PolEval Machine Translation Task

Results and Summary

Questions? : kwolk@pjatk.edu.pl

Translation directions

- English to Polish
- Russian to Polish (low resourced)
- Polish to Russian (low resourced)

- Using any machine translation technology or automatic pre- and post-processing.

Training corpora

- In-domain data – multidisciplinary lecture transcriptions.

	no. of segments		no. of unique tokens			
	TEST	TRAIN	TEST		TRAIN	
			INPUT	OUTPUT	INPUT	OUTPUT
EN to PL	10,000	129,254	9,834	16,978	49,324	100,119
PL to RU	3,000	20,000	6,519	7,249	31,534	32,491
RU to PL	3,000	20,000	6,640	6,385	32,491	31,534

Table 1: Task 4 corpora statistics.

- Permissible out of domain data. – any data from Opus project.

Evaluation

- BLEU
- NIST
- TER
- METEOR (without language adaptation)

Baseline system results

- Baseline systems were trained with ModernMT – basic settings (neural engline and BPE)
- Also compared to Google Translate – note this system is not constrained.

- For EN to PL translation ModernMT obtained 16.29 BLEU points whereas Google engine scored 16.83. For PL to RU we obtained 12.71 versus 15.78 of the Google, in RU to PL the scores were 11.45 and 13.54 respectively

Winners - Marcin Chochowski, Paweł Przybysz Samsung R&D

- The competition winner team was from National Information Processing Institute. They proposed translation solutions to all three translation tasks using only in-domain data. Those systems were better than in-domain baseline systems and obviously worse than system using additionally out of domain data. The best system in English to Polish task was prepared by the Samsung research team. The best scoring systems were neural-based and utilized the Transformer architecture. In the results we can observe big disproportion in scores between rule-based (SIMPLE_SYSTEMS) and neural systems (DeepIf and SRPOL). The results for EN to PL task are given in the Table 2, for PL to RU in the Table 3 and for RU to PL in the Table 4. Please note that Google results cannot be compared directly. Google Translate was trained with bigger and unknown amount of data.

PolEval competition results

Table 2: EN-PL Results

System name	BLEU	NIST	TER	METEOR
SRPOL	28.23	6.60	62.13	47.53
Google Translate	16.83			
ModernMT	16.29			
ModernMT (in-domain)	14.42			
DeepIf (in-domain)	4.92	2.27	86.56	21.74
SIMPLE_SYSTEMS	0.94	1.12	97.94	9.81

Table 3: PL-RU Results

System name	BLEU	NIST	TER	METEOR
Google Translate	15.78			
ModernMT	12.71			
DeepIf (in-domain)	5.38	2.53	83.02	53.54
SIMPLE_SYSTEMS	0.69	0.85	102.75	41.06

Table 4: RU-PL Results

System name	BLEU	NIST	TER	METEOR
Google Translate	13.54			
ModernMT	11.45			
ModernMT (in-domain)	5.73			
DeepIf (in-domain)	5.51	2.97	85.27	24.08
SIMPLE_SYSTEMS	0.57	1.29	109.43	8.35

Thank you for the participation.

- Congratulations to all of the competitors.