

# Multilingual Pre-training with Language and Task Adaptation for Multilingual Text Style Transfer

Huiyuan Lai  
[h.lai@rug.nl](mailto:h.lai@rug.nl)

Antonio Toral  
[a.toral.ruiz@rug.nl](mailto:a.toral.ruiz@rug.nl)

Malvina Nissim  
[m.nissim@rug.nl](mailto:m.nissim@rug.nl)

CLCG, University of Groningen / The Netherlands



# Text Style Transfer

- Reformulate a sentence in the **desired style** while preserving the original **content**

**Informal:** *that is just my gut feeling*



**Formal:** *That is my personal opinion.*

# Text Style Transfer: Task and Data

## ➤ Formality Transfer

- GYAFC [Rao & Tetreault, 2018] — [English](#) informal-formal pairs

# Multilingual Text Style Transfer

## ➤ Formality Transfer

- GYAFC [Rao & Tetreault, 2018] — **English** informal-formal pairs
- XFORMAL [Briakou et al., 2021] — **Multilingual** informal-formal pairs  
(Italian, Brazilian Portuguese, French)

# Multilingual Text Style Transfer

## ➤ Formality Transfer

- GYAFC [Rao & Tetreault, 2018] — **English** informal-formal pairs
- XFORMAL [Briakou et al., 2021] — **Multilingual** informal-formal pairs

(Italian, Brazilian Portuguese, French)

### - EVALUATION

1,000 informal sentences (paired with four formal human rewrites)

### - TRAINING

Pseudo-parallel data in each language, obtained via machine translating GYAFC

# Multilingual Text Style Transfer

## ➤ Formality Transfer

- GYAFC [Rao & Tetreault, 2018] — **English** informal-formal pairs
- XFORMAL [Briakou et al., 2021] — **Multilingual** informal-formal pairs

(Italian, Brazilian Portuguese, French)

### - EVALUATION

1,000 informal sentences (paired with four formal human rewrites)

### - TRAINING

Pseudo-parallel data in each language, obtained via machine translating GYAFC

- Models trained on translated parallel data do **not** outperform a simple rule-based system.

# Multilingual Formality Transfer

- Multilingual Pre-trained Model — mBART-50
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer
- Language Adaptation Training
  - Language underrepresentation in pre-training
- Task Adaptation Training
  - Make the model adapt to the task of formality transfer

# Multilingual Formality Transfer

- **Multilingual Pre-trained Model — mBART-50**
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer
- **Language Adaptation Training**
  - Language underrepresentation in pre-training
- **Task Adaptation Training**
  - Make the model adapt to the task of formality transfer



# Multilingual Formality Transfer

- **Multilingual Pre-trained Model** — mBART-50
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer [Johnson et al., 2017; Hasan et al., 2021]
- **Language Adaptation Training**
  - Language underrepresentation in pre-training
- **Task Adaptation Training**
  - Make the model adapt to the task of formality transfer

# Multilingual Formality Transfer

- **Multilingual Pre-trained Model** — mBART-50
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer [Johnson et al., 2017; Hasan et al., 2021]
- **Language Adaptation Training**
  - Language underrepresentation in pre-training
- **Task Adaptation Training**
  - Make the model adapt to the task of formality transfer

# Multilingual Formality Transfer

- Multilingual Pre-trained Model — **mBART-50**
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer [Johnson et al., 2017; Hasan et al., 2021]
- Language Adaptation Training
  - Language underrepresentation in pre-training
- Task Adaptation Training
  - Make the model adapt to the task of formality transfer

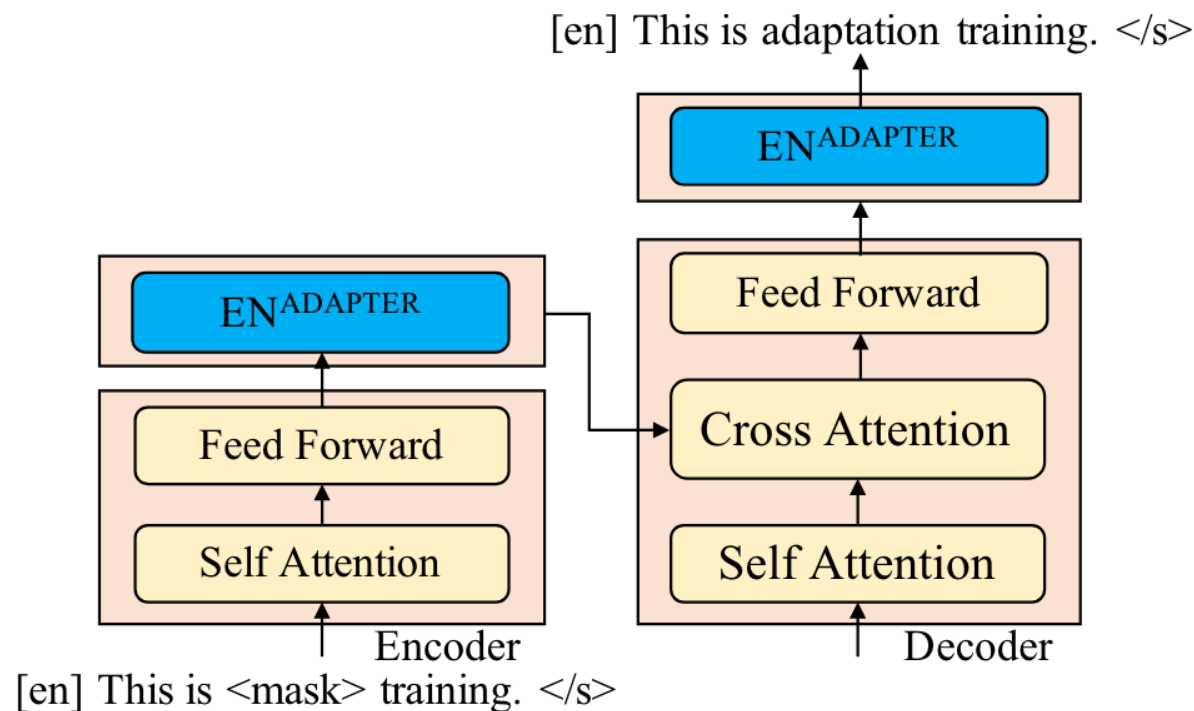
# Multilingual Formality Transfer

- **Multilingual Pre-trained Model — mBART-50**
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer [Johnson et al., 2017; Hasan et al., 2021]
- **Language Adaptation Training**
  - Language underrepresentation in pre-training
- **Task Adaptation Training**
  - Make the model adapt to the task of formality transfer

# Multilingual Formality Transfer

- **Multilingual Pre-trained Model — mBART-50**
  - A large pre-trained model helps preserve content [Lai et al., 2021]
  - Cross-lingual knowledge transfer [Johnson et al., 2017; Hasan et al., 2021]
- **Language Adaptation Training**
  - Language underrepresentation in pre-training
- **Task Adaptation Training**
  - Make the model adapt to the task of formality transfer

# Multilingual Formality Transfer: Language Adaptation

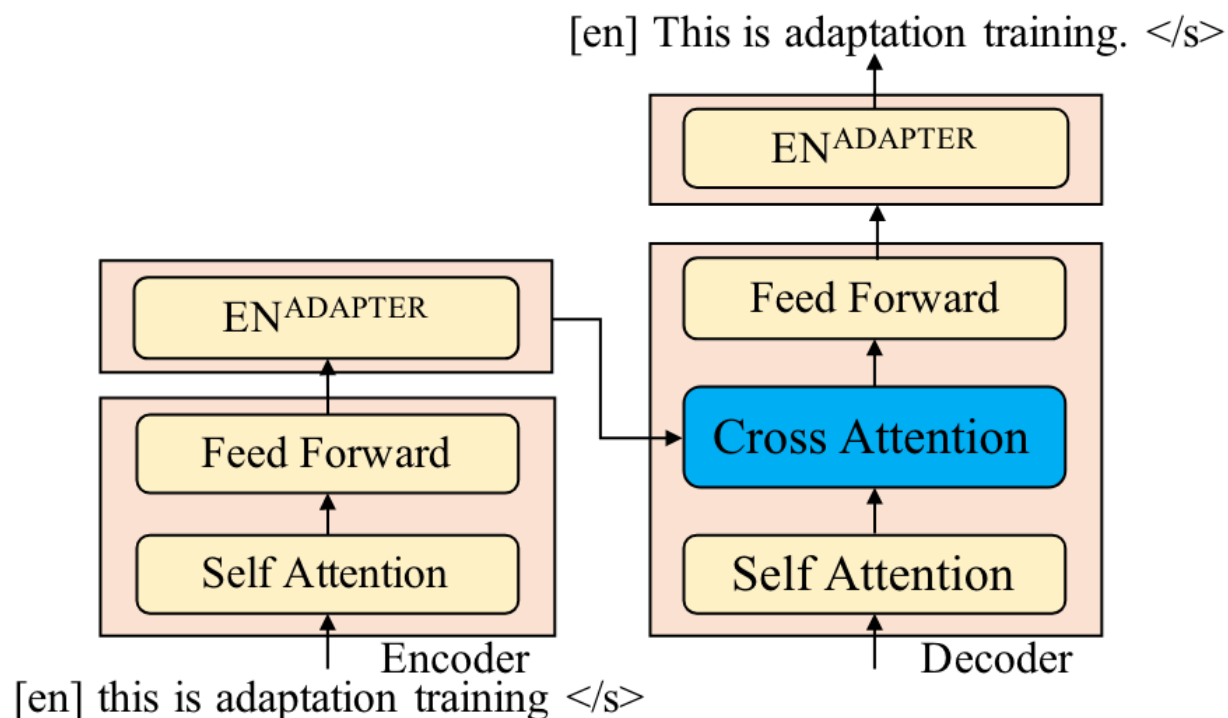


- Adapter layer trained with monolingual data to adapt to target language
- Noise function  $g$  masks 30% of the words
- Only the parameters of the adaptation module are updated

Language adaptation training with monolingual data<sup>[1]</sup>

[1] <http://data.statmt.org/news-crawl/>

# Multilingual Formality Transfer: Task Adaptation

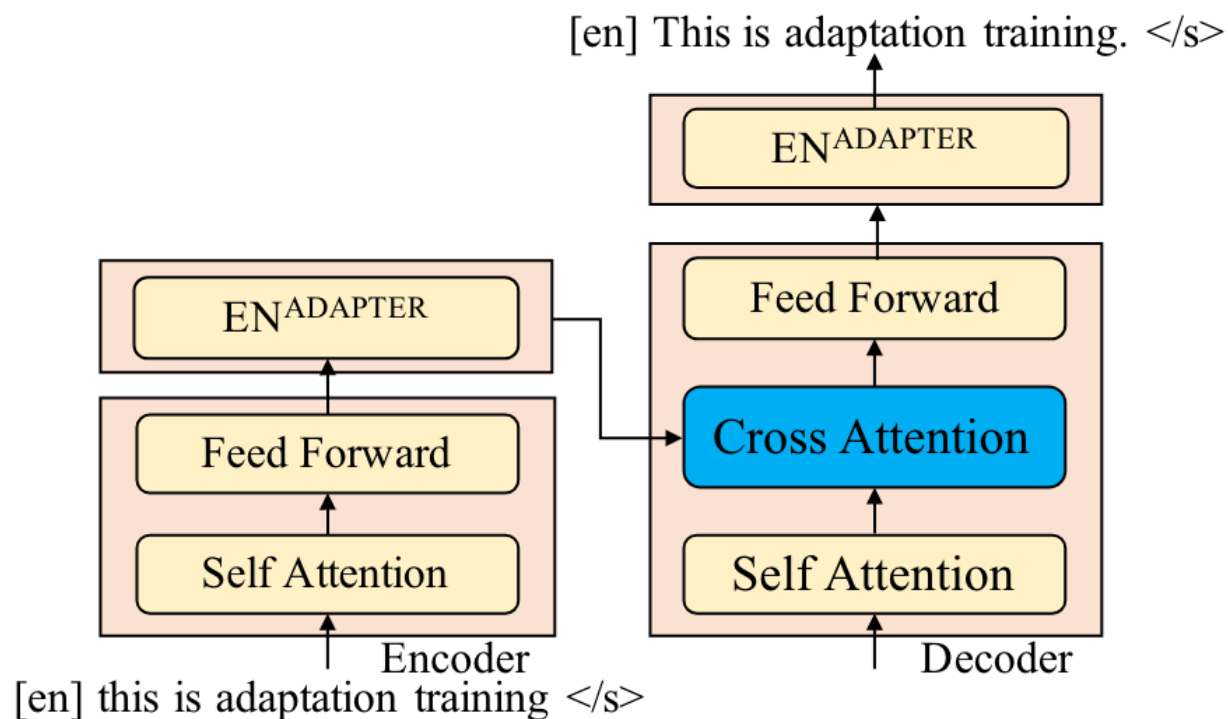


- Fine-tune model on the auxiliary English parallel data
- Only the parameters of the adaptation module (Cross Attention) are updated

Language adaptation training with monolingual data<sup>[1]</sup>

[1] <http://data.statmt.org/news-crawl/>

# Multilingual Formality Transfer: Task Adaptation



- The task adaptation module is from **English model** (EN cross-attn)
- Fine-tune the model of target language with **English parallel data** (EN data)

Language adaptation training with monolingual data<sup>[1]</sup>

[1] <http://data.statmt.org/news-crawl/>



# Experiments: Evaluation Metrics

## ➤ Content Preservation

- BLEU
- COMET [Rei et al., 2020] (check it in our paper)

## ➤ Style Strength

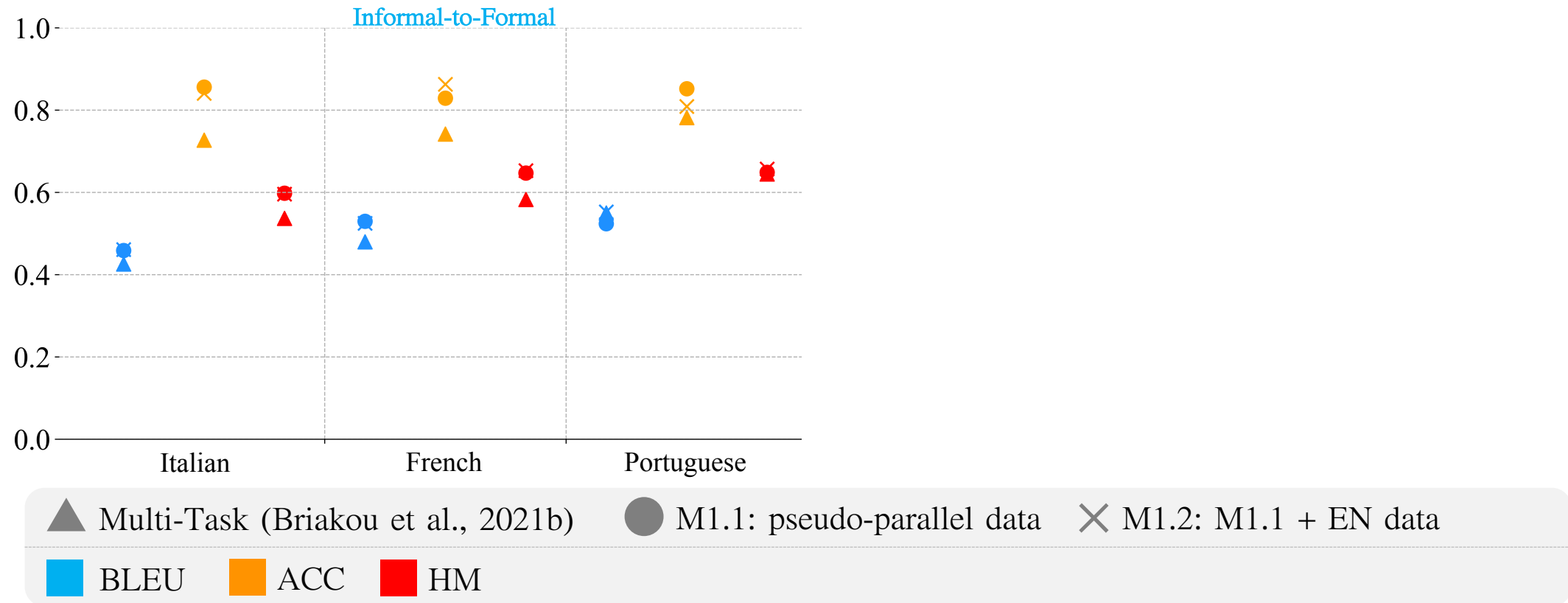
- mBERT based style classifier (ACC)
- XLM-R based style regressor [Briakou et al., 2021] (check it in our paper)

## ➤ Overall Score

- The harmonic mean (HM) of style accuracy and BLEU

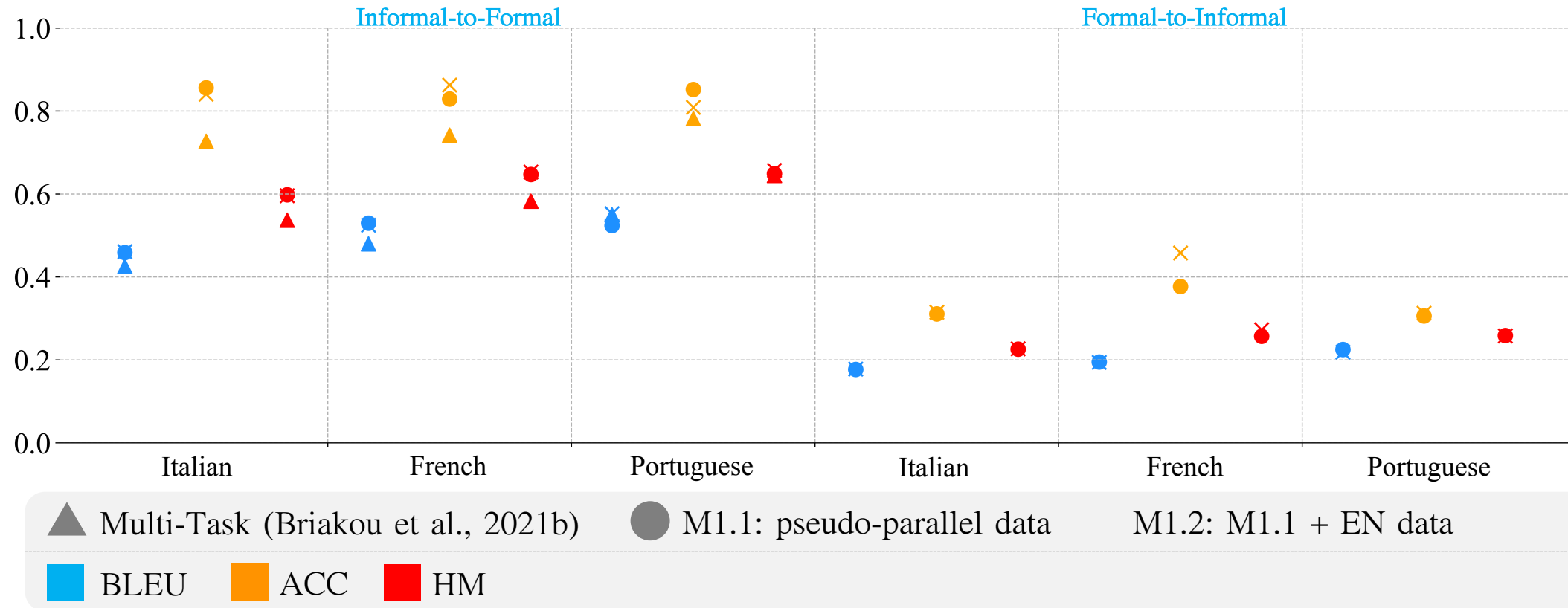
# Experiments: Results

➤ Data in target language machine translated from English



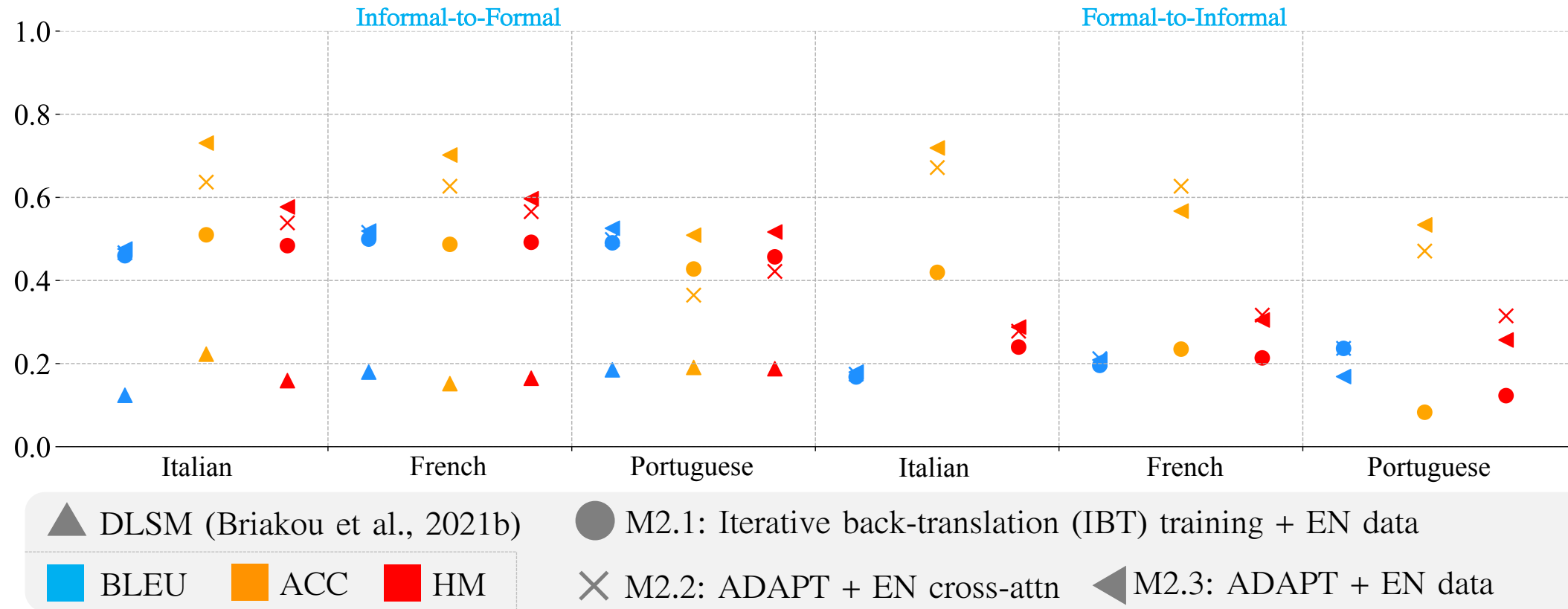
# Experiments: Results

➤ Data in target language machine translated from English



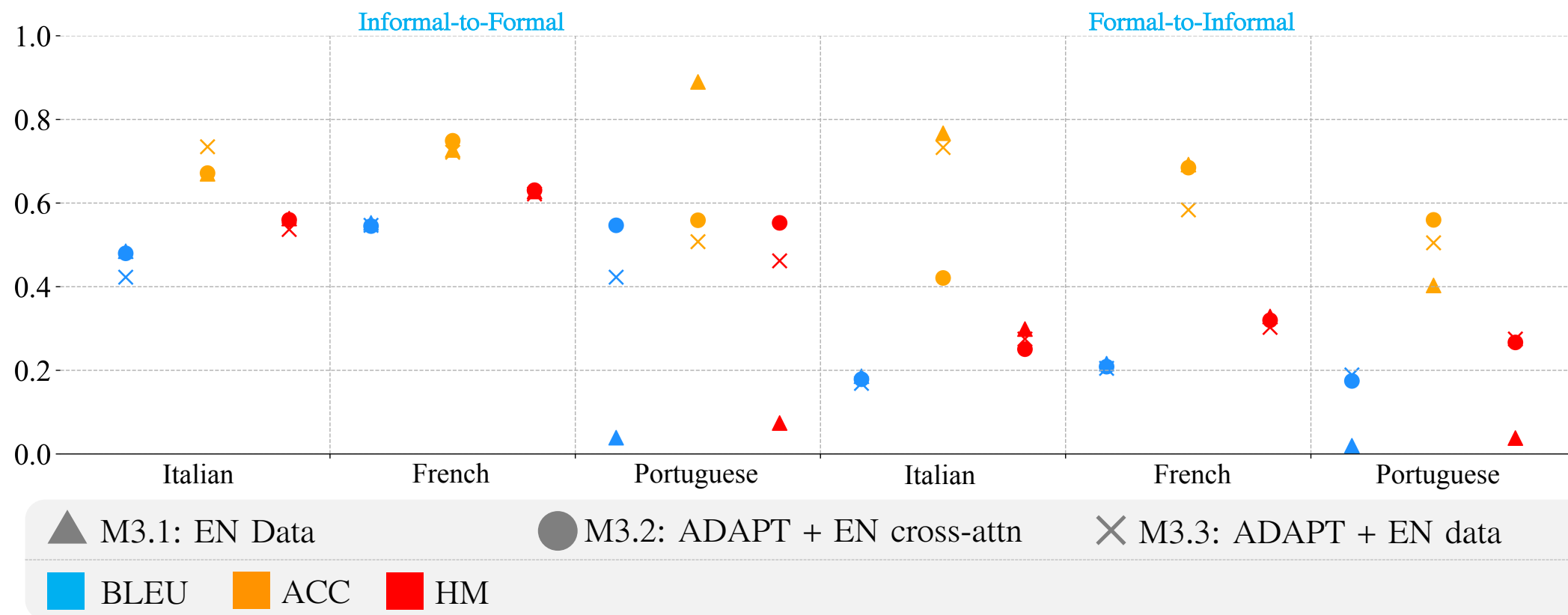
# Experiments: Results

## ➤ Non-parallel style data in target language



# Experiments: Results

➤ No style data in target language



# Summary of Results: Informal-to-Formal

## ➤ Machine Translated Data

- Fine-tuning mBART yields the best overall performance in informal-to-formal
- The results for the formal-to-informal are considerably worse in style accuracy

## ➤ Adaptation Strategies

- Achieve the highest score of content for Italian and French in informal-to-formal
- Achieve the best overall performance in formal-to-informal
- This can be applied to other style transfer tasks as well as to other languages

# Summary of Results: Formal-to-Informal

## ➤ Machine Translated Data

- Fine-tuning mBART yields the best overall performance in informal-to-formal
- The results for the formal-to-informal are considerably worse in style accuracy

## ➤ Adaptation Strategies

- Achieve the highest score of content for Italian and French in informal-to-formal
- Achieve the best overall performance in formal-to-informal
- This can be applied to other style transfer tasks as well as to other languages

# Summary of Results

## ➤ Machine Translated Data

- Fine-tuning mBART yields the best overall performance in informal-to-formal
- The results for the formal-to-informal are considerably worse in style accuracy

## ➤ Adaptation Strategies

- Achieve the highest score of content for Italian and French in informal-to-formal
- Achieve the best overall performance in formal-to-informal
- This can be applied to other style transfer tasks as well as to other languages



What's the Difference Between the Two Transfer Directions?

# What's the Difference Between the Two Transfer Directions?

## ➤ Setting (a)

Source (inf.): *that is just my gut feeling*



Reference (for.): *That is my personal opinion.*



Swapping

## ➤ Setting (b)

Source (for.): *That is my personal opinion.*



Reference (inf.): *that is just my gut feeling*

# What's the Difference Between the Two Transfer Directions?

MODEL	ITALIAN		FRENCH		PORTUGUESE	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	INFORMAL→FORMAL (setting (a))					
INPUT	0.176	0.078	0.198	-0.019	0.244	0.217
M1.1	0.196	0.170	0.234	0.133	0.269	0.282
	FORMAL→INFORMAL (setting (b))					
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	FORMAL→INFORMAL (setting (b))					
INPUT	0.174	0.364	0.196	0.277	0.243	0.463
M1.1	0.194	0.326	0.201	0.239	0.226	0.371

Multilingual formality transfer on content preservation. Setting (a) uses the original test set for each direction; (b) uses the test set of the opposite direction, swapping sources and references.

## ➤ Setting (a)

Source (inf.): *that is just my gut feeling*



Reference (for.): *That is my personal opinion.*



Swapping

## ➤ Setting (b)

Source (for.): *That is my personal opinion.*

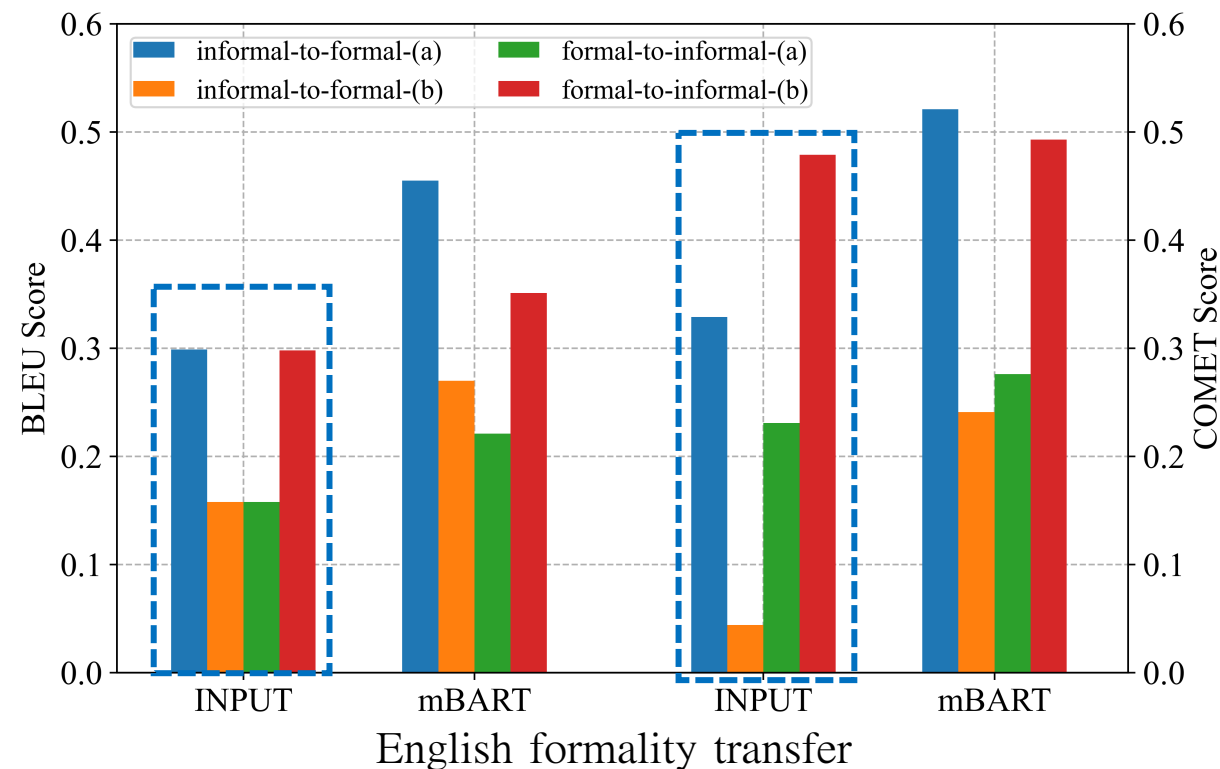


Reference (inf.): *that is just my gut feeling*

# What's the Difference Between the Two Transfer Directions?

MODEL	ITALIAN		FRENCH		PORTUGUESE	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	INFORMAL→FORMAL (setting (a))					
INPUT	0.176	0.078	0.198	-0.019	0.244	0.217
M1.1	0.196	0.170	0.234	0.133	0.269	0.282
MODEL	ITALIAN		FRENCH		PORTUGUESE	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	FORMAL→INFORMAL (setting (b))					
INPUT	0.174	0.364	0.196	0.277	0.243	0.463
M1.1	0.194	0.326	0.201	0.239	0.226	0.371

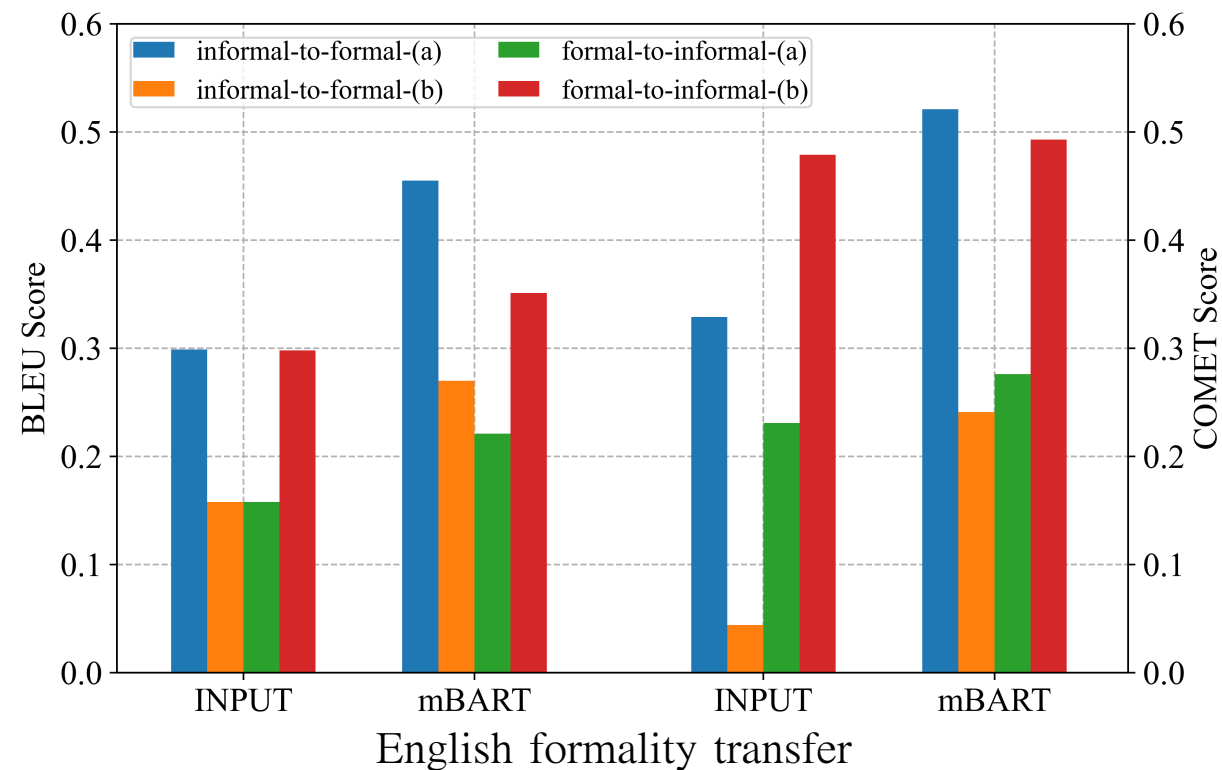
Multilingual formality transfer on content preservation. Setting (a) uses the original test set for each direction; (b) uses the test set of the opposite direction, swapping sources and references.



# What's the Difference Between the Two Transfer Directions?

MODEL	ITALIAN		FRENCH		PORTUGUESE	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	INFORMAL→FORMAL (setting (a))					
INPUT	0.176	0.078	0.198	-0.019	0.244	0.217
M1.1	0.196	0.170	0.234	0.133	0.269	0.282
	FORMAL→INFORMAL (setting (b))					
	BLEU	COMET	BLEU	COMET	BLEU	COMET
INPUT	0.174	0.364	0.196	0.277	0.243	0.463
M1.1	0.194	0.326	0.201	0.239	0.226	0.371

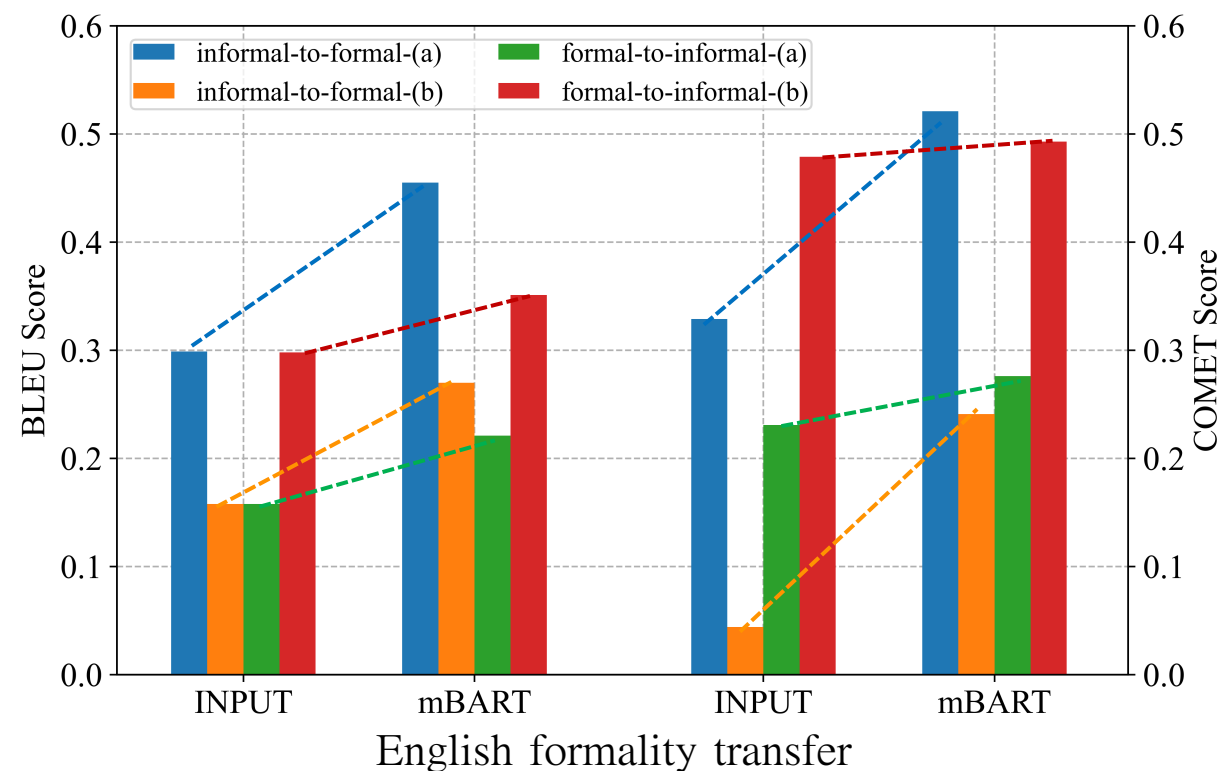
Multilingual formality transfer on content preservation. Setting (a) uses the original test set for each direction; (b) uses the test set of the opposite direction, swapping sources and references.



# What's the Difference Between the Two Transfer Directions?

MODEL	ITALIAN		FRENCH		PORTUGUESE	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	INFORMAL→FORMAL (setting (a))					
INPUT	0.176	0.078	0.198	-0.019	0.244	0.217
M1.1	0.196	0.170	0.234	0.133	0.269	0.282
	FORMAL→INFORMAL (setting (b))					
	BLEU	COMET	BLEU	COMET	BLEU	COMET
INPUT	0.174	0.364	0.196	0.277	0.243	0.463
M1.1	0.194	0.326	0.201	0.239	0.226	0.371

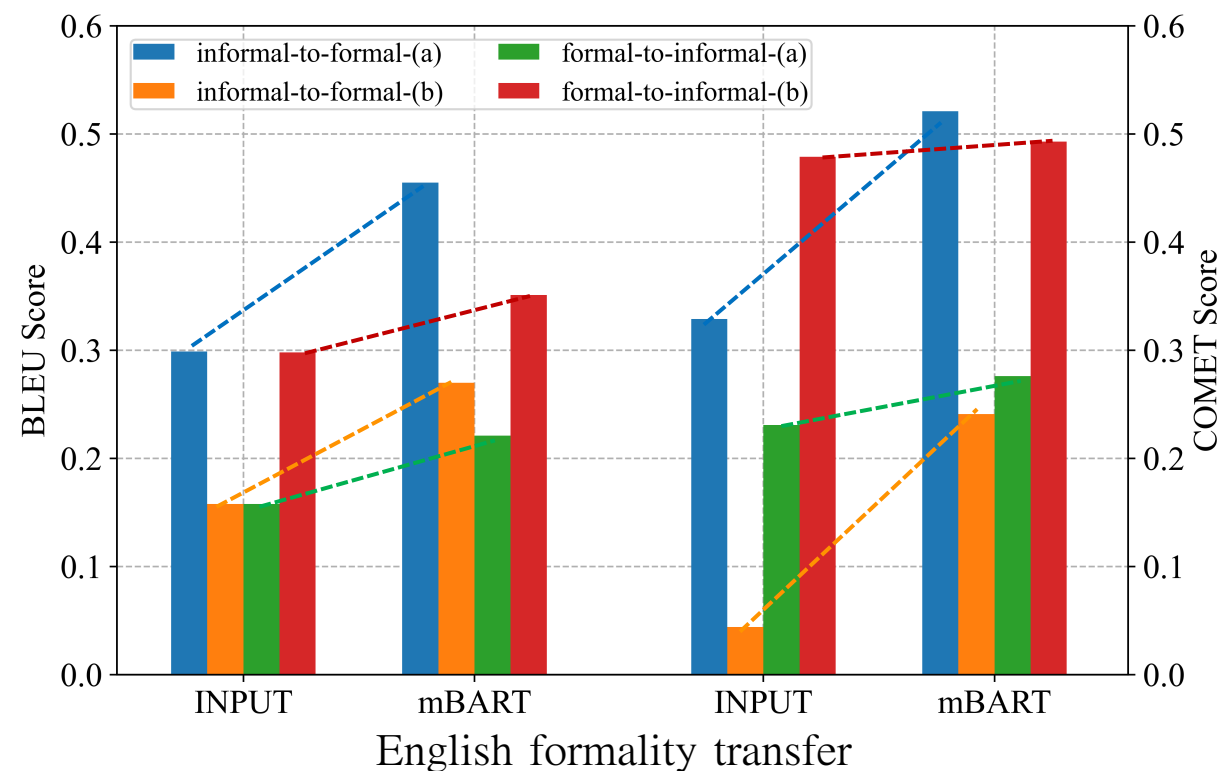
Multilingual formality transfer on content preservation. Setting (a) uses the original test set for each direction; (b) uses the test set of the opposite direction, swapping sources and references.



# What's the Difference Between the Two Transfer Directions?

MODEL	ITALIAN		FRENCH		PORTUGUESE	
	BLEU	COMET	BLEU	COMET	BLEU	COMET
	INFORMAL→FORMAL (setting (a))					
INPUT	0.176	0.078	0.198	-0.019	0.244	0.217
M1.1	0.196	0.170	0.234	0.133	0.269	0.282
	FORMAL→INFORMAL (setting (b))					
	BLEU	COMET	BLEU	COMET	BLEU	COMET
INPUT	0.174	0.364	0.196	0.277	0.243	0.463
M1.1	0.194	0.326	0.201	0.239	0.226	0.371

Multilingual formality transfer on content preservation. Setting (a) uses the original test set for each direction; (b) uses the test set of the opposite direction, swapping sources and references.



Formal-to-informal transformation is harder than the opposite direction!

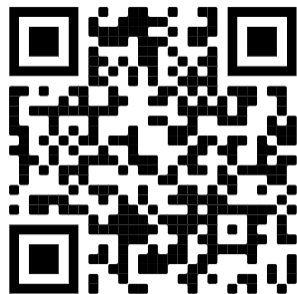
# Thanks for Your Attention!



@HuiyuanLai @\_atoral @MalvinaNissim @GroNlp



huiyuanlai.l@gmail.com



<https://arxiv.org/abs/2203.08552>



[github.com/laihuiyuan/multilingual-tst](https://github.com/laihuiyuan/multilingual-tst)