



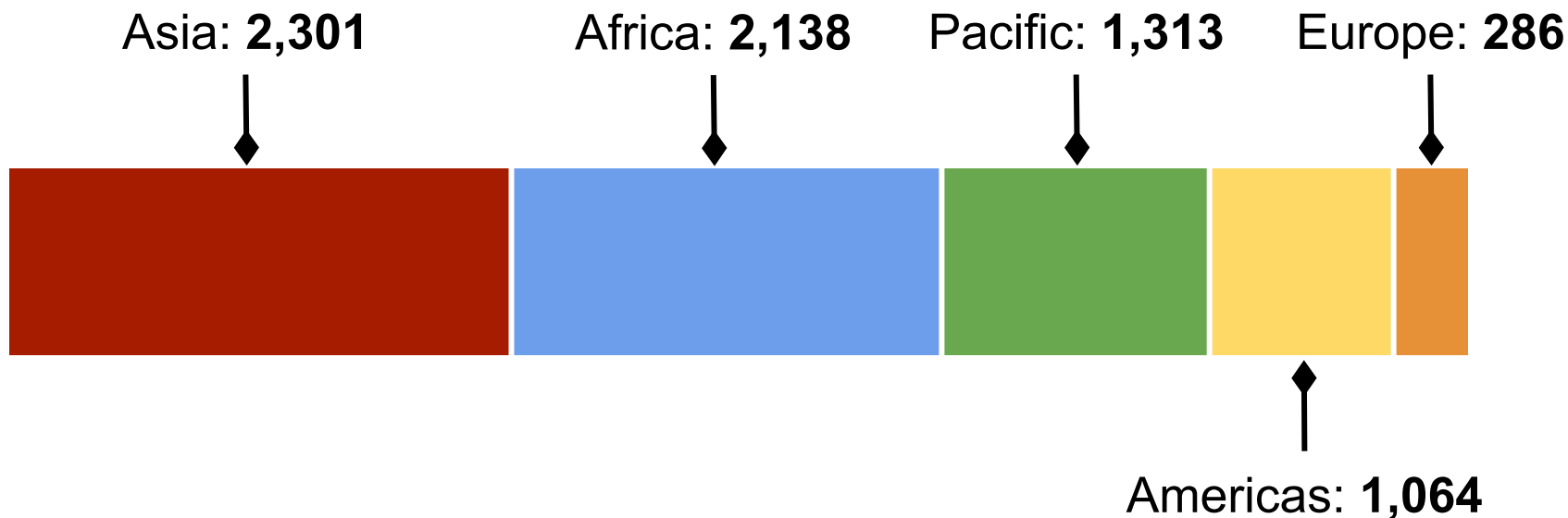
# Improving Zero-shot Translation of Low-Resource Languages

Surafel M. Lakew<sup>1,2</sup>, Quintino F. Lotito<sup>2</sup>, Matteo Negri<sup>1</sup>, Marco Turchi<sup>1</sup>, Marcello Federico<sup>1</sup>


<sup>1</sup>Fondazione Bruno Kessler | <sup>2</sup>University of Trento, Trento, Italy

# Machine Translation: why low-resource & zero-shot

There are at least **7,102** living languages in the world




# Machine Translation: why low-resource & zero-shot




**A very short  
Story of  
mine**

# Machine Translation: why low-resource & zero-shot

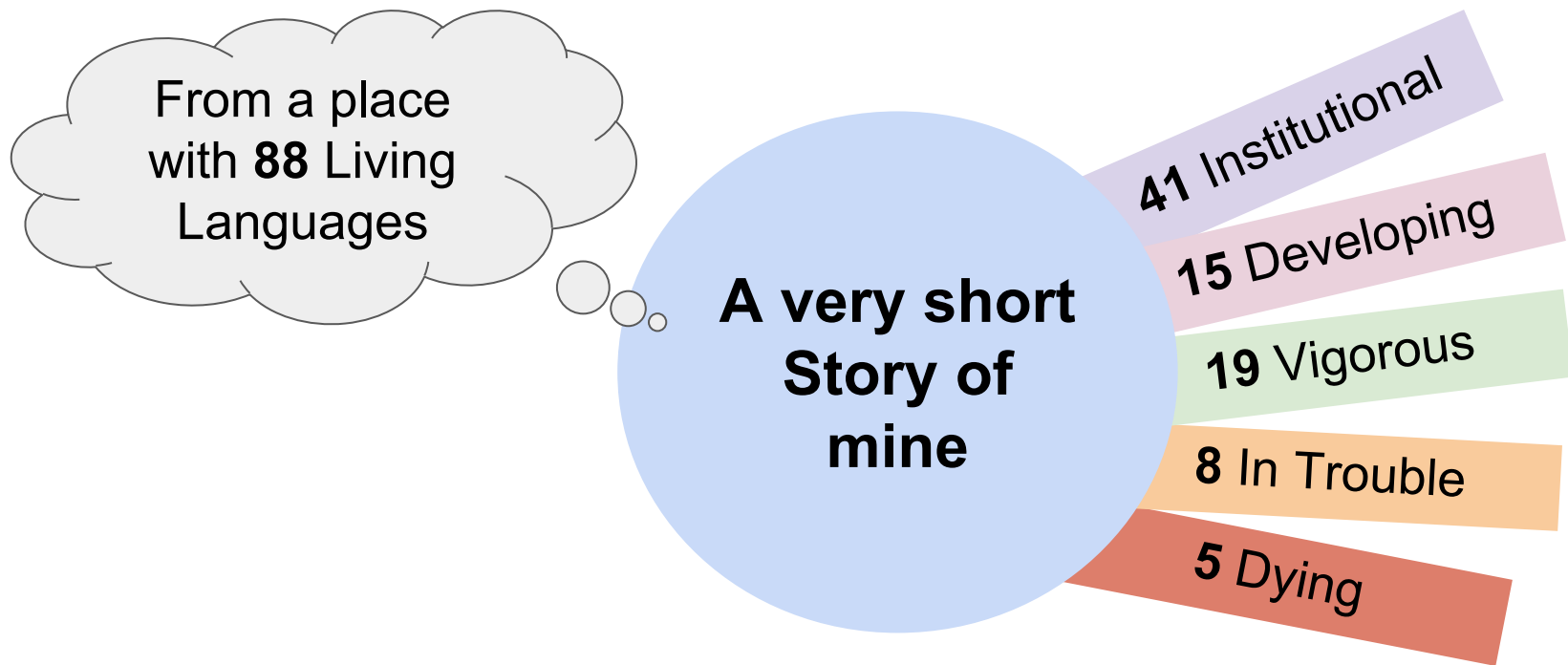


From a place  
with **88** Living  
Languages




**A very short  
Story of  
mine**

# Machine Translation: why low-resource & zero-shot

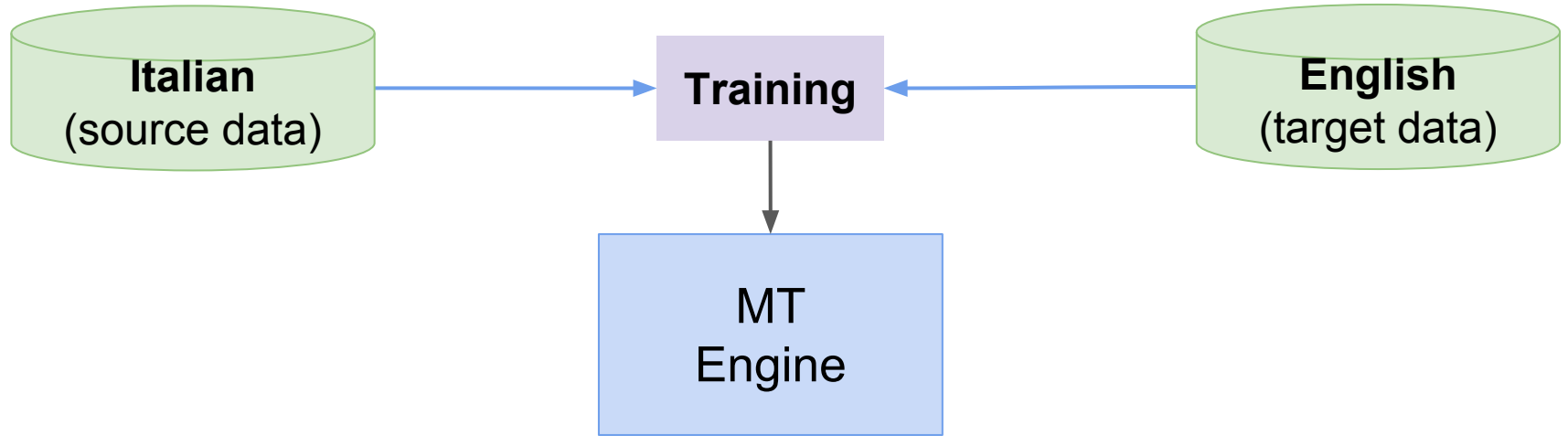


# Machine Translation

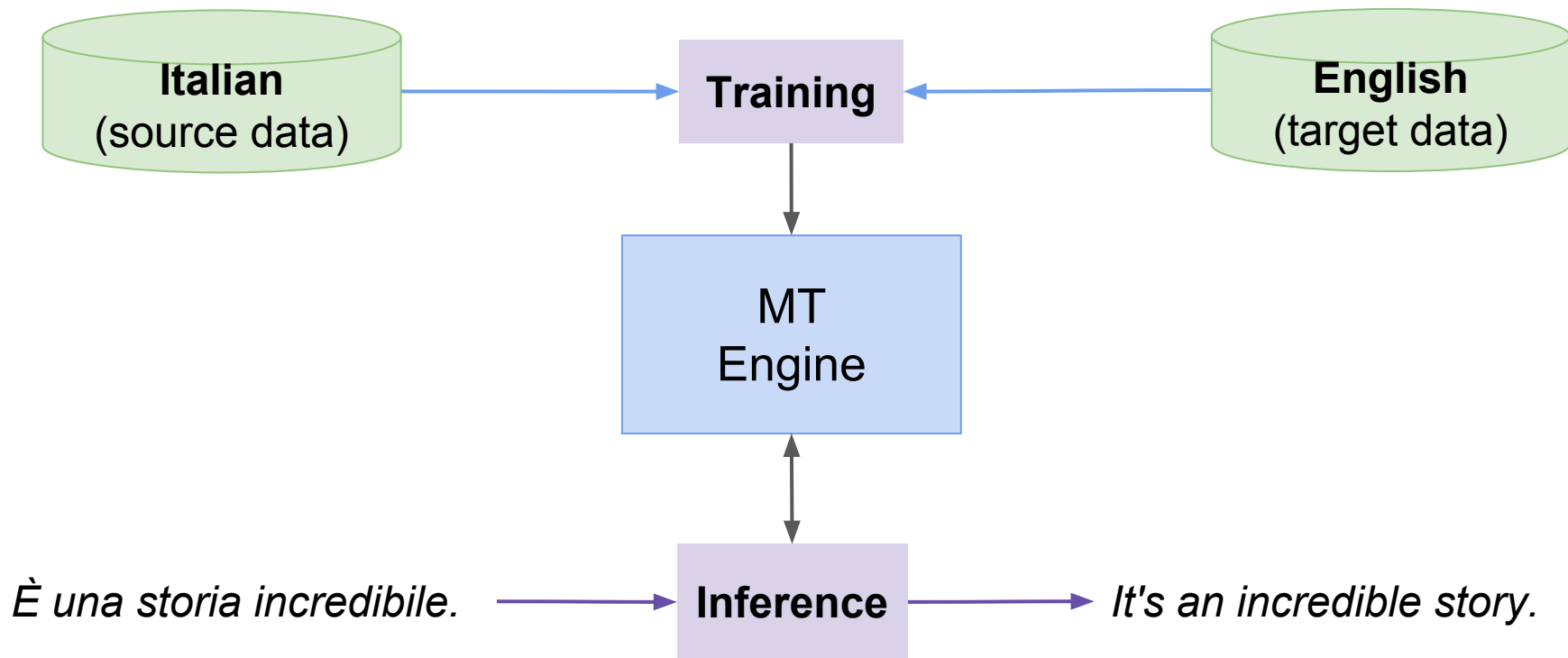


MT  
Engine

# Machine Translation

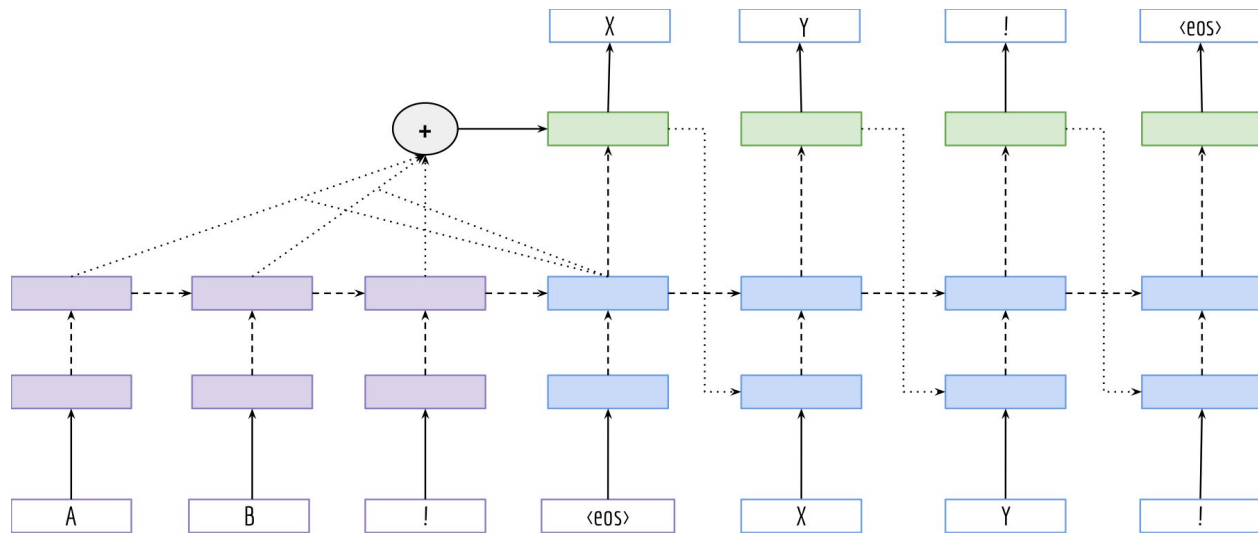


# Machine Translation

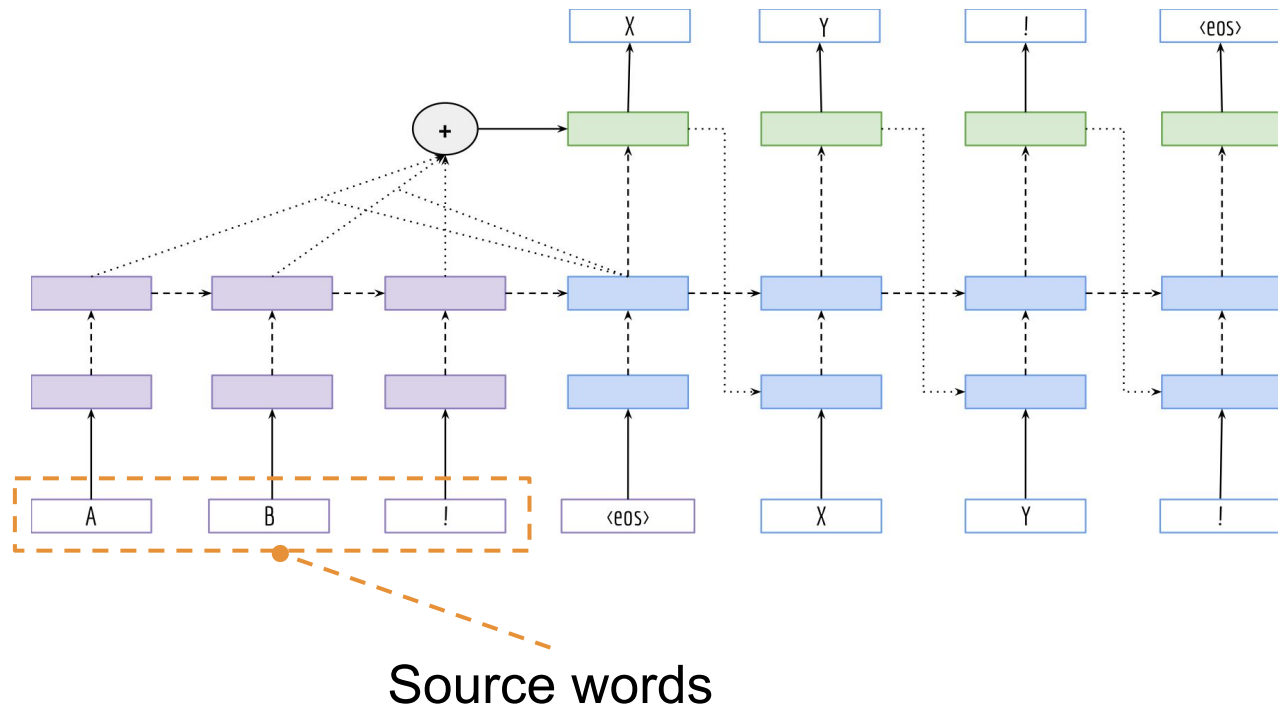




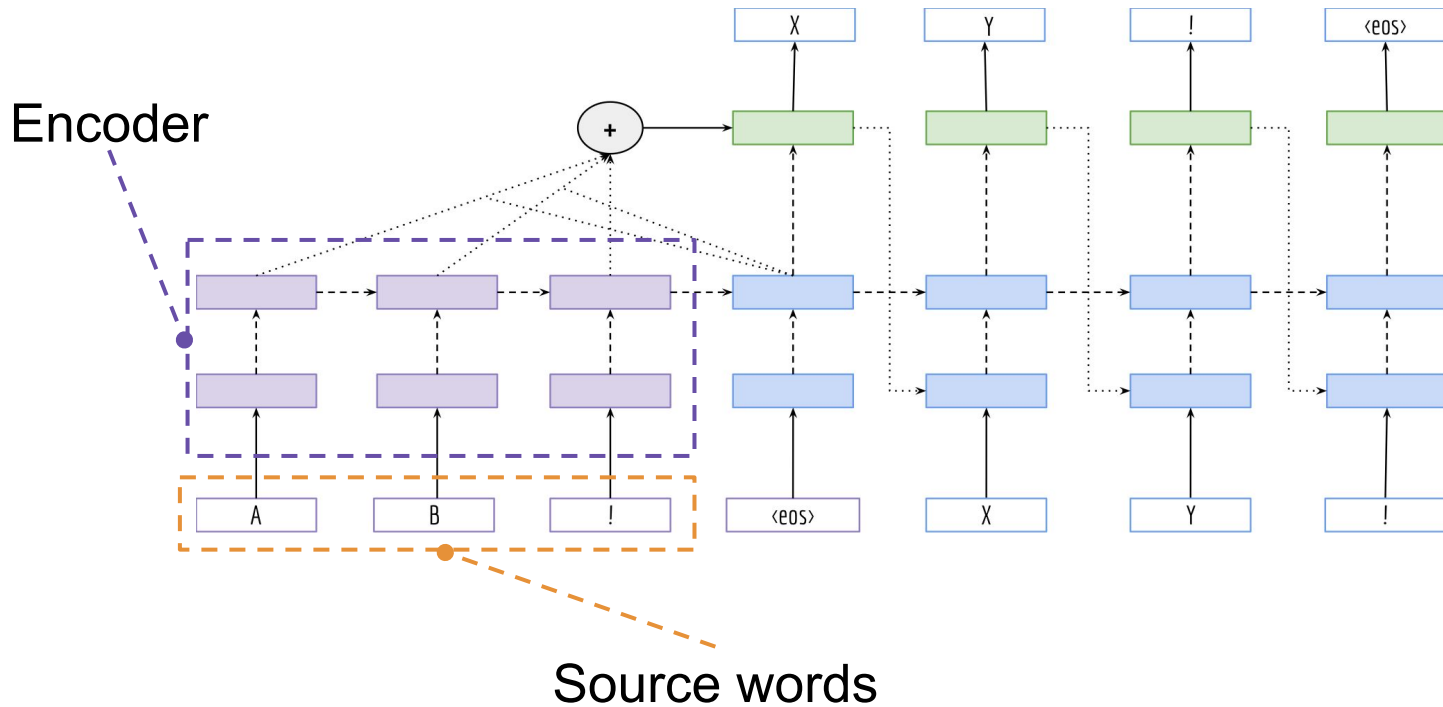
# Neural Machine Translation: working mechanism



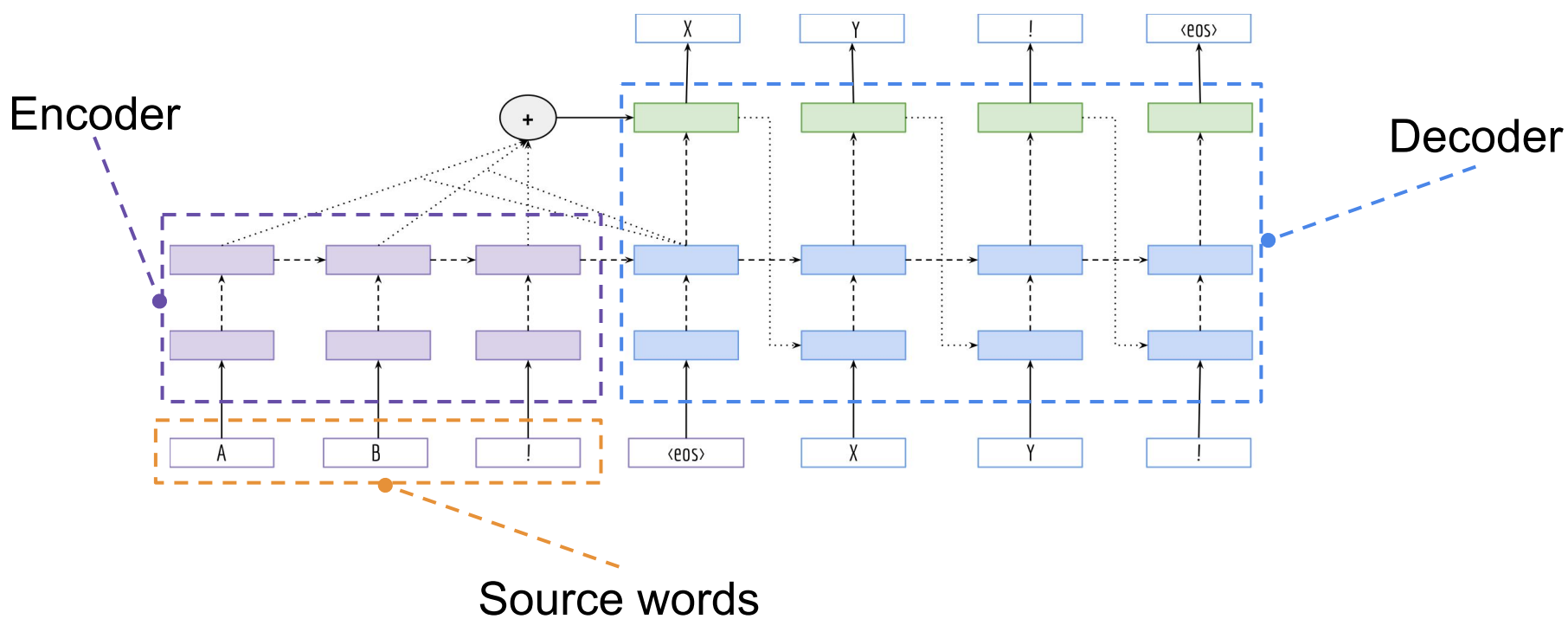
# Neural Machine Translation: working mechanism



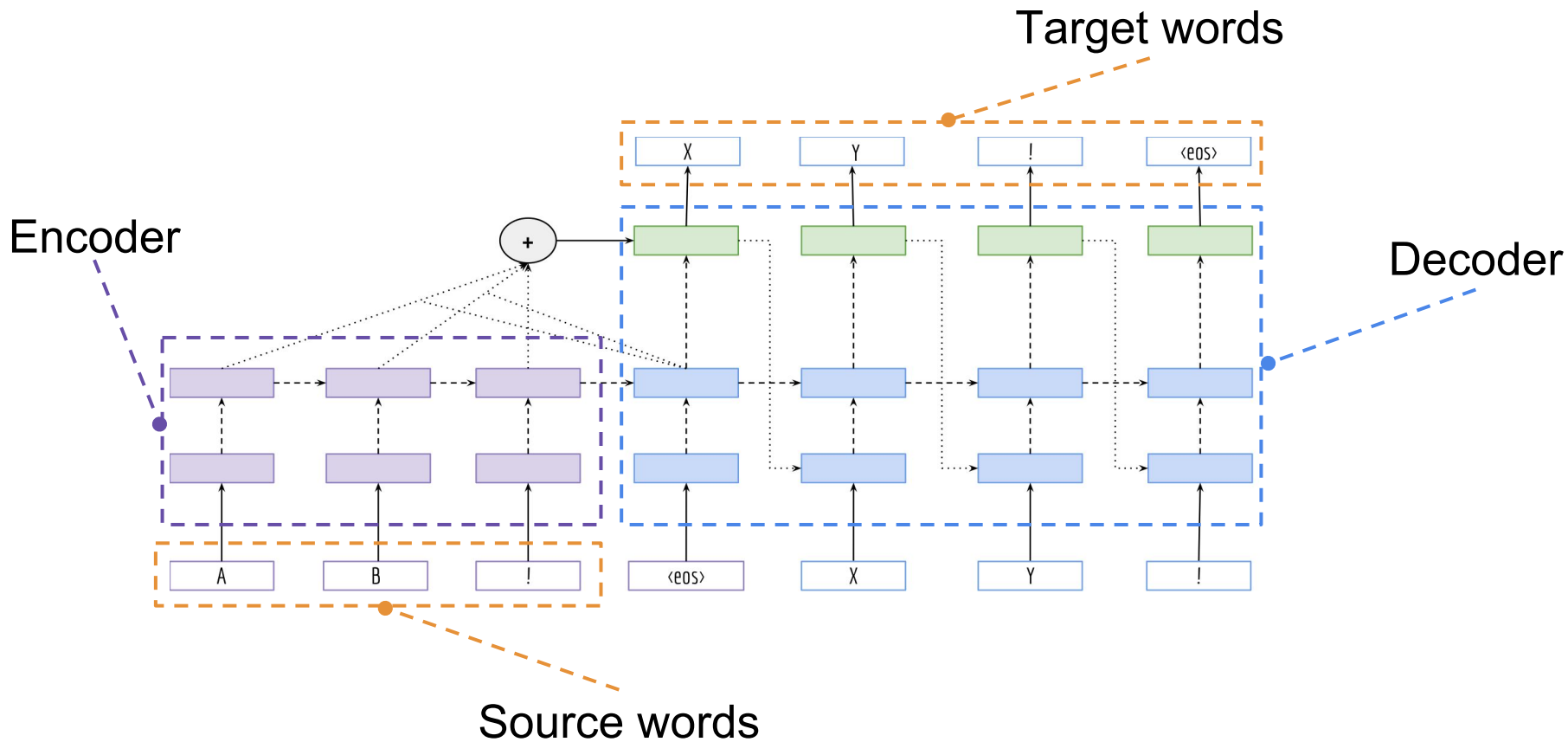
# Neural Machine Translation: working mechanism



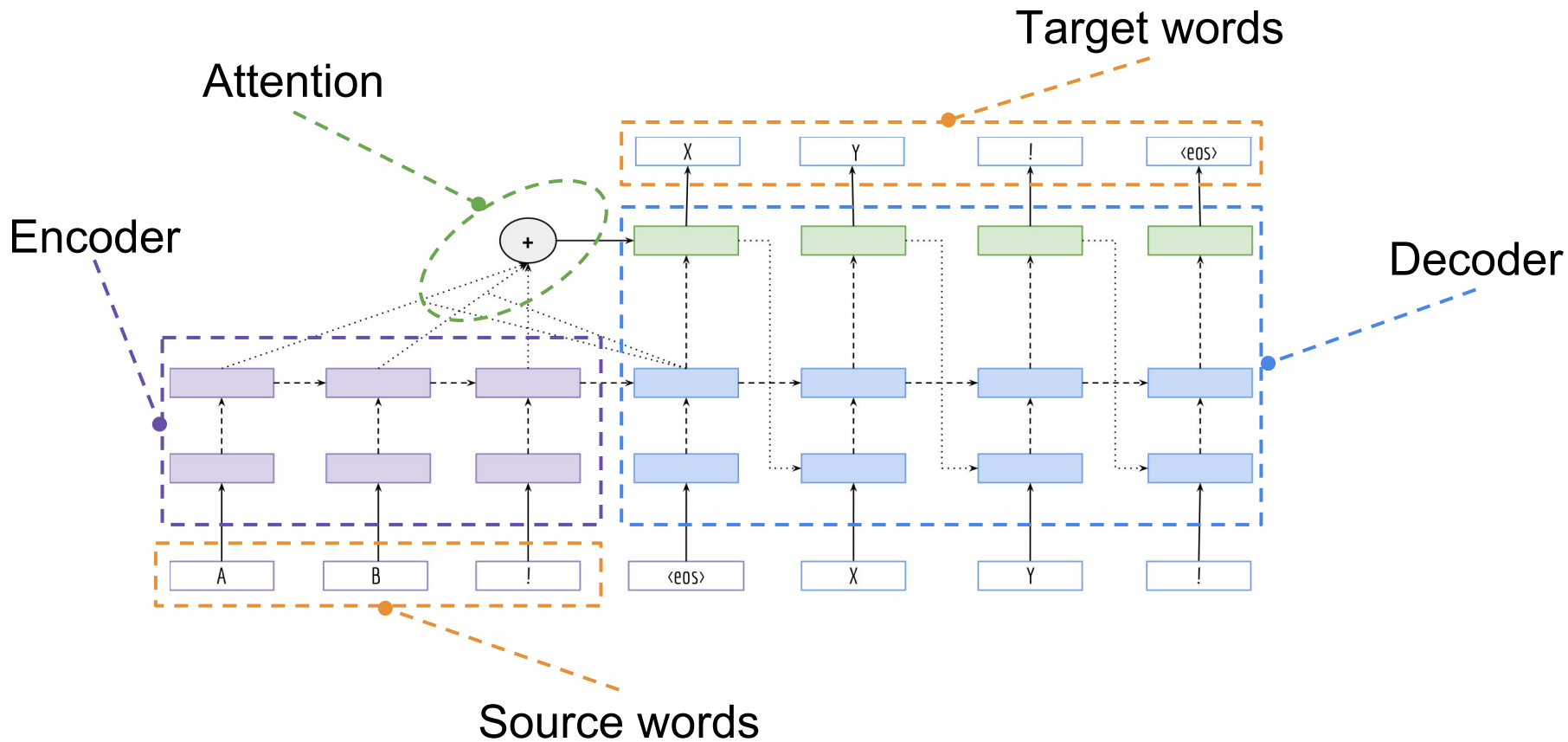
# Neural Machine Translation: working mechanism



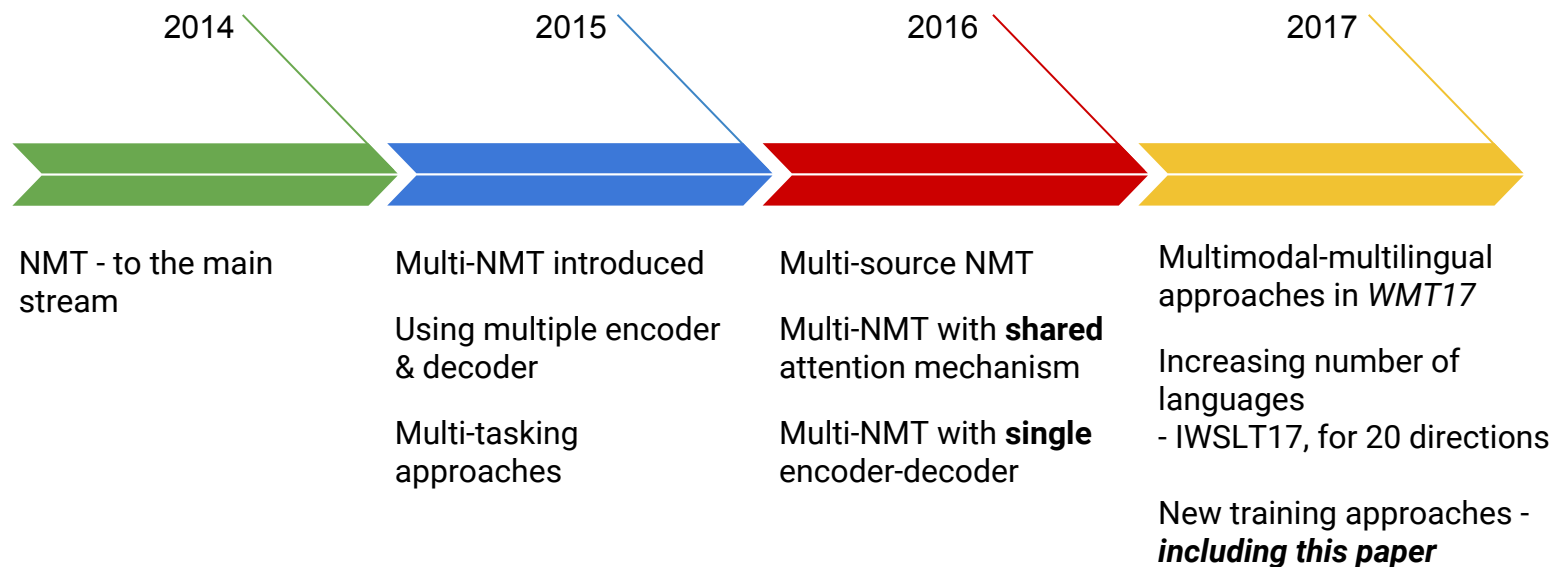
# Neural Machine Translation: working mechanism



# Neural Machine Translation: working mechanism

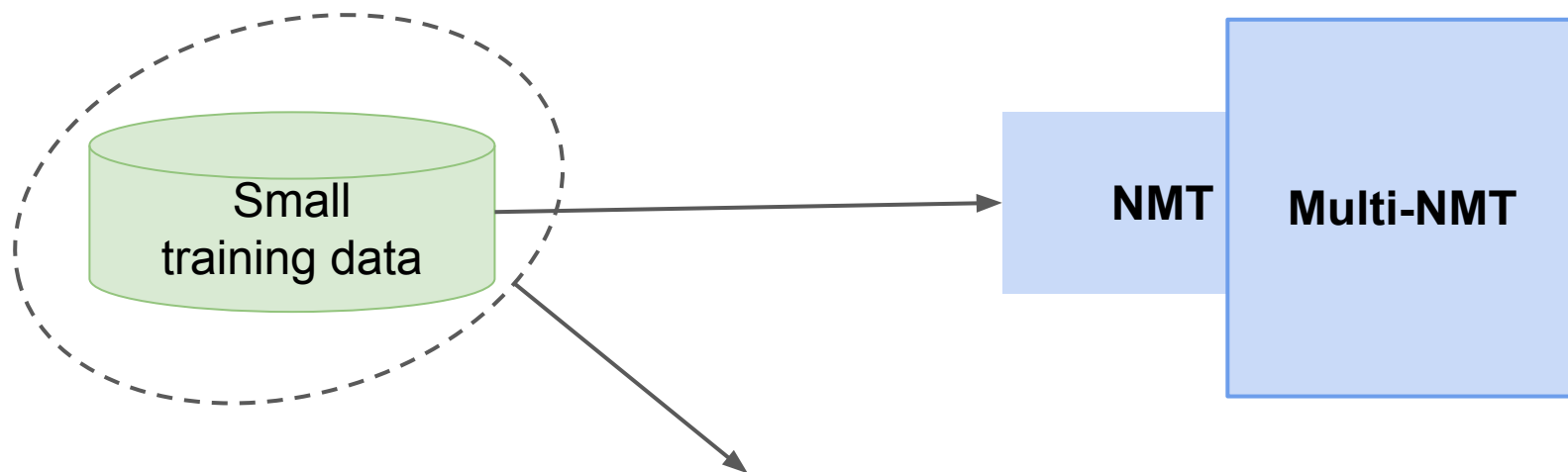


# Neural Machine Translation: a walk to Multilingual NMT



Multilingual-NMT (Multi-NMT) permits *Zero-Shot Translation (ZST)*

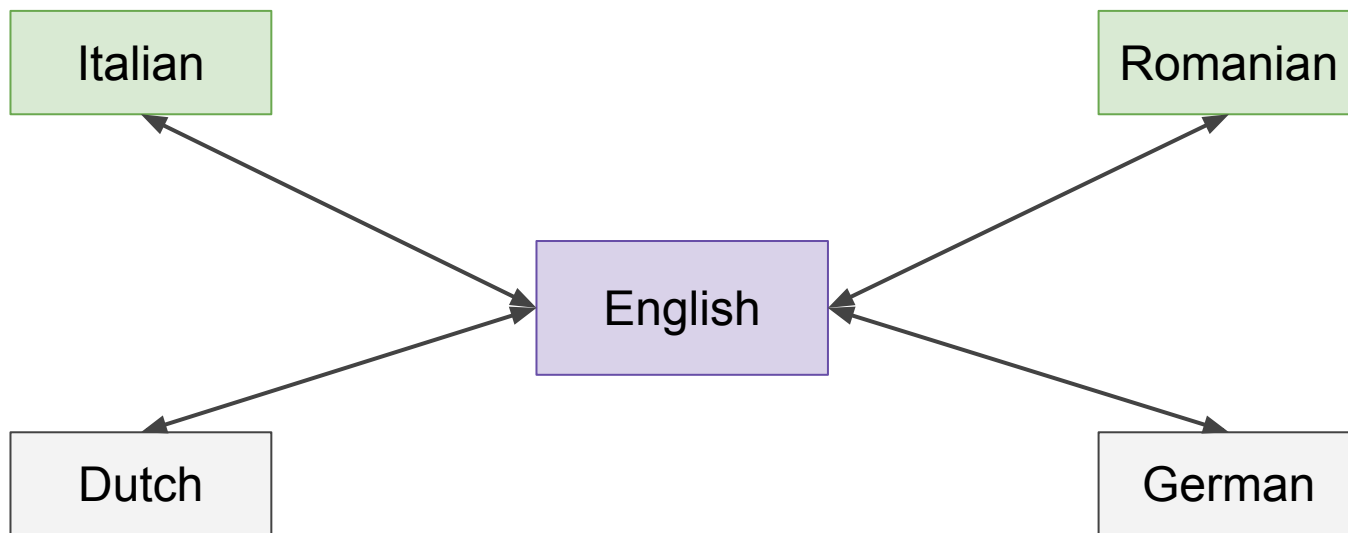
# Multilingual-NMT & ZST: challenge



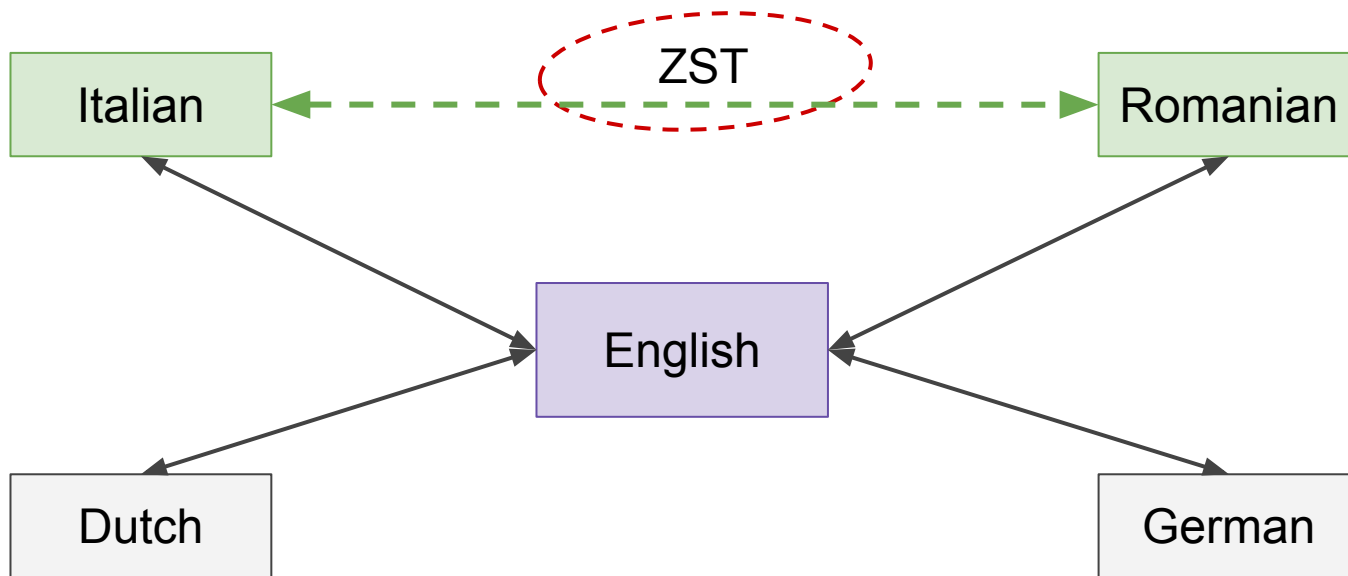
*“NMT systems have a steeper learning curve with respect to the amount of training data, **resulting** in worse quality in low-resource settings...” P. Koehn et al [2017].*



# Multilingual NMT & ZST: scenario & our hypothesis

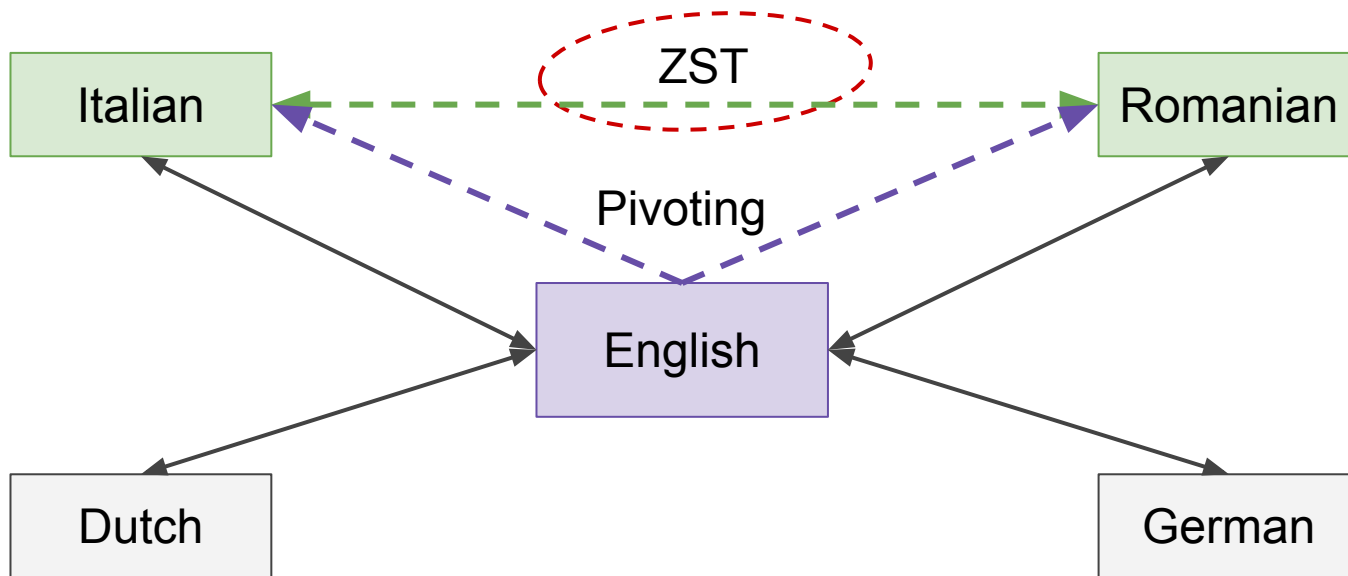


# Multilingual NMT & ZST: scenario & our hypothesis



**ZST** results in worse translation performance in such low-resource setting.

# Multilingual NMT & ZST: scenario & our hypothesis



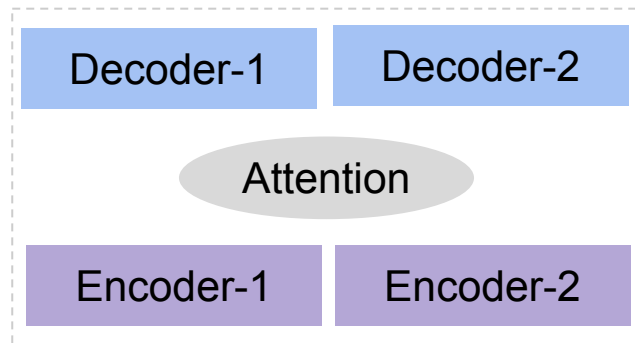
**ZST** results in worse translation performance in such low-resource setting.

**Pivoting** (*x-step translation*) is an alternative approach to a direct ZST.

# Multilingual NMT & ZST: related work

**Firat et al.**, [2016a]; Multi-way, multilingual neural machine translation with a shared attention mechanism.

**Firat et al.**, [2016b]; Zero-resource translation with multilingual neural machine translation.



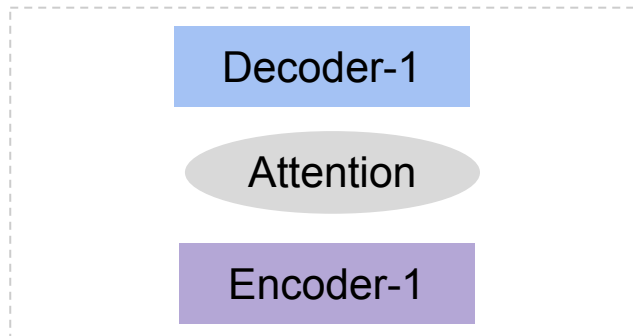
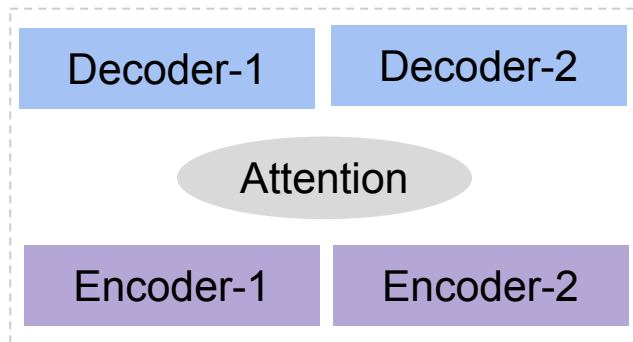
# Multilingual NMT & ZST: related work

**Firat et al.**, [2016a]; Multi-way, multilingual neural machine translation with a shared attention mechanism.

**Firat et al.**, [2016b]; Zero-resource translation with multilingual neural machine translation.

**Ha et al.**, [2016]; Toward Multilingual Neural Machine Translation with Universal Encoder and Decoder

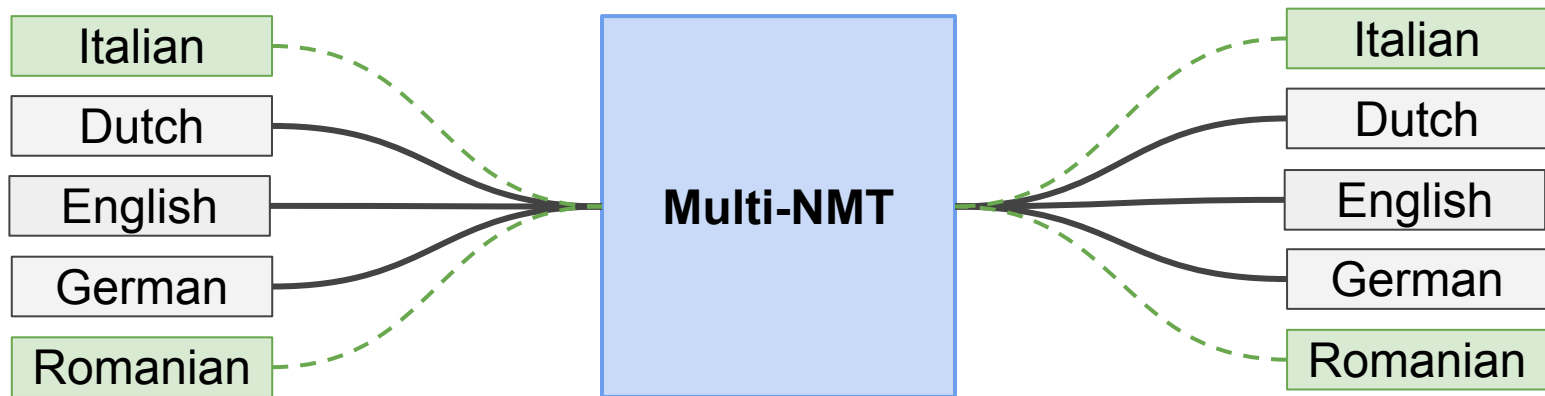
**Melvin et al.**, [2016]; Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation.



# Multilingual NMT & ZST: current limitations

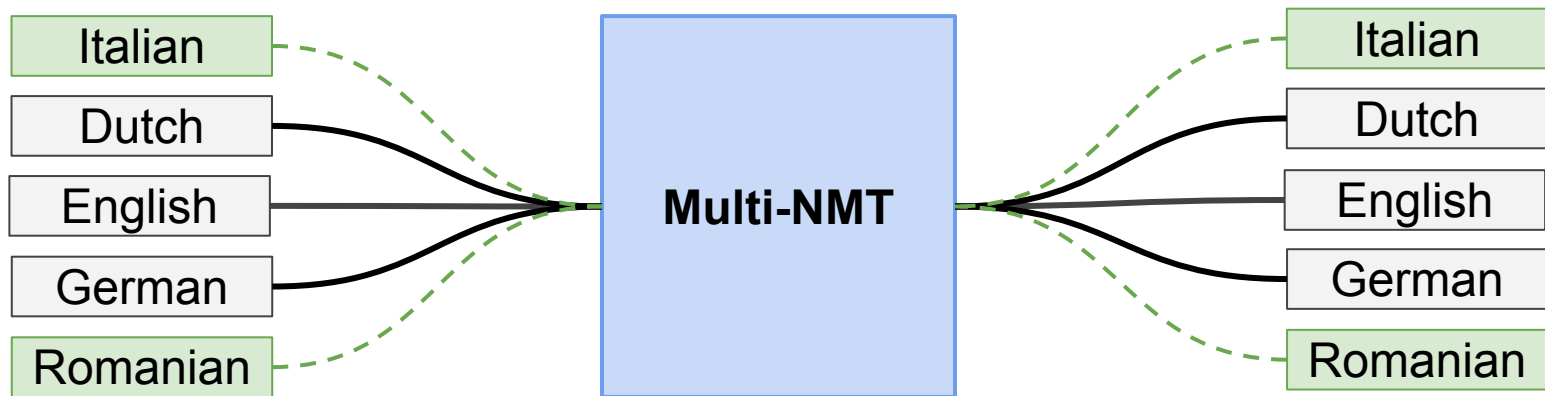
- Pre-assumption to use an *already available* parallel data
- A not so efficient creation and usage of synthetic data
- Weaker target language ID in low-resource scenario

# Multilingual-NMT & ZST: our setup



No available parallel data for **Italian** $\leftrightarrow$ **Romanian** pairs.

# Multilingual-NMT & ZST: our setup



**Training**

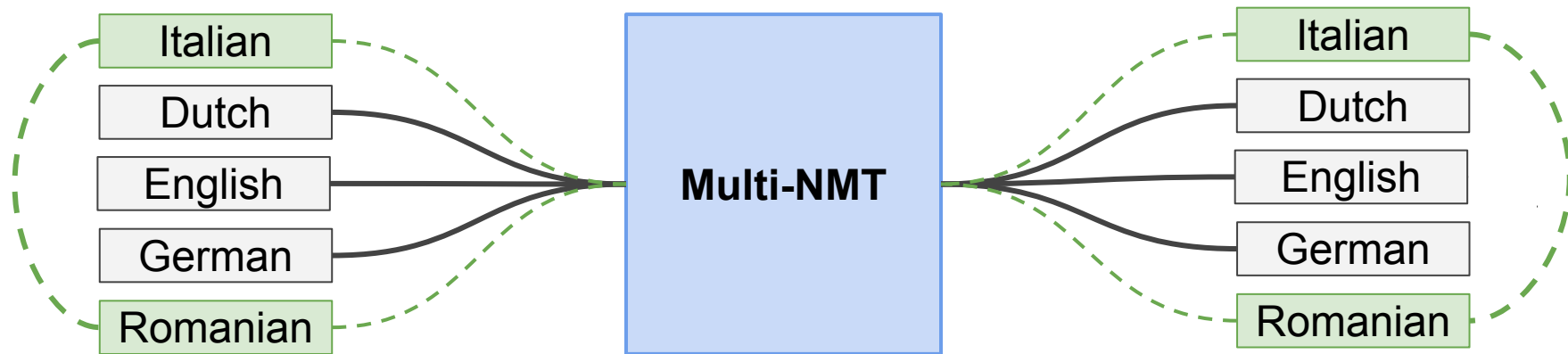
**8 directions**

**Inference**

**8 directions**  
+  
**2 zero-shot**

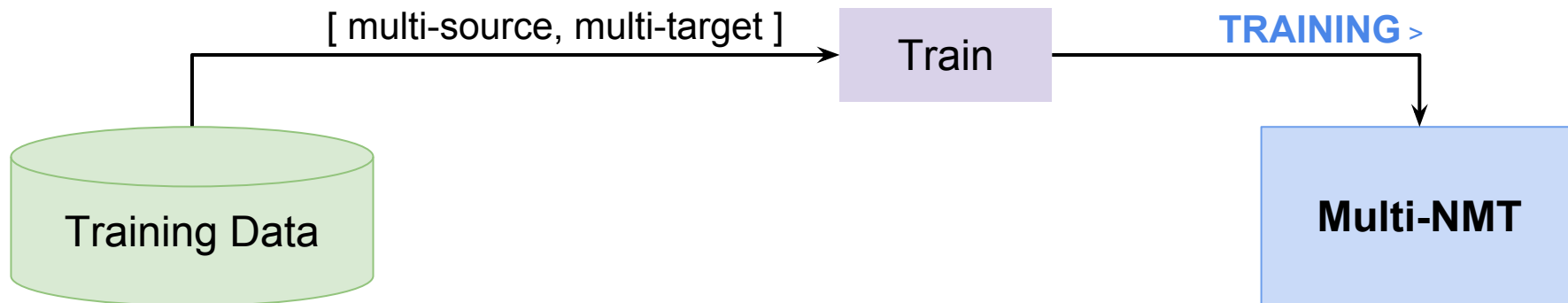


# Multilingual-NMT & ZST: our setup

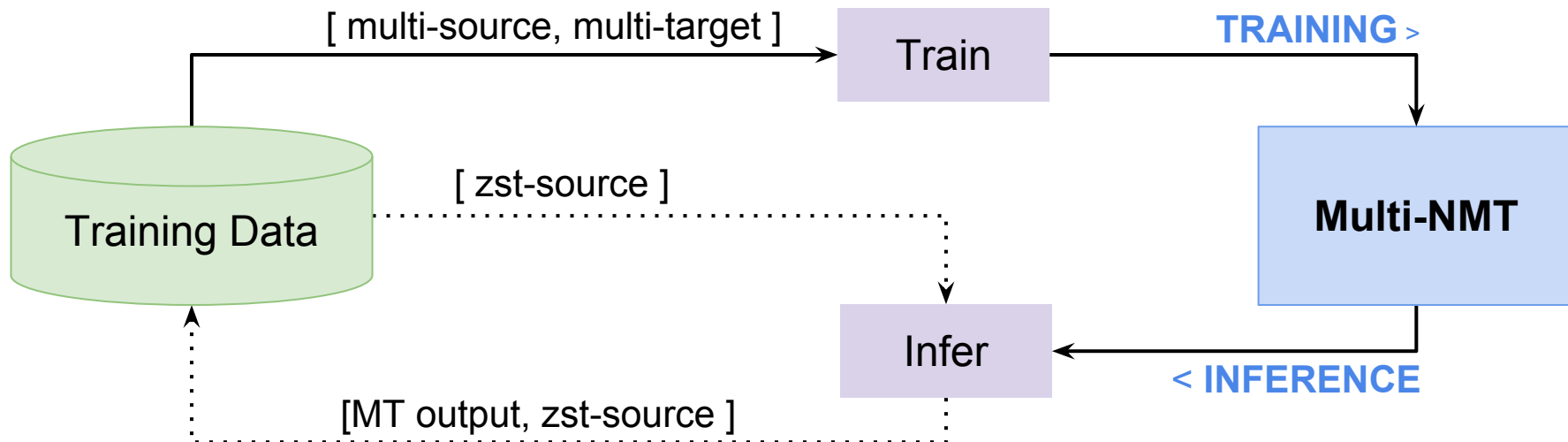


Notice how the **Italian $\leftrightarrow$ Romanian** ZST directions creates a ***dual-translation*** loop.

# Iterative Learning

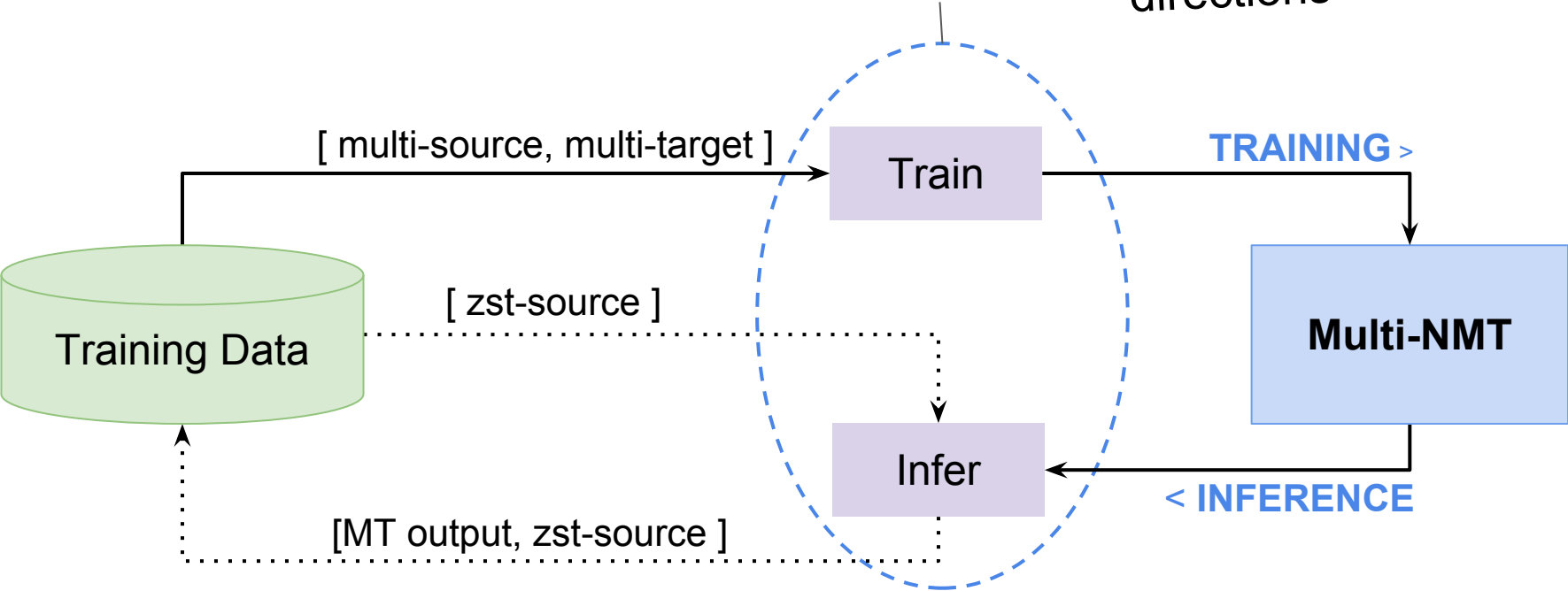


# Iterative Learning

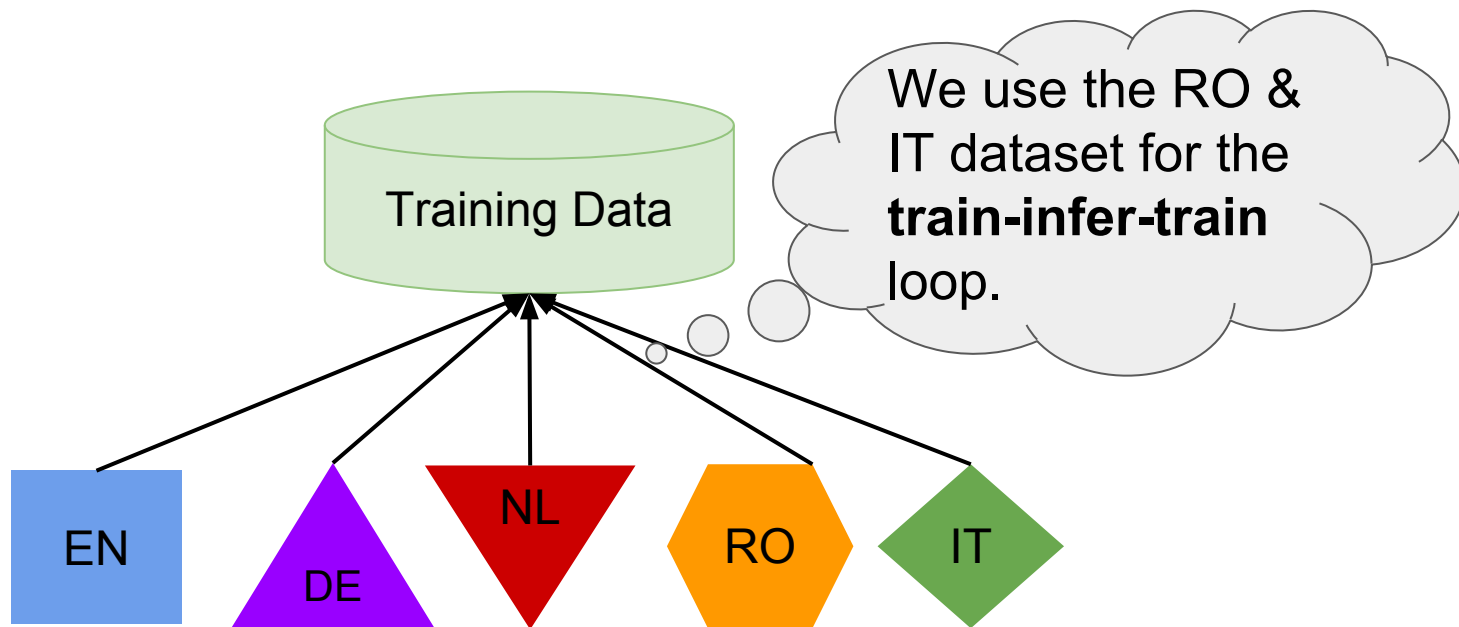


# Iterative Learning

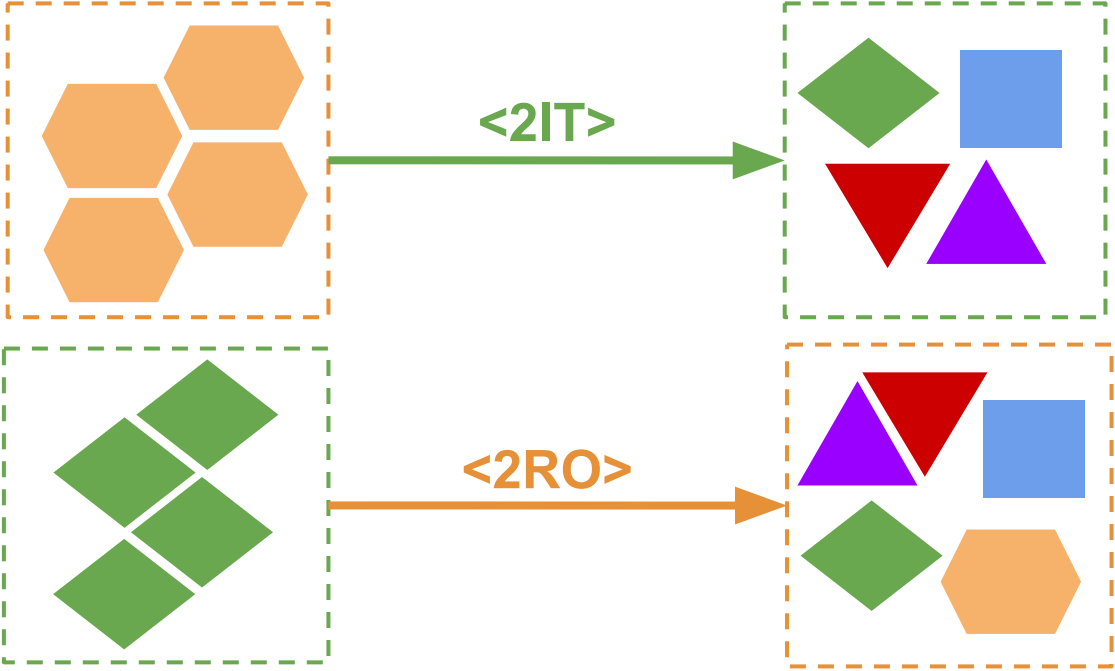
Creates a *train-Infer-train* cycle for a dual translation directions



# Iterative Learning: training with self-generated data

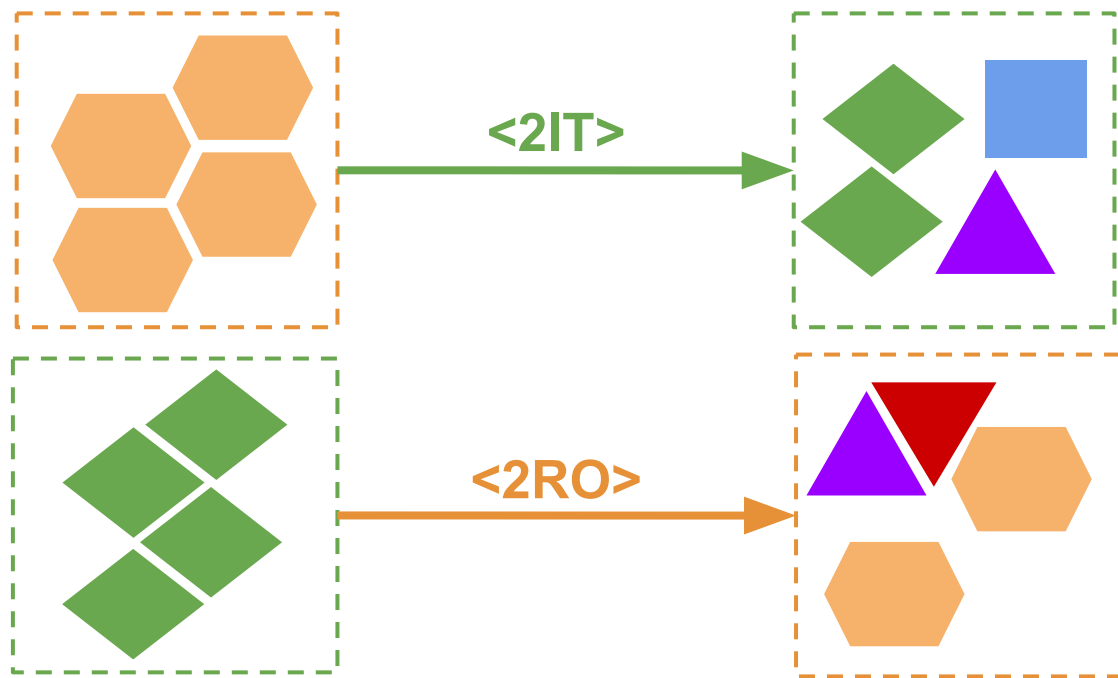


# Iterative Learning: training with self-generated data



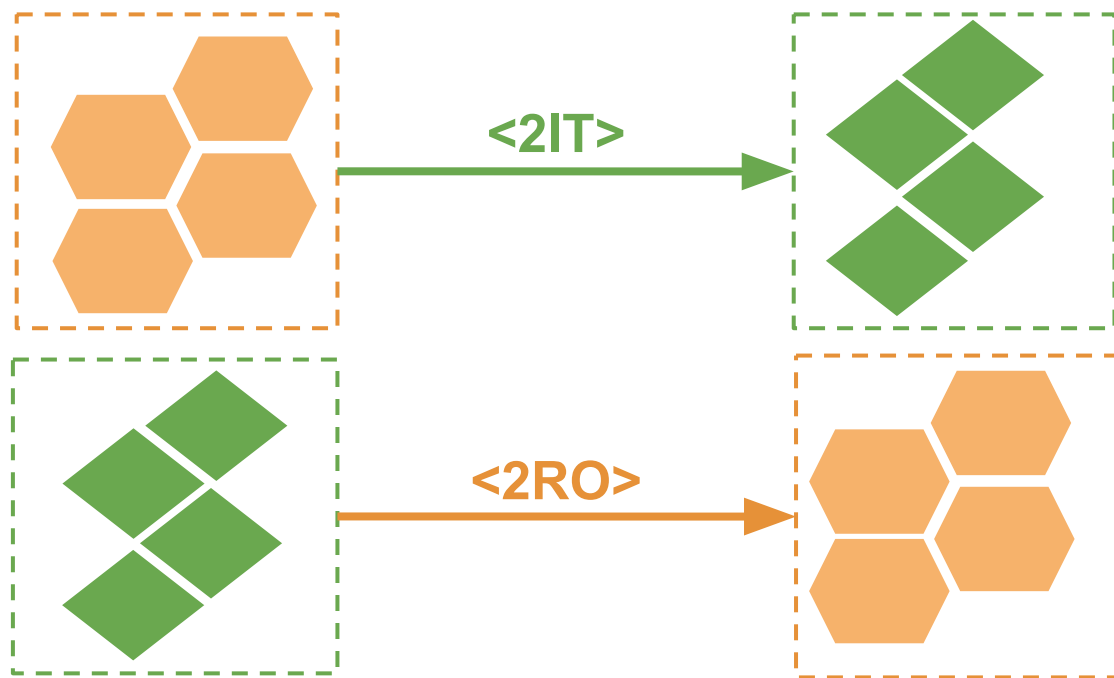
Inference at round **ZERO** - before applying “**train-infer-train**”

# Iterative Learning: training with self-generated data



Inference at round **ONE**

# Iterative Learning: training with self-generated data

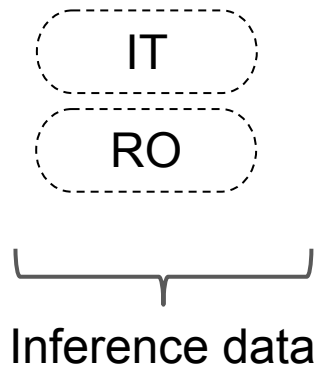


Inference at round ***N*** - ***convergence***



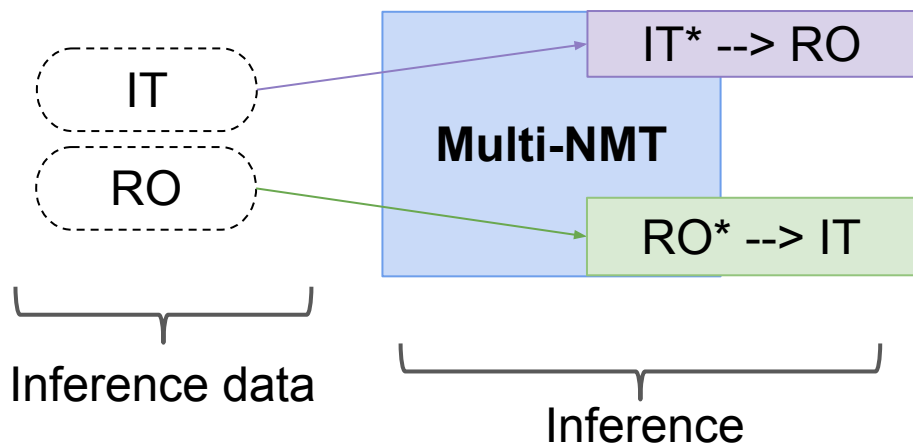
# Iterative Learning: training with self-generated data

How does *translation-duality* helps to improve the ZST directions?



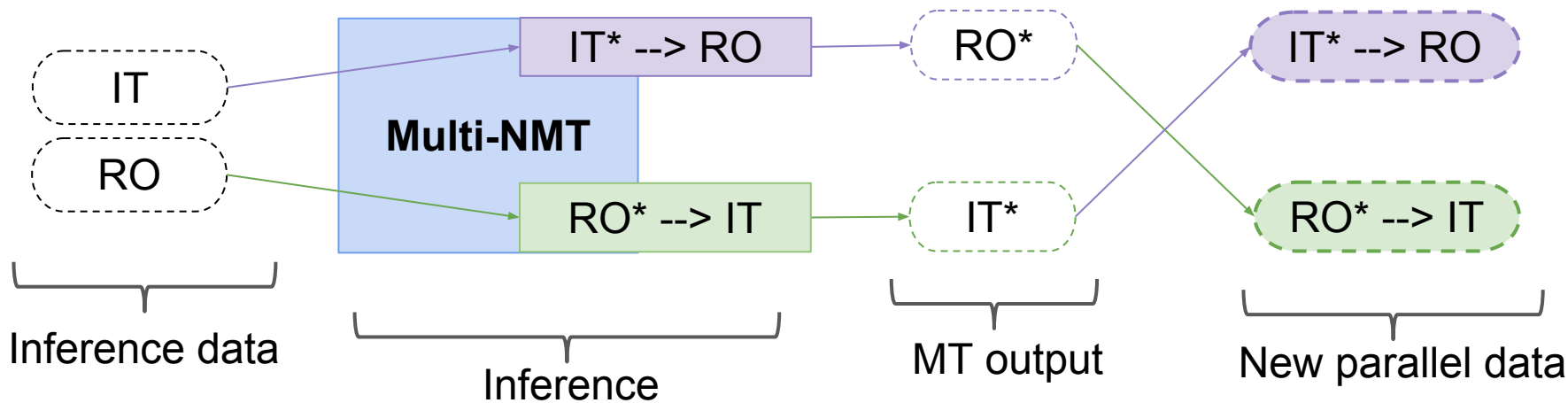
# Iterative Learning: training with self-generated data

How does *translation-duality* helps to improve the ZST directions?



# Iterative Learning: training with self-generated data

How does *translation-duality* helps to improve the ZST directions?

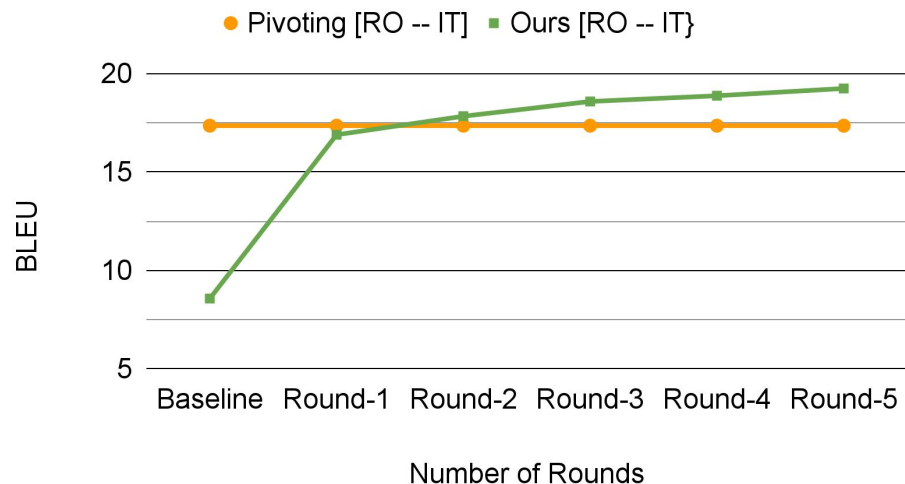
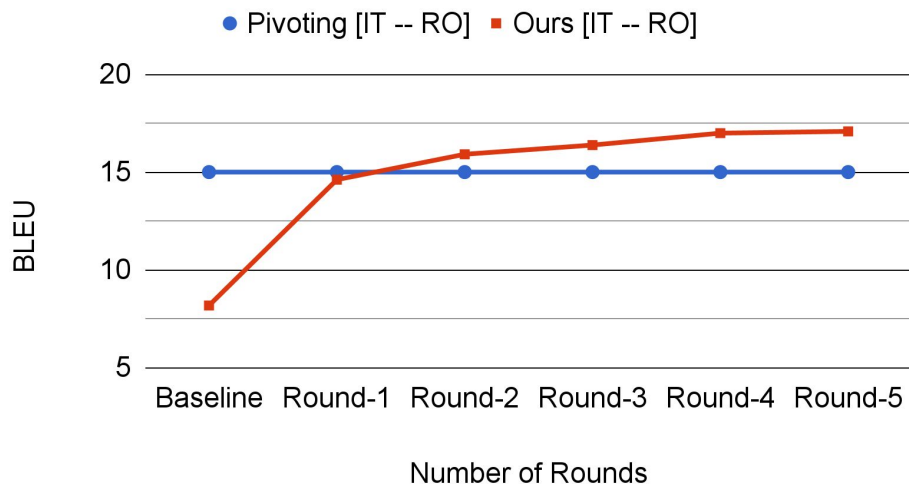


# Experiments: dataset

IWSLT-2017 Multilingual Dataset	
Language Direction	Training Size
EN<-->DE	197,489
EN<-->IT	221,688
EN<-->NL	231,669
EN<-->RO	211,508
IT<-->RO	209,668

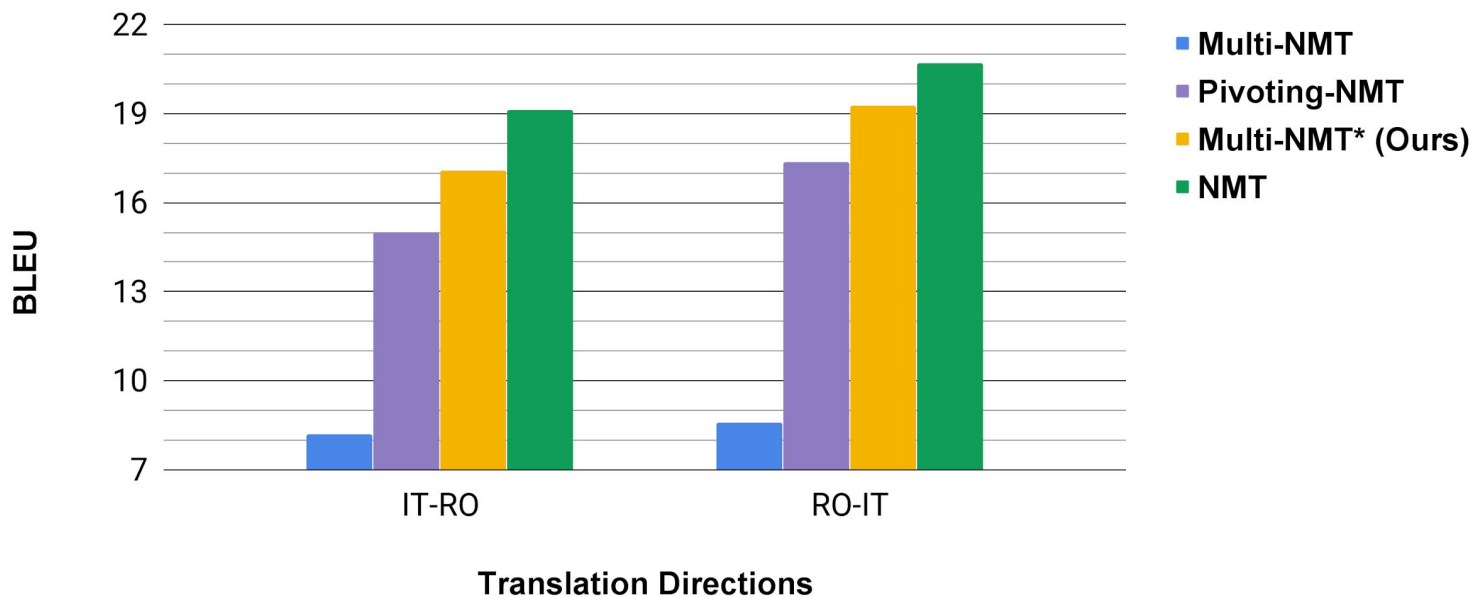
Dataset used to train only the single language pair models.

# Experiments: improvements on each round



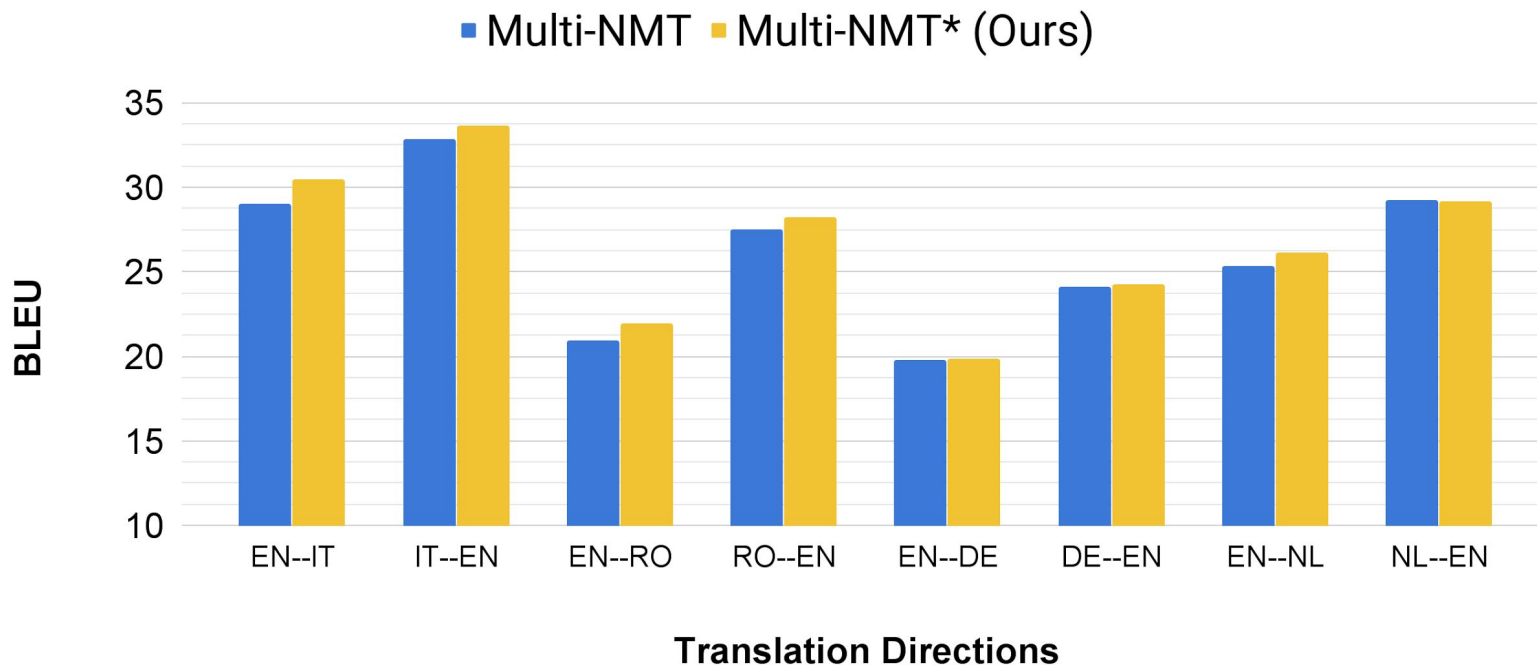
Results of the [Italian <--> Romanian] zero-shot directions on test2017

# Experiments: zero-shot comparison



Our proposed “*train-infer-train*” approach outperformed the baseline **Multi-NMT** and the **Pivoting** mechanism on *test2017*

# Experiments: non-zero-shot comparison



Our proposed “*train-infer-train*” approach slightly improves the baseline **Multi-NMT** on *test2017*

# Translation Examples

## Zero-shot: Italian --> Romanian

Source	... che rafforza la corruzione, l'evasione fiscale, la povertà, l'instabilità.
Pivot	... poarta de bază, evazia <b>fiscală</b> , <b>sărăcia</b> , <b>instabilitatea</b> .
Multi-NMT	... restrânge corupția, fiscale de <b>evasion</b> , <b>poverty</b> , instabilitate.
Multi-NMT*	... <b>care</b> rafinează <b>corupția</b> , evasarea <b>fiscală</b> , <b>sărăcia</b> , <b>instabilitatea</b> .
Reference	... <b>care</b> protejează <b>corupția</b> , evaziunea <b>fiscală</b> , <b>sărăcia</b> și <b>instabilitatea</b> .



# Translation Examples

## Non-zero-shot: English --> Italian

Source

We can't use them to make simple images of things out in the Universe.

Multi-NMT

Non possiamo usarli per **creare immagini semplici** di cose nell'universo.

Multi-NMT\*

Non possiamo **usarle** per **fare semplici immagini** di cose nell'universo.

Reference

Non possiamo **usarle** per **fare semplici immagini** di cose nell'universo

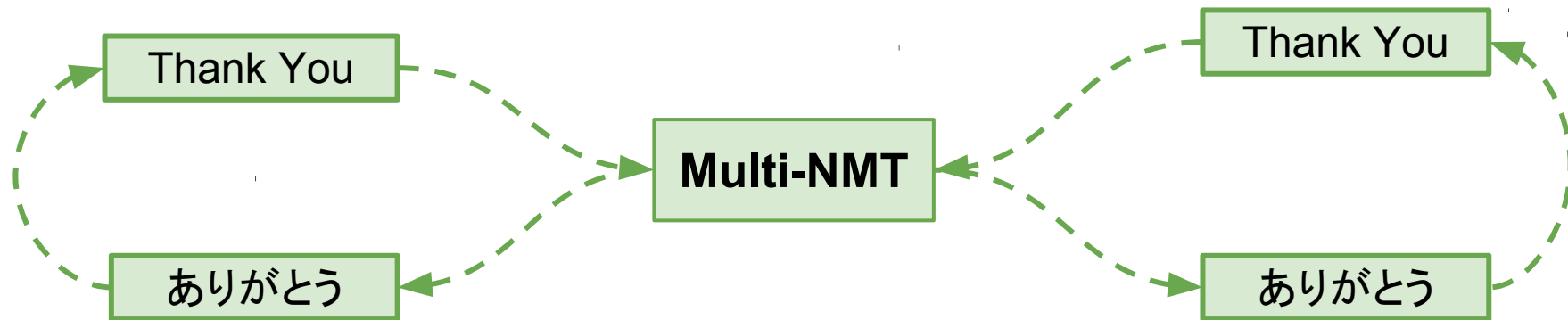
# Conclusion

We introduced “*train-infer-train*”, an approach for improving ZST:

- Efficiently leverages dual-translation directions
- Achieved a significant improvements over a Multi-NMT baseline
- Outperformed a pivoting based approach for ZST

Future work:

- A more efficient training and inference steps
- Including additional monolingual data for the ZST directions



Surafel M. Lakew | [lakew@fbk.eu](mailto:lakew@fbk.eu)  
Fondazione Bruno Kessler & University of Trento

