NAIST's Machine Translation Systems
for IWSLT 2020 Conversational Speech Translation Task

Ryo Fukuda[1], Katsuhito Sudoh[1], and Satoshi Nakamura[1,2]

[1]Nara Institute of Science and Technology
[2]AIP Center, RIKEN, Japan

Brief Overview

Challenge track: Conversational Speech Translation

Translation task from disfluent Spanish to fluent English

- Includes speech-to-text and text-to-text translation subtask

Motivation: Tackle two problems on text-to-text NMT

1. Low-resource translation
2. Noisy input sentences
   - fillers, hesitations, self-corrections, ASR errors, ...

Proposal: Domain adaptation using style transfer

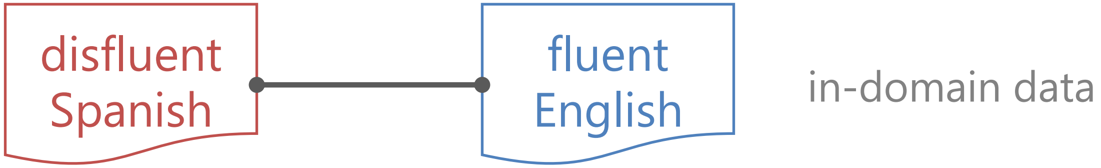
- transfer the **styles** of out-of-domain data to be like in-domain data, and them performed domain adaptation

Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Summary

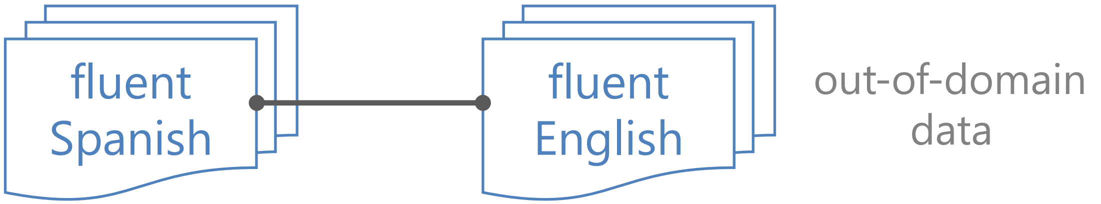
Motivation

The “style” of task data (in-domain):



→ Ideally, augment data by using large corpus same style

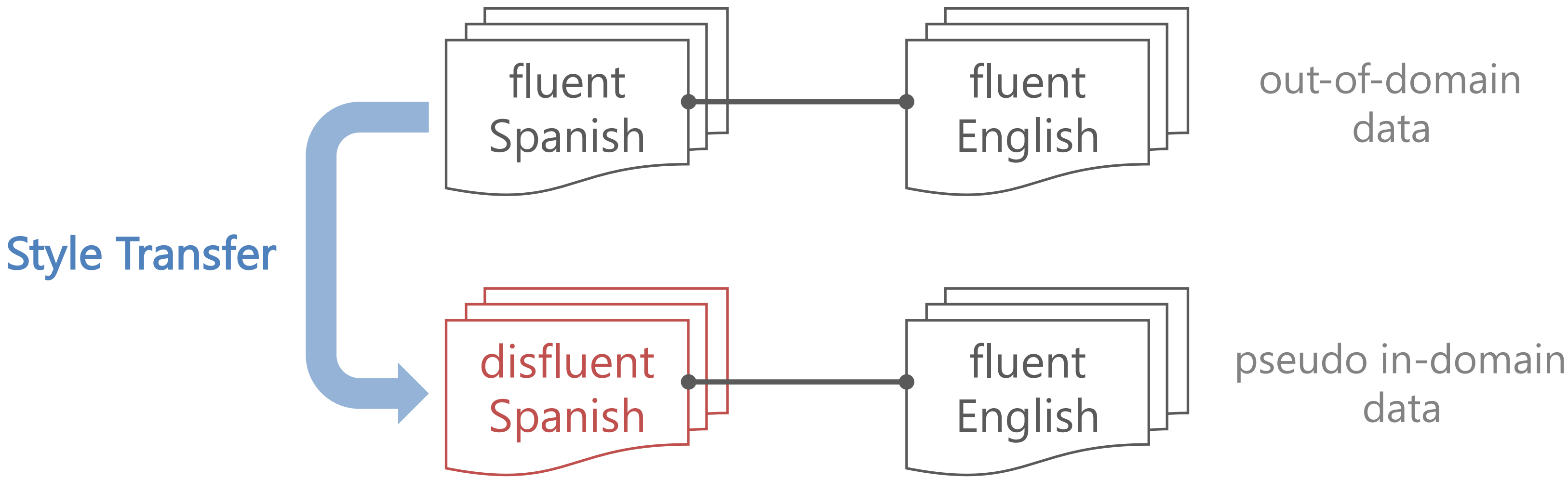
Large corpus available (out-of-domain):



→ Effects of training with them are limited

Motivation

Style transfer model: fluent to disfluent



- increase the the similarity between out-of-domain and in-domain data

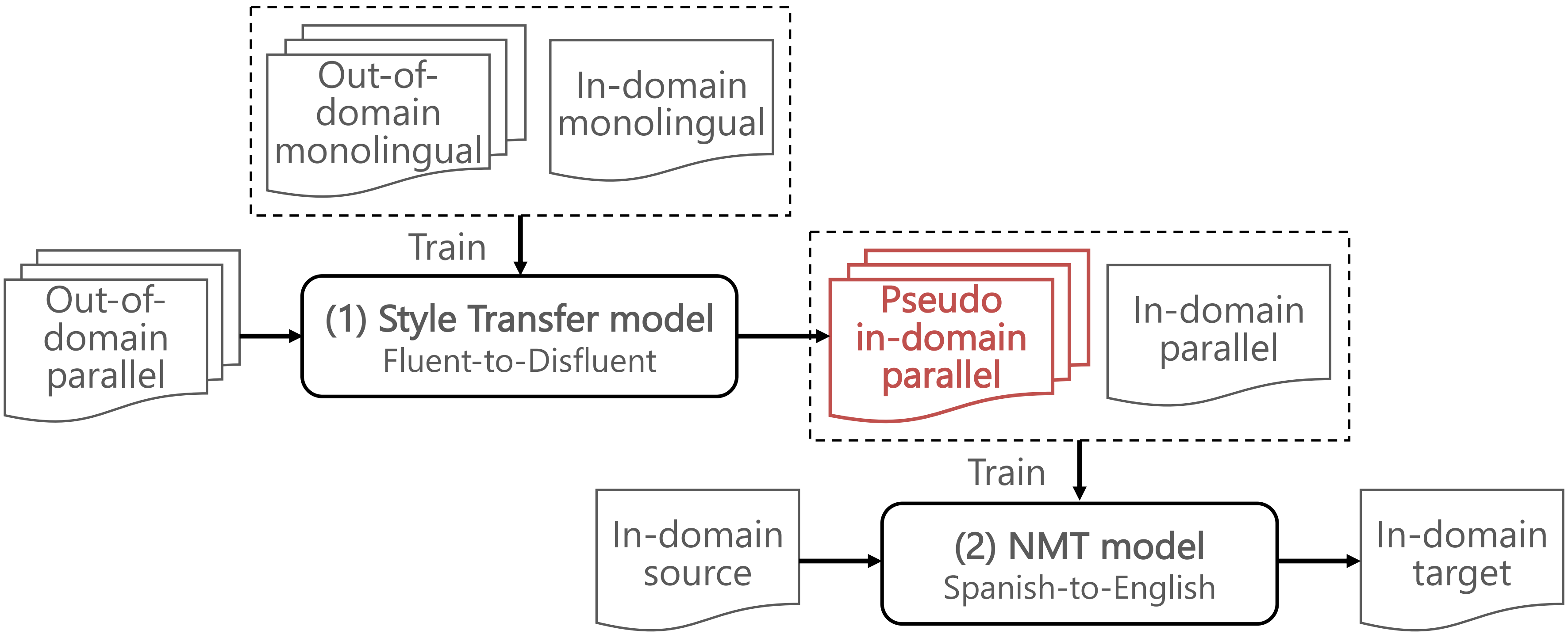
→ Enables effective domain adaptive training

Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Summary

Overview

Generate pseudo in-domain data and adapt it for NMT



(1) Style Transfer model

Transfer fluent input sentences of out-of-domain parallel data into disfluent styles

He estado durmiendo casi once horas — He's been sleeping for almost 11 hours

Style Transfer model
Fluent-to-Disfluent

Ya, ya, so, duerme casi once horas — He's been sleeping for almost 11 hours

Style Transfer model:

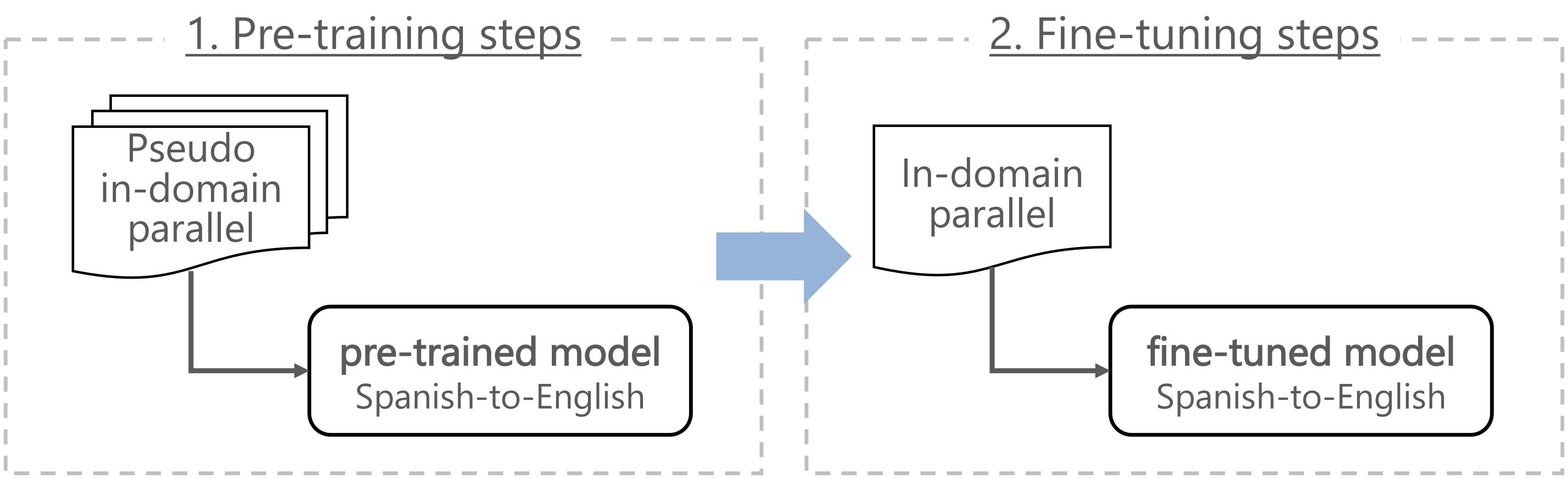
- based on Unsupervised NMT (Artetxe et al., 2018; Lample et al., 2018) with out-of-domain fluent data and in-domain disfluent data

(2) NMT model

Apply fine-tuning

- conventional domain adaptation methods of MT
- greatly improves the accuracy of low-resource domain-specific translation (Dakwale and Monz, 2017)

Learning steps for fine-tuning:



Outline

Datasets

- LDC Fisher Spanish speech with English translations (**Fisher**)
  - parallel in-domain data
  - disfluent Spanish to (fluent/disfluent) English
- United Nations Parallel Corpus (**UNCorpus**)
  - parallel out-of-domain data
  - fluent Spanish to fluent English

Data statistics

	# sentences
Fisher (in-domain)/Train	138,720
Dev	3,977
Test	3,641
UNCorpus (out-of-domain)/Train	1,000,000
Dev	4,000
Test	4,000

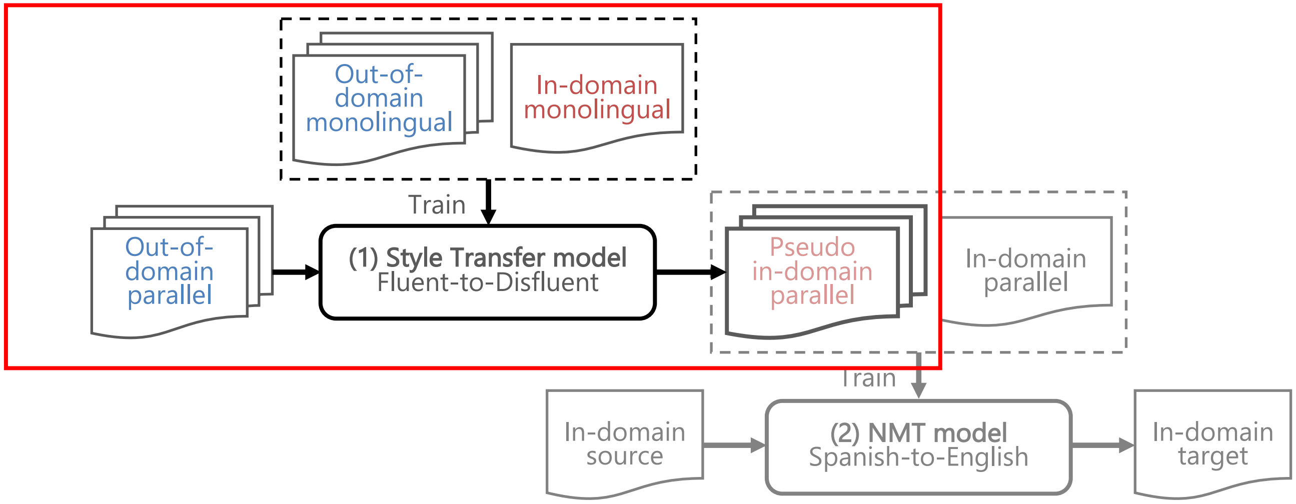
(1) Spanish Style Transfer

Data: Fisher (disfluent) and UNCorpus (fluent) Spanish data

Model: Unsupervised NMT (UNMT) based on Transformer

Evaluation:

- Estimate the **similarity** between domains by measuring the **perplexity** of 3-gram language model



(1) Spanish Style Transfer

Results

- reduced perplexity and number of unknown words by style transfer

Training data	perplexity	unknow words
Fisher	72.46	0
UNCorpus	589.81	5,173,539
Fisher-like UNCorpus	474.47	4,217,819

Examples of pseudo in-domain data (**Fisher-like UNCorpus**)

UNCorpus	Fisher-like UNCorpus
d conducta y disciplina	eh conducta y disciplina
c lista amplia de verificación para la autoevaluación	mhm lista amplia de verificación para la **la tele**

- Delete paragraph symbol
- Insert "Disfluency" (filler, repetition, missing words, ASR error, ..)

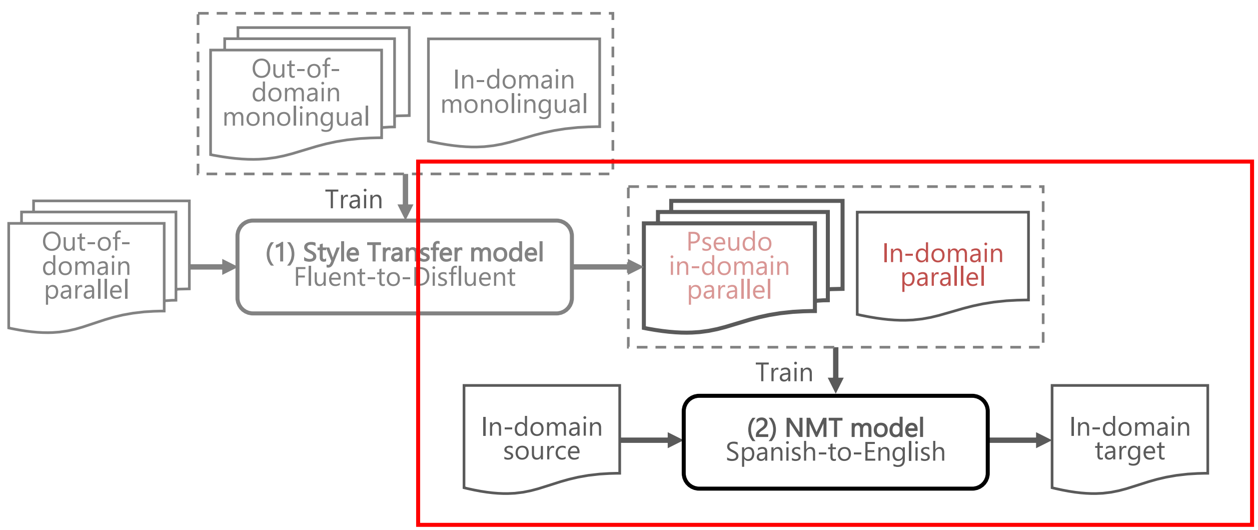
(2) NMT with Domain Adaptation

Data

- in-domain: 130K bilingual pairs of Fisher
- out-of-domain: 1M of UNCorpus or Fisher-like UNCorpus

Model: Transformer (almost follow the *transformer_base* settings)

Evaluation: calculated the BLEU scores with sacreBLEU



(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

BLEU scores of trained NMT models
for Disfluent Spanish to Fluent English

System		Fisher/test
Single Training	Fisher	14.8
	UNCorpus	7.8
	Fisher-like UNCorpus	6.7
Fine-tuning	UNCorpus + Fisher	18.3
	Fisher-like UNCorpus + Fisher	18.5

- Domain adaptation training outperformed the baseline
- slightly improved by using the pseudo in-domain data

(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

BLEU scores of trained NMT models
for Disfluent Spanish to Fluent English

System		Fisher/test
Single Training	Fisher	14.8
	UNCorpus	7.8
	Fisher-like UNCorpus	6.7
Fine-tuning	UNCorpus + Fisher	**18.3**
	Fisher-like UNCorpus + Fisher	18.5

+3.5

- Domain adaptation training outperformed the baseline
- slightly improved by using the pseudo in-domain data

(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

BLEU scores of trained NMT models
for Disfluent Spanish to Fluent English

System		Fisher/test
Single Training	Fisher	14.8
	UNCorpus	7.8
	Fisher-like UNCorpus	6.7
Fine-tuning	UNCorpus + Fisher	18.3
	Fisher-like UNCorpus + Fisher	**18.5**

+0.2

- Domain adaptation training outperformed the baseline
- slightly improved by using the pseudo in-domain data

(2) NMT with Domain Adaptation

Results (1/2) – Effect of Style Transfer

BLEU scores of trained NMT models
for Disfluent Spanish to Fluent English

System		Fisher/test
Single Training	Fisher	14.8
	UNCorpus	7.8
	Fisher-like UNCorpus	6.7
Fine-tuning	UNCorpus + Fisher	18.3
	Fisher-like UNCorpus + Fisher	18.5

-1.1

- Domain adaptation training outperformed the baseline
- slightly improved by using the pseudo in-domain data

(2) NMT with Domain Adaptation

Results (2/2) – Fluent vs Disfluent references

"**Fisher (disfluent)**" did not use Fisher's fluent references but instead used disfluent references

System	Fisher/test
Fisher (fluent)	14.8
UNCorpus + Fisher (fluent)	18.3
Fisher-like UNCorpus + Fisher (fluent)	**18.5**
Fisher (disfluent)	11.6
UNCorpus + Fisher (disfluent)	15.2
Fisher-like UNCorpus + Fisher (disfluent)	15.6

-3.2

-3.1

-2.9

- models trained with Fisher's original disfluent references had about 3 points lower BLEU

Outline

1. Introduction
2. System Description
3. Experiments
4. Discussion
5. Summary

Effect of Style Transfer

The use of pseudo in-domain data improved accuracy, but

- there was no significant improvement
- was worse in the pre-training phase

An example of style transferred sentence:

nueva york 1 a 12 de junio de 2015 (original)
nueva york oh a mi eh de de de de (generated)

- some sentences lost the meaning of the sentence
- style transfer constrains may be too strong

→ **This problem may be mitigated by a model that can control the trade-off between style transfer and content preservation**

Fluent vs Disfluent References

The model trained using Fisher's original disfluent data had a BLEU score of about 3 points lower than the model trained using fluent data.

→ **by removing the disfluency of reference sentences improves the BLEU by about three points for all the learning strategies we tried**

- the use of large out-of-domain data with **fluent reference sentences** did not mitigate this problem

Style of the sentence has an impact on the translation accuracy

Summary

Translation accuracy was improved

- by domain adaptation (+3.7)
- by style transfer of out-of-domain (+0.4)
  - effect was limited due to parallel data quality degradation

Future work

pursue a style transfer that does not reduce the quality of the parallel data