

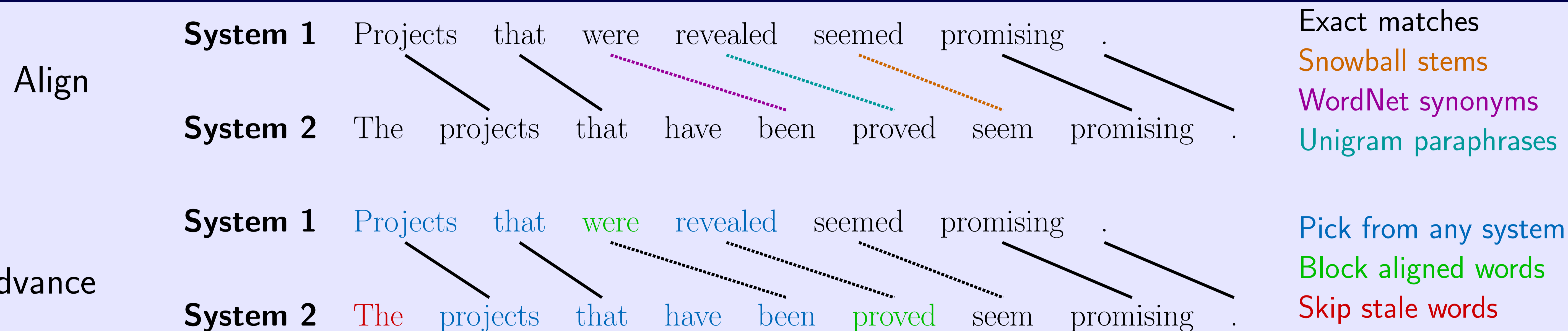
CMU System Combination in WMT 2011



Kenneth Heafield and Alon Lavie
 {heafield,alavie}@cs.cmu.edu
 kheafield.com/code

Carnegie Mellon University

Search



Partial Hypothesis

Projects that have been revealed

Score

Features	Exact Matches			Approximate Matches			Length	Language Model
	Unigram	Bigram	Trigram	Unigram	Bigram	Trigram		
System 1	3	1	0	1	1	0	5	-23.8955, OOV 0
System 2	4	3	2	1	1	1		

German Compounder

Recognize that “klar gemacht” aligns to “klargemacht”.
 Answer: segment → align → combine → **compound**

Compounding:

1. Run every word through cdec's segmenter
2. Count each form in normal text
3. Rejoin if the compound is more frequent
4. Score using sum of squares

Big LMs in 5 Languages

Data: Gigaword, News Crawl, News Commentary, UN, Europarl, Giga-FrEn, CzEng, and SMS (for ht-en)

Building: BigFatLM without pruning

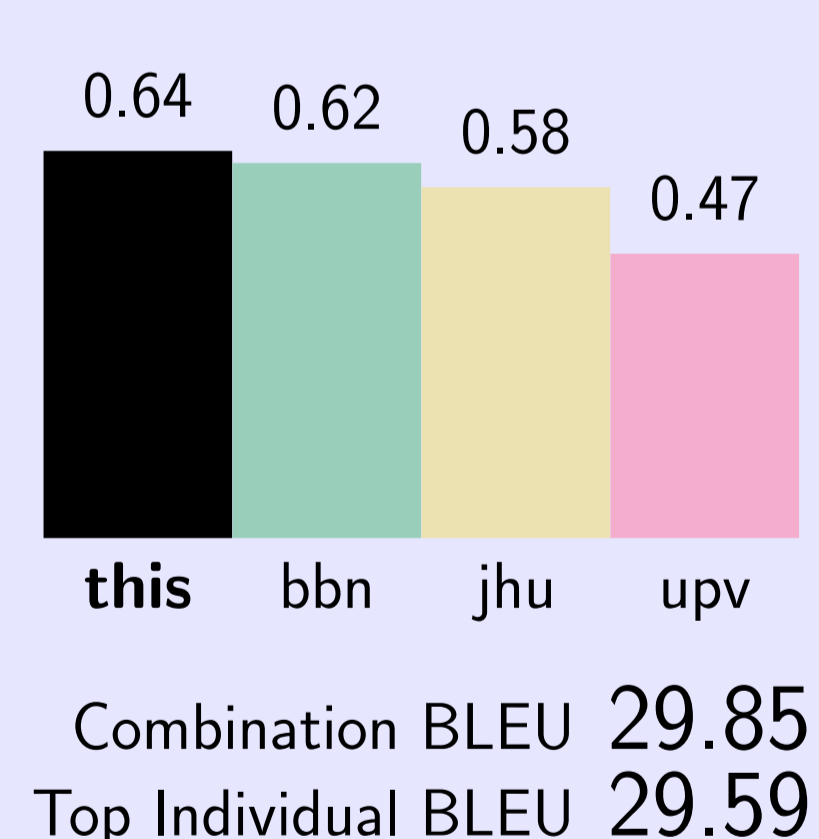
Filtering: Union of sentence-filtered models

Inference: KenLM

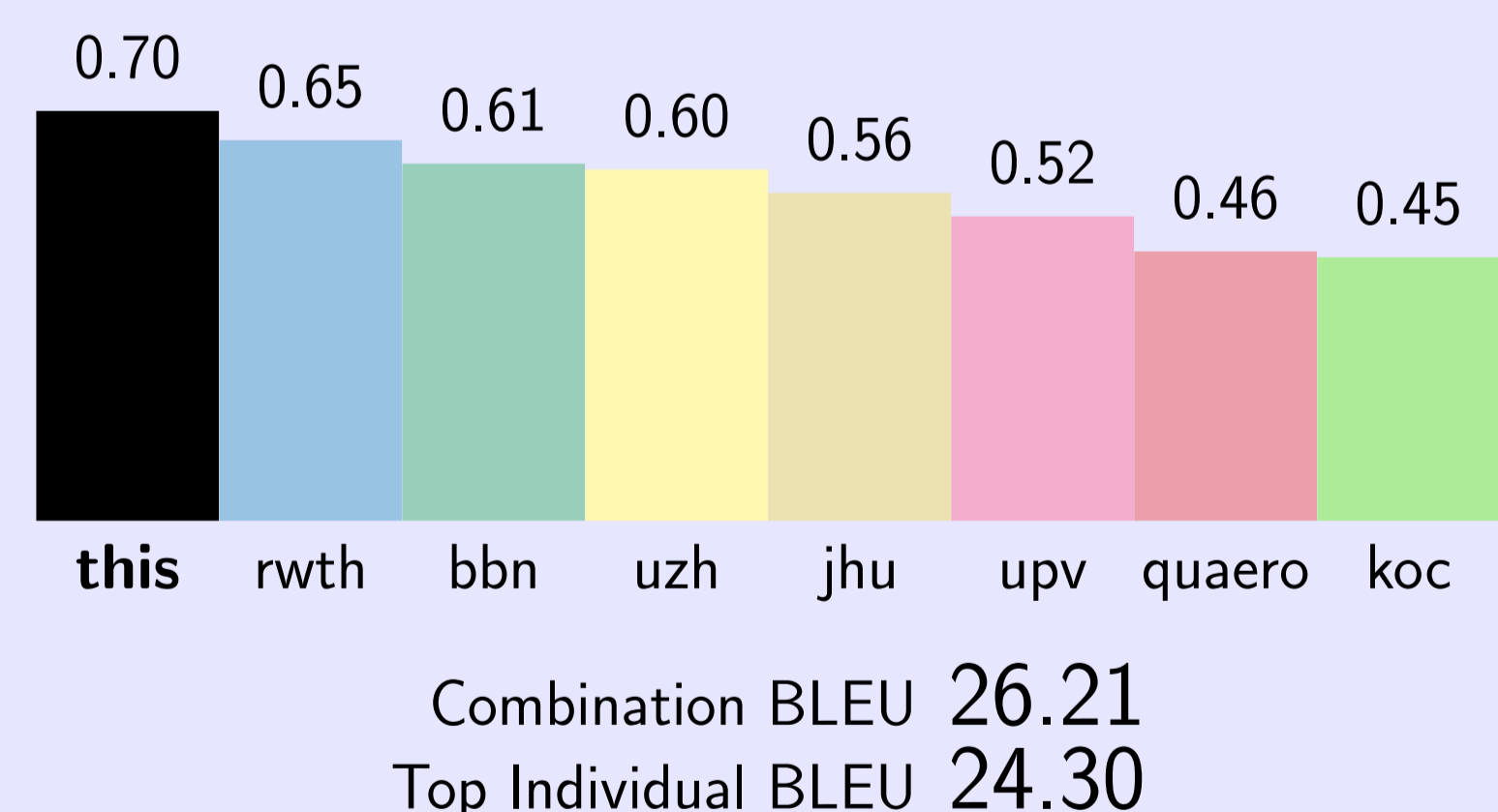
github.com/jhclark and kheafield.com/code

Human Evaluation Results

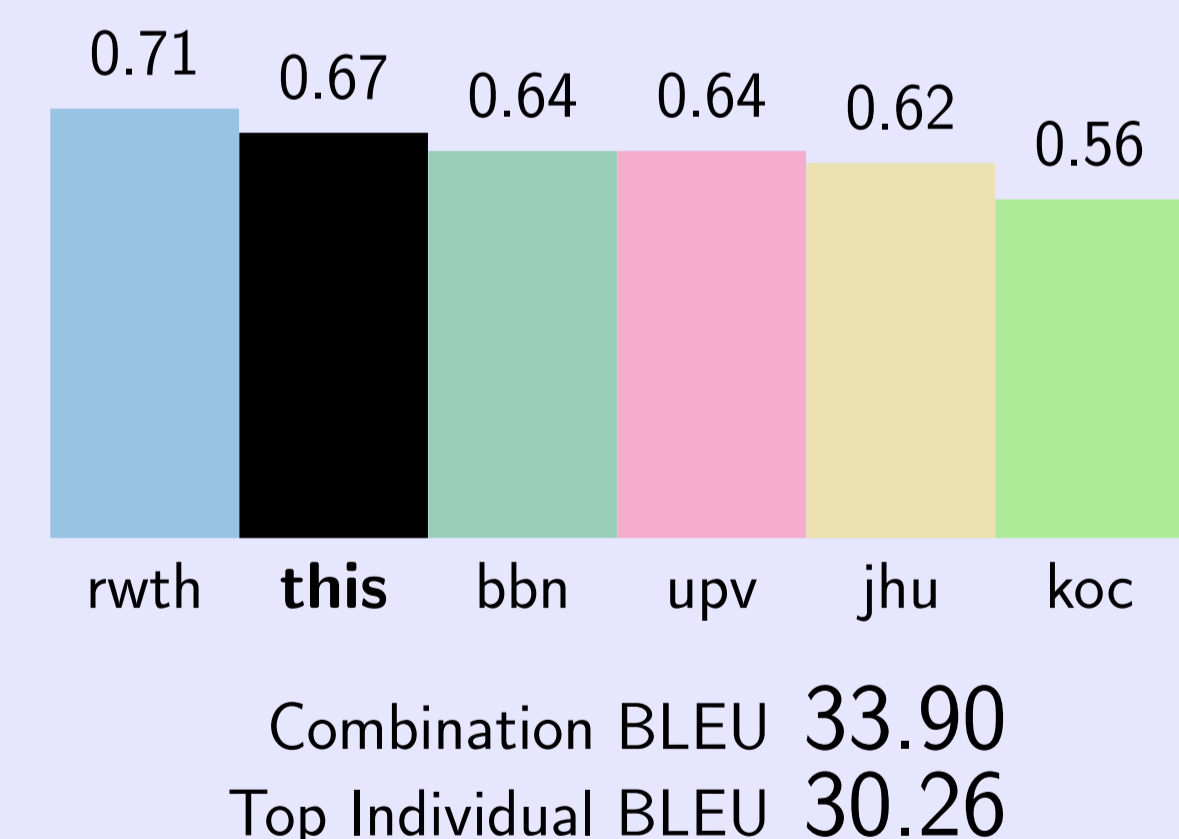
Czech-English



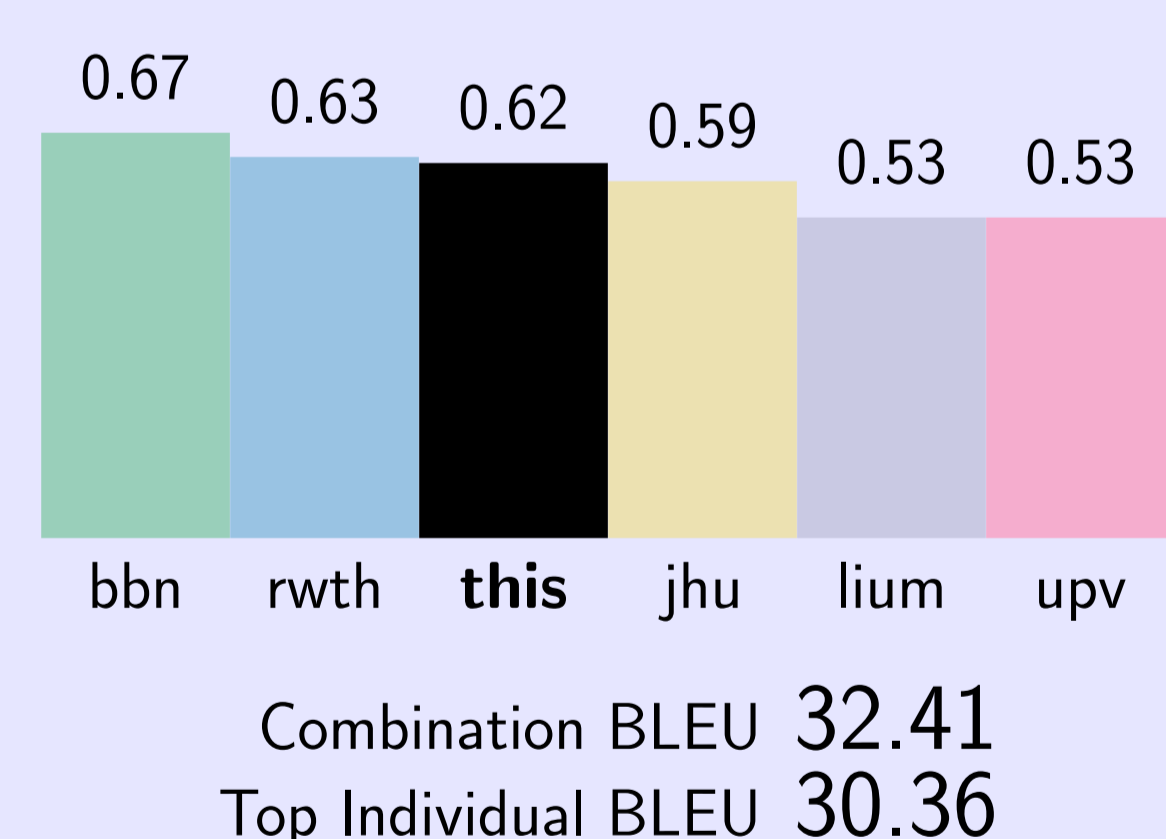
German-English



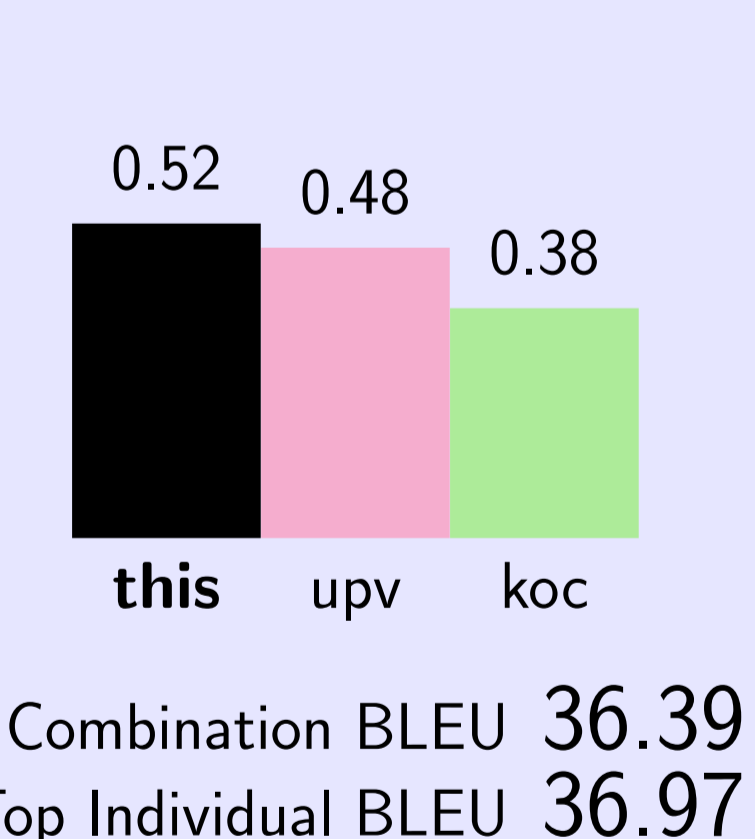
Spanish-English



French-English



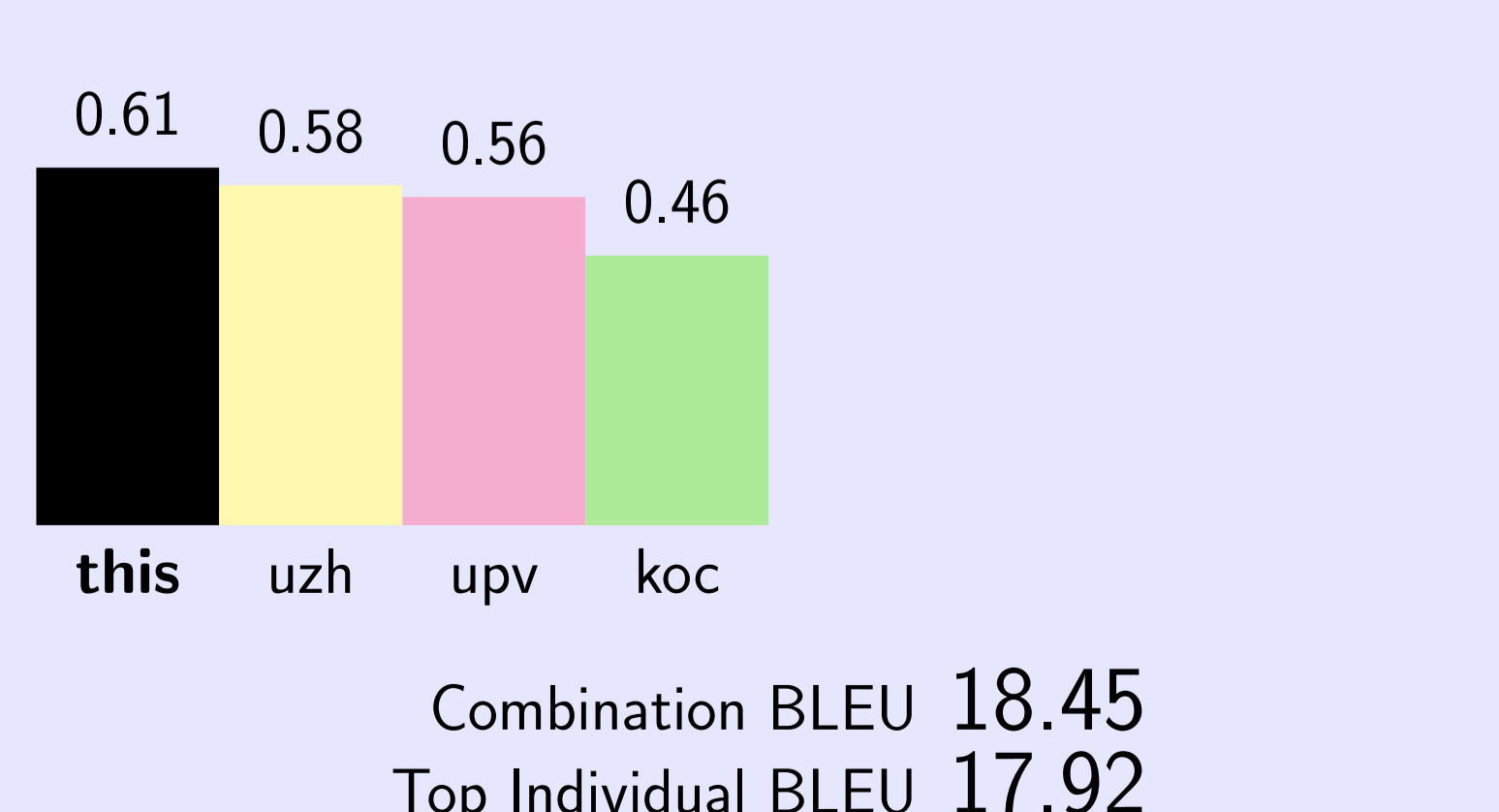
Clean Haitian-English



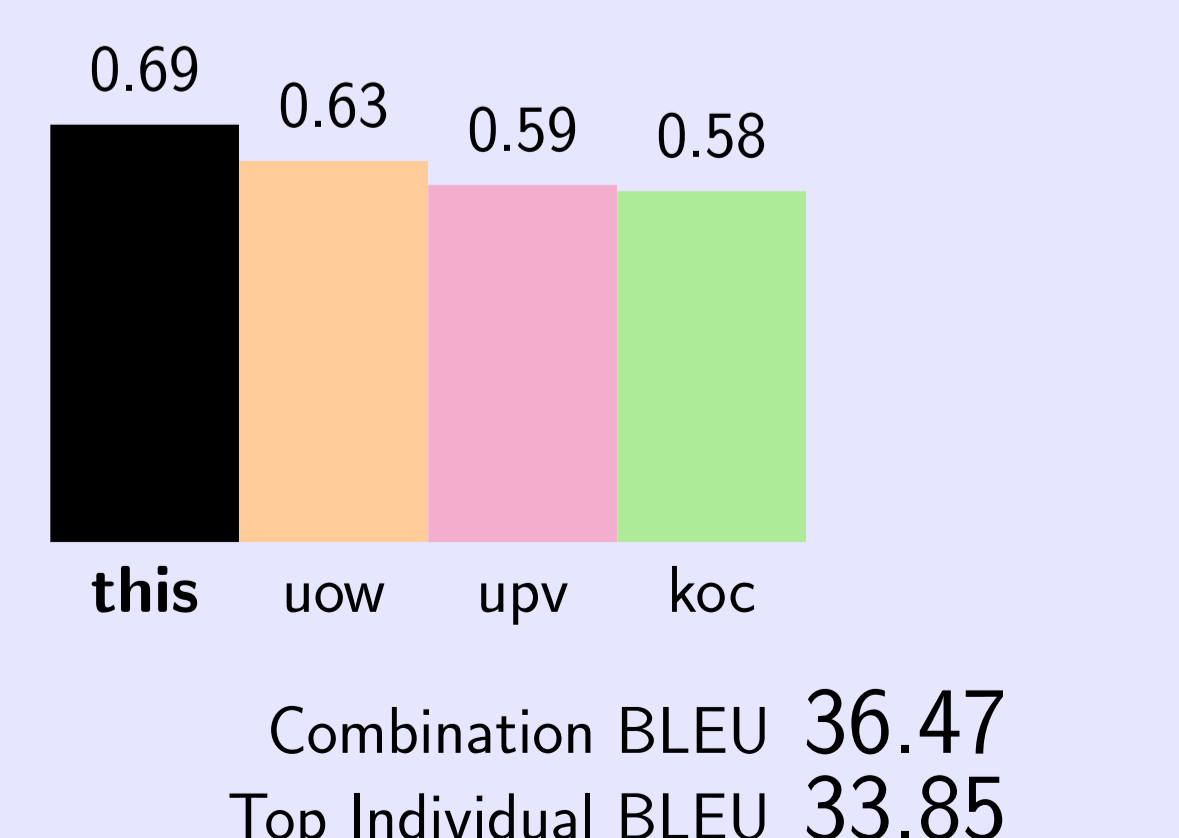
English-Czech



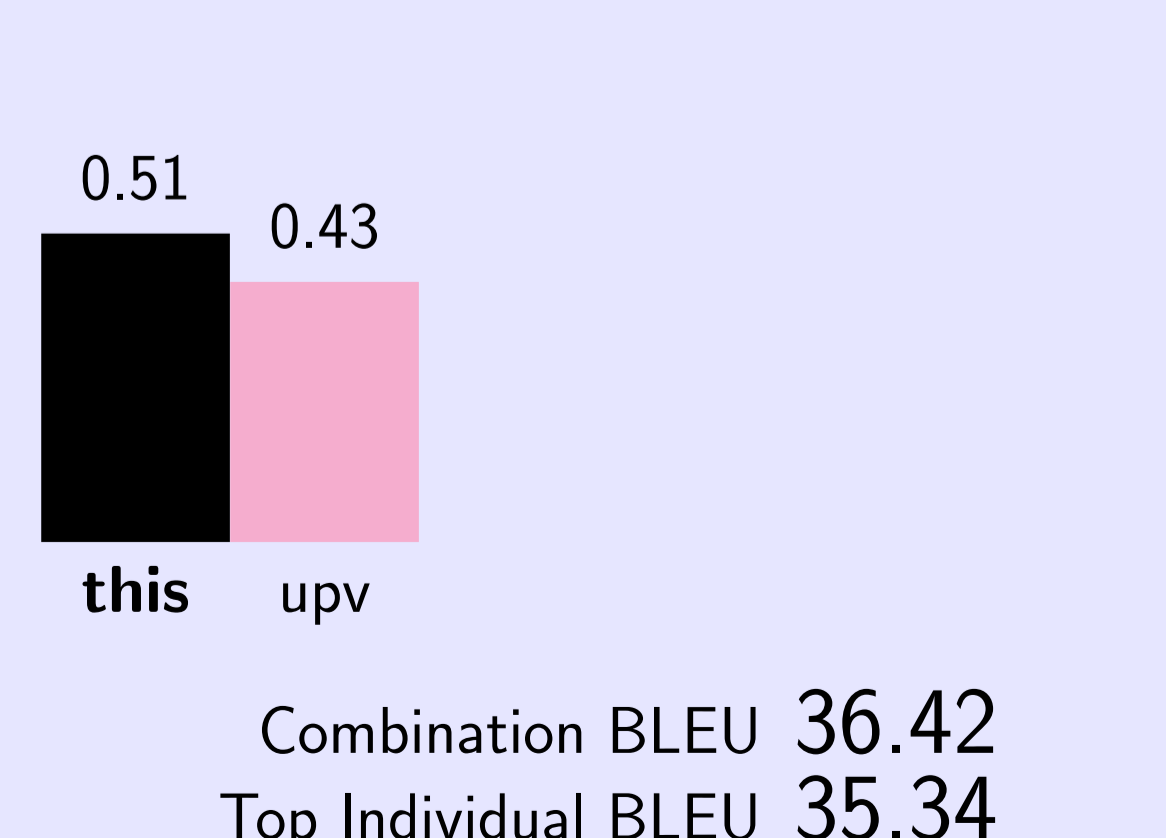
English-German



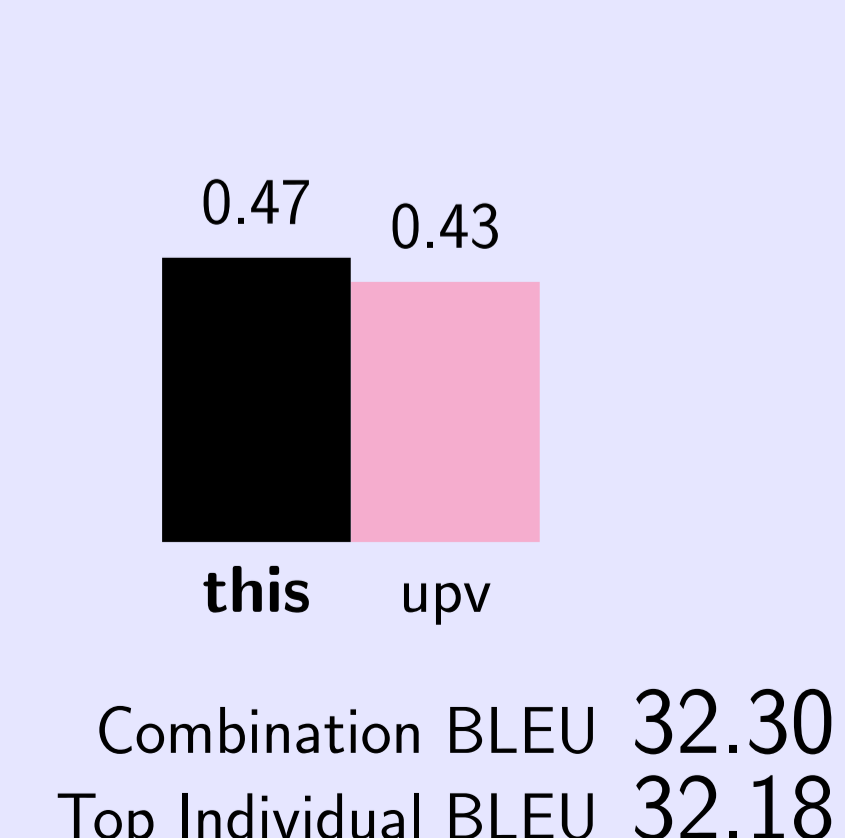
English-Spanish



English-French



Raw Haitian-English



Discussion

Win on 8/10 tasks.

In **Czech-English** and **German-English**, online systems lead by > 1 BLEU. This work wins.

In **French-English** and **Spanish-English**, many systems scored near the top. RWTH and BBN perform better.

In **Haitian-English**, bm-i2r won by a large margin. Combination did little.

Out of English, there are fewer participants. Language-specific processing matters more.

Compared with other participants, this work does better when one system has a clear lead.