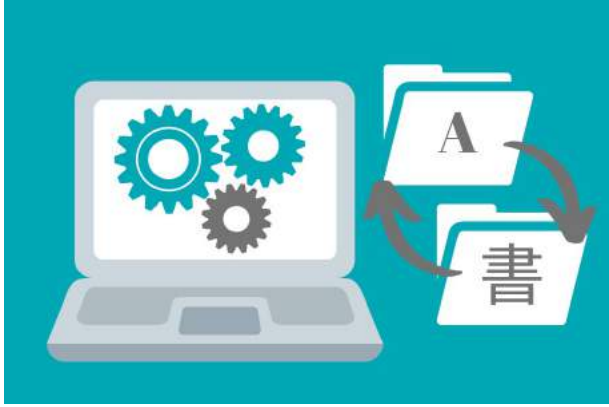
Kumamoto University



Pivot Approaches in Machine Translation for Low-Resource Languages

A thesis defense presentation by Sari Dewi Budiwati
Supervised by Prof. Masayoshi Aritsugi

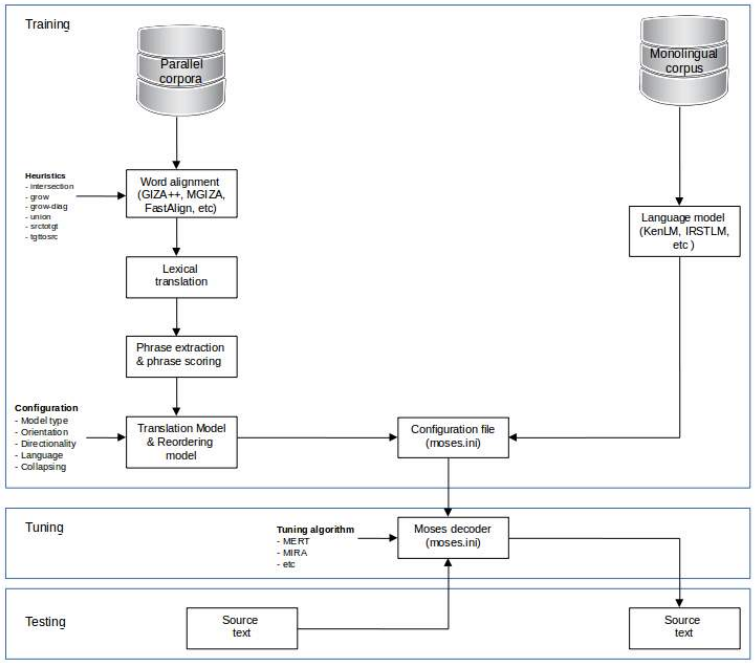
Outline

Kumamoto University

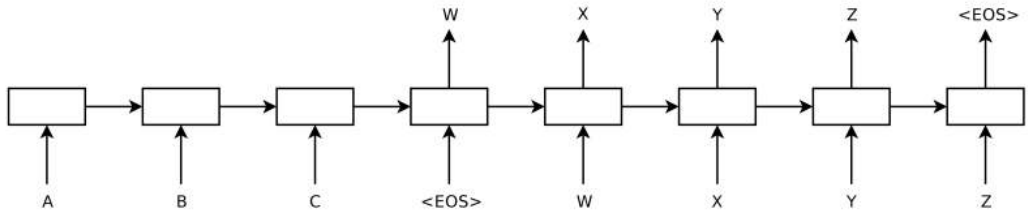
- Background
- Basic theory and related work
- Preliminary work
  - DBMS-KU Interpolation for WMT19 News Translation Task.
  - Multiple pivots in Statistical Machine Translation for Low-Resource Languages
- Proposed approach
  - Word reordering in multiple pivots for Japanese to Indonesian (Ja-Id)
  - Phrase Table Combination based on Symmetrization of Word Alignment for Low-Resource Language
- Conclusion and future work

Background

- Machine Translation (MT) is a task of automatically translate a text from one natural language, e.g., English, to another language, e.g., Japanese
- MT Model



a. Statistical Machine Translation (SMT)



b. Neural Machine Translation (NMT)

This picture is a basic of encoder-decoder model of NMT (Sutskever et al., 2014)

Background

- Both models need parallel corpora and monolingual corpora

Example of parallel corpora, taken from ALT dataset.

File 1: English corpus

Line 1 Italy have defeated Portugal 31-5 in Pool C of the 2007 Rugby World Cup at Parc des Princes, Paris, France.
Line 2 Andrea Masi opened the scoring in the fourth minute with a try for Italy.
Line 3 Despite controlling the game for much of the first half, Italy could not score any other tries before the interval but David Bortolussi kicked three penalties to extend their lead.
Line 4 Portugal never gave up and David Penalva scored a try in the 33rd minute, providing their only points of the match.
Line 5 Italy led 16-5 at half time but were matched by Portugal for much of the second half.

File 2: Indonesian corpus

Line 1 Italia berhasil mengalahkan Portugal 31-5 di grup C dalam Piala Dunia Rugby 2007 di Parc des Princes, Paris, Perancis.
Line 2 Andrea Masi membuka skor di menit keempat dengan satu try untuk Italia.
Line 3 Meskipun mengontrol jalannya pertandingan untuk sebagian besar dari setengah permainan, Italia tidak dapat menambah skor melalui try lainnya sebelum istirahat, namun David Bortolussi menendang tiga penalti untuk memperlebar keunggulan mereka.
Line 4 Portugal tidak pernah menyerah dan David Penalva mencetak skor melalui try di menit ke-33 yang menjadi skor akhir mereka.
Line 5 Italia memimpin 16-5 di babak pertama namun mendapat perlawanan dari portugal di babak kedua.

File 3: Japanese corpus

Line 1 フランスのパリ、パルク・デ・プランスで行われた2007年ラグビーワールドカップのプールCで、イタリアは31対5でポルトガルを下した。
Line 2 アンドレア・マージが開始4分後のトライでイタリアにとって最初の得点を入れた。
Line 3 イタリアは試合前半のほとんどをコントロールしていながら、その後インターバルまでのトライでは得点できなかったが、デビッド・ボルトルッシがペナルティキックで3点を入れリードを広げた。
Line 4 ポルトガルはあきらめることなく、デビッド・ペナルバが33分でトライを決め、試合中唯一の得点をあげた。
Line 5 イタリアはハーフタイムの時点で16対5とリードしていたが、後半ではポルトガルと互角になった。

Building parallel corpora is (Wolk and Wolk, 2018):
- Challenging
- Time consuming
- Expensive
- Needs expert

Background

- Both models need parallel corpora and monolingual corpora

```
1       $1.5 million is needed this year alone to purchase medical equipment for patient care, each SAH program and department with its own specific equipment needs.
2       $2700/m inclusive - 12x14 bay door.
3       "36-Hours" DVD Now Available for PurchaseGovernor Bruce Dinwiddy offered his strong support, saying: "I hope the community will continue to support the Recovery Fund through the challenging times
4       $6 plus $5 parking (per vehicle).
5       A 15-year-old who lost his wallet on Iceland is now the subject of a search by the man who found it.
6       A 1 Territory government would start, again using world's best practice to deal with intractable, damaged and lost youth to turn them back into productive citizens and repay their victims for the
7       A $20,000 Commercial/Industrial Property Study, to be funded by the Eastern Ontario Development fund, was brought before Council for their approval.
8       A 2016 Police crime report says violent crimes, which take various forms such as carjackings, and criminal attacks with some turning fatal were on the rise in the country.
9       A 42-year-old man has been arrested for killing his coworker.
10      "AAR decisions are specific to the case, but they do have some precedence value," the official said.
11      A: Assuming the debaters of is equal skill.
12      A Baby Ratty cameo, Justin Bieber ringtones, surf missions and a trip to the big vibe small town Morant Bay to get some ingredients for a PROPER ITAL.
13      A badlands milkvetch that he first identified bears his name.
14      A bee delivers a stop signal by giving another bee a brief, vibrational pulse, usually through a head-butt.
15      A Bendals man was fined $45,000 on drug charges in the St. John's Magistrate's Court on Monday.
```

English monolingual corpora of newscrawl, taken from Leipzig corpora collection.

- Building a monolingual corpora
  - Relatively easy because it only contains one language.

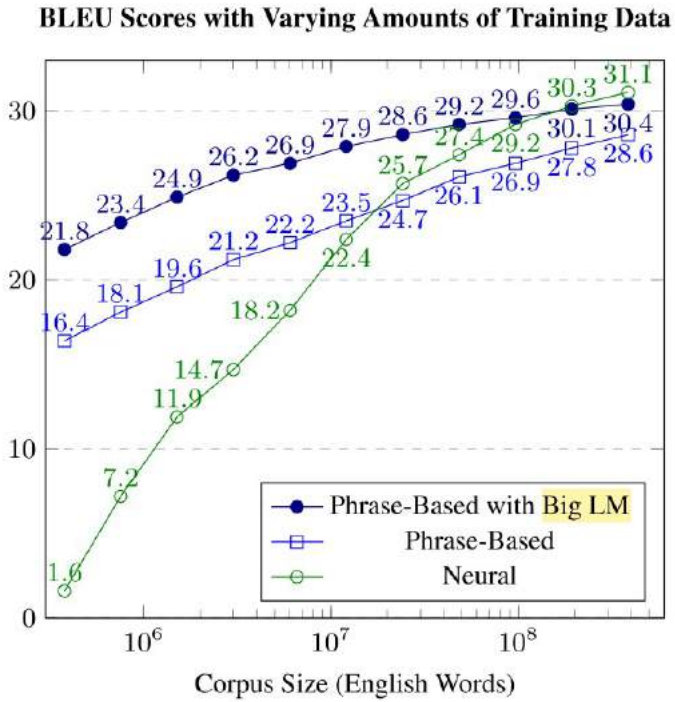
Background

Kumamoto University

- Languages in MT are divided into
  - Rich-resource language, ex: English, Spanish, German, Japanese, Chinese.
  - Low-resource language, ex: African languages, most Asian languages (Indonesian, Myanmar, etc)
    - Zero-shot resource language
- Low-resource language
  - Limited corpora (parallel or monolingual)
  - Limited linguistic tools

Background

- NMT still has to overcome various challenges (Koehn and Knowles, 2017)
  - Performance of out-of-domain
  - Amount of training data
    - Worse quality in low-resource, but better performance in high-resource settings



Koehn and Knowles (2017)

Background

Kumamoto University

- There are two strategies to achieve high-quality in low-resource (Trieu, 2017)
  - Building parallel corpora
  - Utilizing existing corpora
- We focused on
  - Two low-resource language pairs : Kk-En and Ja-Id
    - Kk-En avalaible parallel corpora : 953,240
    - Ja-Id available parallel corpora : 1,468,155
  - SMT model
  - Utilizing existing corpora, i.e., pivot approaches
  - Additional experiments: NMT

Objectives

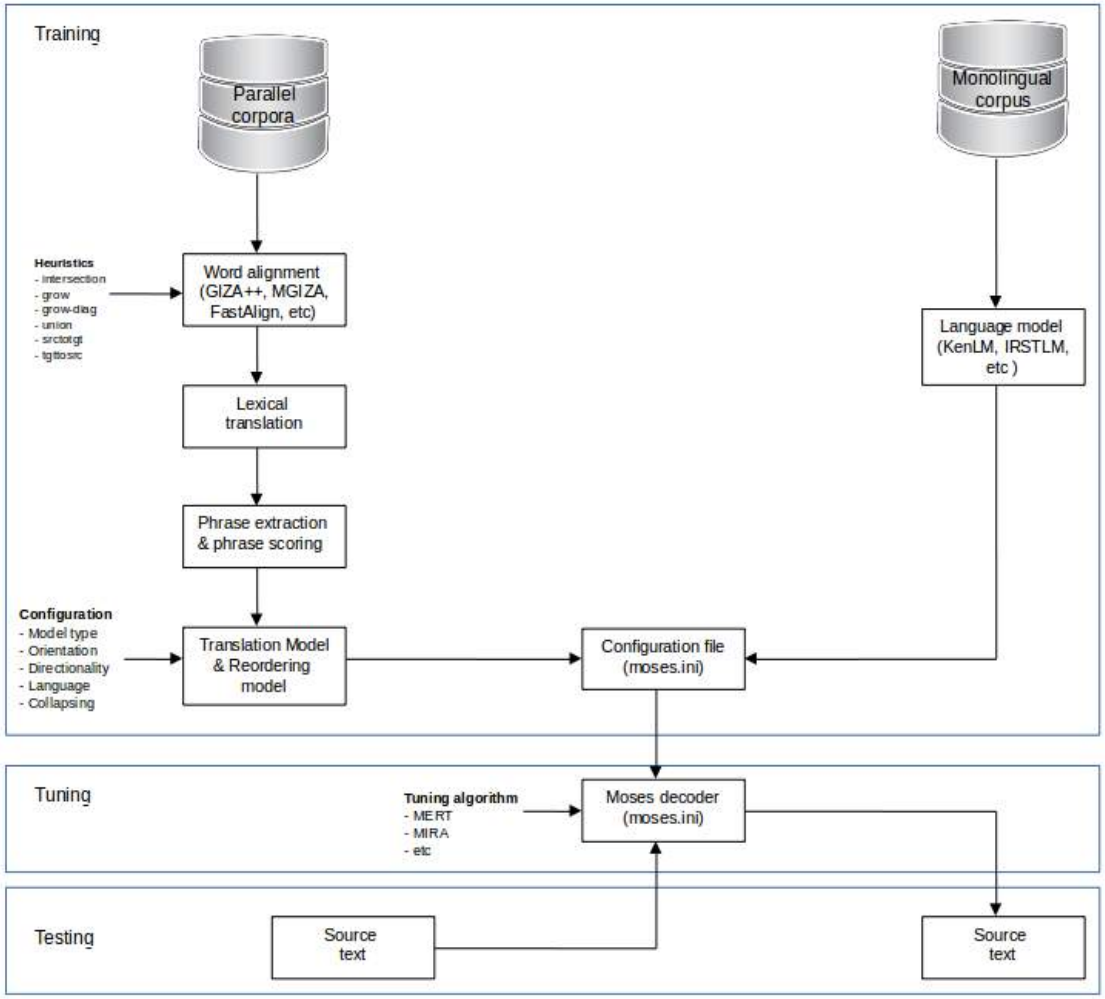
Kumamoto University

- To apply pivot approaches and examine issues in two low-resource language pairs: Kk-En and Ja-Id.
- To propose a technique that could improve the translation quality compare to the direct translation.

Theoretical

Kumamoto University

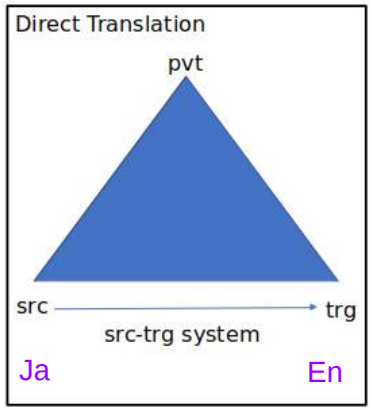
Statistical Machine Translation model



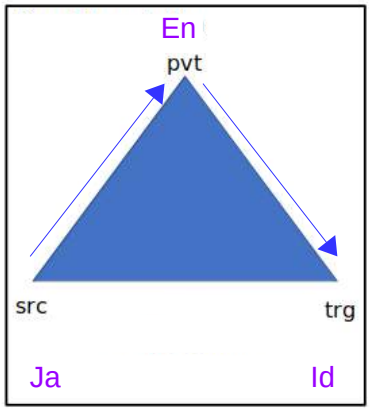
- Components of SMT
  - Parallel corpora
  - Monolingual corpora
  - Translation model (or phrase table)
  - Language model
  - Reordering model
  - Decoder
- SMT process
  - Training
  - Tuning
  - Testing
  - Evaluation

Theoretical

- MT direction
  - Direct translation
    - Between source (src) and target (trg) languages
  - Pivot translation
    - Through intermediate language (pvt): English
    - Non-English as a pivot language could improve translation quality for specific language pairs
      - French to Spanish, using Greek as a pivot (Wu and Wang, 2007)
      - From 420 experiments in Indo-European and Asian languages, 54.8% is preferable using non-English (Paul et al. 2013)



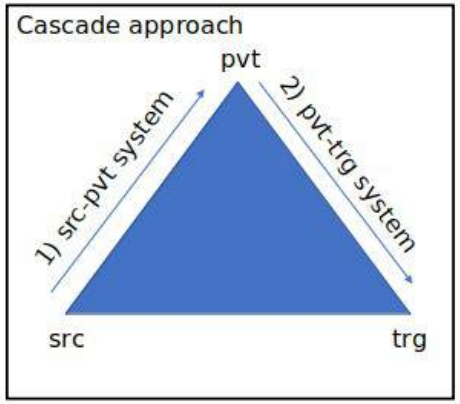
a. Direct translation



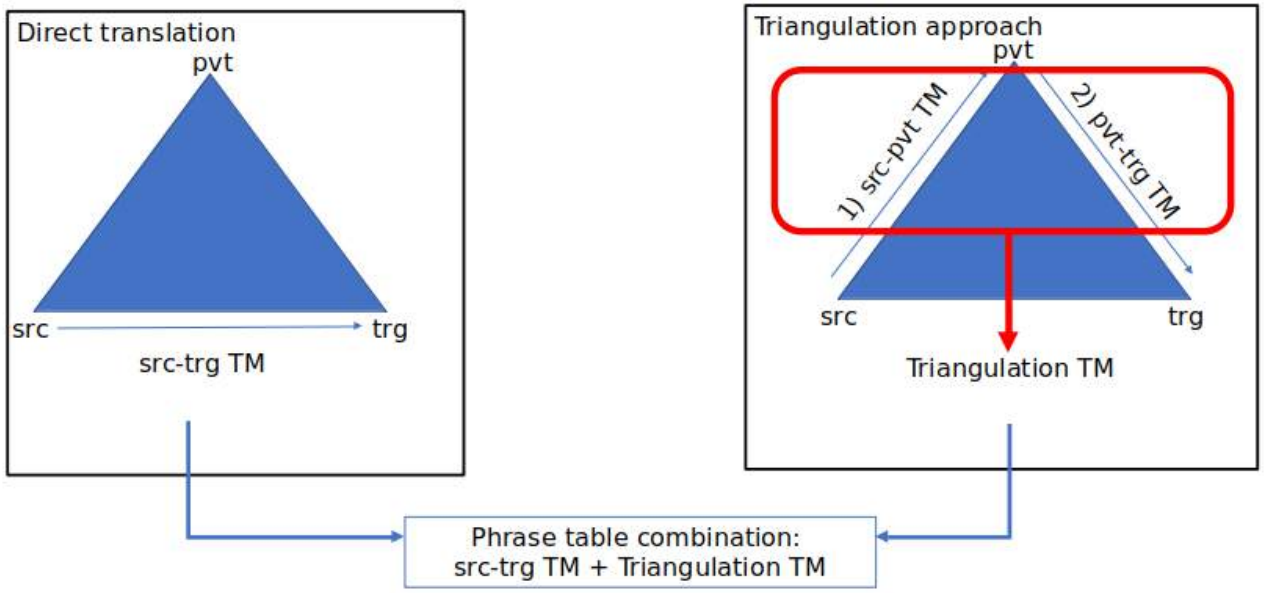
a. Pivot translation

Theoretical

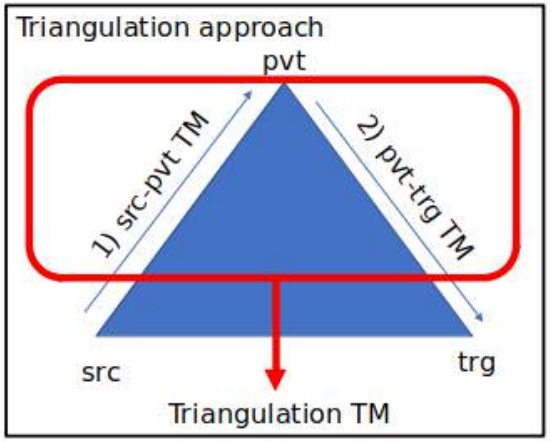
- Pivot approaches
  - Cascade
  - Triangulation
  - Interpolation (phrase table combination)



a. Cascade



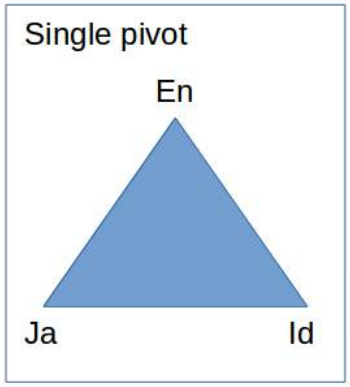
c. Interpolation

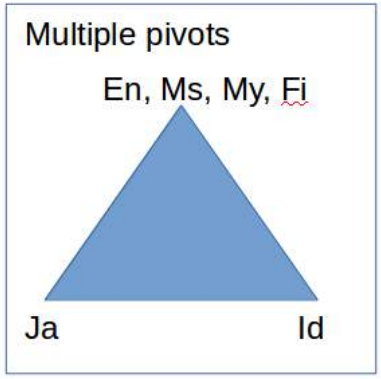


b. Triangulation

Theoretical

- Pivot types
  - Single pivot (Ahmadnia et al., 2017; Trieu and Nguyen, 2017; Dabre et al., 2015, Utiyama and Isahara, 2007, Wu and Wang, 2007)
  - Multiple pivots
    - French to Spanish translation using four languages as multiple pivots: Greek, Portuguese, English, and Finnish (Wu and Wang, 2007).
    - Japanese to Hindi translation using seven languages as multiple pivots: Chinese, Korean, Marathi, Kannada, Telugu, Paite, and Esperanto (Dabre et al, 2015)





Preliminary works

1. DBMS-KU Interpolation for WMT19 News Translation Task. WMT 2019
2. Multiple pivots in Statistical Machine Translation for Low-Resource Languages. The 33rd PACLIC 2019.

Preliminary works

- We explore
  - SMT model
  - Three pivot approaches: Cascade, Triangulation, Interpolation
  - Two pivot types: single and multiple pivots
- We used
  - Two low-resource language pairs
    - Kazakh to English (and vice versa)
    - Japanese to Indonesian (and vice versa)
  - Two datasets
    - Asian Language Treebank (ALT) (Riza et al., 2016)
    - News-Commentary (Barrault et al., 2019)
  - MOSES SMT framework (Koehn et al., 2007)
  - Giza++ word alignment for translation model
  - Ken-LM for language model (Heafield, 2011)
  - MERT for tuning (Och, 2003)
  - BLEU for evaluation score

Preliminary works (1): DBMS-KU Interpolation

- WMT2019 is a first of Kk-En language exploration task. There is no experience system description from previous WMT.
- Previous research
  - Myrzakhmetov and Kozhirbayev (2018)
    - They conducted LM experiments using n-gram, neural LSTM, and sub-word units.
    - They created Kazakh datasets from the web-pages.
    - They showed that neural LSTM outperforms the n-gram model and morphological segmentation has an advantage compared to the word-based models.
    - The performance evaluated using perplexity score.
  - Kartbayev (2015)
    - They described a word alignment by incorporating a morpheme information to the alignment.
    - They constructed three experiments: Baseline (SMT), Morfessor and Rule-base with 30.47, 31.90, and 33.89 of BLEU score, respectively.

Preliminary works (1): DBMS-KU Interpolation

- Previous research
  - Assylbekov and Nurkas (2014)
    - They used the SMT model and a morphological segmentation.
    - They constructed the Kk-En dataset of email messages.
    - They showed a selective morphological segmentation improves the SMT system by 18.74 and 24.05 of BLEU score for small (5,241) and large corpus (27,792), respectively.
  - Kuandykova et al., (2014)
    - They constructed the English-Kazakh dataset of legal text.
    - They used the SMT model to train and evaluate the dataset.
    - They used 3-gram to 5-gram LM orders and obtained 10.03 of BLEU score.

Preliminary works (1): DBMS-KU Interpolation

- We experimented
  - Kazakh to English (and vice versa)
  - Two LM orders: 3-gram and 5-gram
  - Russian as a pivot language
  - Pivot approach: Linear Interpolation (LI)
  - Pivot type: Single pivot
- We build eight systems
  - Four direct translation.
  - Four pivot translation using LI.

Preliminary works (1): DBMS-KU Interpolation

- BLEU score results
  - A higher BLEU score indicates a better system, while lower BLEU scores indicate poor system
  - The interpolation system obtained high BLEU score compare to Baseline.
  - The improvement of the BLEU score from Baseline to Interpolation for Kk-En is 0.1 and 0.5 for 3-gram and 5-gram, respectively.
  - The improvement of the BLEU score from Baseline to Interpolation for En-Kk is 0.1 for 3-gram and 5-gram.
- Finding
  - The use of pivot language in the interpolation system combined with a longer LM model: 5-gram, had a significant influence on the BLEU score.
  - The number of target LM dataset might improve the BLEU score
    - Kk-En LM dataset:532,560
    - En-Kk LM dataset: 9,619

Language Pair	3-gram LM	5-gram LM
KK-EN		
1. Baseline system	2.6	2.9
2. Interpolation system	2.7	3.4
EN-KK		
1. Baseline system	0.8	0.8
2. Interpolation system	0.9	0.9

Preliminary works (1): DBMS-KU Interpolation

- Perplexity score result
  - A lower perplexity score indicates better LM, while high perplexity scores indicate poor LM.
  - For Kk-En: 5-gram of Baseline and Interpolation system.
  - For En-Kk: 5-gram of Baseline, 3-gram of Interpolation system.
- Finding
  - Most of the lowest pp score obtained by the 5-gram LM order
  - A longer LM order and interpolation approach could be a primary option to improve the BLEU score and minimize perplexity score.

Language pair	3-gram LM	5-gram LM
KK-EN		
1. Baseline system	- Incl OOVs: 829.59 - Excl OOVs: 77.79	- Incl OOVs: 617.36 - Excl OOVs: 45.51
2. Interpolation system	- Incl OOVs: 1034.50 - Excl OOVs: 94.72	- Incl OOVs: 762.79 - Excl OOVs: 50.93
EN-KK		
1. Baseline system	- Incl OOVs: 328.940 - Excl OOVs: 103.27	- Incl OOVs: 256.138 - Excl OOVs: 77.185
2. Interpolation system	- Incl OOVs: 256.13 - Excl OOVs: 79.34	- Incl OOVs: 276.85 - Excl OOVs: 85.40

Preliminary works (2): Multiple Pivots in SMT

- We experimented
  - Japanese to Indonesian (and vice versa)
  - One LM orders: 3-gram
  - Pivot approach: Cascade, Triangulation, Interpolation (Linear interpolation, Fillup Interpolation)
  - Pivot type: Single pivot and multiple pivots
  - Pivot languages: English, Myanmar, Malay, Filipino
- We build 132 systems
  - 4 baseline
  - 32 src-pvt and pvt-trg
  - 64 single pivot
  - 32 multiple pivots
- Dataset
  - 8.5K of Asian Language Treebank (ALT) (Riza et al., 2016)

Preliminary works (2): Multiple Pivots in SMT

- BLEU and Perplexity score results of **Ja-Id**
  - Baseline random obtained higher BLEU score, i.e., 12.17 compare to Baseline sequential, i.e, 11.96, however
  - Baseline random obtained higher pp score compare to Baseline sequential

Table 3.5: Ja-Id BLEU score on sequential data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		11.96		
Single pivot system				
JaId (English)	10.89	9.71	11.97	12.07
JaId (Myanmar)	9.37	8.71	11.91	12.27
JaId (Malay)	12.01	8.37	11.71	12.09
JaId (Filipino)	9.95	9.41	12.23	12.19

Table 3.6: Ja-Id BLEU score on random data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		12.17		
Single pivot system				
JaId (English)	10.81	9.10	12.18	12.22
JaId (Myanmar)	9.60	8.60	11.91	12.29
JaId (Malay)	11.81	9.25	12.22	12.05
JaId (Filipino)	9.68	9.62	12.09	11.99

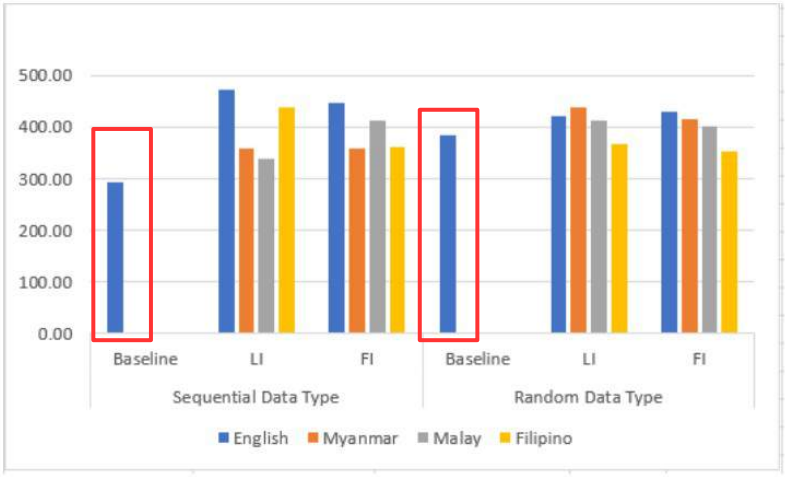


Figure 3.1: Perplexity Score of Ja-Id single pivot for and FI approaches.

Preliminary works (2): Multiple Pivots in SMT

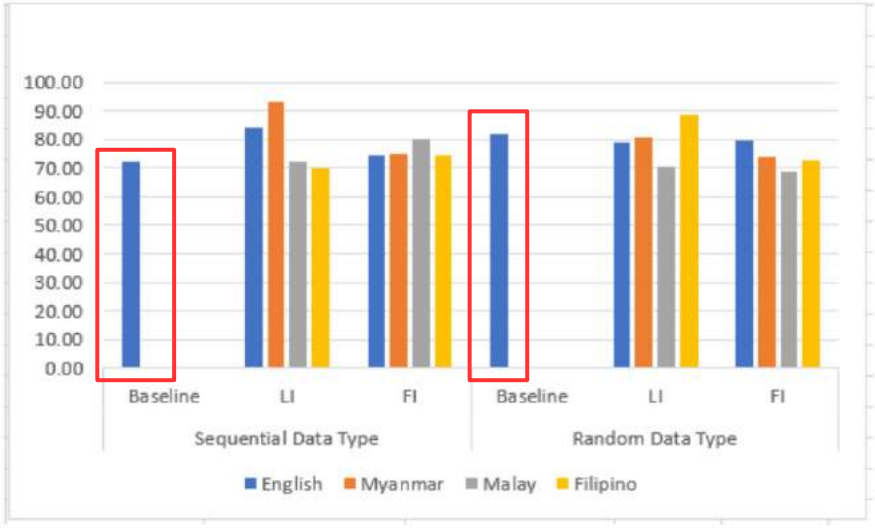
- BLEU and Perplexity score results of **Id-Ja**
  - Baseline random obtained higher BLEU score, i.e., 12.00 compare to Baseline sequential, i.e, 11.00, however
  - Baseline random obtained higher pp score compare to Baseline sequential.

Table 3.7: Id-Ja BLEU score on sequence data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		11.00		
Single pivot system				
JaId (English)	12.07	8.26	12.65	12.05
JaId (Myanmar)	9.97	6.76	10.89	12.4
JaId (Malay)	12.18	6.76	12.2	11.87
JaId (Filipino)	10.36	7.28	12.06	12.2

Table 3.8: Id-Ja BLEU score on random data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		12.00		
Single pivot system				
JaId (English)	7.58	7.96	12.10	11.99
JaId (Myanmar)	10.32	6.51	12.84	12.88
JaId (Malay)	11.13	9.17	12.52	11.82
JaId (Filipino)	10.46	7.97	12.25	12.68



[Figure 3.2: Perplexity Score of Id-Ja single pivot for LI and FI approaches.

Preliminary works (2): Multiple Pivots in SMT

- BLEU and Perplexity score results of **Single pivot of Ja-Id**
  - The worst approach: Triangulation. A potential approach: Interpolation (Linear/Fillup)
  - Pivot language options: My, Fi, Ms
  - Pp scores of interpolation system increased

Table 3.5: Ja-Id BLEU score on sequential data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		11.96		
Single pivot system				
JaId (English)	10.89	9.71	11.97	12.07
JaId (Myanmar)	9.37	8.71	11.91	12.27
JaId (Malay)	12.01	8.37	11.71	12.09
JaId (Filipino)	9.95	9.41	12.23	12.19

Table 3.6: Ja-Id BLEU score on random data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		12.17		
Single pivot system				
JaId (English)	10.81	9.10	12.18	12.22
JaId (Myanmar)	9.60	8.60	11.91	12.29
JaId (Malay)	11.81	9.25	12.22	12.05
JaId (Filipino)	9.68	9.62	12.09	11.99

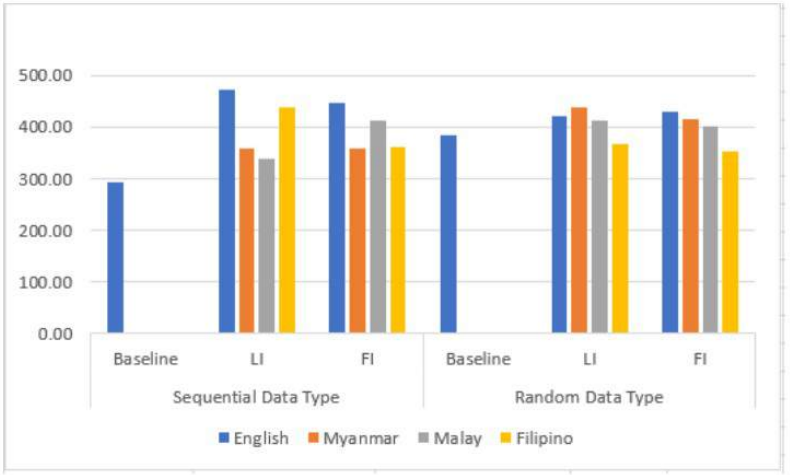


Figure 3.1: Perplexity Score of Ja-Id single pivot for and FI approaches.

Preliminary works (2): Multiple Pivots in SMT

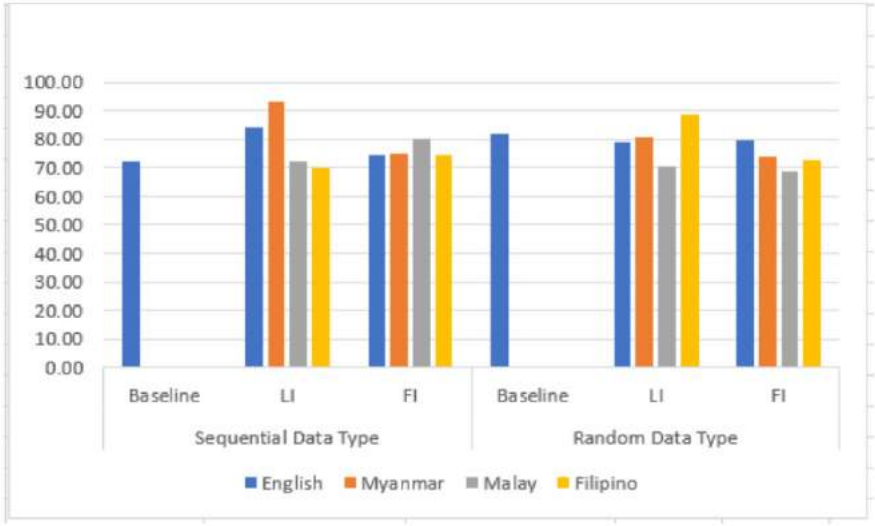
- BLEU and Perplexity score results of **Single pivot Id-Ja**
  - The worst approach: Triangulation. A potential approach: Interpolation (Linear/Fillup)
  - Pivot language options: My, En
  - Pp scores of interpolation system increased.

Table 3.7: Id-Ja BLEU score on sequence data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		11.00		
Single pivot system				
JaId (English)	12.07	8.26	12.65	12.05
JaId (Myanmar)	9.97	6.76	10.89	12.4
JaId (Malay)	12.18	6.76	12.2	11.87
JaId (Filipino)	10.36	7.28	12.06	12.2

Table 3.8: Id-Ja BLEU score on random data type

Systems	Cascade	Triangulation	LI	FI
Direct translation system				
Baseline		12.00		
Single pivot system				
JaId (English)	7.58	7.96	12.10	11.99
JaId (Myanmar)	10.32	6.51	12.84	12.88
JaId (Malay)	11.13	9.17	12.52	11.82
JaId (Filipino)	10.46	7.97	12.25	12.68



[Figure 3.2: Perplexity Score of Id-Ja single pivot for LI and FI approaches.

Preliminary works (2): Multiple Pivots in SMT

- Multiple pivots: combination of four phrase tables using LI and FI of single pivot
  - Without src-trg phrase table: All-LI, All-FI
  - With src-trg phrase table: Base-LI, Base-FI
- The combination of All-LI and All-FI arranged by ascending and descending orders
  - Example: LI results from single pivot, i.e., 11.34 for EnPT, 12.21 for MyPT, 12.11 for MsPT, and 12.15 for FiPT.
  - All-LI ascending order: EnPT, MsPT, FiPT, MyPT,
  - All-LI descending order: MyPT, FiPT, MsPT, and EnPT

Preliminary works (2): Multiple Pivots in SMT

- Multiple pivots experiment results

	Multiple pivots of Ja-Id	Multiple pivots of Id-Ja
Approaches that improve the BLEU score	All-LI, All-FI, Base-LI	All-LI, All-FI, Base-LI, Base-FI
Phrase table orders	Descending	Ascending
Data type	Base-LI Random: +0.23 point	All-FI Sequence: +1.84 point
Perplexity score	Base-LI Random could reduce perplexity score	All-LI, All-FI, Base-LI, Base-FI could reduce perplexity score

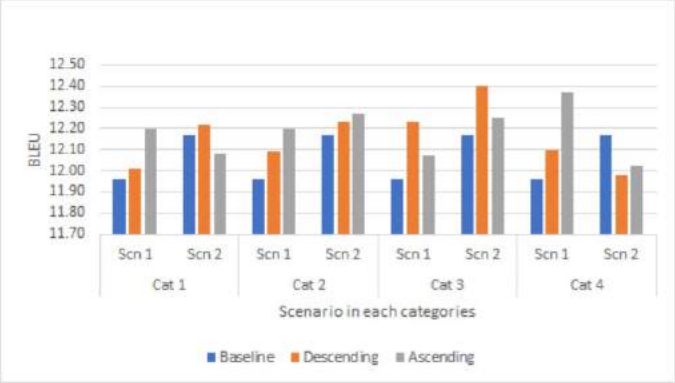


Figure 3.5: BLEU score for Ja-Id in multiple pivots.

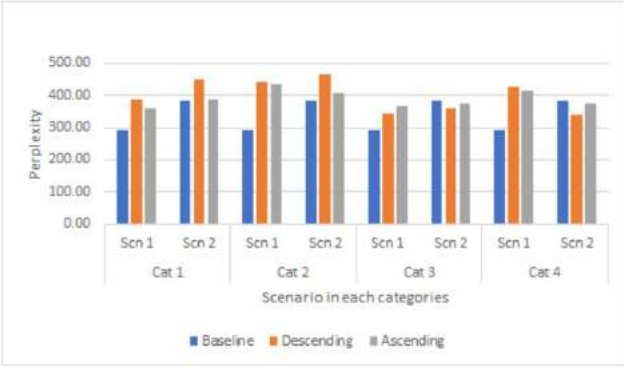


Figure 3.7: Perplexity score for Ja-Id in multiple pivots.

Preliminary works (2): Multiple Pivots in SMT

Table 3.9: Best BLEU score in Baseline, single and multiple pivots for Ja-Id

Scenario No	Baseline	Single Pivot				Multiple Pivots	
		Cascade	Triangulate	Interpolate	Fillup Interpolation	Desc	Asc
Scenario 1	11.96	12.01 (MS)	9.71 (EN)	12.21 (MY)	12.27 (MY)	12.23 (Cat 3)	12.37 (Cat 4)
Scenario 2	12.17	11.81 (MS)	9.62 (FI)	12.22 (MS)	12.29 (MY)	12.40 (Cat 3)	12.27 (Cat 2)

Table 3.10: Best BLEU score in baseline, single and multiple pivots for Indonesia to Japanese

Scenario No	Baseline	Single Pivot				Multiple Pivots	
		Cascade	Triangulate	Interpolate	Fillup Interpolation	Desc	Asc
Scenario 1	11.00	12.18 (MS)	8.26 (EN)	12.03 (MY)	12.40 (MY)	12.15 (Cat 3)	12.84 (Cat 2)
Scenario 2	12.00	11.13 (MS)	9.17 (MS)	12.84 (MY)	12.88 (MY)	12.74 (Cat 2)	13.02 (Cat 2)

Proposed approach

1. Word reordering in multiple pivots for Japanese to Indonesian (Ja-Id). Machine Translation Journal. Special Issue on Machine Translation for Low-Resources Languages. Submitted on February 2020, second review on May 2021.

2. Phrase Table Combination based on Symmetrization of Word Alignment for Low-Resource Language. MDPI Applied Science Journal. 2021.

Proposed approach (1): Word reordering (Ext PT)

- The SMT model is known as it does not work for language pairs that have different word order. (Bisazza and Federico, 2016; Isozaki et al., 2012; Simon and Purwarianti, 2013).
- We also find this issue in our previous experiment, i.e., the generated text of Indonesian from our multiple pivots experiment followed the Japanese sentence structure.
- Approaches for word order issue, i.e., pre-ordering, post-ordering, and word ordering as part of decoding process

Figure 4.3: Translation output example of Ja-Id in without reordering (WoR) experiment.

Proposed approach (1): Word reordering (Ext PT)

- Experiments
  - Without Reordering (WoR)
  - With Reordering (WR)
  - NMT : 1) encoder-decoder, 2) Transformer
- Without reordering (WoR)
  - Taken from single and multiple pivots Ja-Id experimental result (Preliminary works: Paclic).
  - Taken from Triangulation, LI, multiple pivots of ascending results.

Table 4.5: Translation output examples of single and multiple pivots in without reordering (WoR) experiments

Source (Ja)	地震 は、南東 アジア を 壊滅 さ せ た 2004 年 の インド洋 大 地震 が 襲っ た 日 から ちょうど 二 年 後 に 起き た。		
System	Approach	BLEU score	Translation output
JaId(En)	Single pivot - Triangulation	9.71	gempa アジア tenggara, dan mereka untuk 壊滅 Samudra Hindia tahun 2004, menghantam gempa besar dari hanya dua tahun setelah 起き.
JaId(My)	Single pivot - LI	12.21	bahwa gempa Tenggara, yang telah menghancurkan Asia tahun 2004 Samudera Hindia gempa melanda besar dari tanggal yang hanya 2 tahun setelah terjadi.
JaId(EnMsFiMy)	Multiple pivots - LI	12.20	bahwa gempa Tenggara, yang telah menghancurkan Asia tahun 2004 Samudera Hindia gempa melanda besar dari tanggal yang hanya 2 tahun setelah terjadi.
Baseline(EnMsFiMy)	Multiple pivots - LI	12.07	gempa SNT.57162.18909 tenggara, yang telah menghancurkan Asia tahun 2004 gempa melanda besar Samudera Hindia, yang hanya dari hari kedkecualiua terjadi pada tahun.

Proposed approach (1): Word reordering (Ext PT)

- WoR experimental result
  - The Triangulation obtained the lowest BLEU score.
  - The multiple pivots of JaId(EnMsFiMy) outperformed the Baseline and LI, except JaId-My
  - The multiple pivots of Baseline+(EnMsFiMy) outperformed the Baseline and Triangulation, however poorly compare to LI
  - We analyze the decline of Baseline+(EnMsFiMy) BLEU score compare to JaId(EnMsFiMy)
    - Phrase translation parameter scores of phrase pairs.
    - Feature functions weight.

Table 4.3: BLEU scores of single and multiple pivots in without reordering (WoR) experiments

Single pivot			Multiple pivot	
JaId			11.96	
System	Triangulation	LI	System	LI
JaId-En	9.71	11.34	JaId(EnMsFiMy)	12.20
JaId-My	8.71	12.21	Baseline(EnMsFiMy)	12.07
JaId-Ms	8.37	12.11		
JaId-Fi	9.41	12.15		

Proposed approach (1): Word reordering (Ext PT)

- Phrase translation parameter scores of phrase pair
  - We compared 2,000 of the same phrase pairs from two phrase tables: Baseline(EnMsFiMy) and JaId(EnMsFiMy).
  - We found that more than 1,795 phrase pairs of Baseline(EnMsFiMy) obtain the same phrase translation parameter score with the phrase pair of JaId(EnMsFiMy)
  - We also found that these two phrase tables have the same size: 1,041,599
  - The result indicate that the baseline phrase table does not significantly affect the combining process when using the LI approaches.

Table 4.4: Example of phrase pairs and their phrase translation parameter scores.

Phrase-pair	Phrase translation parameters	Score	
		JaId(EnMsFiMy)	Baseline(EnMsFiMy)
は 民主党 で |||nya adalah Demokrat	Inverse phrase translation probability (p(f|e))	0.886859	0.886859
	Inverse lexical weighting (lex(f|e))	0.00138704	0.00138704
	Direct phrase translation probability (p(e|f))	0.888217	0.888217
	Direct lexical weighting (lex(e|f))	0.000010435	0.000010435

Proposed approach (1): Word reordering (Ext PT)

- Feature function weights
  - Feature function weights are parameter settings consisting of lexical reordering, distortion, LM, word penalty, phrase penalty, and translation model.
  - We found that particular feature functions of Baseline(EnMsFiMy) obtain lower weights: distortion, LM, and TM, compared with the JaId(EnMsFiMy).
  - We argue that a lower feature function weight might affect the final translation score.

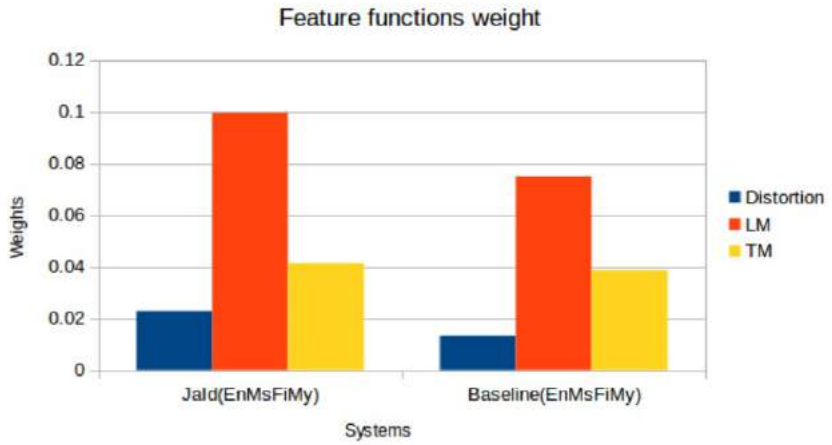


Figure 4.2: Feature functions weight of JaId(EnMsFiMy) and Baseline(EnMsFiMy).

Proposed approach (1): Word reordering (Ext PT)

- With reordering (WR)
  - Pre-ordering Japanese dataset by Lader (Neubig et al., 2012)
  - Systems: direct translation as baseline, one pivot system (single pivot), two pivot systems, three pivots systems, four pivot systems.

Figure 4.1: Example of reordering of Japanese sentence into Indonesian word order. The same colors indicate the same word positions.

Proposed approach (1): Word reordering (Ext PT)

- With Reordering (WR)
  - We found that the unknown words of *gdfand* have candidate phrase pairs in the *tgttosrc* phrase table.
  - We propose a strategy to merge the phrase tables from two symmetrizations: gdfand and tgttosrc.
    1) We construct two pivot systems, viz., src-pvt *gdfand* and src-pvt *tgttosrc*
    2) We sorted the unknown word of generated text from src-pvt *gdfand*
    3) We query the unknown word from src-pvt *tgttosrc*
    4) We merged the src-pvt *gdfand* and src-pvt *filtered* phrase table

Language pairs	src-pvt of WoR experiments	src-pvt of WR experiments
Ja-En	13.49	Preliminary: 8.11 Extend PT: 8.14
Ja-Ms	12.95	Preliminary: 7.60 Extend PT: 7.56
Ja-Fil	11.22	Preliminary: 7.95 Extend PT: 8.01
Ja-My	9.75	Preliminary: 4.50 Extend PT: 4.45

Proposed approach (1): Word reordering (Ext PT)

- WR experimental result
  - The Triangulation approach obtained the lowest BLEU score, however the generated text of one pivot WR experiment significantly change by means that it become more understand compared to one pivot WoR experiment.
  - The result shows that by combining more numbers of pivot languages, then the BLEU score gradually improved.
  - The generated text become more understandable, however the generated text of each system has the same result (text).

Proposed approach (1): Word reordering (Ext PT)

Table 4.7: BLEU scores of single and multiple pivots in With Reordering (WR) experiment

One pivot language		Two pivot languages		Three pivot languages		Four pivot languages	
System	Triangulation	System	LI	System	LI	System	LI
JaId				6.75			
JaId-En	5.99	JaId(EnMs)	6.92	JaId(EnMsFi)	6.94	JaId(MsEnFiMy)	7.15
JaId-Ms	6.30	JaId(EnFi)	6.29	JaId(EnMsMy)	6.98		
JaId-Fi	5.05	JaId(EnMy)	6.49	JaId(EnFiMy)	6.59		
JaId-My	3.16	JaId(MsFi)	6.73	JaId(MsFiMy)	6.85		
		JaId(MsMy)	6.46				
		JaId(FiMy)	5.57				

Table 4.8: Example of translation output of single and multiple pivots in with reordering (WR) experiment.

Source (Ja)	地震 は、ちょうど た 起き 二 年 に 後 から 日 たっ 襲 インド 洋 大 地震 が の 年 2004 た せ さ 壊滅 南東 アジア を。		
System	Approaches	BLEU score	Translation Output
JaId-Ms	One pivot –Triangulation	6.30	gempa bumi tersebut terjadi hanya 2 tahun setelah dari hari melanda India besar gempa bumi pada tahun 2004 telah menghancurkan Tenggara Asia.
JaId(EnMs)	Two pivot –LI	6.92	gempa terjadi hanya 2 tahun setelah dari hari melanda India gempa besar pada tahun 2004 yang telah menghancurkan Asia selatan.
JaId(EnMsMy)	Three pivot –LI	6.98	gempa terjadi hanya 2 tahun setelah dari hari melanda India gempa besar pada tahun 2004 yang telah menghancurkan Asia Selatan.
JaId(MsEnFiMy)	Four pivot –LI	7.15	gempa terjadi hanya 2 tahun setelah dari hari melanda India gempa besar pada tahun 2004 yang telah menghancurkan Asia Selatan.

Proposed approach (1): Word reordering (Ext PT)

- NMT experiment: 1) Encoder-decoder model
  - 2-layer LSTM, with 500 hidden units, dropout 0.3, running by 200,000 steps.
  - We constructed two systems:
    1) LSTM-8.5K : ALT dataset
    2) LSTM-100K: ALT + TEDTalk dataset

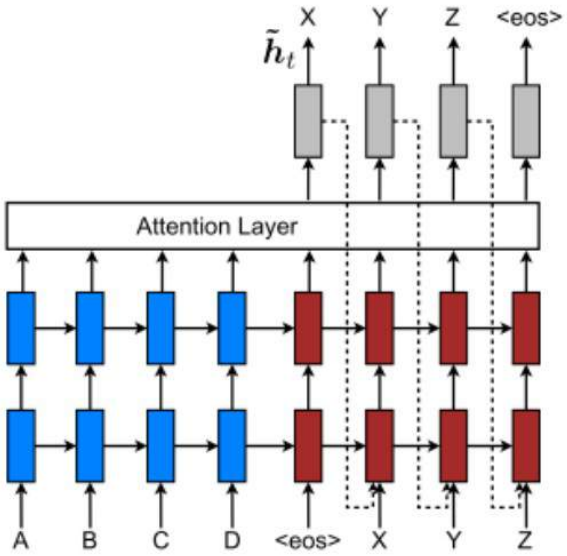


Figure 2.7: A stacking recurrent encoder-decoder model (Luong et al., 2015).

Proposed approach (1): Word reordering (Ext PT)

- NMT experiment: 2) Transformer model
  - 2-layers, 2.0 LR, 4.096 batch size, dropout 0.1, running by 100,000 steps.
  - We constructed one system: Transformer-100K

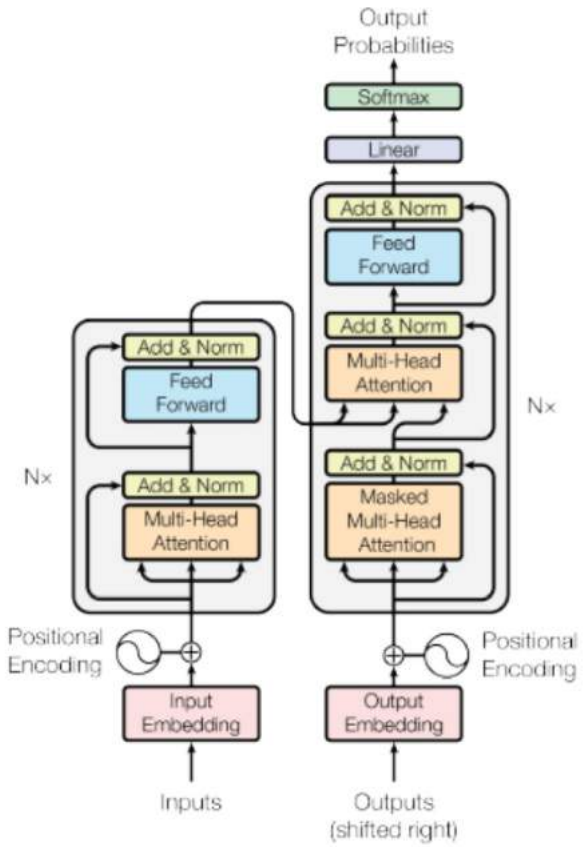


Figure 2.8: The transformer model (Vaswani et al., 2017).

Proposed approach (1): Word reordering (Ext PT)

- NMT experimental result
  - The LSTM-100K system obtained a higher BLEU score than the LSTM-8.5K: 7.8.
  - The Transformer-100K system obtained a lower BLEU score than the LSTM-100K, that is, 3.8.
  - The SMT WoR system obtained the highest BLEU score, that is, 11.96, even with small dataset, i.e., 8.5K of ALT, while the LSTM-100K system obtained a BLEU score of 7.8 with an additional dataset, i.e., 100K of TEDTalk.

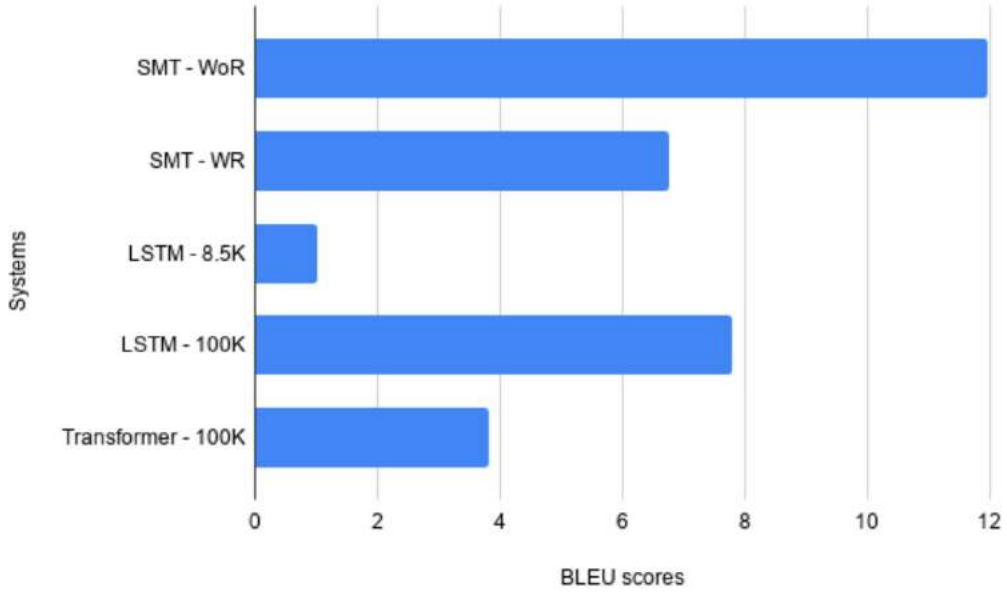


Figure 4.5: BLEU scores of the Ja-Id language pair in the SMT and NMT systems.

Proposed approach (1): Word reordering (Ext PT)

- NMT experimental result
  - The LSTM-8.5K system produced translation output unrelated to the input and did not match the reference file.
  - Other systems produced a better translation output after added additional dataset, however some phrases are missing: “the year 2004, “Indian Ocean”.

Table 4.9: Example of translation output of the SMT and NMT systems.

Source	地震は、南東アジアを壊滅させた2004年のインド洋大地震が襲った日からちょうど二年後に起きた。 (The quake occurred just two years after the 2004 Indian Ocean quake that devastated Southeast Asia.)
Reference	Gempa itu terjadi tepat dua tahun sejak gempa tahun 2004 yang terjadi di Samudra Hindia, yang meluluhlantakkan Asia Tenggara. (The quake occurred exactly two years since the 2004 earthquake that struck the Indian Ocean, which devastated Southeast Asia.)
SMT-WoR	bahwa gempa tenggara, yang telah menghancurkan Asia tahun 2004 Samudera Hindia gempa melanda besar dari tanggal yang hanya 2 tahun setelah terjadi (that the Southeastern earthquake, which destroyed Asia in the 2004 Indian Ocean earthquake struck a large data from only 2 years after it occurred)
SMT-WR	gempa bumi tersebut terjadi hanya 2 tahun setelah dari hari melanda India besar gempa bumi pada tahun 2004 telah menghancurkan Tenggara Asia (The earthquake occurred just 2 years after from the day India struck the massive earthquake in 2004 that devastated Southeast Asia)
LSTM-8.5K	Gempa terjadi pada pukul 11:00 a.m. waktu setempat (03.00 UTC), dan saat berjalan lancar (The earthquake occurred at 11:00 a.m. local time (03.00 UTC), and as it went well)
LSTM-100K	Dua tahun kemudian, gempa bumi terjadi setelah gempa bumi yang menghancurkan gempa bumi barat dan menghancurkan Asia Tenggara (Two years later, the earthquake occurred after an earthquake that destroyed the western earthquake and devastated Southeast Asia)
Transformer-100K	Gempa bumi sungguh terjadi setelah gempa bumi dan tsunami yang saya rasakan pada hari 2004 setelah gempa tahun 2004 (The earthquake really happened after the earthquake and tsunami that I felt on the day of 2004 after the earthquake in 2004)

Proposed approach (2): Phrase table combination (H-ISA)

- Background
  - The common techniques to improve translation quality is using a phrase table combination in pivot approaches (Budiwati and Aritsugi, 2019; Ahmadnia et al., 2017; Trieu and Nguyen, 2017; Dabre et al., 2015, Utiyama and Isahara, 2007, Wu and Wang, 2007)
  - The phrase table contains of phrase pairs that is extracted from word alignment by using symmetrization techniques such as grow-diag-final-and (gdfand), grow, final-grow, grow-diag, intersection, union, srctotgt, and tgttosrc.
  - The standard symmetrization is gdfand. Prior studies have shown that non-standard symmetrization, i.e., intersection, could obtain higher BLEU scores (Singh, 2015; Stymne et al., 2014, Koehn et al., 2005).
  - The non-standard symmetrizaton has not been commonly used in pivot approaches.
  - We propose a strategy, i.e., phrase table combination, that uses symmetrization of word alignment, which obtains the highest BLEU scores.

Proposed approach (2): Phrase table combination (H-ISA)

- Experiments
  - Two language pairs : Kk-En and Ja-Id
  - Pivot type : Single pivot
  - Two LM orders : 3-gram and 5-gram
  - Three experiments :
    - Direct System Approach (DSA)
    - Standard-Interpolation System Approach (Std-ISA)
    - Highest-Interpolation System Approach (H-ISA)

Proposed approach (2): Phrase table combination (H-ISA)

- DSA experimental results
  - We explore five symmetrization: gdfand, intersection, union, srctotgt, and tgttosrc.
  - Of 30 experiments, 20 showed that translation systems using LM05 produced higher BLEU scores.
  - We found that the highest BLEU scores were not always generated by gdfand, showing that non-standard could be an alternative option to improving the BLEU scores of pivot appoaches.

Language pair	BLEU scores				
system	gdfand	intersection	union	srctotgt	tgttosrc
Kk–En LM03	3.08	2.05	3.07	2.51	**3.36**
Kk–En LM05	3.42	2.26	3.28	2.77	**3.56**
Kk-Ru LM03	**6.22**	4.98	4.31	5.41	5.10
Kk-Ru LM05	**6.49**	5.17	4.35	5.64	5.56
Ru-En LM03	**4.77**	0	2.92	4.09	3.12
Ru-En LM05	**4.63**	0	2.73	3.80	2.85
Ja–Id LM03	**11.96**	10.54	9.55	9.79	11.63
Ja–Id LM05	**12.20**	10.47	9.43	9.82	12.04
Ja-Ms LM03	**12.95**	10.09	10.23	10.46	12.65
Ja-Ms LM05	**13.24**	11.06	10.17	10.54	12.93
Ms-Id LM03	**35.07**	34.66	34.90	34.52	34.99
Ms-Id LM05	35.04	34.75	34.89	34.62	**35.14**

Table 4.12: The obtained bilingual evaluation understudy (BLEU) scores of direct system approach (DSA). Results in bold indicate the highest translation quality.

Proposed approach (2): Phrase table combination (H-ISA)

- ISA experimental results
  - Std-ISA: interpolation system that uses gdfand as symmetrization of word alignment. Example of Kk-En LM05 of Std-ISA:
    - Src-trg: gdfand
    - Src-pvt: gdfand
    - Pvt-trg: gdfand
  - H-ISA: interpolation system that uses the symmetrization of word alignment that obtained the highest BLEU score (from DSA). Example of Kk-En LM05 of H-ISA:
    - Src-trg: tgttosrc
    - Src-pvt: gdfand
    - Pvt-trg: gdfand

	Kk–En			Ja–Id	
LM Order	**Lang Pair**	**H-ISA**	**LM Order**	**Lang Pair**	**H-ISA**
LM03	Kk–En	tgttosrc	LM03	Ja–Id	gdfand
	Kk-Ru	gdfand		Ja-Ms	gdfand
	Ru-En	gdfand		Ms-Id	gdfand
LM05	Kk–En	tgttosrc	LM05	Ja–Id	gdfand
	Kk-Ru	gdfand		Ja-Ms	gdfand
	Ru-En	gdfand		Ms-Id	tgttosrc

Table 4.15: Candidates of symmetrization of word alignment for highest-interpolation system approach (H-ISA).

Proposed approach (2): Phrase table combination (H-ISA)

- ISA experimental results
  - We found that all the translation system of LM05 obtained higher BLEU scores than LM03
  - For Kk-En, H-ISA is a competitive approach because it provided absolute improvements of 0.35 and 0.22 BLEU points over Baseline and Std-ISA in LM03 and LM05.
  - For Ja-Id, H-ISA obtained an absolute improvements of 0.11 BLEU points over Baseline in LM03, and absolute drop of -0.12 BLEU points over Baseline in LM05
    - It could be because a lower phrase translation parameter scores of H-ISA LM05 phrase pair.

Language Model	Baseline	Std-ISA	H-ISA
Kk–En			
LM03	3.08	3.08	**3.43**
LM05	3.42	3.42	**3.64**
Ja–Id			
LM03	11.96	**12.07**	**12.07**
LM05	**12.20**	12.08	12.08

Table 4.16: The obtained BLEU scores of direct translation of src–trg (baseline) and interpolation system approach (ISA). Results in bold indicate the highest translation quality.

Proposed approach (2): Phrase table combination (H-ISA)

- Phrase translation parameter scores of phrase pair
  - We compared 2,000 of the same phrase pairs from two phrase tables: Baseline and H-ISA.
  - We found that more than 1,900 phrase pairs of H-ISA in LM05 obtained lower score than Baseline.

Phrase-Pair	Phrase Translation Parameter	Phrase Table		
		Baseline	Std-ISA	H-ISA
この 建物 の ||| bangunan	Inverse phrase translation probability ($p(t/s)$)	0.00952381	*0.00841842*	*0.00849199*
	Inverse lexical weighting ($lex(t/s)$)	0.00000289	*0.00000024*	*0.00000024*
	Direct phrase translation probability ($p(s/t)$)	0.333333	*0.294644*	*0.294915*
	Direct lexical weighting ($lex(s/t)$)	0.488889	*0.431259*	*0.432099*

Table 4.17: Example of phrase translation parameter scores of Ja–Id LM05. Results in italic indicate the lowest score.

Proposed approach (2): Phrase table combination (H-ISA)

- Perplexity score
  - For Kk-En, the longer LM order: 5-gram, obtained a lower pp score, contrast with Ja-Id.
  - It could be because: target monolingual corpus and feature function weight.
  - The target monolingual corpus: Kk-En (532,560) and Ja-Id (8,500).

Language Model	Direct Translation	Std-ISA	H-ISA
Kk–En			
LM03	148.21	148.18	284.05
LM05	93.41	115.90	206.15
Ja–Id			
LM03	309.32	310.25	310.25
LM05	403.13	411.48	414.46

Table 4.20: Perplexity scores for the direct translation of src–trg (baseline) and ISA.

Proposed approach (2): Phrase table combination (H-ISA)

- Perplexity score
  - The feature function weight
    - We found that the feature function of LM for Kk-En LM05 was higher, in contrast with Ja-Id that was lower.

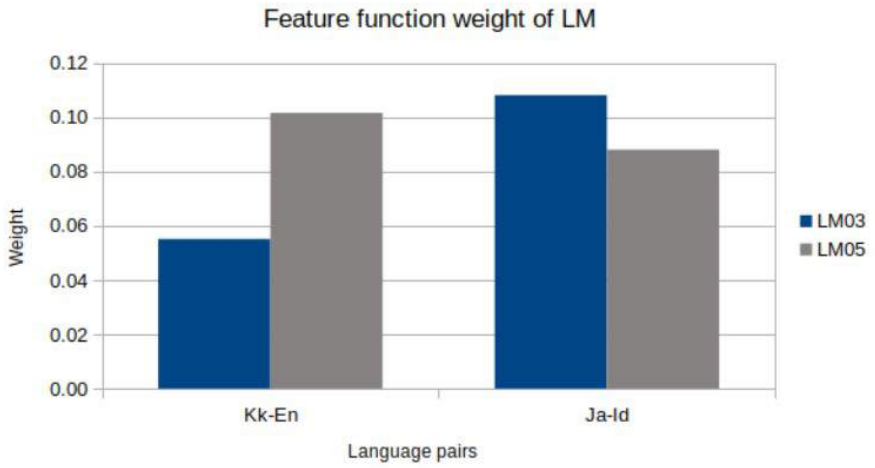


Figure 4.6: Feature function weight of LM for Kk–En and Ja–Id.

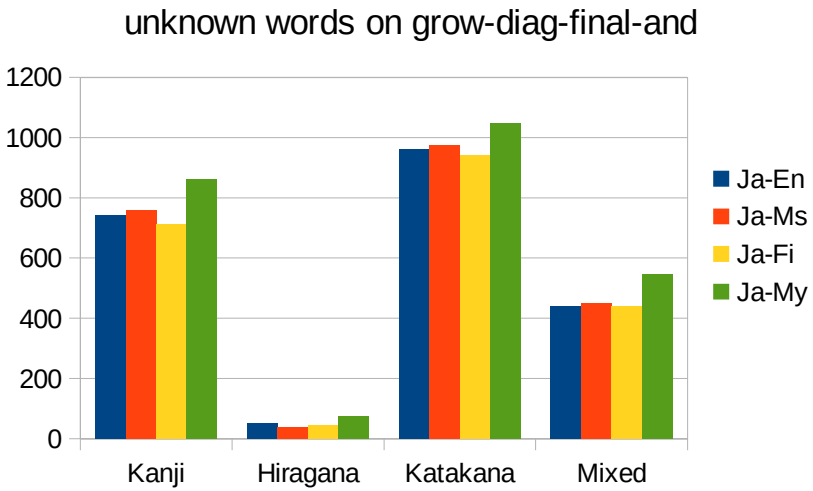
Conclusion and Future Work

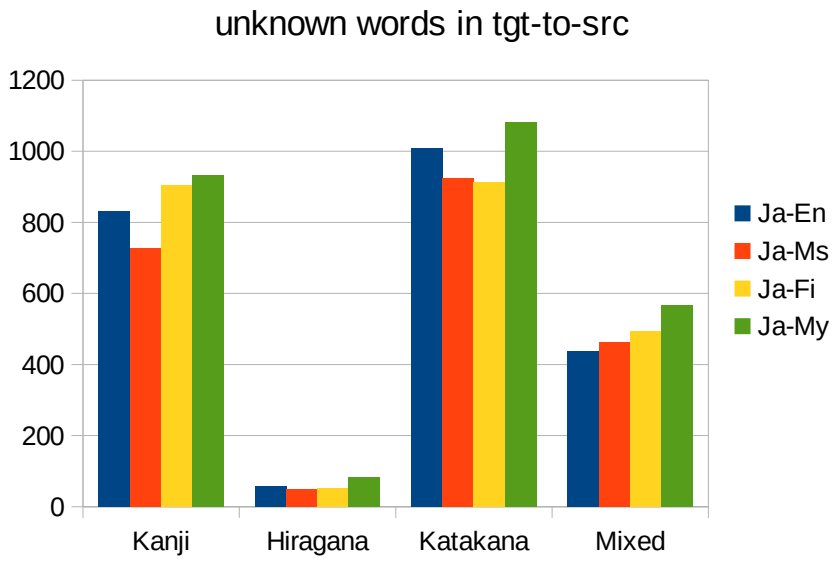
- We propose two strategies to improve the translation quality of Kk-En and Ja-Id
  - Extending phrase table
    - Produce an absolute gain of up to 0.06 BLEU points for Ja-En and Ja-Fil.
    - Obtain an absolute drop of -0.05 BLEU points for Ja-Ms and Ja-My
  - Phrase table combination based on symmetrization of word alignment
    - Outperforms direct translation in Kk-En with absolute improvements of 0.35 and 0.22 BLEU points for 3-gram and 5-gram
    - Outperforms the direct translation of 3-gram in Ja-Id with an absolute improvement of 0.11 BLEU points
- We also present empirical results to show that SMT outperformed the NMT for Ja-Id as low-resource language pair. SMT system obtained the highest BLEU score, that is, 11.96, even with a small dataset, that is, 8.5K ALT dataset, while the NMT system obtained a BLEU score of 7.8 with an additional dataset, that is, 100K of the TEDTalk. Our results indicates that the SMT obtained better results than NMT, even with a small dataset.

Conclusion and Future Work

- Future work
  - We consider implementing the extending phrase table of the pvt–trg system, which did not employ in this study.
  - We plan to combine the src–pvt and pvt–trg phrase tables using the LI approach in a single pivot of the WR.
  - We aim to investigate the similarity of translation outputs in two pivots, three pivots, and four pivots of multiple pivots of Ja–Id.
  - We plan to investigate the optimization of hyper-parameters, pre-processing of the NMT experiment. We also aim to apply other NMT models, i.e., transfer learning and multilingual, to compare with multiple pivots of SMT results.
  - We will increase our Indonesian target monolingual corpus size of Ja–Id to be as large as Kk–En. Then, we will re-evaluate the parameter of the target monolingual corpus as a factor for decreasing the perplexity scores.
  - The applicability of our proposed strategies was demonstrated on limited language pairs. Thus, another direction is investigating other language pairs and datasets.

Thank you





	Grow-diag-final-and				Tgt-to-src			
	Ja-En	Ja-Ms	Ja-Fi	Ja-My	Ja-En	Ja-Ms	Ja-Fi	Ja-My
Kanji	742	760	714	863	832	728	906	932
Hiragana	51	39	45	75	58	49	52	82
Katakana	961	975	942	1048	1008	925	914	1082
Mixed	440	450	439	545	438	463	494	568