

# LLMs for translation



Charles University  
Faculty of Mathematics and Physics  
Institute of Formal and Applied Linguistics



unless otherwise stated

# Introduction

# Introduction

- The task is to translate two types of texts:
  - Phone conversations
  - Interviews with refugees
- in two Arabic dialects:
  - Tunisian
  - North Levantine
- Into English

# Introduction

- Traditional neural machine translation: trained on sentence pairs (source sentence, target sentence)
- LLMs: trained mostly just on unrelated text in multiple languages (in larger chunks than sentences)
- Challenges:
  - The dialects itself (not very common in the parallel data) -> advantage LLMs
  - Code-switching (e.g. French words in Tunisian sentence) -> advantage LLMs
  - Context is important, one sentence does not contain enough info -> advantage LLMs

# Problems

# Problems

- Not that many LLMs are trained on Arabic (finetuning is possible, but complicated bc. of different script)
- Not that many LLMs are trained for translation -> finetuning
- Finetuning requires large GPU memory
- We tested:
  - Baseline: Google Translate, NLLB 1.3B
  - Mistral
  - Gemma
  - Aya-101
  - Cohere4AI/Command-R
  - Jais
  - Falcon
  - Bloom-z
  - ALMA

# Forcing LLMs to translate

- Some LLMs fail completely, some fail in the scenario with context
- Few-shot for some reason harms the output quality in our scenario?
- Context-less prompt:

Translate this line from Tunisian Arabic into English. Arabic: {source\_example}  
English: {target\_example}

Translate this line from Tunisian Arabic into English. Arabic: {src\_line}  
English:

- *Source\_example, target\_example* is a sentence pair from some data similar to our test set (1-shot learning)
- *Src\_line* is our input sentence

# Forcing LLMs to translate

Model	MADLAD 10B	eTranslation	Google translate	NLLB 1.3B	Bloomz	aya-101	Command-R	jais-13B-chat
BLEU	18.6	15.1	<b>28.7</b>	20.7	7.2	17.6	<b>22.1</b>	20.7

- Sentence-level results are quite promising, we should try to use the context



# Forcing LLMs to translate

- We want to use context, but the evaluation is still sentence-level -> same number of sentences in the output
- We need to split the translation of a larger chunk back into sentences:
  - Sentence splitter - unreliable, introducing another tool
  - Asking the model to keep the same number of lines: only works for few lines and it still moves the content between the sentences
  - Using separator token, e.g. `[s]`: a bit better, especially after finetuning, but still not 100% reliable
  - **Only translating one sentence at a time, but providing the whole context (source document and previously translated prefix) in the prompt**

# Forcing LLMs to translate

- Context prompt:

"We need to translate a single line from conversation in Tunisian Arabic into English. This is the conversation {src\_context}

The start of the conversation is already translated into English: {prev\_context}

Translate the following line from Arabic to English. Be very literal, and only translate the content of the line, do not add any explanations: {src\_line}"

- The only “document-level” prompt that work reliably at least for some models (Command-R)
- BLEU 22.1 -> 26.1
- We finetuned with this prompt (maybe the other approaches would start to work after finetuning as well)

# Forcing LLMs to translate

We need to translate the following conversation in Levantine Arabic into English. It is very important to keep the same number of lines in the translation:

ماشي  
بالبدايه \* بالشت \* الربيع العربي متل ما بيقولوا بعتقد بالالفيز وأر بعطاعش إذا مانو كفتن غلطاد بتونس  
بعديد امتدت على \* دول ثانية  
ف وقت اللي وصلت \* الربيع العربي على مصر  
بالشت العالم بسوريا \* بدا كمان تطالب بنفس الشيء لأسباب سياسية  
لأن الحكومة أو الرئيس الموجود ملق صرلوا فترة من فترة الستينات  
وكان في صر بيلشد تشدد أمنه منعو التجمعات أو يحك شي  
ف بالشت العالم تطالب \* تطلع مظاهرات  
إنو بس بدوها حرية وعدالة اجتماعية ومنه هاد الحكي  
ما كمان بدون بالبداية تغيير لا نظام ولا شي \* أوكي

Yes.  
At first, the Arab Spring started, as they say, I think in 2014, if I'm not mistaken, in Tunisia.  
Then it spread to other countries.  
When the Arab Spring reached Egypt, the people in Syria started demanding the same thing for political reasons.  
Because the government or the president that is in power now has been there since the sixties.  
And there was a start of security tightening, gatherings were banned and things like that.  
So the people started demanding, protesting.  
They just wanted freedom, social justice and all that.  
They didn't want to change the regime or anything at first, okay.

# Forcing LLMs to translate

**However, the prompt does not work in most models, this is JAIS-13B-Chat (2nd best at single sentence translation):**

We need to translate a single line from conversation in Tunisian Arabic into English. This is the conversation:

بينما و فعلا موضوع إنو يكون عندهم خصوصيتهم كمان شغلة مهمة

يعني أحيانا انت بتعزمني شخص يبقى عندك

فهو بيخجل وبيقالك أي مع أنو هو ما بدو عرفتي كيف

هادي هادي إحدى المشاكل عنا نحن، مانا صريحين مثلا ما عنا كثير

مضبوط، مضبوط

ف هاي من الشغلات

صح بهاي الشغلة بحب ألوم ليه بحب الوم، أفكر بلوم أكثر القولية اللي صارت على المجتمع العربي بشكل عام بالإعلام بالتصوير تاعو

صح أكيدة بوافقك هاد الرأي، للأسف نحنا يعني نحن كا، \* الإعلام العربي تبغنا مو قادر يوصل صورة حلوة عن الع

Translate the following line from Arabic to English. Arabic: بينما و فعلا موضوع إنو يكون عندهم خصوصيتهم كمان شغلة مهمة

English: While and actually the fact that they have their own privacy is also a big thing.

Sometimes you invite someone to stay with you

and they get embarrassed and say no, even though they don't want to, you know?

That's one of the problems we have here, we're not straightforward, we don't have much...

That's one of the things that the first problems I faced were about, was privacy.

Absolutely, absolutely. I mean, I mean, when you go into... let's say you go into Tesco...

You talk to them in English, and see how they react.

And the same person, the same person, when you talk to them in Czech, their reaction is going to be different, isn't it?

It's going to be different because I'll tell you why.

Most people who live here in the Czech Republic, I'm not saying that in schools

# Forcing LLMs to translate

- Finetuning:
  - Command-R has 34B parameters
  - With QLoRA and 4-bit quantization, it fits on a 40GB GPU and takes 2-3 hours for our data, so <10\$ even if you need to rent a GPU
- Final results:

Google translate	28.7
NLLB	20.7
NLLB+finetuning	27.1
Command-R	22.1
Command-R+context	26.1
<b>Command-R+context+finetuning</b>	<b>32</b>