BiVert: Bidirectional Vocabulary Evaluation using Relations for Machine Translation

Carinne Cherf, Yuval Pinter

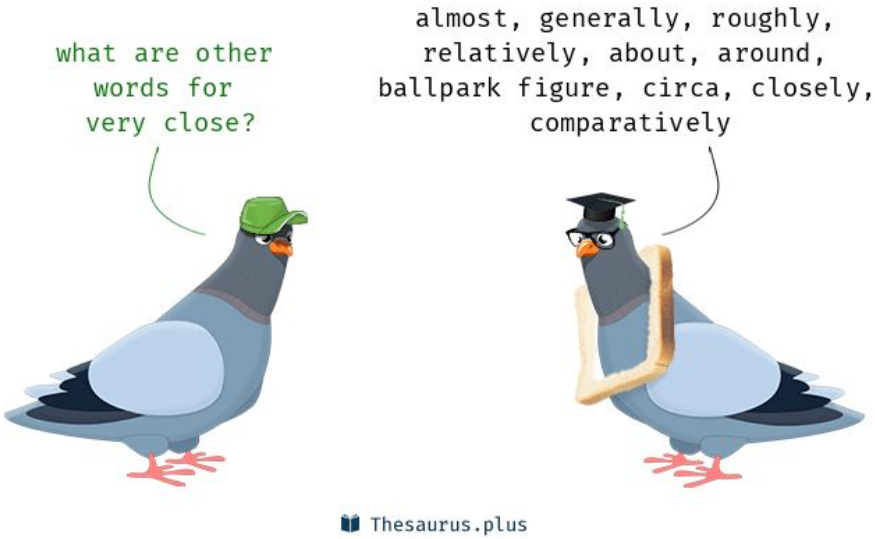
LREC-Coling 2024

Initial Research

Searching for translation mistakes through homonyms.

What are homonyms?

What were we looking for?



Evaluation Methods Today

Reference Based

- BLEU, ROUGE

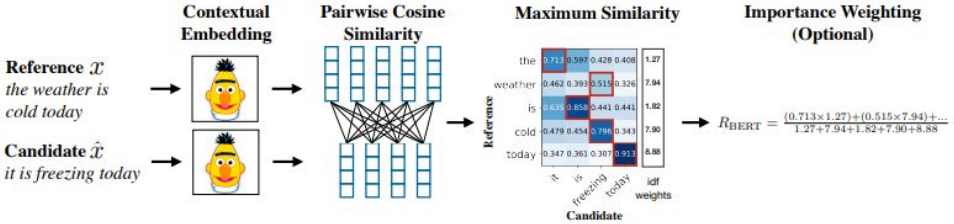
reference → The way to make people trustworthy is to trust them

$\ell_{ref}^{unigram} = 10$

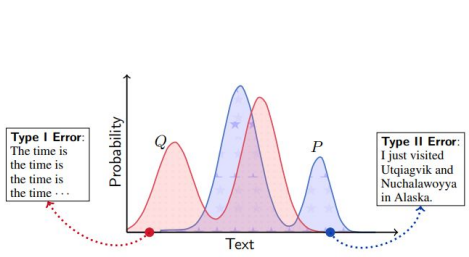
hypothesis → To make people trustworthy, you need to trust them

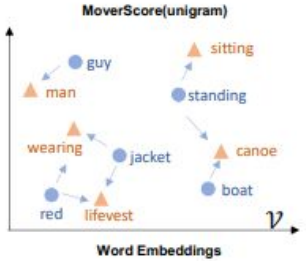
$\ell_{hyp}^{unigram} = 9$

- BertScore



- MAUVE, MoverScore

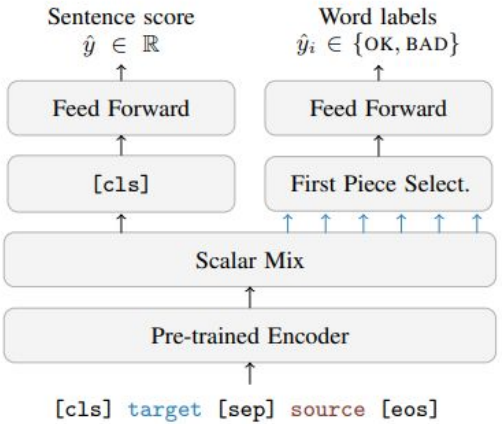




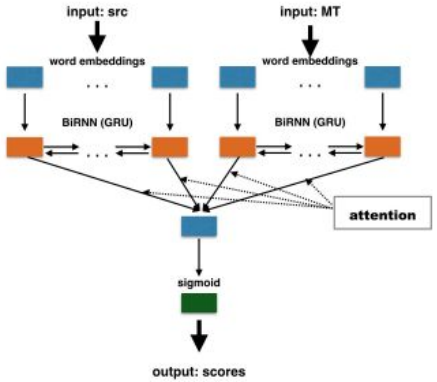
Evaluation Methods Today

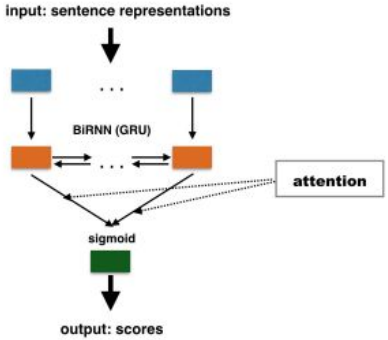
Quality Estimation

- CometKiwi

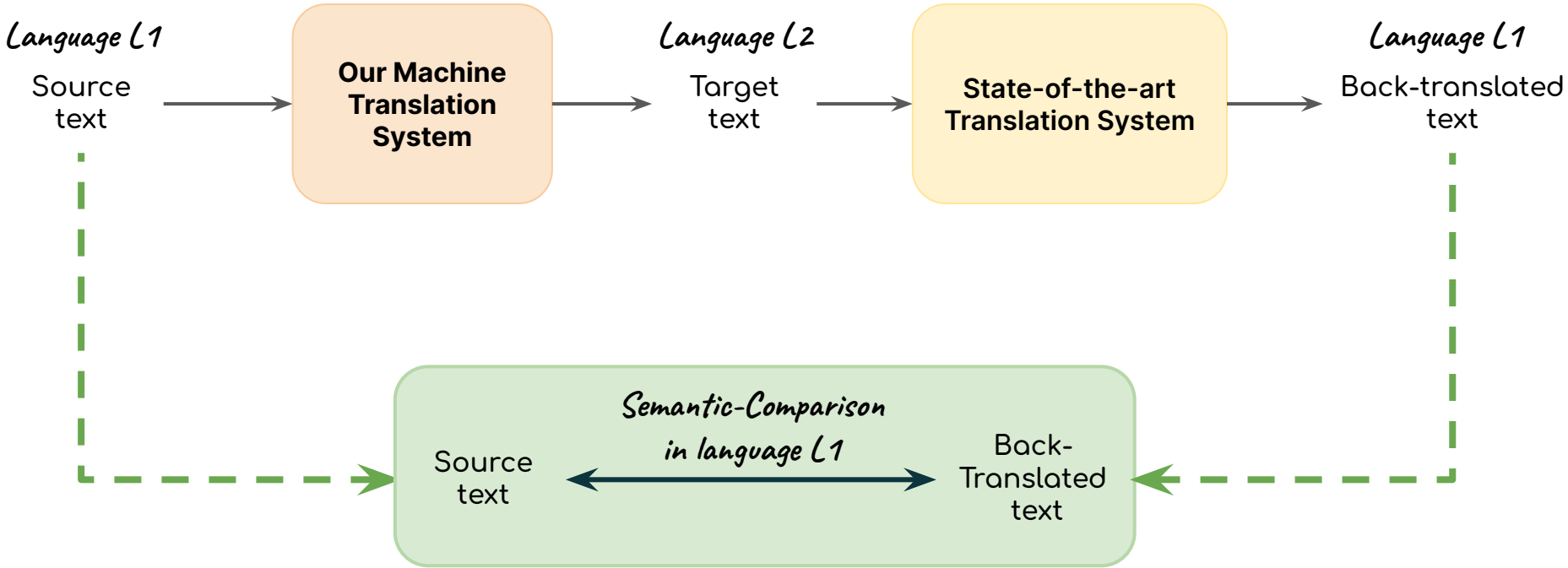


- DeepQuest





BiVert


Language L1
Source text
Our Machine Translation System
Language L2
Target text
State-of-the-art Translation System
Language L1
Back-translated text
Semantic-Comparison in language L1
Source text
Back-Translated text

BiVert

Back-translation

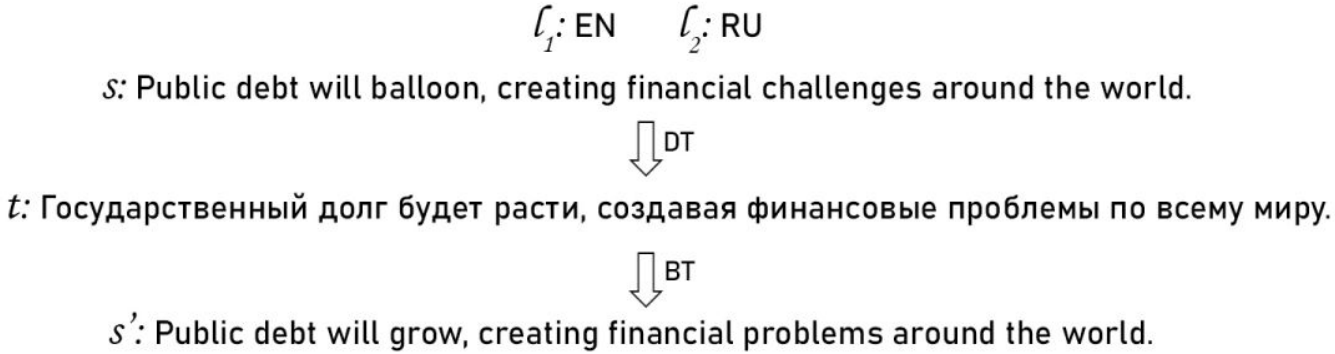


Figure 1: Example of a direct translation from English to Russian using the system we wish to evaluate, and its back-translation using a state-of-the-art translation system suitable for BiVert.

BiVert

Word Pair Relations

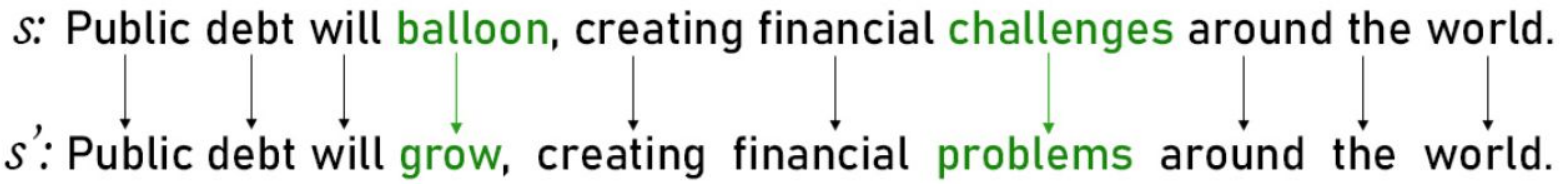


Figure 2: An example of final words alignment using the linear sum assignment problem algorithm.

Categories:

1. Same
2. Extra
3. Missing
4. Stop Words
5. Inflection: eat - ate (same lemma)
6. Derivation: happy - happiness
7. Sense

BiVert

Word Pair Alignment

The students read **quietly in the library**

They **quietly** studied books **in the library**

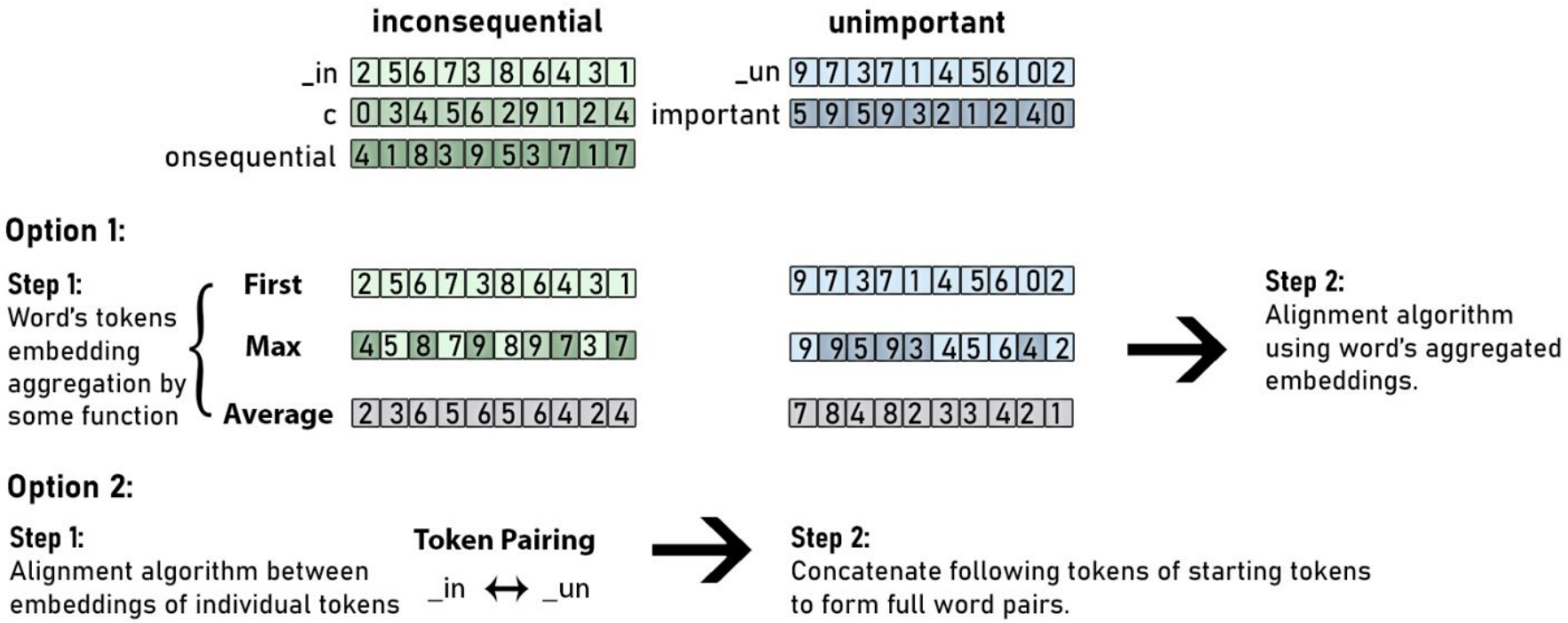
	The	students	read
They	**1-**0.3507	**1-**0.4420	**1-**0.4078
studied	**1-**0.3738	**1-**0.2949	**1-**0.3170
books	**1-**0.3866	**1-**0.3187	**1-**0.3178

Categories:

1. Same
2. Extra
3. Missing
4. Stop Words
5. Inflection: eat - ate (same lemma)
6. Derivation: happy - happiness
7. Sense

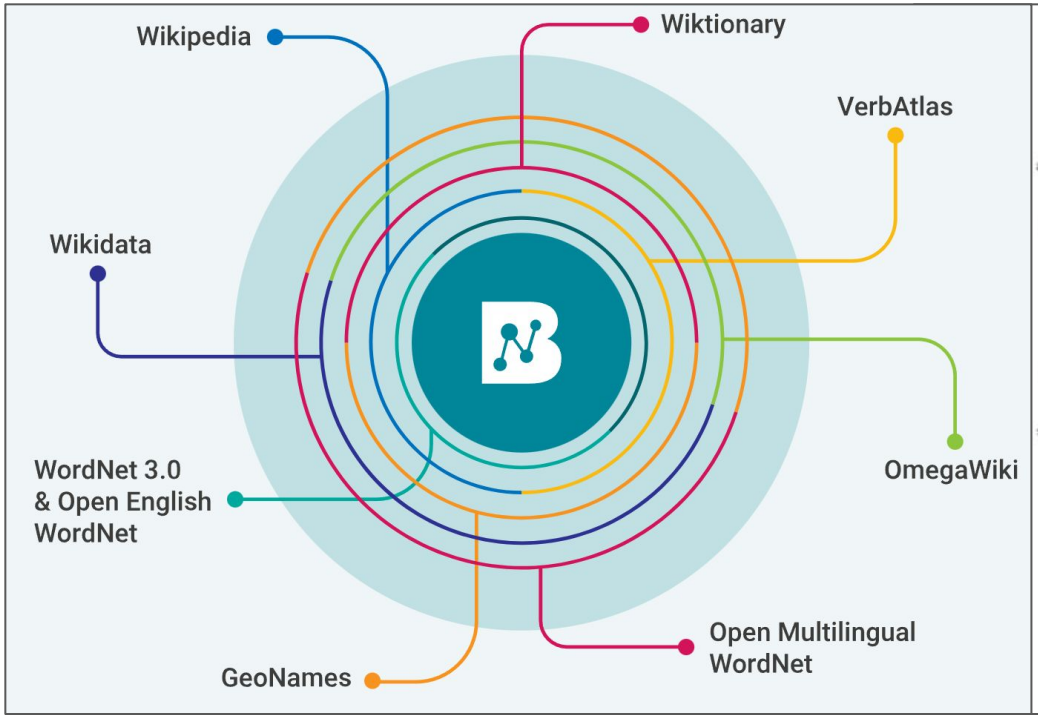
BiVert

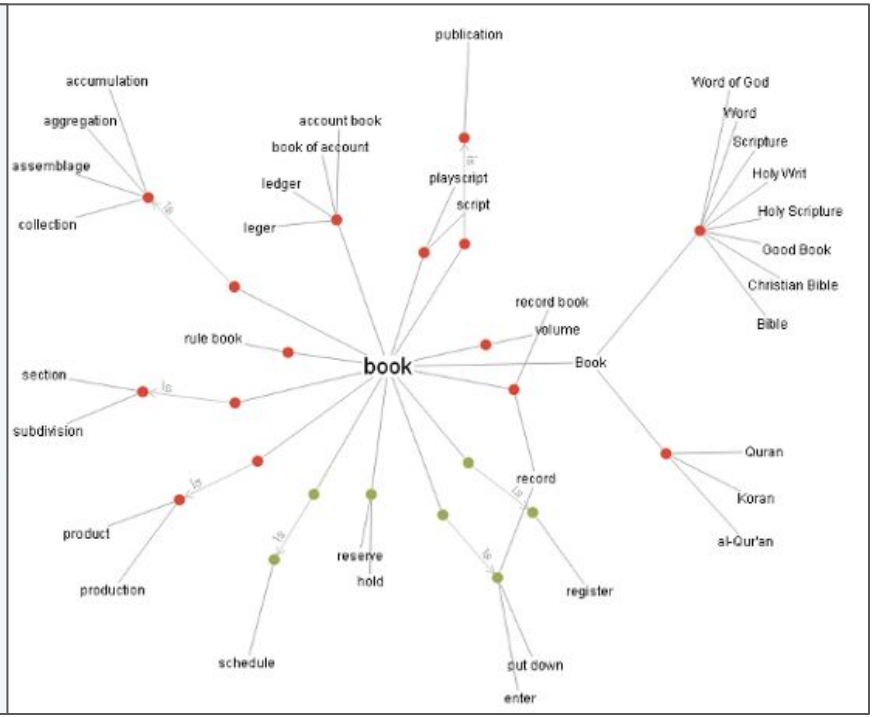
Word Pair Alignment



BiVert

Sense Relation - BabelNet





BiVert

Sense Relation

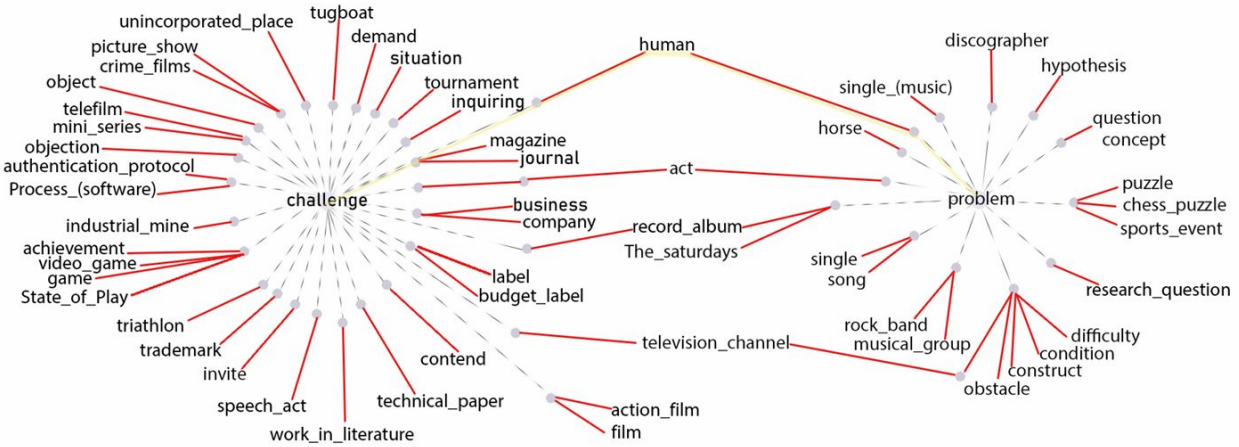


Figure 4.3: Fragment of a semantic graph between the two words *challenge* and *problem*. The hatched grey edges connect roots to their senses, and the red edges represent hypernym relations between the nodes contents. Shortest path highlighted.

BiVert

Sense Relation

$$w(a,b) = \frac{w\left(a \rightarrow_r b\right) + w\left(b \rightarrow_{r^{-1}} a\right)}{2d}$$

$$w\left(x \rightarrow_r y\right) = \max_r - \frac{\max_r - \min_r}{n_r(X)}$$

where $\rightarrow_r$ is a relation of type r and r^{-1} is its inverse; d is the depth of the deeper of the two nodes; $\max_r$ and $\min_r$ are the maximum and minimum weights possible for a relation of type r; and $n_r(X)$ is the number of relations of type r leaving node X. The final graph score $S(a,b)$ from root a to b is given by the normalized sum of the edge weights along the path between them:

$$S(a,b) = 2 \times \left(0.5 - \frac{1}{\sum_{e \in P(a \rightsquigarrow b)} w(e)}\right)$$

(By Michael Sussna,1993)

BiVert

Final Score

1. Same = 0
2. Extra = $1/len(s)$
3. Missing = $1/len(s)$
4. Stop Words = $1/len(s)$
5. Inflection = cosine similarity
6. Derivation = cosine similarity
7. Sense = $S(a,b) = 2 \times \left(0.5 - \frac{1}{\sum_{e \in P(a \rightsquigarrow b)} w(e)}\right)$

$$score_t = \sum_{(a,b) \in t} f_t(a,b).$$

Where t is a category in set:
{Same, Inflection, Extra, Derivation, Missing, Sense, Stop Words}

$$bivert\_similarity(s, s') = \sum_{t \in Types} \alpha_t \times score_t$$

Experiments and Results

- WMT Metrics Task 2021 (train), 2022 (test)
- State-of-the-art MT system: Marian MT
- Gradient Boosting Regression

	Train	Predict
English-German	19,501	19,725
English-Russian	12,000	19,725
Chinese-English	16,124	28,124

	Extra	Missing	Stopword	Inflection	Derivation	Sense
English-German	0.121	0.134	0.188	0.101	0.092	0.360
English-Russian	0.112	0.164	0.196	0.087	0.063	0.375
Chinese-English	0.172	0.203	0.126	0.000	0.000	0.497

Experiments and Results

	System	Human	BiVert
English → German	bleu_bestmbr	9.615	0.614
	bleurt_bestmbr	9.555	0.609
	comet_bestmbr	9.567	0.627
	JDExploreAcademy	9.581	0.620
	Lan-Bridge	9.435	0.608
	M2M100_1.2B-B4	8.872	0.598
	Online-A	9.514	0.608
	Online-B	9.585	0.626
	Online-G	9.510	0.605
	Online-W	9.684	0.624
	Online-Y	9.480	0.615
	OpenNMT	9.329	0.603
	PROMT	9.297	0.596
	QUARTZ_TuneReranking	9.462	0.622
	refB	9.634	0.614
English → Russian	bleu_bestmbr	9.715	0.296
	comet_bestmbr	9.677	0.286
	eTranslation	9.417	0.295
	HuaweiTSC	9.476	0.292
	JDExploreAcademy	9.679	0.279
	Lan-Bridge	9.639	0.236
	M2M100_1.2B-B4	9.298	0.272
	Online-A	9.561	0.292
	Online-B	9.701	0.310
	Online-G	9.687	0.333
	Online-W	9.789	0.320
	Online-Y	9.608	0.263
	PROMT	9.548	0.296
	QUARTZ_TuneReranking	9.375	0.310
	SRPOL	9.434	0.305
Chinese → English	AISP-SJTU	9.682	0.443
	bleu_bestmbr	9.701	0.435
	bleurt_bestmbr	9.749	0.452
	comet_bestmbr	9.714	0.459
	HuaweiTSC	9.692	0.443
	JDExploreAcademy	9.718	0.446
	Lan-Bridge	9.753	0.460
	LanguageX	9.727	0.434

Language pair Human Translation Included	**eng-deu** yes	**eng-deu** no	**eng-rus** no	**zho-eng** yes	**zho-eng** no
BERTScore	0.338	0.428	0.811	0.843	0.924
Cross-QE	0.643	0.661	**0.806**	**0.817**	0.870
COMETKiwi	0.592	0.674	0.763	0.795	0.866
MS-COMET-QE-22	0.417	0.539	0.672	0.799	**0.897**
UniTE-src	0.509	0.509	0.779	0.791	0.874
MATESE-QE	0.363	0.337	0.637	0.741	0.767
COMET-QE	0.480	0.502	0.468	0.544	0.569
KG-BERTScore	0.369	0.400	0.612	0.617	0.743
HWTSC-TLM	0.311	0.428	0.597	0.368	0.460
HWTSC-Teacher-Sim	0.290	0.385	0.675	0.294	0.356
BiVert	**0.694**	**0.703**	0.657	0.376	0.239

Table 5.4: System-level Pearson correlation between human scores and BiVert scores, compared to other evaluation metrics. "Human Translation Included" refers to one other system, refB which may be included or excluded from the total correlation calculation. See system-level scores in Table 5.2. Highest reference-free scores are **bolded**.

Future Work

- Idioms and phrases
- Language categories
- Switching between system roles


GRAZAS
DANKIE
QUJAN
THANK YOU
QUJANARSUAQ
TODA
ASANTE
TAKK
MERCI
XIE XIE
THANK YOU
DANKIE
MITELA
THANKS
TAKK
ASANTE
GRAZAS
MAHALO
THANK YOU
GRACIAS
OBRIGADO
APACHE
TODA
SPASIBO
MERCI
OBRIGADA
THANK YOU
GRACIAS
OBRIGADO
MUCHAS GRACIAS
YEWO
DANKIE
THANK YOU
SUKRIA
TODA
TACK
TAKK
PAXMAT
СПАСИБО
TESEKKÜR EDERIM
THANK YOU
SPASIBO
MITELA
TÄNAN
KIITOS
ДЯКУЮ
TAKK
DANKE
ДЗЯКУЙ
HVALA
PAXMET!
多謝
GRAZIE
DZIEKUJE
DEKUJI
ЕНАРИСТО
TESEKKÜR EDERIM
THANKS
KIITOS
MULTUMESC
MERCI
唔該
DO JEH
SPAS
XIE XIE
FOOLE TUNK
THANK YOU
CHOKRAN
SHUKRAN
SPAS
DANKIE
ASANTE
WILIWNI
BAIE DANKIE
NGIYABONGA
FAKAAUE
NGIYABONGA
HVALA
SUKRIA
SUKRIA
ASANTE
KOP KUN
SPASIB
AITEH
TAK
GRAZIE
SUKRIA
XIE XIE
TAK
TODA
THANK YOU
多謝
TA
SALAMAT PO