

Findings of the WMT 2021 Shared Tasks in Unsupervised MT and Very Low Resource Supervised MT

Jindřich Libovický and Alexander Fraser
{libovicky,fraser}@cis.lmu.de

Center of Information and Language Processing
LMU Munich

Presented online at WMT 2021, November 11, 2021

Task descriptions

1 Task descriptions

2 Evaluation

3 Results

Supervised MT: German ↔ Upper Sorbian

Translation between **German** and **Upper Sorbian**

- Minority language spoken in Saxony, **30k speakers**
- Slavic language, related to Czech (and Polish)
- 147k parallel sentences, 700k monolingual
very likely all that exists digitally

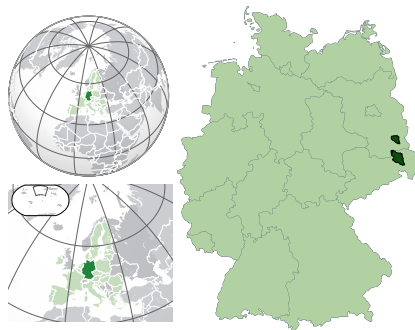


Image from Wikipedia
https://en.wikipedia.org/wiki/Sorbian_languages
<https://en.wikipedia.org/wiki/Germany>

Unsupervised MT: German ↔ Lower Sorbian

Translation between **German** and **Lower Sorbian**

- Minority language spoken in Brandenburg, **7k speakers**
- Slavic language, similar to Upper Sorbian, more similar to Polish than Czech
- No parallel data exist, except for our development and test data

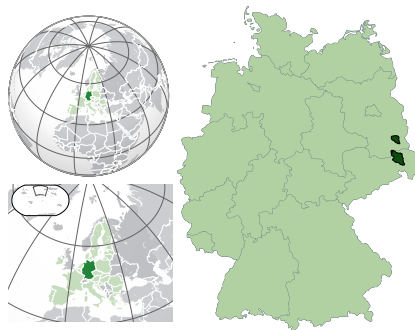


Image from Wikipedia
https://en.wikipedia.org/wiki/Sorbian_languages
<https://en.wikipedia.org/wiki/Germany>

Data come from the local communities:



Witaj Language Center
<https://www.witaj-sprachzentrum.de>

Sorbian Institue
<https://www.serbski-institut.de>

- Upper Sorbian: 147k parallel (87k more than last year), 340k monolingual
 - Lower Sorbian: 145k monolingual only
 - All participants used additional German-Czech data
- + validation, devtest and blindtest data for each of the task

Supervised MT: Russian ↔ Chuvash

Translation between **Russian** and **Chuvash**

- Minority language spoken in south-west Russia,
1M speakers
- Turkic language, Cyrillic script
- Relatively isolated in the Turkic family



Image from Wikipedia
<https://en.wikipedia.org/wiki/Chuvashia>
<https://en.wikipedia.org/wiki/Russia>

Data provided by the **Chuvash Language Laboratory**

<https://en.corpus.chv.su> <https://en.chuvash.org>

- 714k parallel sentences
- 5.6M monolingual sentences

Evaluation

1 Task descriptions

2 Evaluation

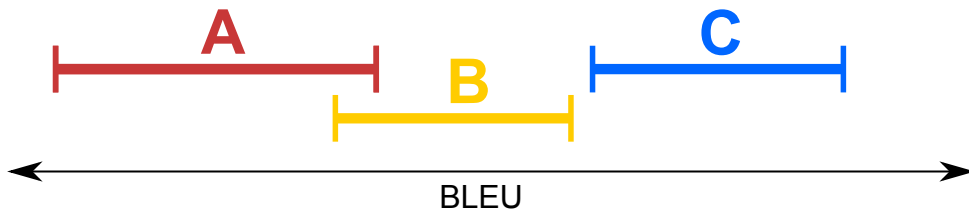
3 Results

Combine **multiple metrics** and compute **confidence intervals** using bootstrap resampling.

- **BLEU** and **chrF** as they are in SacreBLEU
- **BERTScore**
 - XLM-R for German and Russian
 - mBERT for Chuvash
 - Skipped for Upper and Lower Sorbian

Awarding Points

A system gets one point for **each** significantly **outperformed** **system** in **each** metric.



A and **B**: Confidence intervals overlap → both get **0 points**.

C does not overlap with any → gets **2 points**.

Results

1 Task descriptions

2 Evaluation

3 Results

German ↔ Upper Sorbian

German → Upper Sorbian

Team	BLEU	chrF	BERTScore	Points
NRC-CNRC	66.3 3	83.7 3	—	6
NoahNMT	65.9 3	83.3 3	—	6
LMU	63.3 1	81.9 2	—	3
CFILT	60.2 0	79.6 0	—	0
IICT-Yverdon	61.6 0	80.6 0	—	0

Upper Sorbian → German

Team	BLEU	chrF	BERTScore	Points
NRC-CNRC	67.3 3	83.6 3	.981 4	10
NoahNMT	67.7 3	83.4 3	.981 3	9
LMU	64.3 2	81.9 2	.979 2	6
IICT-Yverdon	61.4 0	80.2 0	.976 1	1
CFILT	60.1 0	79.2 0	.975 0	0

German ↔ Upper Sorbian: Comments

- Typical approach: pre-trained on German-Czech, finetune on Sorbian, iterate back-translation
- Two winners, different strengths:
 - **NoahNMT** used large data
 - **NRC-CNRC** smaller data, but use many tricks (clever BT filtering, language tags, BPE Dropout)
- LMU Munich and IICT-Yverndron did basically the same, but neither did the tricks nor used large data

German → Lower Sorbian

Team	BLEU	chrF	BERTScore	Points
NRC-CNRC	29.9 3	59.9 3	—	6
LMU	27.5 3	57.9 3	—	6
CL_RUG	24.1 2	54.2 2	—	4
IICT-Yverdon	8.0 0	32.1 1	—	1
CFILT	6.4 0	29.0 0	—	0

Lower Sorbian → German

Team	BLEU	chrF	BERTScore	Points
NRC-CNRC	33.5 1	63.8 1	.953 2	4
CL_RUG	32.4 1	62.2 1	.953 2	4
LMU	33.3 1	62.0 1	.952 1	3
CFILT	5.9 0	31.6 0	.884 0	0

- Typically **initialize with Upper Sorbian** and iterate back-translation
- Winner again: **NRC-CNRC**

Russian ↔ Chuvash

Russian → Chuvash

Team	BLEU	chrF	BERTScore	Points
NoahNMT	22.1 0	51.3 0	.857 1	1
LMU	20.9 0	50.1 0	.856 0	0

Chuvash → Russian

Team	BLEU	chrF	BERTScore	Points
NoahNMT	23.4 0	47.6 0	.944 1	1
LMU	22.0 0	46.3 0	.942 0	0

- Winner: NoahNMT – initialize with **English-Russian**
- LMU Munich initialize with Kazakh-Russian, smaller data

Thank you for participating in the task!

Thanks to our coorganizers:

- **Hauke Bartels, Marcin Szczepanski** – Sorbian Institute
- **Olaf Langner** – Witaj Sprachzentrum
- **Alexander Antonov** – Chuvash Language Laboratory

Many thanks also to **Tom Kocmi** and **Christian Federmann**
from Microsoft for running Ocelot.