# Learning How to Translate North Korean through South Korean

Hwichan Kim, Sangwhan Moon, Naoaki Okazaki, Mamoru Komachi
kim-hwichan@ed.tmu.ac.jp, sangwhan@iki.fi, okazaki@c.titech.ac.jp, komachi@tmu.ac.jp

## Introduction

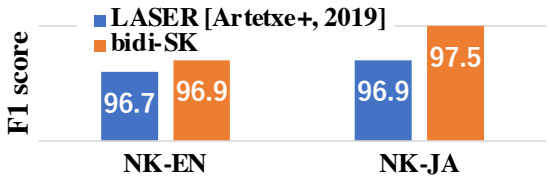
- There are some differences between South Korean (SK) and North Korean (NK) in their vocabulary and orthography.

| | SK | NK | English |
|---|---|---|---|
| Synonym | 옥수수 | 강냉이 | Corn |
| Orthography | 농구 | 롱구 | Basketball |

- SK NMT systems cannot translate North Korean inputs well because they were not trained by NK data.
- We create NK-English (EN) and NK-Japanese (JA) parallel corpora using North Korean articles.
  - Articles and sentences are not aligned between each language.
  - We manually and automatically align them.
- Our experiments show that our NK data can significantly enhance the translation quality when used in conjunction with SK datasets.

## Parallel Data Construction

- Evaluation data construction
  - Manually align all articles and randomly extracted 1,000 sentences.
- Comparison of automatic alignment methods in NK alignment




  - The bidi-SK, a method using SK NMT model, achieves the highest scores.
- Training data construction
  - Automatically align the sentences using the bidi-SK.

Summary of our NK corpora

| | dev | test | train |
|---|---|---|---|
| NK-EN | 500 | 500 | 4,109 |
| NK-JA | 500 | 500 | 3,739 |

## North Korean Translation

- The fine-tuned SK model by NK data (SK→NK) significantly outperforms the SK and NK models in BLEU score.

| | NK-EN translation | |
|---|---|---|
| model | dev | test |
| SK | 11.4 ± .17 | 11.9 ± .21 |
| NK | 21.4 ± .15 | 20.4 ± .09 |
| SK→NK | **36.7 ± .12** | **35.6 ± .12** |

- The SK→NK can translate NK specific words such as "**울라지보스또크** (ul-la-ji-bo-seu-tto-keu)" which means *Vladivostok*.

| | |
|---|---|
| NK | … 로씨야련방 **울라지보스또크** 시에 … |
| Ref. | … **Vladivostok**, the Russian Federation … |
| SK | … at the city of **Ulazibosto** … |
| SK→NK | … **Vladivostok**, the Russian Federation … |
| NAVER | … **Ulajibos Tok** City, a training room for RoC. |

- Our paper presents more details about the alignment and translation experiments.