Multi-dialect Neural Machine Translation and Dialectometry

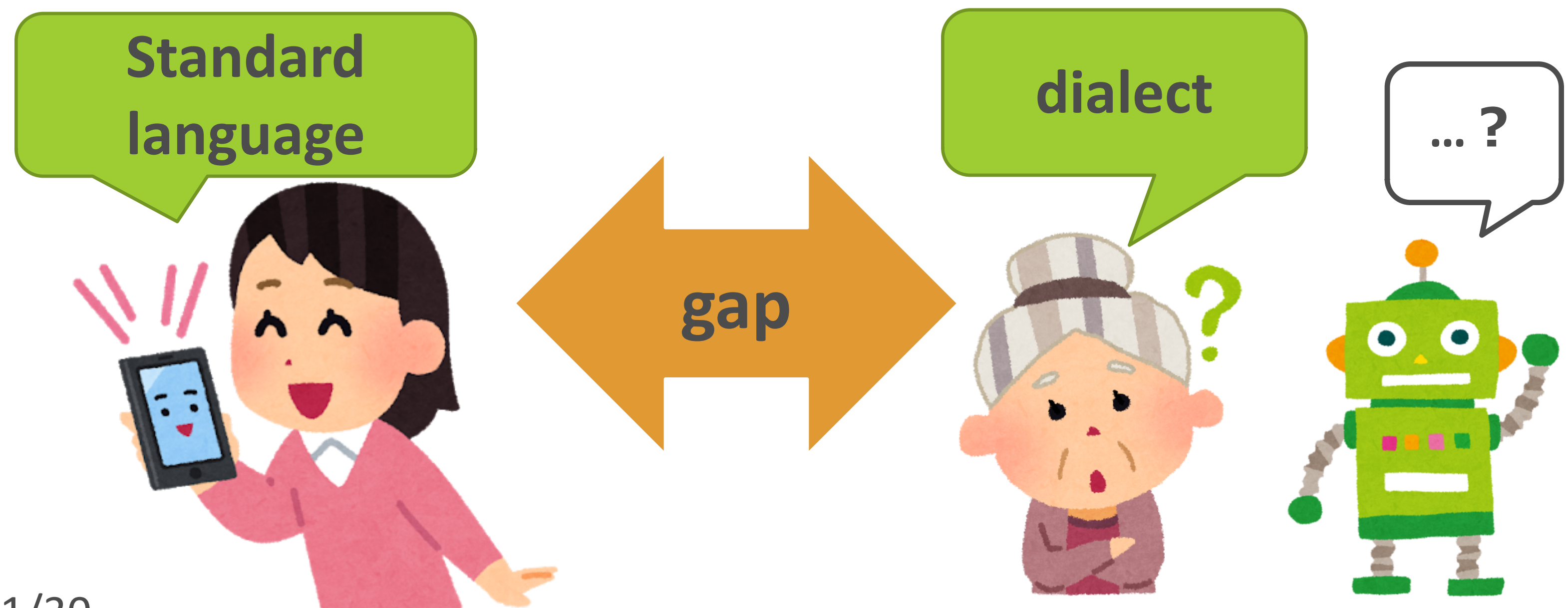
Kaori Abe[1]
Yuichiroh Matsubayashi[2], Naoaki Okazaki[3], Kentaro Inui[1,2]

1. Graduate School of Information Sciences, Tohoku University
2. RIKEN Center for Advanced Intelligence Project (AIP)
3. School of Computing, Tokyo Institute of Technology

2018/12/1

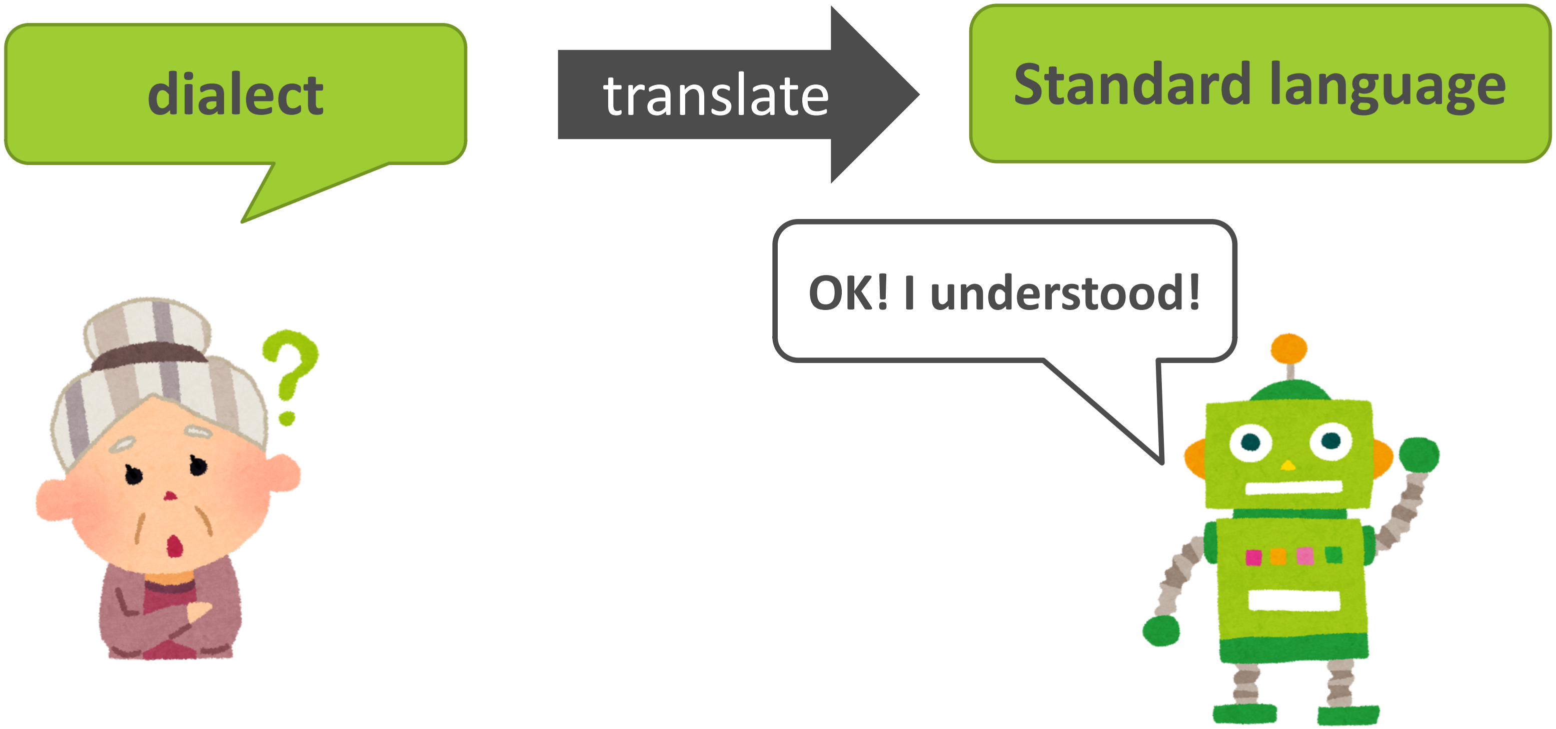
Background

- Recently, smart speakers(e.g. Siri) can understand standard languages in the world
- However, (especially in Japan,) humans sometimes use ***regional dialects*** to talk



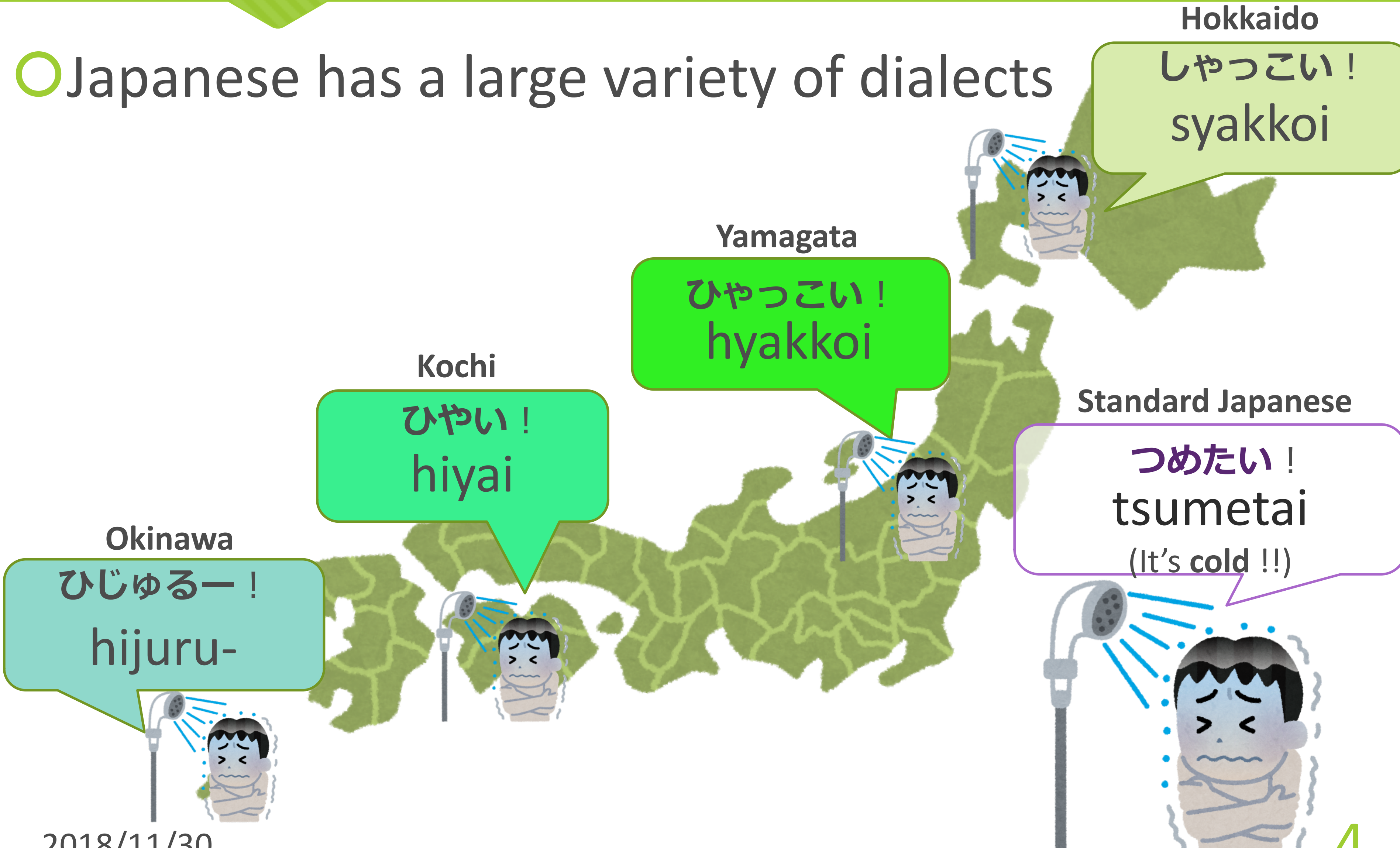
Motivation

- ***Dialect translation*** enables smart speakers to understand a request in a dialect



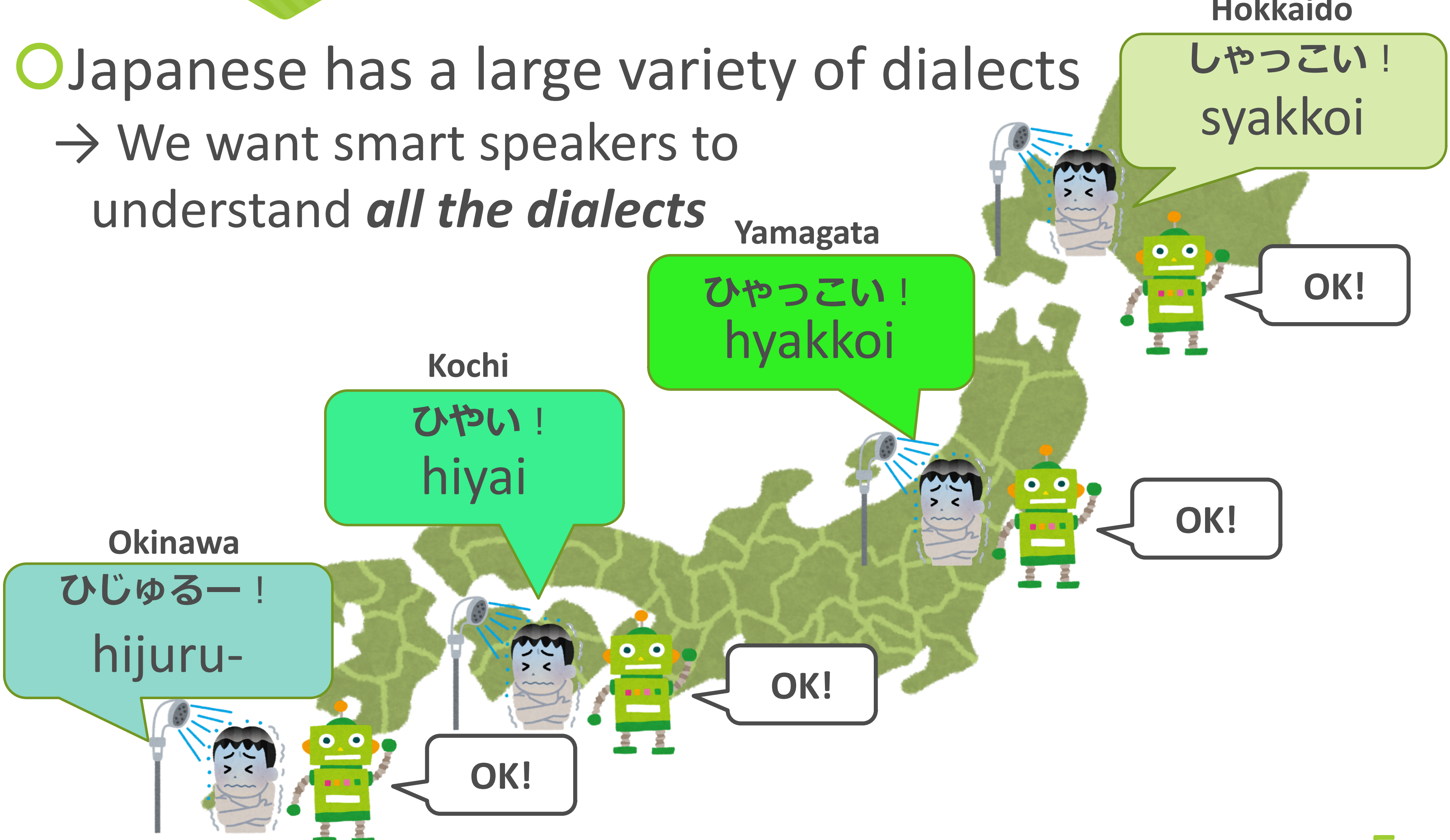
Problem: A variety of Japanese Dialects

- Japanese has a large variety of dialects



Problem: A variety of Japanese Dialects

- Japanese has a large variety of dialects
  → We want smart speakers to understand ***all the dialects***



Problem: Lack of Dialect Corpus

- Japanese Dialect Corpus:
  - 48 dialects × 30 minutes **dialog**

    → **34,117 sentence pairs of Transcript**

    (**718 sentence pairs** per dialect)

Dialects are spoken rather than written

It's too small !!!

Problem: Lack of Dialect Corpus

- Japanese Dialect Corpus:
  - 48 dialects × 30 minutes **dialog**
    → **34,117 sentence pairs of Transcript**
    (**718 sentence pairs** per dialect)

Dialects are spoken rather than written

It's too small !!!

- **We should consider how to make the best use of this small language resource**

Relation between a Japanese Dialect and Standard Japanese

- From a Japanese dialect to standard Japanese,
  - **Change : vocabulary and particular syllables**
  - **Unchange : word order**

Example : My mother said "It's cold !"

Aomori Dialect	おが oga	が ga	"しゃっこい！" "syakkoi !"	と to	ゆった yutta
Standard Japanese	はは haha	が ga	"つめたい！" "tsumetai !"	と to	いった itta
	My mother		"It's cold"		said

Relation between a Japanese Dialect and Standard Japanese

- From a Japanese dialect to standard Japanese,
  - **Change : vocabulary and particular syllables**
  - **Unchange : word order**

In our study, we use a phonetic letter system named "kana" (not "kanji")

Example : My mother said "It's cold !"

Aomori Dialect	おが oga	が ga	"しゃっこい！" "syakkoi !"	と to	ゆった yutta
Standard Japanese	はは haha	が ga	"つめたい！" "tsumetai !"	と to	いった itta
	My mother		"It's cold"		said

Relation between a Japanese Dialect and Standard Japanese

- From a Japanese dialect to standard Japanese,
  - Change : vocabulary and particular syllables
  - Unchange : word order

In our study, we use a phonetic letter system named "kana" (not "kanji")

Example : My mother said "It's cold !"

Aomori Dialect	おが oga	が ga	"しゃっこい！" "syakkoi !"	と to	ゆった yutta
Standard Japanese	はは haha	が ga	"つめたい！" "tsumetai !"	と to	いった itta
	My mother		"It's cold"		said

Relation between a Japanese Dialect and Standard Japanese

- From a Japanese dialect to standard Japanese,
  - Change : vocabulary and particular syllables
  - Unchange : word order

In our study, we use a phonetic letter system named “kana” (not “kanji”)

Example : My mother said “It’s cold !”

Aomori Dialect	おが oga	が ga	“しゃっこい！” “syakkoi !”	と to	ゆった yutta
Standard Japanese	はは haha	が ga	“つめたい！” “tsumetai !”	と to	いった itta
	My mother		“It’s cold”		said

Relation between a Japanese Dialect and Standard Japanese

- From a Japanese dialect to standard Japanese,
  - **Change : vocabulary and particular syllables**
  - **Unchange : word order**

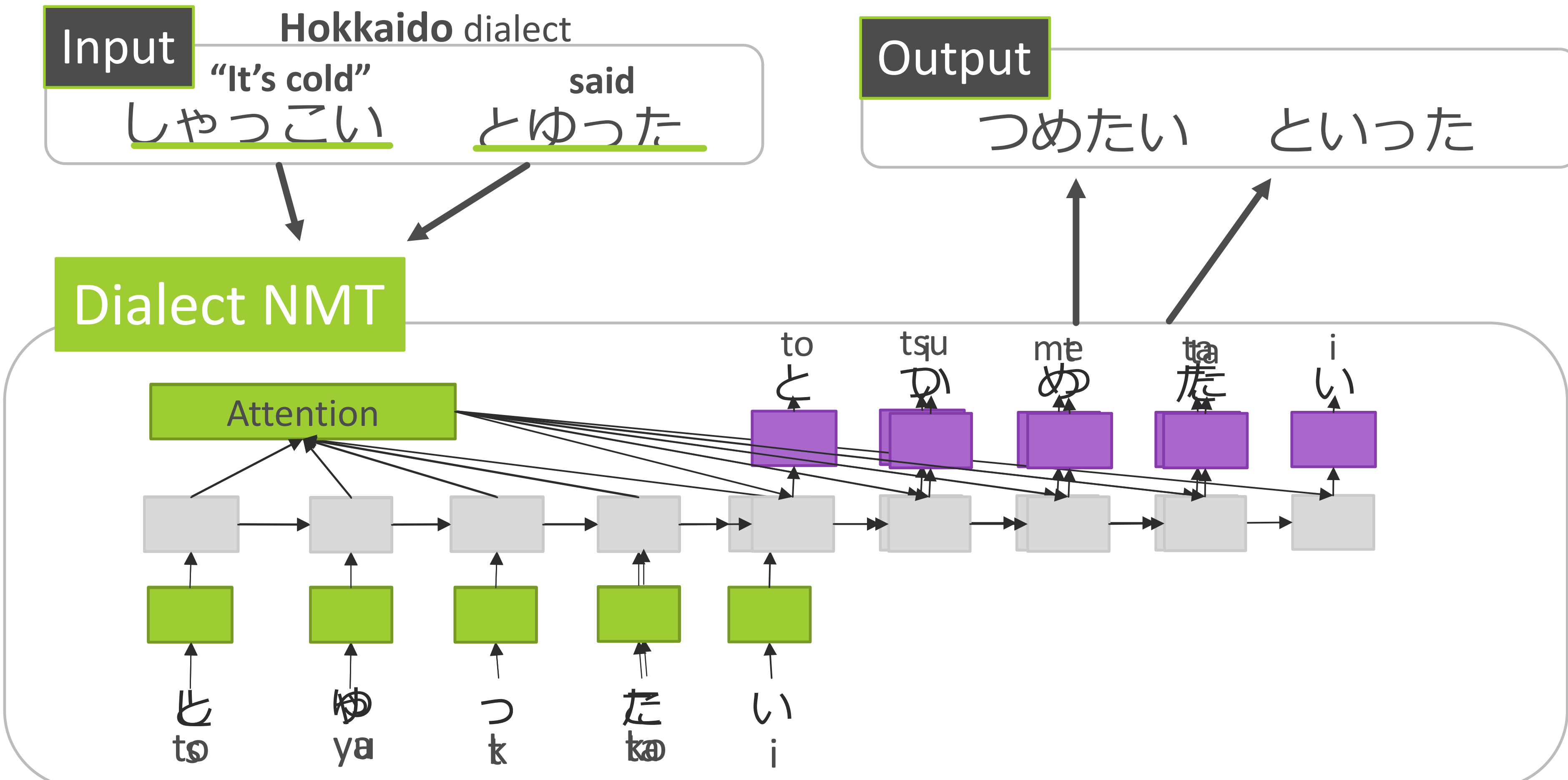
In our study, we use a phonetic letter system named “kana” (not “kanji”)

Example : My mother said “It’s cold !”

Aomori Dialect	おが oga	が ga	“しゃっこい！” “syakkoi !”	と to	ゆった yutta
Standard Japanese	はは haha	が ga	“つめたい！” “tsumetai !”	と to	いった itta
	My mother		“It’s cold”		said

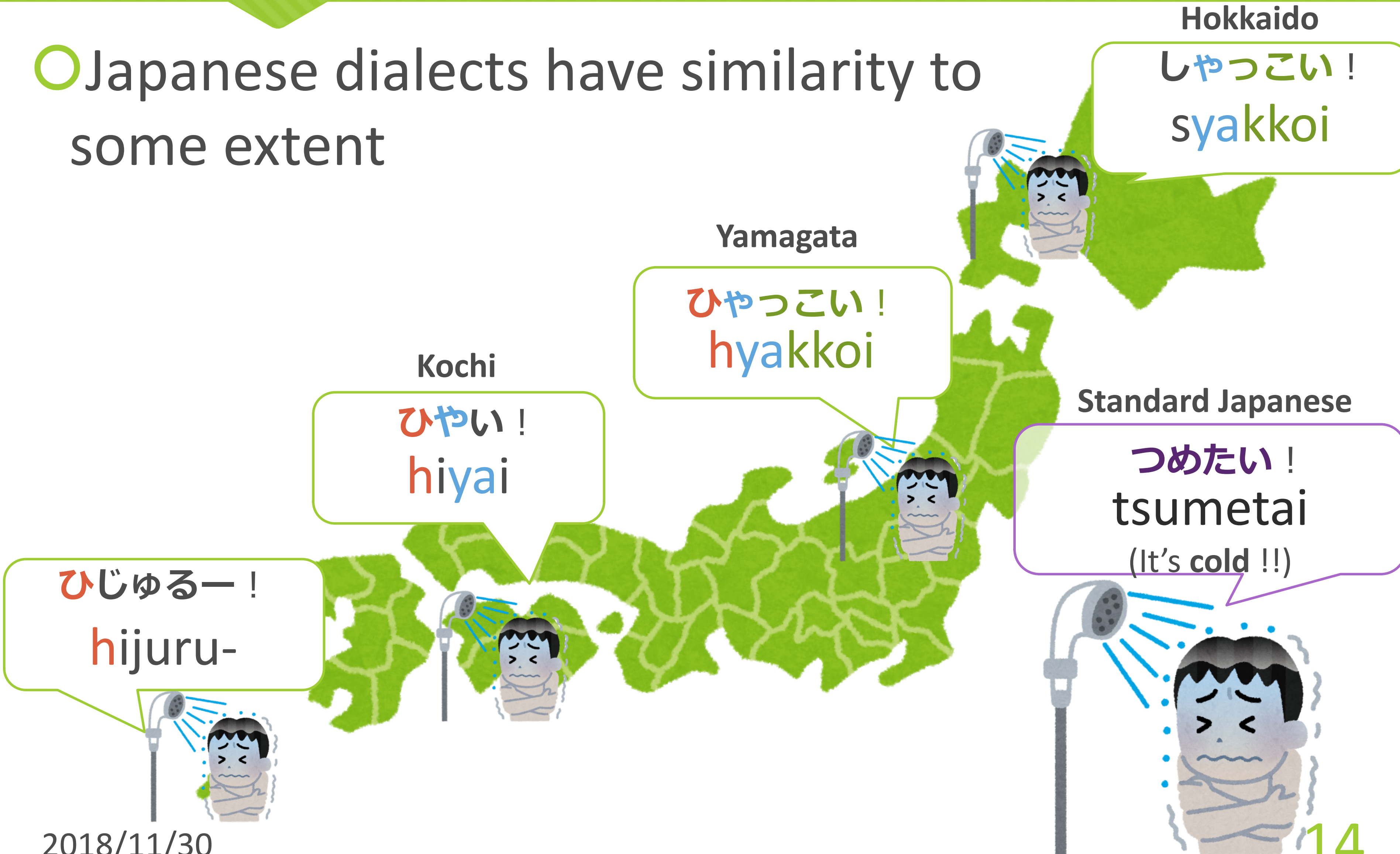
Model (Fixed-order + Character)

- **Fixed-order + Character** NMT



Japanese Dialects

- Japanese dialects have similarity to some extent



Similarity between Japanese Dialects

- Most of the dialects share fundamental properties
  - **Same or Similar Vocabulary**

Hokkaido dialect この　みず　は　しゃっこい！ (syakkoi)

Same Vocabulary　**Similar** Vocabulary

Yamagata dialect この　みず　は　ひゃっこい！ (hyakkoi)

Same Vocabulary　Different Vocabulary

Standard Japanese この　みず　は　つめたい！

This　water　is　cold!

Similarity between Japanese Dialects

- Most of the dialects share fundamental properties
  - **Phonetic correspondence rules**

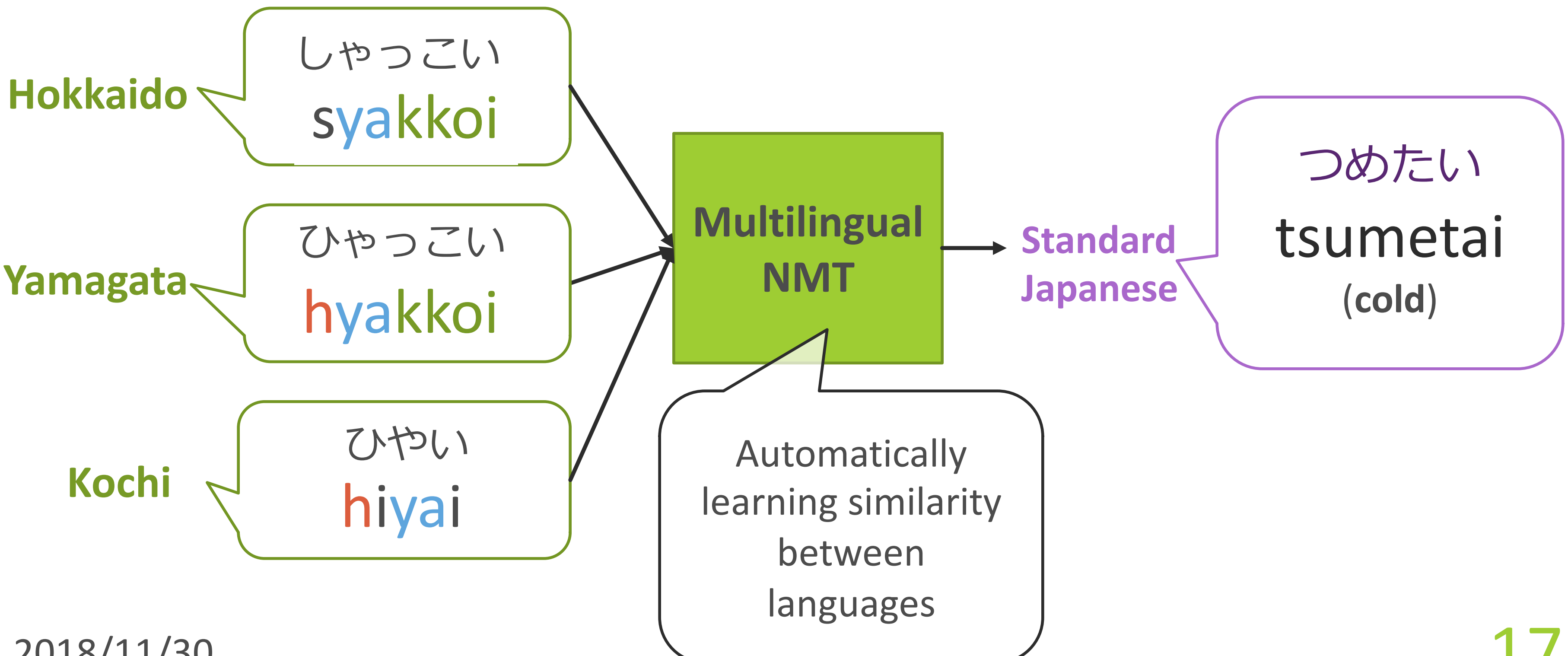
Hokkaido dialect	imawa いまは	soudewa そうでは	nai**ge**dona ない**げ**どな
Standard Japanese	imawa いまは Now	soudewa そうでは it is	nai**ke**dona ない**け**どな

Same rule

Yamagata dialect	sakana さかな	ga が	ya**ge**da や**げ**だ
Standard Japanese	sakana さかな Fish	ga が	ya**ke**ta や**け**た grilled

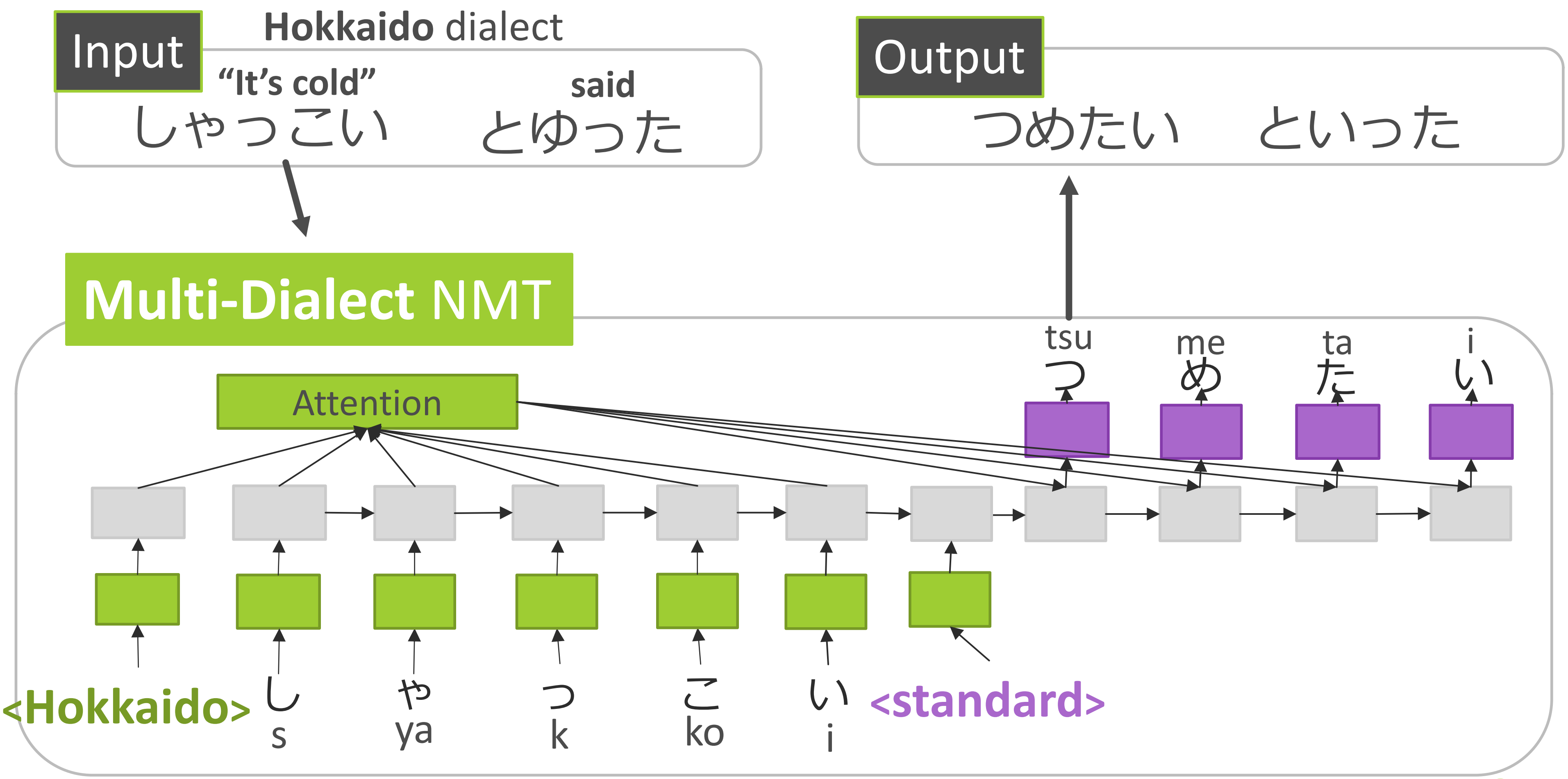
Approach (Multilingual NMT)

- **A multilingual NMT**[Johnson+, 2017] utilizes shared properties between dialects by way of a unified model that learns multiple languages jointly



Model (Character + Fixed-order + Multilingual)

- **Character-level** + **Fixed-order** + **Multilingual** NMT



Experiments

Setting

- Corpus :
  - 48 dialects × 30 minutes dialog
    → **34,117 sentence pairs** (116,928 “bunsetsu” pairs)
    (718 sentence, 2436 “bunsetsu” pairs per dialect)
    - Train : Validation : Test = 8 : 1 : 1
- Evaluation : BLEU (**character-level**)
- System
  - NMT : OpenNMT-py
  - SMT (baseline system) : Moses

Evaluation

- **Our model** archived the best performance

System	Character-level	Fixed-order	Dialect labels	jointly learning	BLEU
Original					35.10
Multi NMT	○	○	○	○	**75.66**
Mono NMT	○	○	×	×	22.45
Sentence-Multi NMT	○	×	○	○	71.29
Multi NMT (w/o labels)	○	○	×	○	69.74
Mono SMT	○	○	×	×	52.98
Multi SMT (w/o labels)	○	○	×	○	73.54

Mono NMT vs. Multi NMT

System	Character-level	Fixed-order	Dialect labels	jointly learning	BLEU
Original					35.10
Multi NMT	○	○	○	○	**75.66**
Mono NMT	○	○	×	×	**22.45**
Sentence-Multi NMT	○	×	○	○	71.29
Multi NMT (w/o labels)	○	○	×	○	69.74
Mono SMT	○	○	×	×	52.98
Multi					73.54

- Multilingual NMT performs significantly better than monolingual NMT

Original vs. Mono NMT

System	Character-level	Fixed-order	Dialect labels	jointly learning	BLEU
Original					35.10
Multi NMT	○	○	○	○	75.66
Mono NMT	○	○	×	×	22.45
Sentence-Multi NMT	○	×	○	○	71.29
Multi NMT (w/o labels)	○	○	×	○	69.74
Mono SMT	○	○	×	×	52.98

- Mono NMT could not learn translation rules because each monolingual dialect-to-standard corpus is too small

Sentence vs. Fixed-order

System	Character-level	Fixed-order	Dialect labels	jointly learning	BLEU
Original					35.10
Multi NMT	○	○	○	○	**75.66**
Mono NMT	○	○	×	×	22.45
Sentence-Multi NMT	○	×	○	○	**71.29**
Multi NMT (w/o labels)	○	○	×	○	69.74
M[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]98
Multi SMT (w/o labels)	○	○	×	○	73.54

- Although it has a weak point that it cannot consider context, the strategy of fixed-order translation is effective

With labels vs. Without labels

System	Character-level	Fixed-order	Dialect labels	jointly learning	BLEU
Original					35.10
Multi NMT	○	○	○	○	**75.66**
Mono NMT	○	○	×	×	22.45
Sentence-Multi NMT	○	×	○	○	71.29
Multi NMT (w/o labels)	○	○	×	○	**69.74**
Mono SMT	○	○	×	×	52.98
Multi SMT (w/o labels)	○	○	×	○	72.54

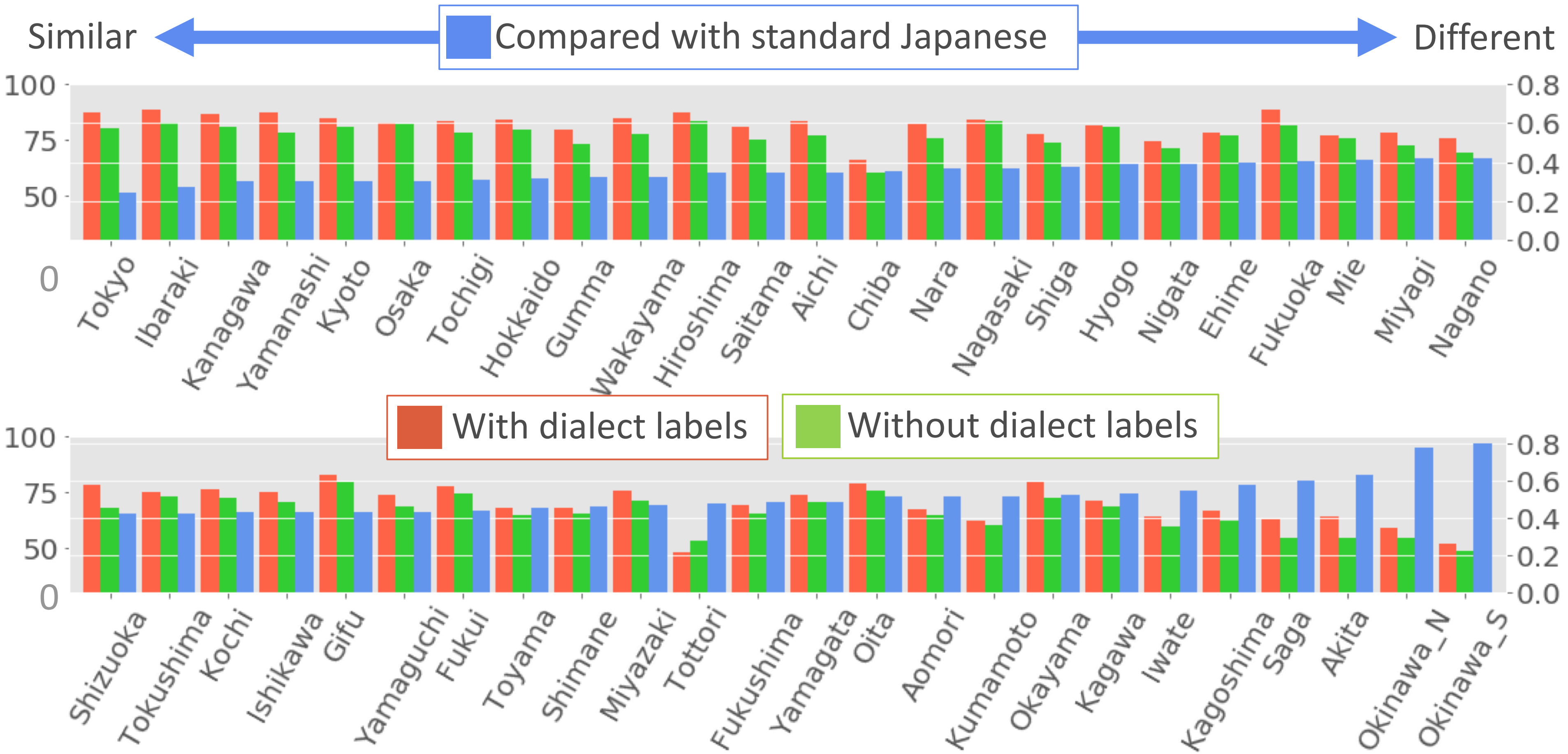
- Dialect labels contribute to increasing the BLEU score because it clearly teach the NMT what the input dialect is

SMT vs. NMT

System	Character-level	Fixed-order	Dialect labels	jointly learning	BLEU
Original					35.10
Multi NMT	○	○	○	○	75.66
Mono NMT	○	○	×	×	22.45
Sent					29
Multi NMT (w/o labels)	○	○	×	○	69.74
Mono SMT	○	○	×	×	52.98
Multi SMT (w/o labels)	○	○	×	○	73.54

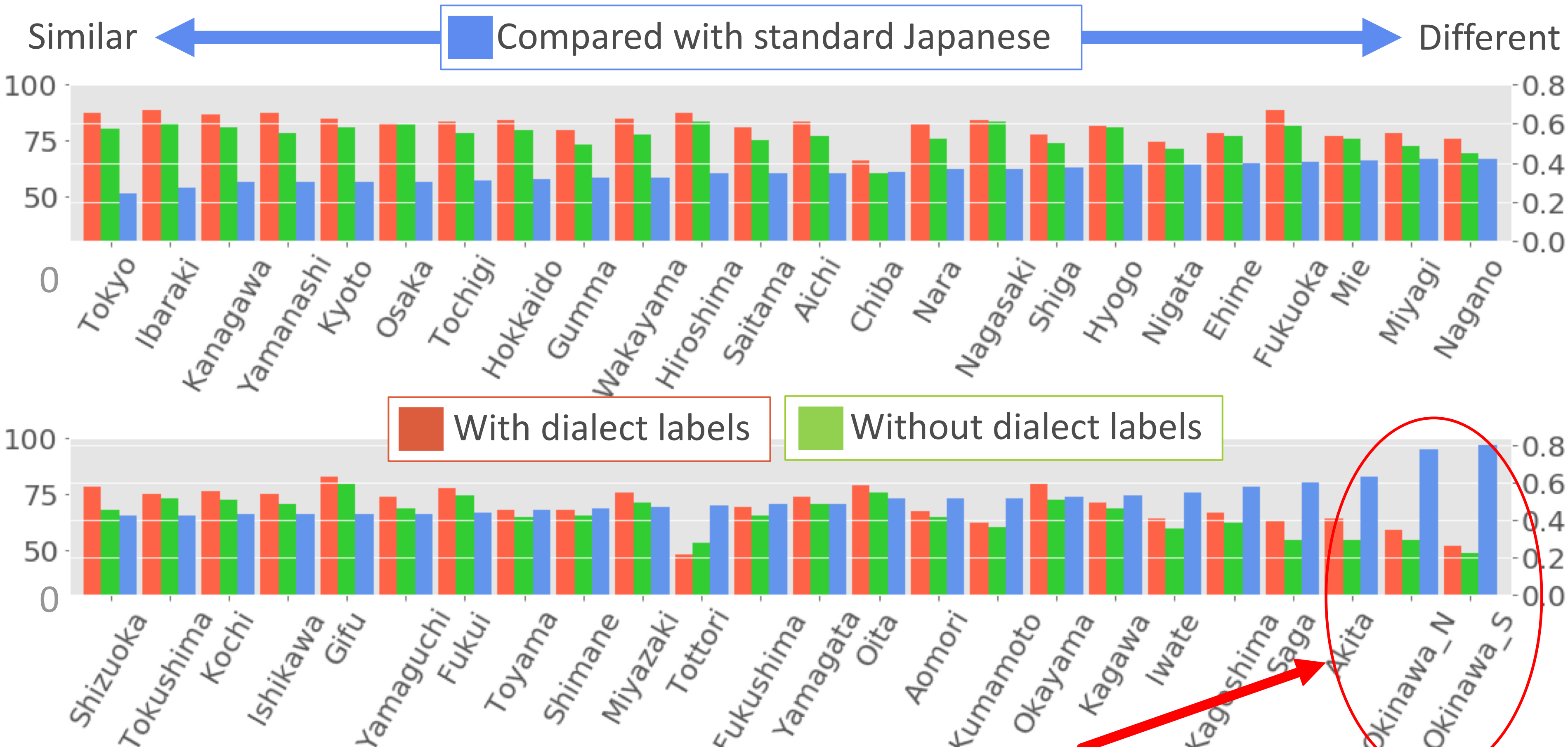
- Our model outperformed standard SMT models!

Effectiveness of Dialect labels



Dialect labels **which teach what the dialect is contribute to increasing BLEU scores in most of the dialects**

Effectiveness of Dialect labels

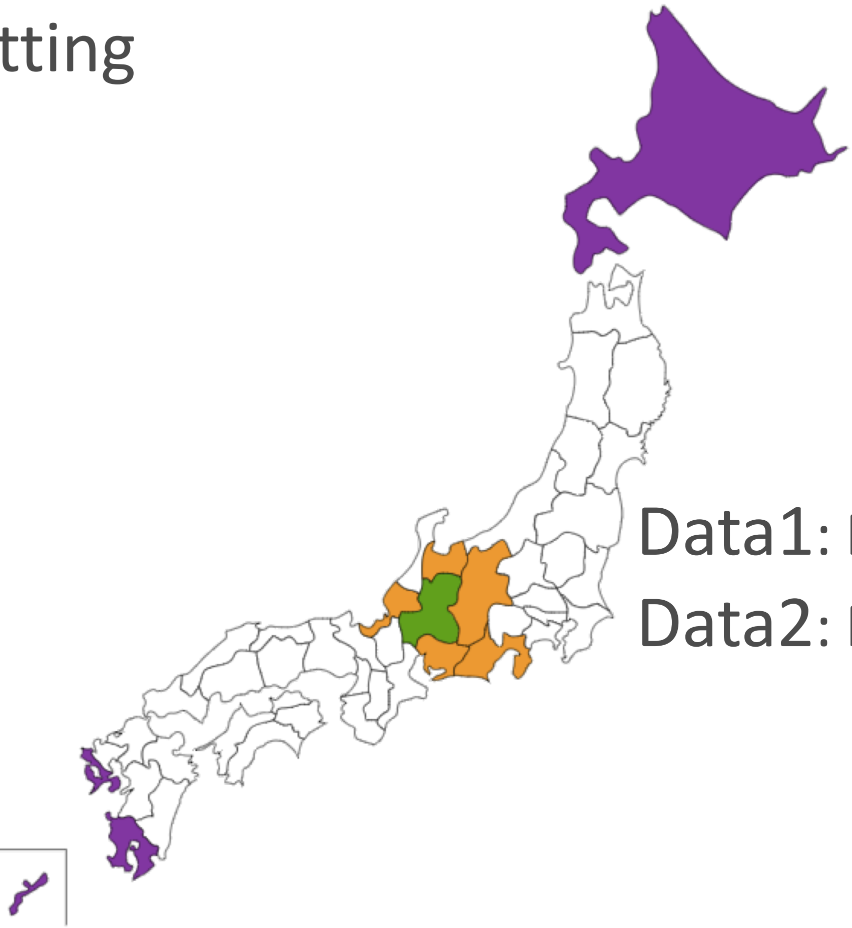


Our model could not much perform in Okinawa dialect because it is quite different from standard Japanese

Analysis

Effect of Nearby Dialects

- Assumption: the data of **nearby dialects** might contribute to the high performance under the multilingual architecture
- Setting



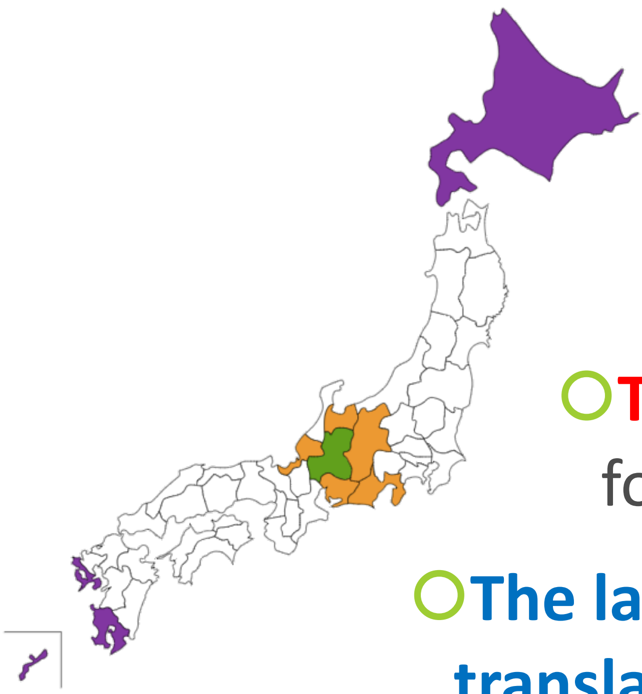
For **"Gifu" dialect** …

Data1: **Removed** the **nearest 5** dialects

Data2: **Removed** the **farthest 5** dialects

Analysis (Effect of Nearby Dialects)

- Evaluating whether **neighbor dialect data** improves a BLEU score

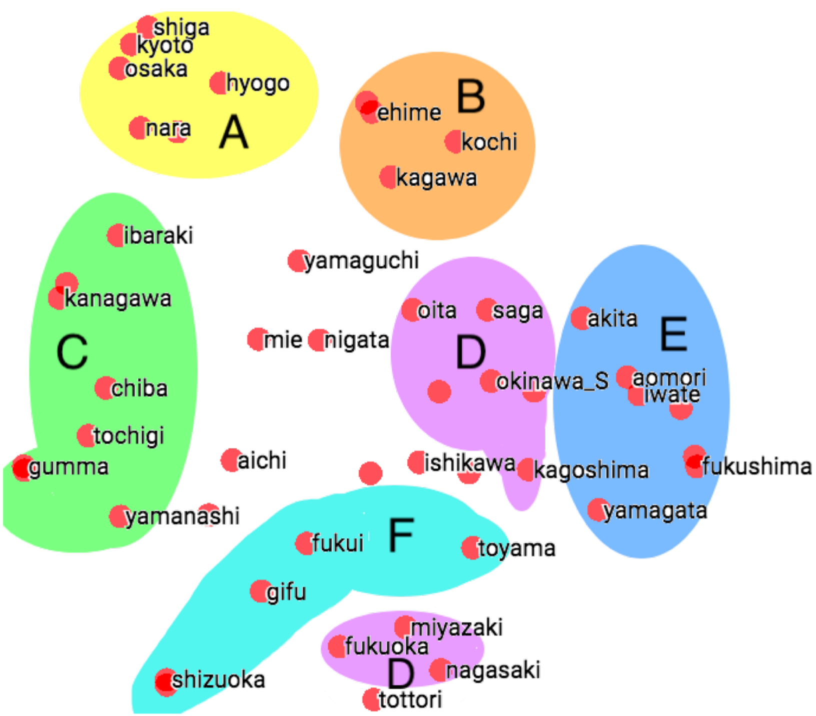


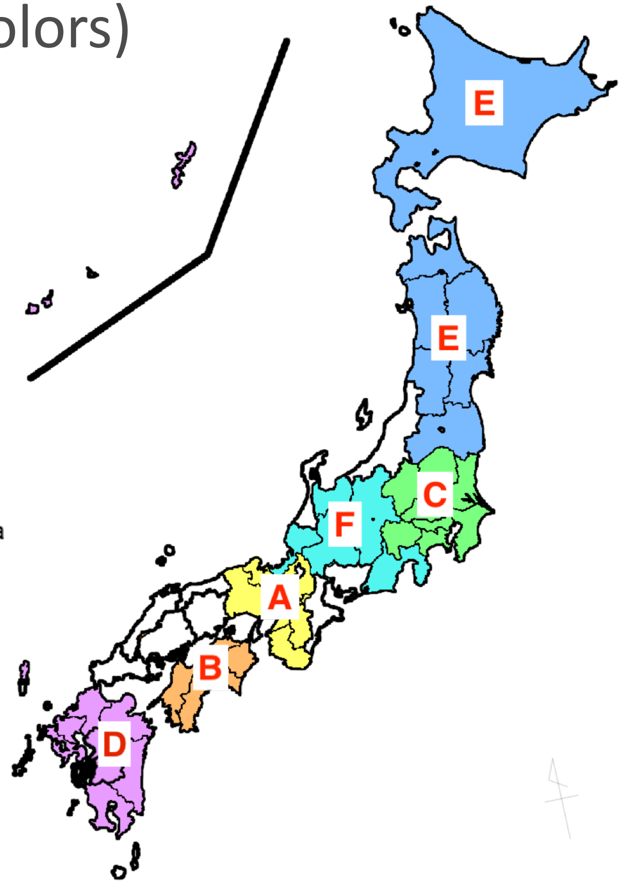
Dataset	Avg. Δ	#Regions BLEU decreased
All -nearest 5	**-0.94**	**34/48(71%)**
All -farthest 5	-0.22	**31/48(65%)**

- **The data of near areas are more effective** for multilingual NMT
- **The lack of 5 dialects** in supervision data **affect translation accuracy** in a low-resource setting

Analysis (Visualize dialect embeddings)

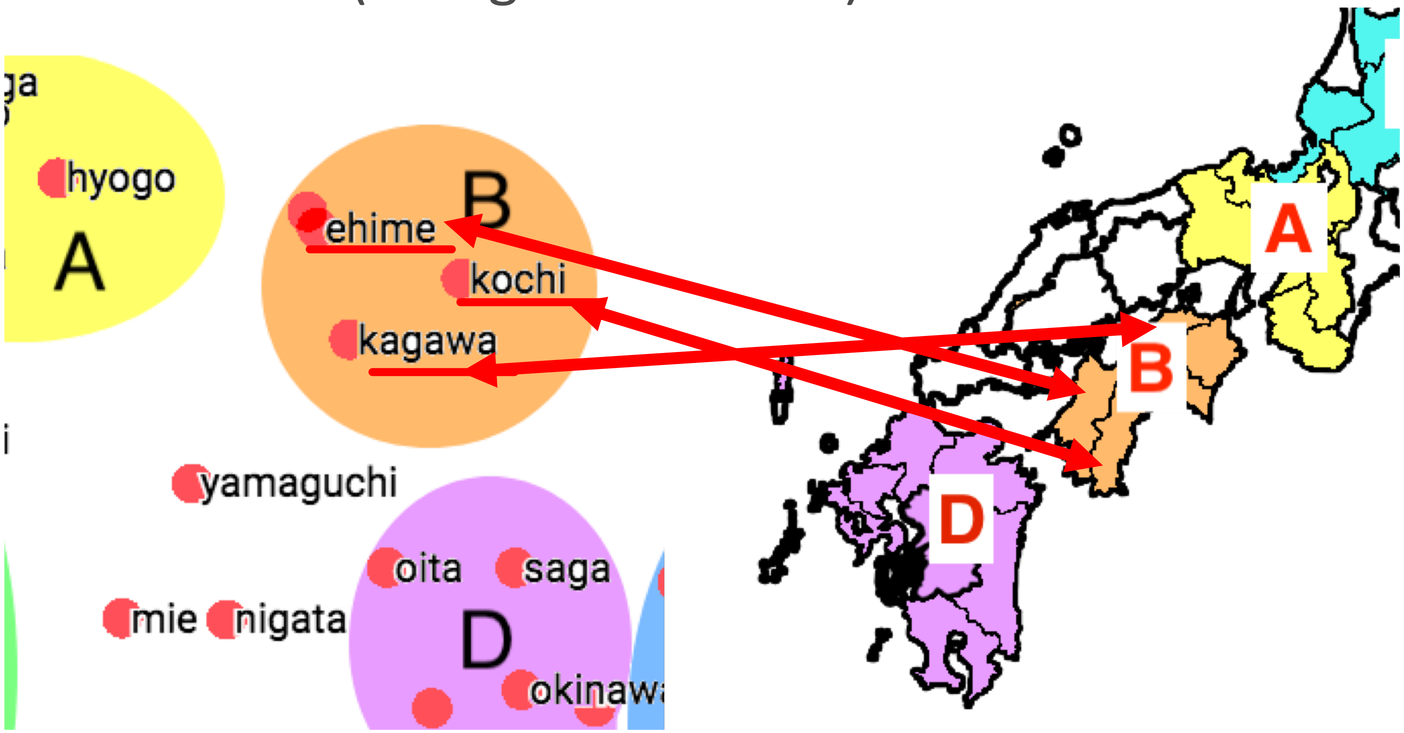
- A t-SNE projection of dialect embeddings follows dialectological typology
  - The nearer the distance between two areas is, the more similar dialects are used (background colors)





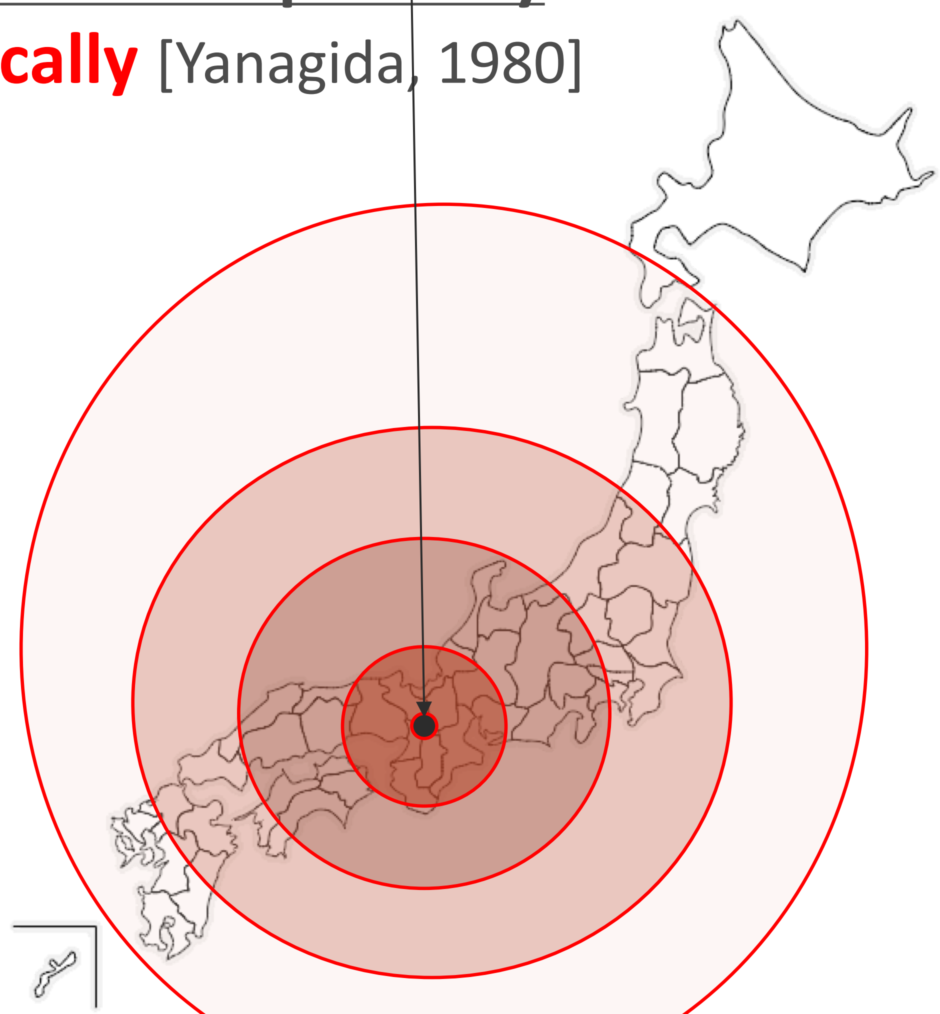
Analysis (Visualize dialect embeddings)

- A t-SNE projection of dialect embeddings follows dialectological typology
  - The nearer the distance between two areas is, the more similar dialects are used (background colors)



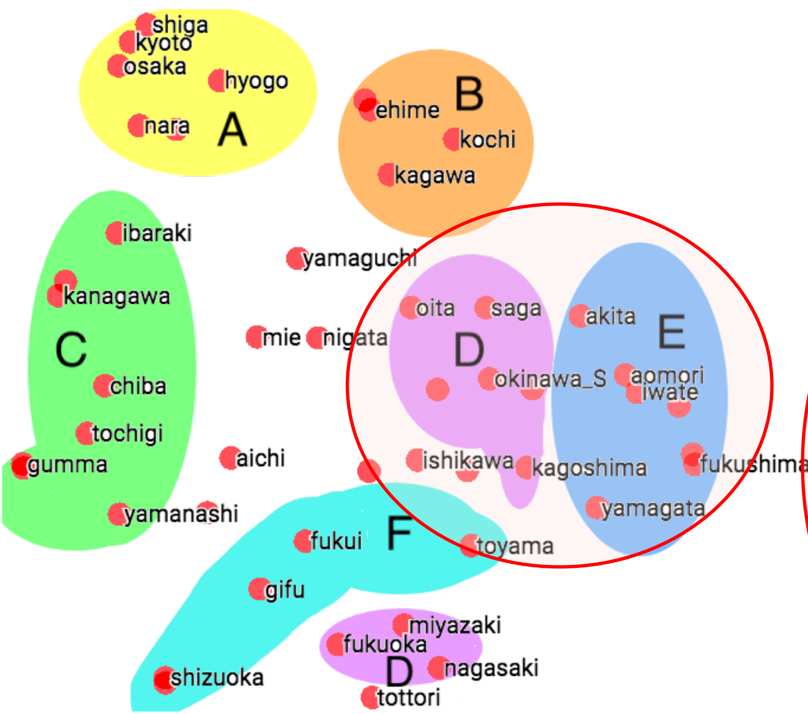
Japanese dialectal typology

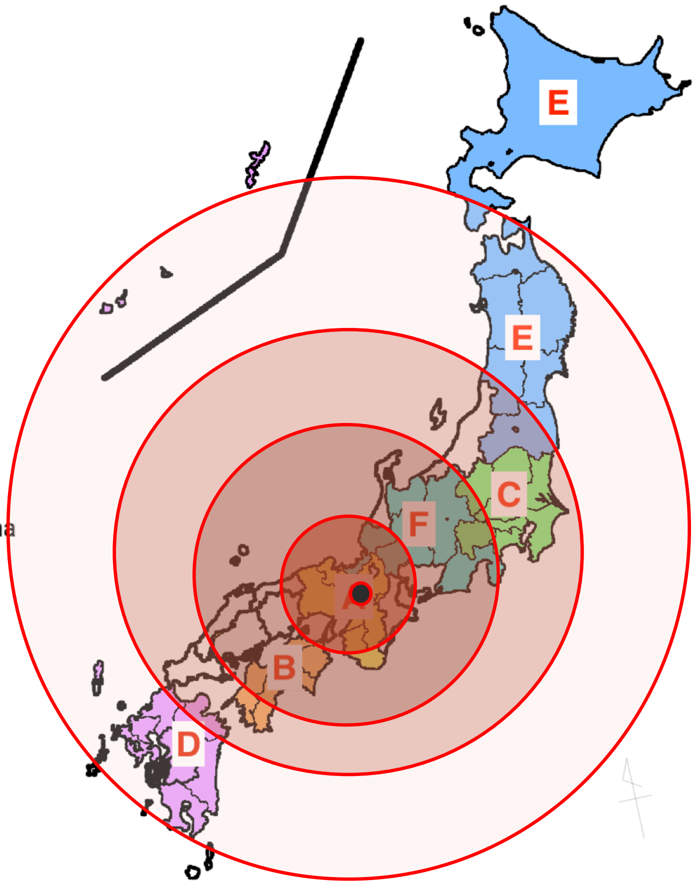
- Dialectological researcher said that dialects spread from an **ancient capital city** to remote areas **concentrically** [Yanagida, 1980]



Analysis (Visualize dialect embeddings)

- A t-SNE projection of dialect embeddings follows dialectological typology
  - Though the distance between **D and E** is far away, similar dialects are used





Conclusions

- We presented Multi-dialect NMT system
  - **character-level** + **fixed-order + multilingual**
- The unified model that learns **similar** multiple dialects jointly is effective for multi-dialect translation
- We can observe similar relationships to the existing dialect typology in some dialects by analyzing similarity of the dialect embeddings

References

- [Johnson+, 2017] Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation, ACL2017
- [Yanagida, 1980] Kunio Yanagida. 1980. *“Kagyuko”*. Iwanami Shoten, Publishers.

Appendix

FAQ

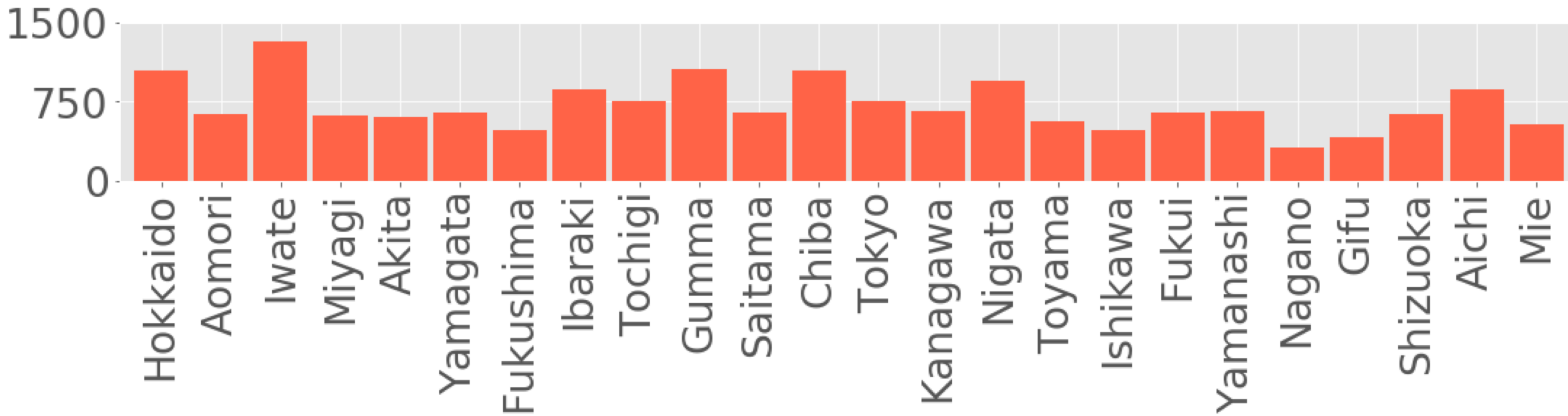
- Can you construct standard-to-dialect Multi NMT with the same model?
  - Yes (But the translation accuracy dropped)
  - Due to a weak language model in each target dialect
- Why is there not a “Multi SMT (w/ label)” setting ?
  - We could not devise an alternative to dialect labels in SMT

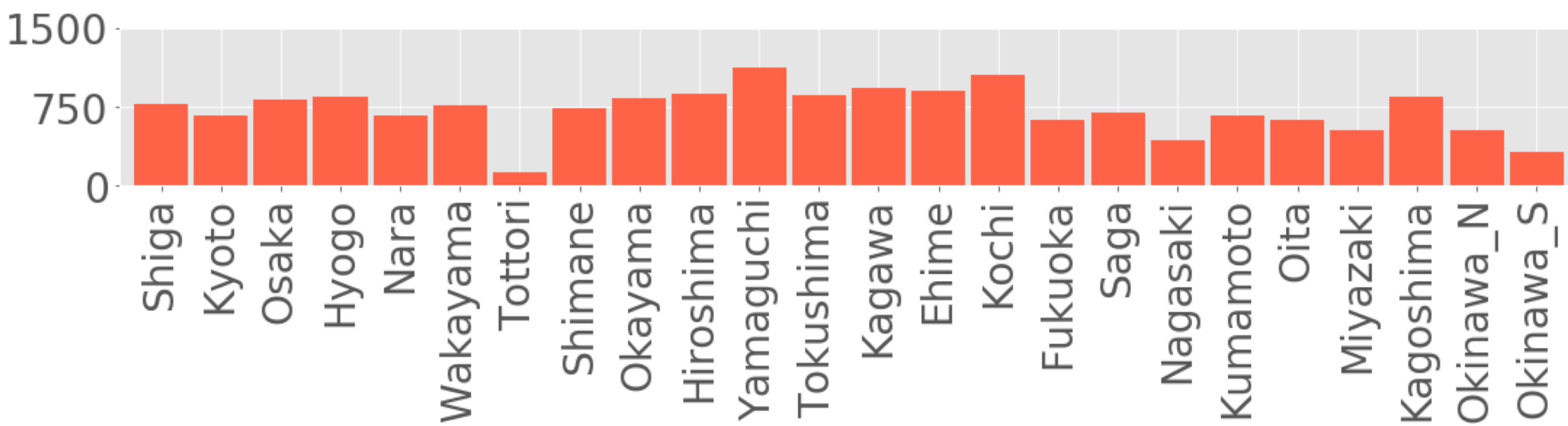
Is Word Order really unchanged?

- We checked 100 dialect-standard sentence pairs in all 48 dialects

  → **All the pairs are unchanged**

The Sentence Pairs of each dialect





Analysis (Translation Examples: Good)

- The output of Multi NMT completely agree with the reference

Example : "(We) had skated, aren't we?"	
Source (Aomori dialect)	shi ke - to / no ri su ta de ba - しけーと / のりすたでばー
Reference	su ke - to / no tsu ta de ha na i de su ka すけーと / のったではないですか
Multi NMT	su ke - to / no tsu ta de ha na i de su ka すけーと / のったではないですか
Sentence-Multi NMT	u ke i to / no ri shi ta de ha na i de su ka うけいと / のりしたではないですか
Multi NMT (w/o labels)	shi ke - to / no tsu ta de ha na i de su ka しけーと / のったではないですか
Multi SMT (w/o labels)	su ke - to / no tsu ta de ha na i de su すけーと / のったではないです

Analysis (Translation Examples: Bad)

- Multi NMT could not translate **too rare word**

Example : “(I) ran a horse”	
Source (Okinawa dialct)	ma - / pa ra - chi ya - まー / **ぱらーちやー**
Reference	u ma / ha shi ra se te ne うま / はしらせてね
Multi NMT	u ma / ha ra de ha うま / はらでは
Sentence-Multi NMT	a a / ha ra u shi ya ああ / はらうしゃ
Multi NMT (w/o labels)	ma a / ha na shi te ha まあ / はなしては
Multi SMT (w/o labels)	ma a / pa ra - to ne まあ / ぱらーとね