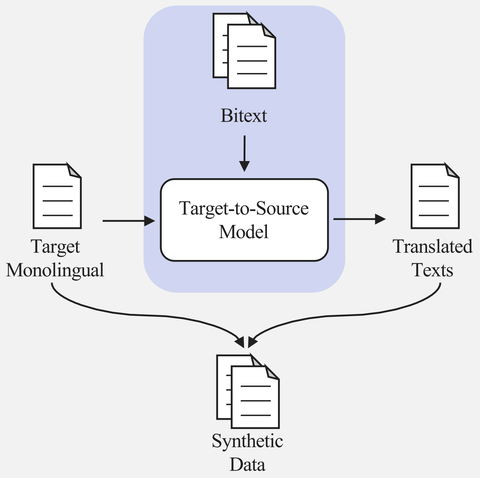
Tohoku-AIP-NTT at WMT 2020 News Translation Task

Shun Kiyono[1,2] Takumi Ito[2,1] Ryuto Konno[2,1] Makoto Morishita[3,2] Jun Suzuki[2,1]

1 AIP 2 TOHOKU UNIVERSITY 3 NTT

Overview of Our System

① Synthetic Data Generation



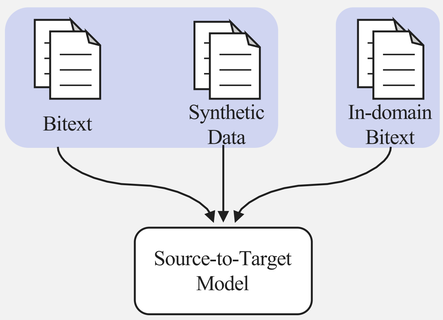
Base Model: Transformer

- Transformer-based Encoder-Decoder model
- **More layers** (6 → 9) and **bigger feed-forward network** (4,096 → 8,192)

Large-scale Back-translation (BT)

- Back-translated NewsCrawl for En↔De and CommonCrawl for En↔Ja
- **200~300M synthetic data** for every language direction

② Training



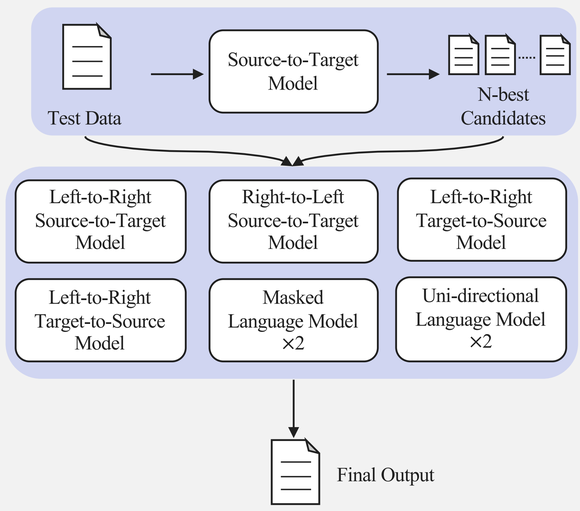
Tagged Back-translation

- **Special tag <BT> is prepended** to every synthetic data
- Use bitext and synthetic data in 1:1 ratio

Fine-tuning

- Systems are fine-tuned using **in-domain news corpus** (e.g., newstest20xx)

③ Decoding



Ensemble w/ Right-to-Left Models

- Ensemble of 4 left-to-right (L2R) models and 4 right-to-left (R2L) models
- Generate N-best candidates for final output

Reranking

- N-best candidates are reranked with external generative models
- Minimum Error Rate Training (MERT)-like module is applied to **maximize BLEU score on development set**

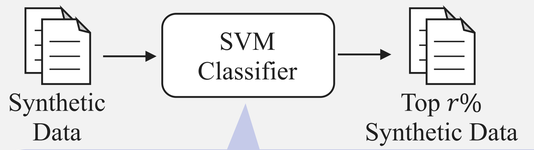
Experimental Results on Development Data

ID	Setting	1st place En→De	1st place De→En	4th place En→Ja	2nd place Ja→En
(a)	Base Model + Synthetic Data	42.7	42.5	22.0	23.9
(b)	(a) + fine-tuning	44.9	42.3	23.1	24.4
(c)	(b) x 4 + R2L Models	45.4	43.6	24.2	25.9
(d)	(c) + reranking	45.7	43.8	24.9	26.2
-	WMT'19 Best	44.9	42.8	-	-

- Participated in 4 language directions: **En→De, De→En, En→Ja, Ja→En**
- Dataset: newstest2019 for En↔De and newsdev2020 for En↔Ja
- Each technique consistently improved the BLEU score
- Strong results compared to top-performing system from last year

Takeaway from Negative Results

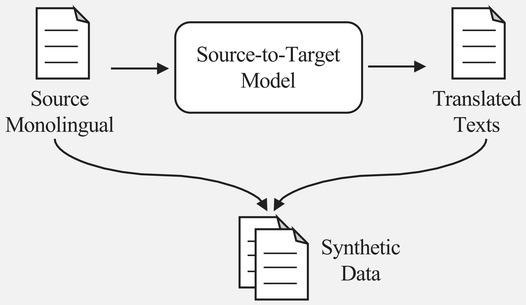
① Synthetic Data Filtering: It Didn't Work



- Pointwise HSIC score [Yokoi et al. 2018] w/ fasttext
- Target-to-Source translation score

Synthetic Data: r%	En→De
100	42.0
50	42.3
33	42.2
25	42.4

② Forward Translation (FT): Improvements were Marginal



Setting	En→De
Base Model	42.2
Base Model + BT	42.0
Base Model + FT	42.1
Base Model + BT + FT	42.4