Linguistically Motivated Neural Machine Translation

Haiyue Song
NICT, Kyoto, Japan

Hour Kaing
NICT, Kyoto, Japan

Raj Dabre
NICT, Kyoto, Japan
IIT Madras, India
IIT Bombay, India

Get access to the slides here (update)

https://github.com/prajdabre/eamt24-linguistic-mt

(under construction)

Self Introduction: Haiyue **Song**

- Technical Researcher at National Institute of Information and Communications Technology (NICT)
- Research
  - Low-Resource Machine Translation
  - Subword Segmentation
  - Large Language Models for Machine Translation

haiyue.song@nict.go.jp

Self Introduction: Hour Kaing (hour.kaing@nict.go.jp)

- Experience
  - 2022 - Present: Researcher at NICT
  - 2018 - 2022 : Technical Researcher at NICT
  - 2018 - 2022 : Ph.D. at Nara Institute of Science and Technology (NAIST), Japan
    - Low-Resource Morphological and Syntax Analysis
  - 2013 - 2014 : M.Sc at University of Grenoble 1, France

- Research interests
  - Machine Translation, Language Modeling, Linguistic Analysis, and Speech Processing
  - Low-Resource and Linguistically-Motivated NLP
  - Cross-Lingual Transfer and Multilingual Learning

Self Introduction: Raj Dabre (raj.dabre@nict.go.jp)

- Experience
  - 2018-present: Researcher at NICT, Japan
    - Adjunct Faculty at IIT Madras
    - Visiting Researcher at IIT Bombay
  - 2014-2018: MEXT Ph.D. scholar at Kyoto University, Japan
  - 2011-2014: M.Tech. Government RA at IIT Bombay, India

- Research
  - Low-Resource Natural Language Processing
    - **Multilingual Machine Translation: 2012-present**
    - **Document Level Machine Translation: 2021-**
    - **Large Scale Pre-training for Generation: 2021-**
  - Efficient Deep Learning:
    - **Compact, flexible and fast models (2018-present)**

Table of Contents

Introduction to Neural Machine Translation

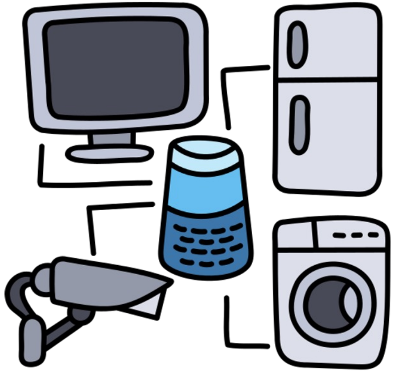
Why Machine Translation is still an important task?

Inclusivity and Accessibility

Bridge gap between low-resource languages (HRL) and high-resource languages (HRL)

Improve language coverage
(only covers ~1K of ~7K in the world)

Data Augmentation for Multilingual Performance Enhancement



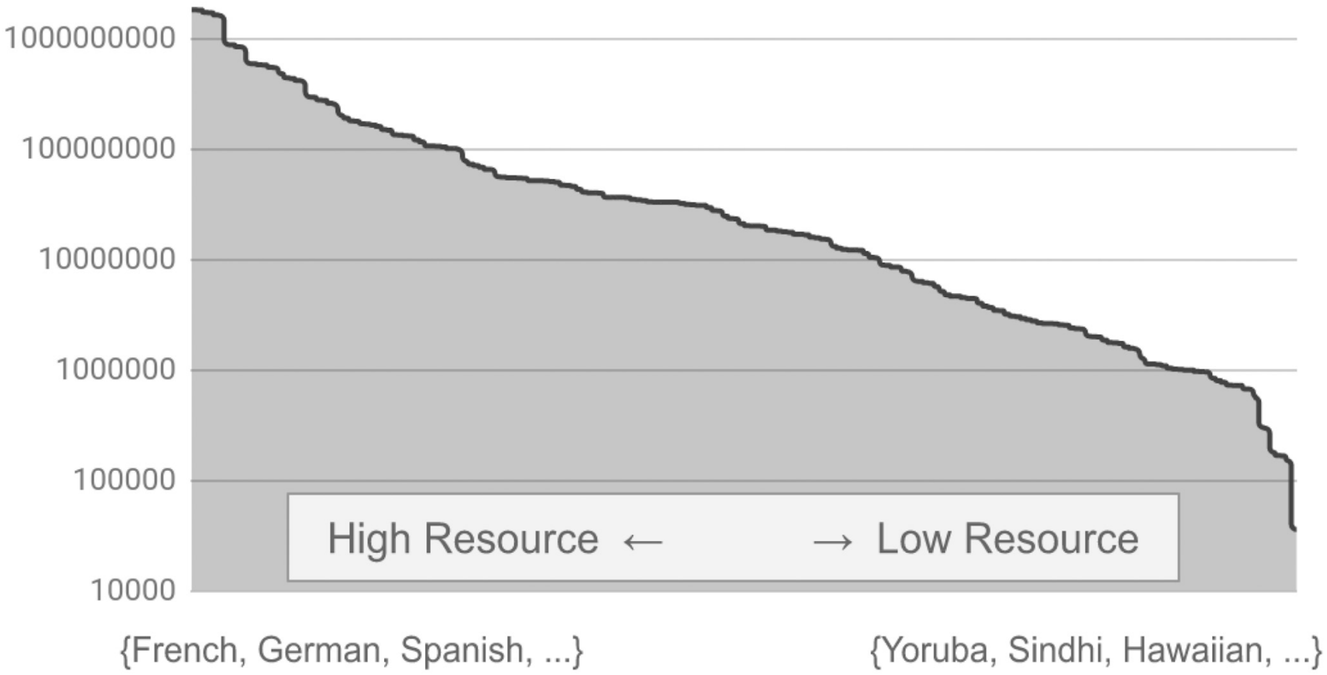
Transfer Learning via Translation

Unlocking Multilingual Capabilities of LLMs

Evolution of Machine Translation

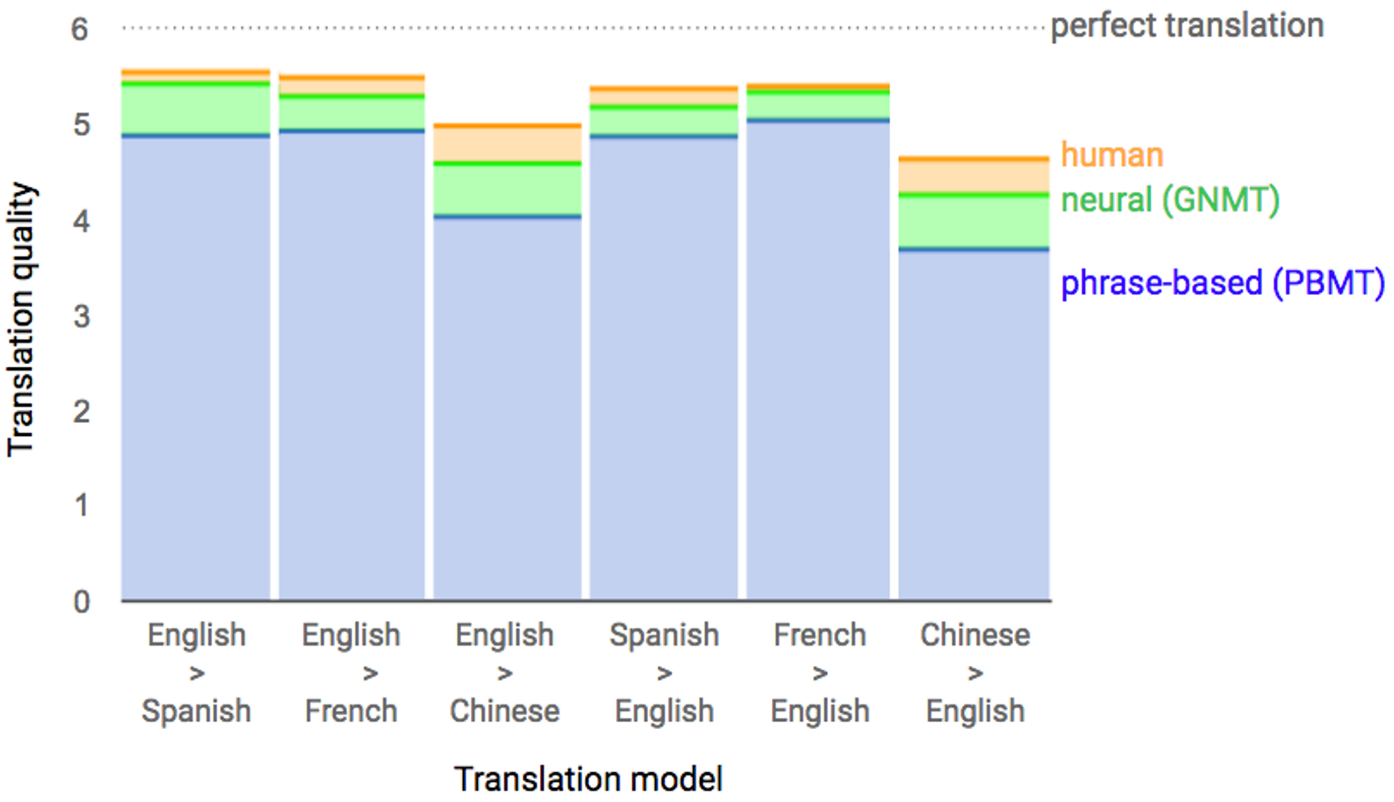
Rule-Based Machine Translation (RBMT)	Example-Based Machine Translation (EBMT)	Statistical Machine Translation (SMT)	Neural Machine Translation (NMT)
• Direct MT • Transfer-based MT • Interlingua MT		• Word-based • Syntax-based • Phrase-based	• RNNs • LSTMs • Transformers
1950 - 1980	1980 - 1990	1990 - 2015	2015 -

Why NMT? (Good for transfer learning)



Data Distribution over language pairs (Arivazhagan et al., 2019)

Quality of NMT Compared to Phrase-based SMT



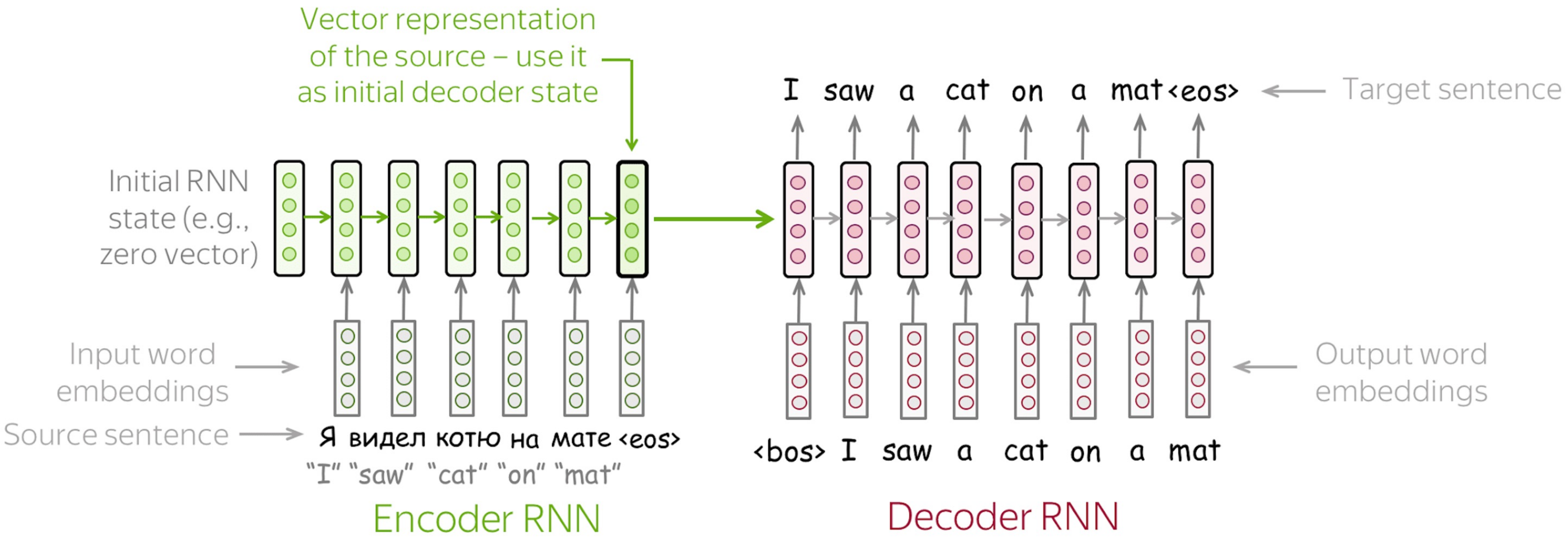
https://research.googleblog.com/2016/09/a-neural-network-for-

Evolution of Machine Translation

Rule-Based Machine Translation (RBMT)	Example-Based Machine Translation (EBMT)	Statistical Machine Translation (SMT)	Neural Machine Translation (NMT)
• Direct MT • Transfer-based MT • Interlingua MT		• Word-based • Syntax-based • Phrase-based	• RNNs • LSTMs • Transformers
1950 - 1980	1980 - 1990	1990 - 2015	2015 -

This Tutorial

Neural MT Basics: **Encoder-Decoder Paradigm**



[image credits]

[Sutskever et al. 2014]

Neural MT Basics: **Encoder-Decoder with *Attention***

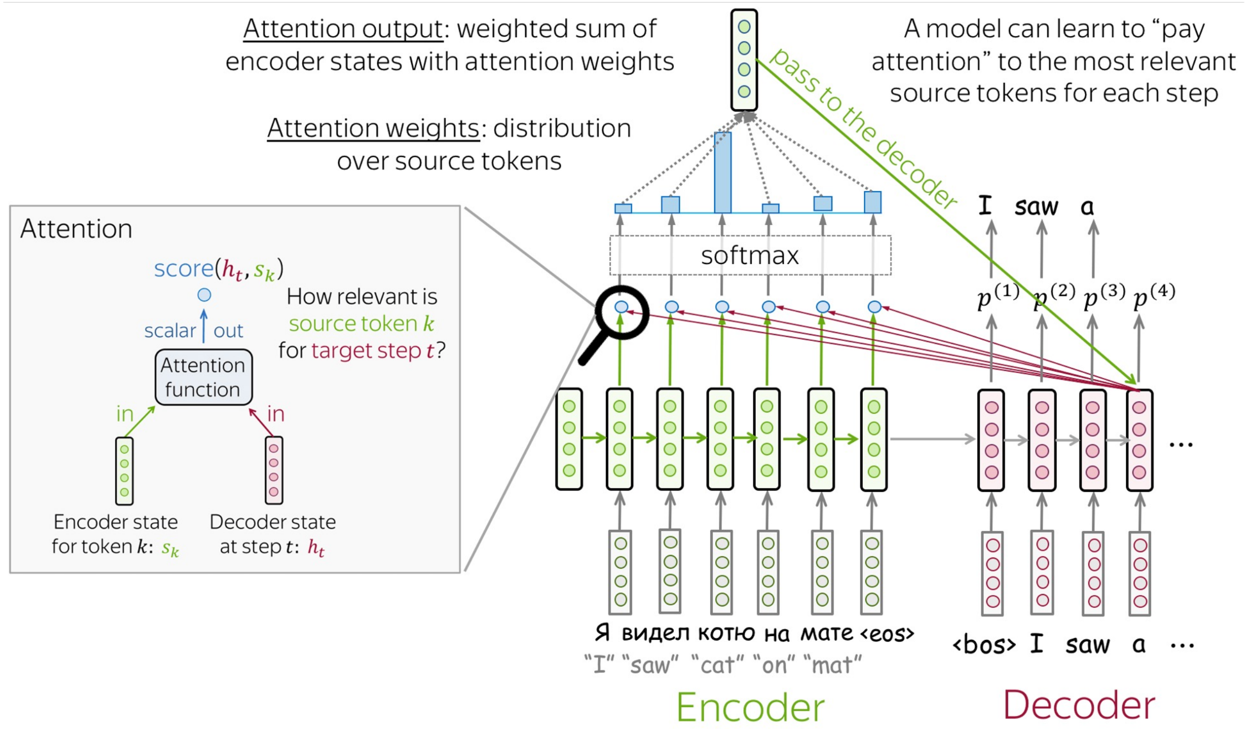


image credits

Bahdanau et al. 2015

Neural MT Basics: **Transformer** Architecture

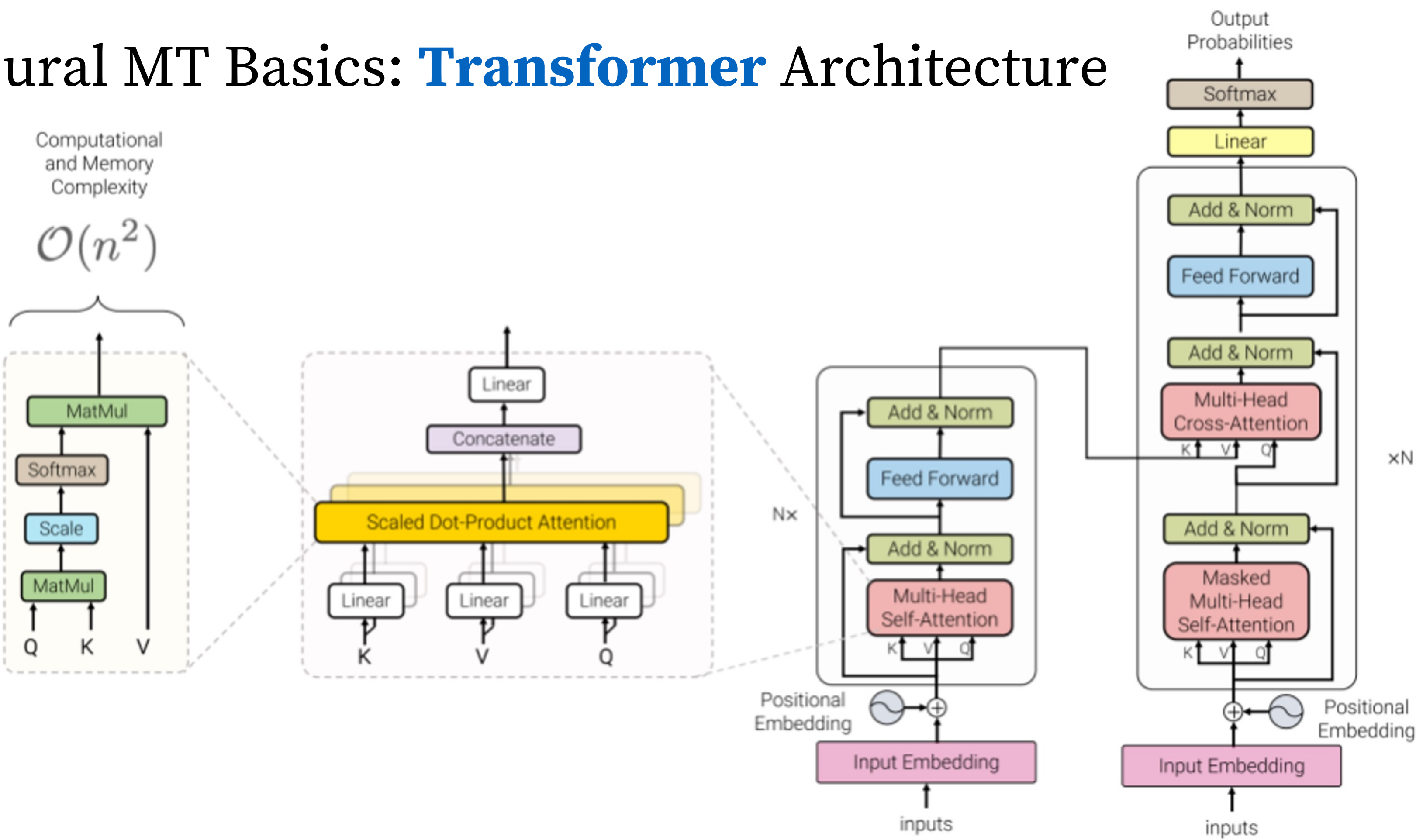


image credits

Vaswani et al. 2017

Is Linguistics **dead**?

No not quite!

Lets re-think!

- Does the model understand
  - **NO!**
- What do they underst
  - **Embedding**
- What are embeddings
  - **Features**

Representing *linguistic info*
as features

- Words: I am a boy
- POS: PRON VERB DET NOUN

Interchangeable?

Why Linguistic Features?

- Supplementary information in low-resource settings

- **Reduce burden** on model to learn *complex features*

- Reuse existing tools rather than waste them

In This Tutorial

- What linguistic features can be leveraged?
- How do we incorporate them in models?
- What is the impact of linguistic features?

Limitations and Future Directions

Limitations

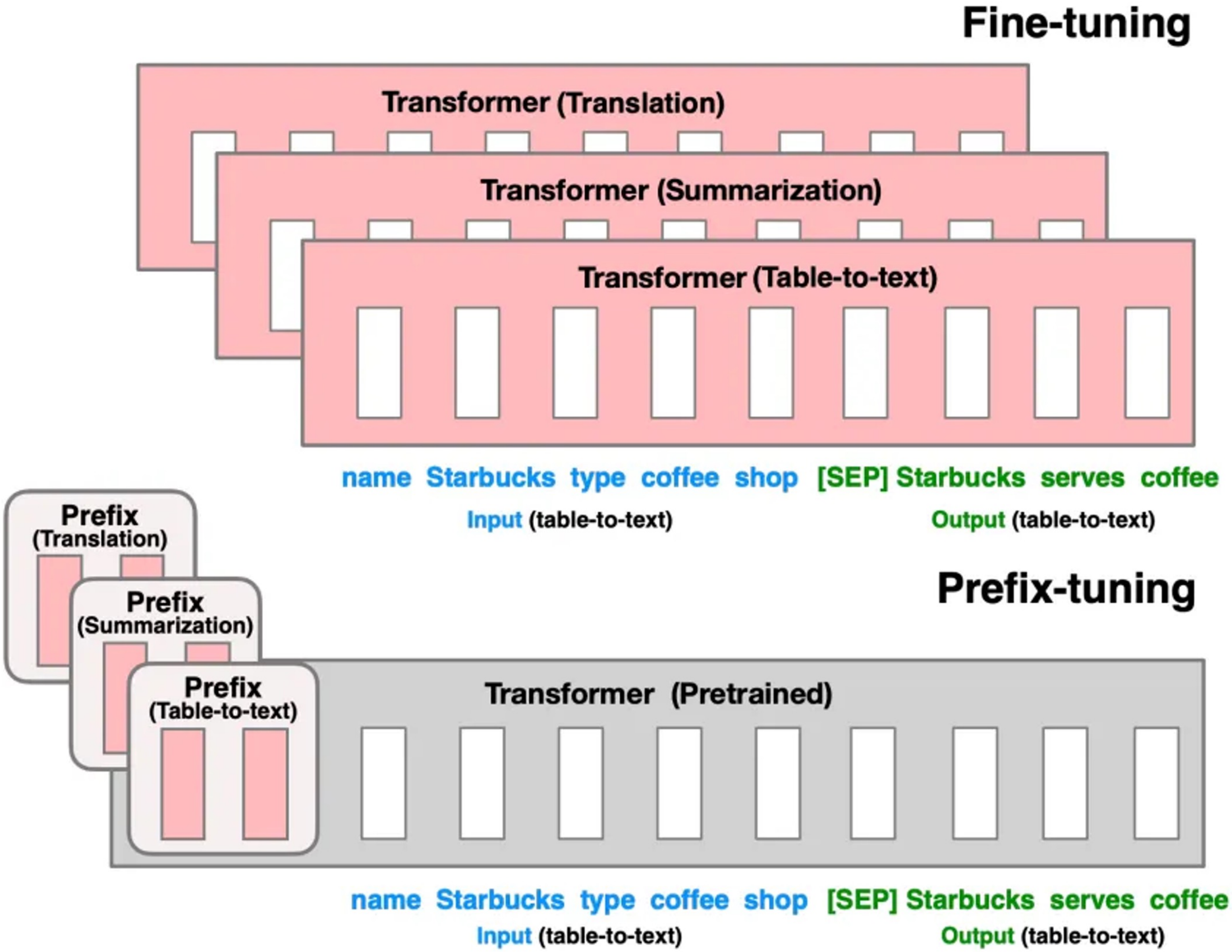
- Mostly impactful in low-resource settings :-(
  - Most languages are low-resource :-)
- Identifying useful features needs exhaustive study :-(
  - In low-resource settings its fine :-)
- Interpretability analysis is hard :-(
  - Extrinsically performance improves :-)
  - But proving it intrinsically is challenging :-(
- Slow speed and error propagation :-(
  - Requires high quality feature extractors (typically available for English)

Future Directions

- Identifying approaches for high-resource settings
    - Also in **multilingual** settings
- Methods to auto-choose features
- Intrinsic analysis of models to show impact of features
- Speed improvement
    - **Latent features** as opposed to explicit features
- Incorporation in LLMs
    - **Mostly open area**

Inco

- P
- F



Fine-tuning
Transformer (Translation)
Transformer (Summarization)
Transformer (Table-to-text)
name Starbucks type coffee shop [SEP] Starbucks serves coffee
Input (table-to-text)
Output (table-to-text)
Prefix-tuning
Prefix (Translation)
Prefix (Summarization)
Prefix (Table-to-text)
Transformer (Pretrained)
name Starbucks type coffee shop [SEP] Starbucks serves coffee
Input (table-to-text)
Output (table-to-text)

Summary

Summary

- Basic overview of NMT and motivation

- Methods to incorporate features at various points in the system
    - Data: tokenization and related languages
    - Model: inputs, encoder, representation, and self-attention
    - Decoding: tree-structure decoding
    - Evaluation: linguistic benchmark

- Comparison of effective practices

- Limitations and Future Work

Q&A

Thank You