 2007–2011 manual rankings for training and development data sets (converted to comparison pairs)
- Input vectors are filled with differences of rel. error frequencies per category
- Addicter- and Hjerson-based analysis vectors are “concatenated”

System-level correlations:

Metric	fr-en	de-en	es-en	cs-en	*-en	en-fr	en-de	en-es	en-cs	en-*
TerrorCat	0.89	0.80	0.80	0.95	0.86	0.85	0.84	0.81	0.90	0.85
BLEU	0.85	0.48	0.90	0.88	0.78	0.86	0.44	0.87	0.65	0.70
mp4ibm1	0.08	0.56	0.12	0.91	0.42	0.61	0.91	0.71	0.76	0.75
MTeRater-Plus	0.93	0.90	0.91	0.95	0.92	-	-	-	-	-
AMBER_ti	0.94	0.63	0.85	0.88	0.83	0.84	0.54	0.88	0.56	0.70
meteor-1.3-rank	0.93	0.71	0.88	0.91	0.86	0.85	0.30	0.74	0.65	0.63

Segment-level correlations:

Metric	fr-en	de-en	es-en	cs-en	*-en	en-fr	en-de	en-es	en-cs	en-*
TerrorCat	0.21	0.30	0.33	0.23	0.27	0.29	0.20	0.23	0.20	0.23
mp4ibm1	0.15	0.16	0.18	0.12	0.15	0.21	0.13	0.13	0.06	0.13
MTeRater-Plus	0.30	0.36	0.45	0.36	0.37	-	-	-	-	-
AMBER_ti	0.24	0.26	0.33	0.27	0.28	0.32	0.22	0.31	0.21	0.27
meteor-1.3-rank	0.23	0.25	0.38	0.28	0.29	0.31	0.14	0.26	0.19	0.23

<https://github.com/fishel/terrorcat>