CSCE 638 Natural Language Processing Foundation and Techniques

Lecture 17: Multilingual Models

Kuan-Hao Huang

Spring 2026

(Some slides adapted from Graham Neubig)

Invited Talk

- **Date:** 4/6 online
- **Talk:** Improving Personalization and Consistency of Large Foundation Models
- **Speaker:** Jindong Wang, Assistant Professor, The College of William & Mary

Invited Talk

Abstract:
Foundation models, including large language models and multimodal models, are becoming indispensable in how we live, work, and communicate. Our research aims to improve these models by enhancing personalization and consistency. First, although global alignment has significantly improved safety, it remains unclear whether and how foundation models can align to support individual safety. Second, large multimodal models can both generate and understand content, but these capabilities can conflict, leading to inconsistencies, such as “what I can understand, I cannot create,” and vice versa. In this talk, I will share our recent work on personalization and consistency in foundation models, with the goal of drawing the community’s attention to these two critical concepts.

Invited Talk

Speaker Bio:

Dr. Jindong Wang is an Assistant Professor at the Department of Data Science, William & Mary, and a faculty member of the Future of Life Institute. Previously, he was a Senior Researcher at Microsoft Research Asia (2019–2024). His research focuses on machine learning, large foundation models, and generative AI. He is recognized as one of the World's Top 2% Highly Cited Scientists and Most Influential AI Scholars. He serves as Associate Editor of IEEE TNNLS and has held area chair roles for ICML, NeurIPS, ICLR, KDD, ACL, ACMMM, and ACML. Dr. Wang has published 60+ papers with 23,000+ citations (H-index 54). His work is supported by awards from Amazon, Google, NVIDIA, and others, and has been deployed in Microsoft healthcare systems and quantitative finance. His research has been featured in Forbes and MIT Technology Review. He is also the author of Introduction to Transfer Learning and has delivered tutorials at major AI conferences including IJCAI, KDD, AAAI, and CVPR.

Final Presentation

- Each team has 7 minutes for presentation
  - You have to stop once you reach 7 minutes
- The presentation should include
  - The topic you choose
  - Novelty/challenges
  - Your approach/design
  - Experiments/evaluation
  - Results, findings, insights/demo
- Clarity is important
  - Teach your classmate about your topic
- Time control is also important

Final Presentation

- Presentation dates
  - 4/20, 4/22, 4/27
- Online
  - Zoom link will be posted later
- Presentation order
  - https://docs.google.com/spreadsheets/d/1qUZPFI4wciToJsXye8-WN4L7xVG38IWdS2GCCzmu84A/edit?usp=sharing

Multilingual NLP

- Different languages have different linguistic properties
  - Scripts
    - Alphabets, symbols, spaces

I like this restaurant

我喜欢这家餐厅

Me gusta este restaurante

我喜歡這家餐廳

J'aime ce restaurant

나는 이 레스토랑을 좋아한다

मुझे यह रेस्तरां पसंद है

このレストランが好きです

Multilingual NLP

- Different languages have different linguistic properties
  - Scripts
    - Alphabets, symbols, spaces
  - Grammatical rules
    - SVO (Subject-Verb-Object): English, Chinese, French
    - SOV (Subject-Object-Verb): Japanese, Hindi, Korean
    - VSO (Verb-Subject-Object): Arabic, Tagalog, Irish

	Value	Representation
●	Subject-object-verb (SOV)	564
●	Subject-verb-object (SVO)	488
●	Verb-subject-object (VSO)	95
◆	Verb-object-subject (VOS)	25
◆	Object-verb-subject (OVS)	11
◆	Object-subject-verb (OSV)	4
●	Lacking a dominant word order	189
	Total:	1376

https://wals.info/chapter/81

Multilingual NLP

- Different languages have different linguistic properties
  - Scripts
    - Alphabets, symbols, spaces
  - Grammatical rules
    - SVO (Subject-Verb-Object): English, Chinese, French
    - SOV (Subject-Object-Verb): Japanese, Hindi, Korean
    - VSO (Verb-Subject-Object): Arabic, Tagalog, Irish
  - Writing systems
    - Left-to-right
    - Right-to-left

I like this restaurant

انا احب هذا المطعم

我喜歡這家餐廳

אני אוהב את המסעדה הזו

More About Language Properties

- World Atlas of Language Structures (WALS) database

https://wals.info/

Lang2Vec

Vector type	#Languages	#Features	#Data points	%Coverage
Syntax (from sources)				
syntax_wals	1808	98	78732	44%
syntax_sswl	230	33	6404	84%
syntax_ethnologue	1336	30	18105	45%
Syntax (averaged over sources)				
syntax_avg	2654	103	94227	34%
Syntax (predicted)				
syntax_knn	7970	103	820910	100%
Phonology (from sources)				
phonology_wals	832	27	14358	64%
phonology_ethnologue	543	8	1017	23%
Phonology (averaged over sources)				
phonology_avg	1296	28	15303	42%
Phonology (predicted)				
phonology_knn	7970	28	223160	100%
Inventory (from sources)				
inventory_phoible_aa	202	158	31916	100%
inventory_phoible_gm	428	158	67624	100%
inventory_phoible_ph	404	158	63832	100%
inventory_phoible_ra	100	158	15800	100%
inventory_phoible_saphon	334	158	52772	100%
inventory_phoible_spa	219	158	34602	100%
inventory_phoible_upsid	334	158	75050	100%
Inventory (averaged over sources)				
inventory_avg	1715	158	270970	100%
Inventory (predicted)				
inventory_knn	7970	158	1259260	100%

Lang2Vec

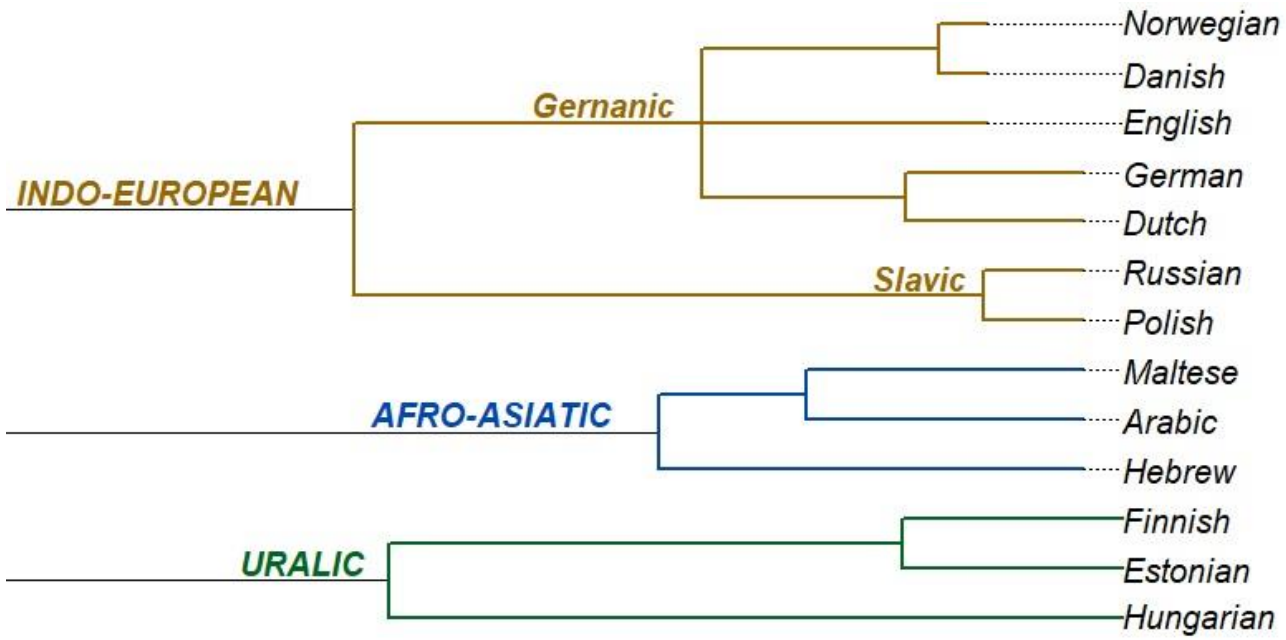
- https://github.com/antonisa/lang2vec

```
>>> features = l2v.get_features(["eng", "fra"], "geo")
>>> features["fra"]
[0.7378000020980835, 0.7682999968528748, 0.7982000112533569, 0.6941999793052673, ...]

>>> features = l2v.get_features("eng fr", "syntax_wals")
>>> features["eng"]
[1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...]
>>> features["fr"]
[1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...]
```

```
>>> import lang2vec.lang2vec as l2v
>>> features = l2v.get_features("eng", "geo")
>>> features["eng"]
[0.7664999961853027, 0.7924000024795532, 0.8277999758720398, 0.7214000225067139,...]
>>>
>>> l2v.distance("syntactic", "eng", "fra")
0.4569
```

Language Hierarchy



http://www.elinguistics.net/

Multilingual Tokenization is Challenging

I like this restaurant

我喜歡這家餐廳 → 我 喜歡 這家 餐廳

一個半小時 → 一個 半小時 (A half hour)

一個半小時 → 一個半 小時 (One and a half hour)

這幾天天天天氣不好 → 這幾天 天天 天氣 不好

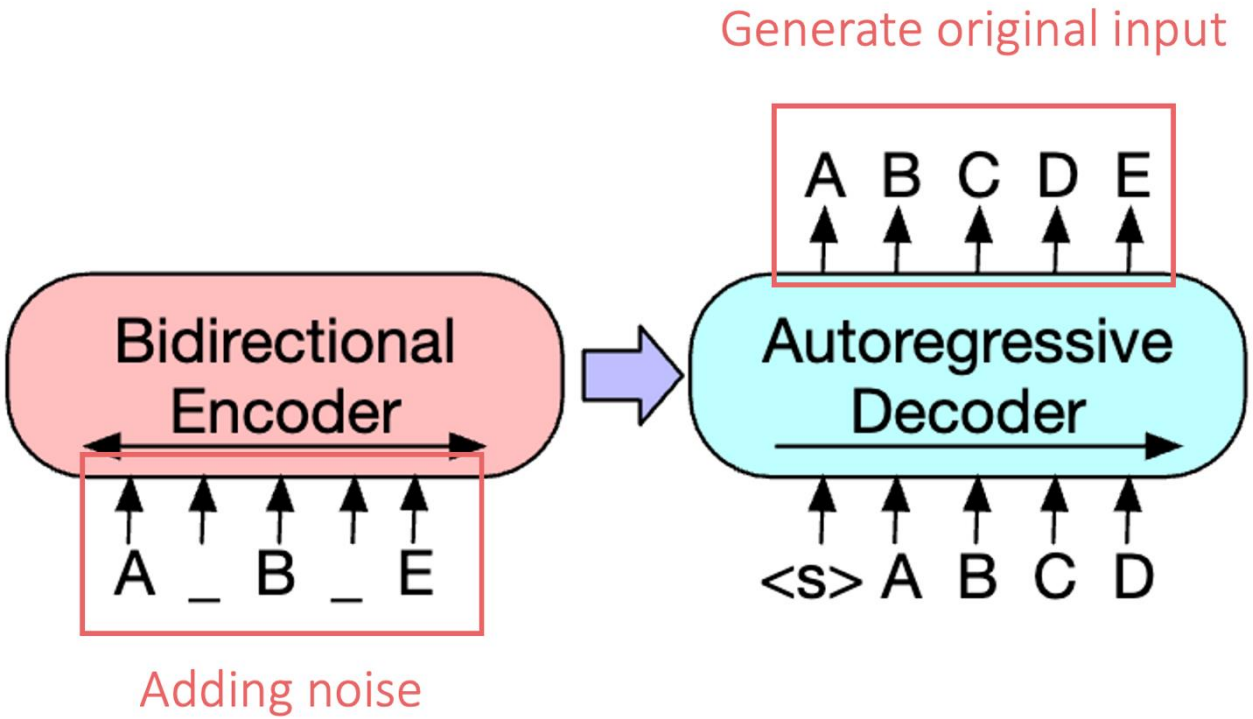
Translation

- A sequence-to-sequence task
- Need parallel data

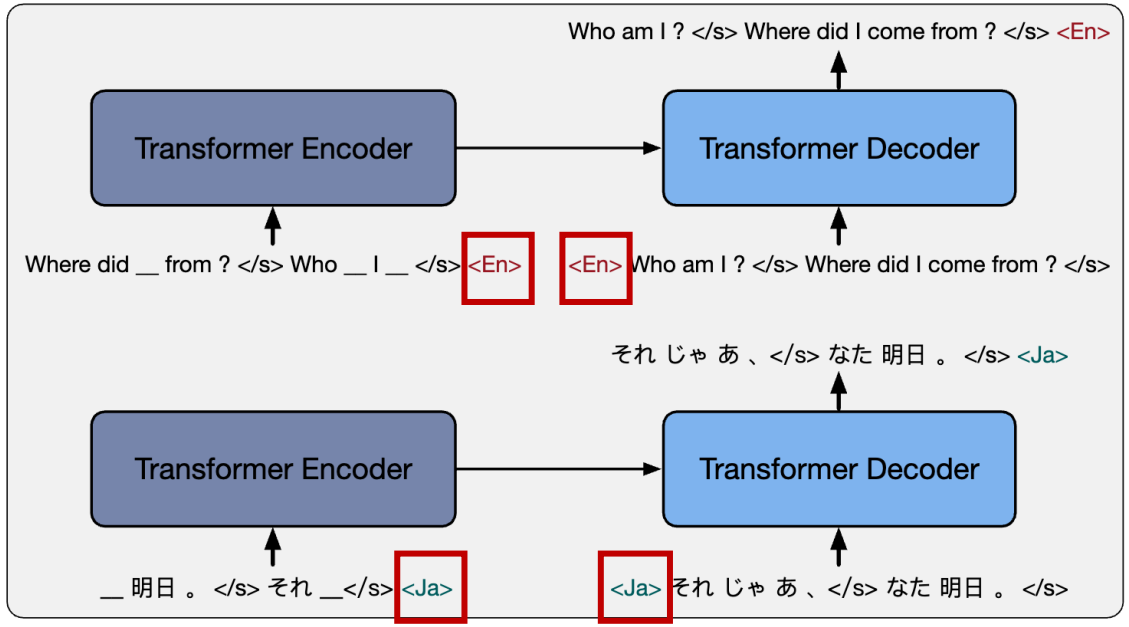
I like this restaurant	我喜欢这家餐厅
Me gusta este restaurante	我喜歡這家餐廳
J'aime ce restaurant	나는 이 레스토랑을 좋아한다
मुझे यह रेस्तरां पसंद है	このレストランが好きです

Recap: BART – Denoising Objective

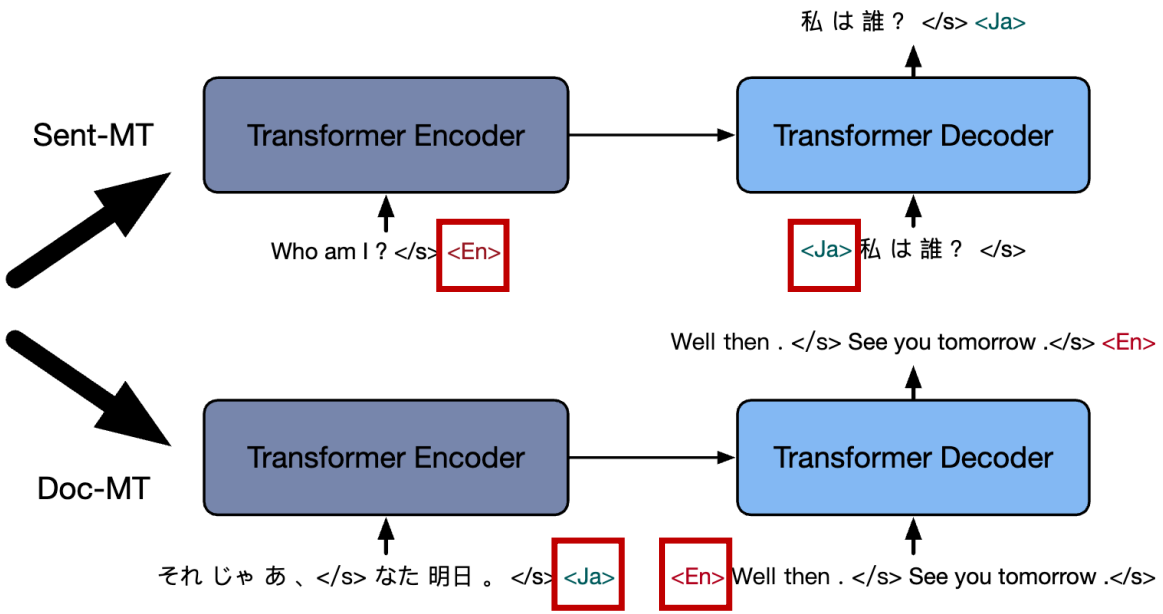
- Token Masking
  - A<mask>CD<mask>F. → ABCDEF
- Token Deletion
  - ACDF. → ABCDEF.
- Text Infilling
  - A<mask>D<mask>F. → ABCDEF.
- Sentence Permutation
  - FG. ABC. DE. → ABC. DE. FG.
- Document Rotation
  - E. FG. ABC. D → ABC. DE. FG.



Multilingual BART (mBART)



Multilingual Denoising Pre-Training (mBART)



Fine-tuning on Machine Translation

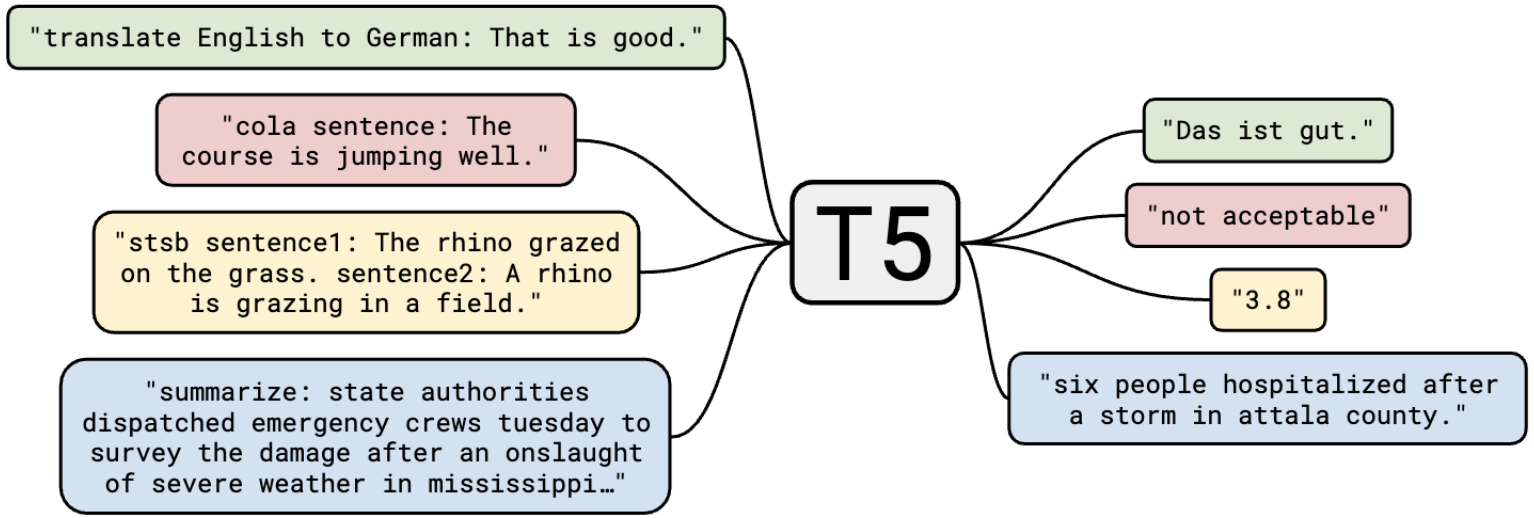
Multilingual BART (mBART)

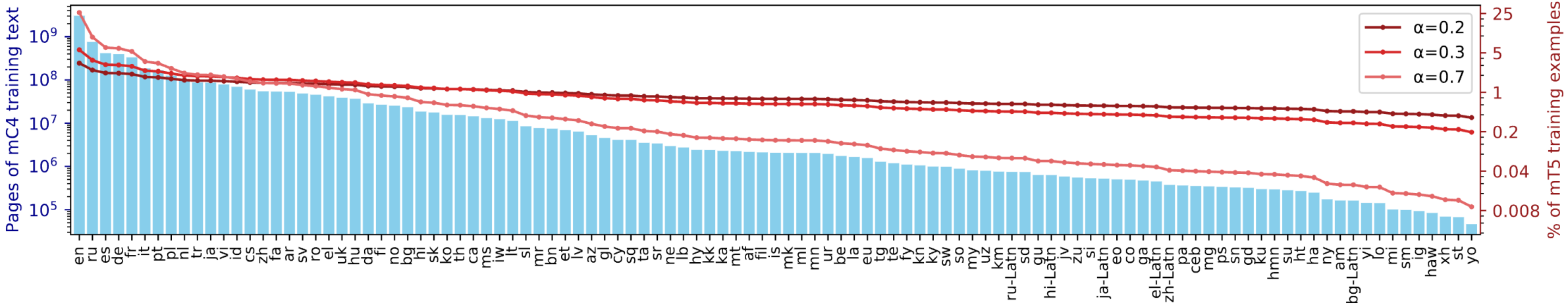
- mBART25 and mBART50

Code	Language	Tokens/M	Size/GB
En	English	55608	300.8
Ru	Russian	23408	278.0
Vi	Vietnamese	24757	137.3
Ja	Japanese	530 (*)	69.3
De	German	10297	66.6
Ro	Romanian	10354	61.4
Fr	French	9780	56.8
Fi	Finnish	6730	54.3
Ko	Korean	5644	54.2
Es	Spanish	9374	53.3
Zh	Chinese (Sim)	259 (*)	46.9
It	Italian	4983	30.2
Nl	Dutch	5025	29.3
Ar	Arabic	2869	28.0
Tr	Turkish	2736	20.9
Hi	Hindi	1715	20.2
Cs	Czech	2498	16.3
Lt	Lithuanian	1835	13.7
Lv	Latvian	1198	8.8
Kk	Kazakh	476	6.4
Et	Estonian	843	6.1
Ne	Nepali	237	3.8
Si	Sinhala	243	3.6
Gu	Gujarati	140	1.9
My	Burmese	56	1.6

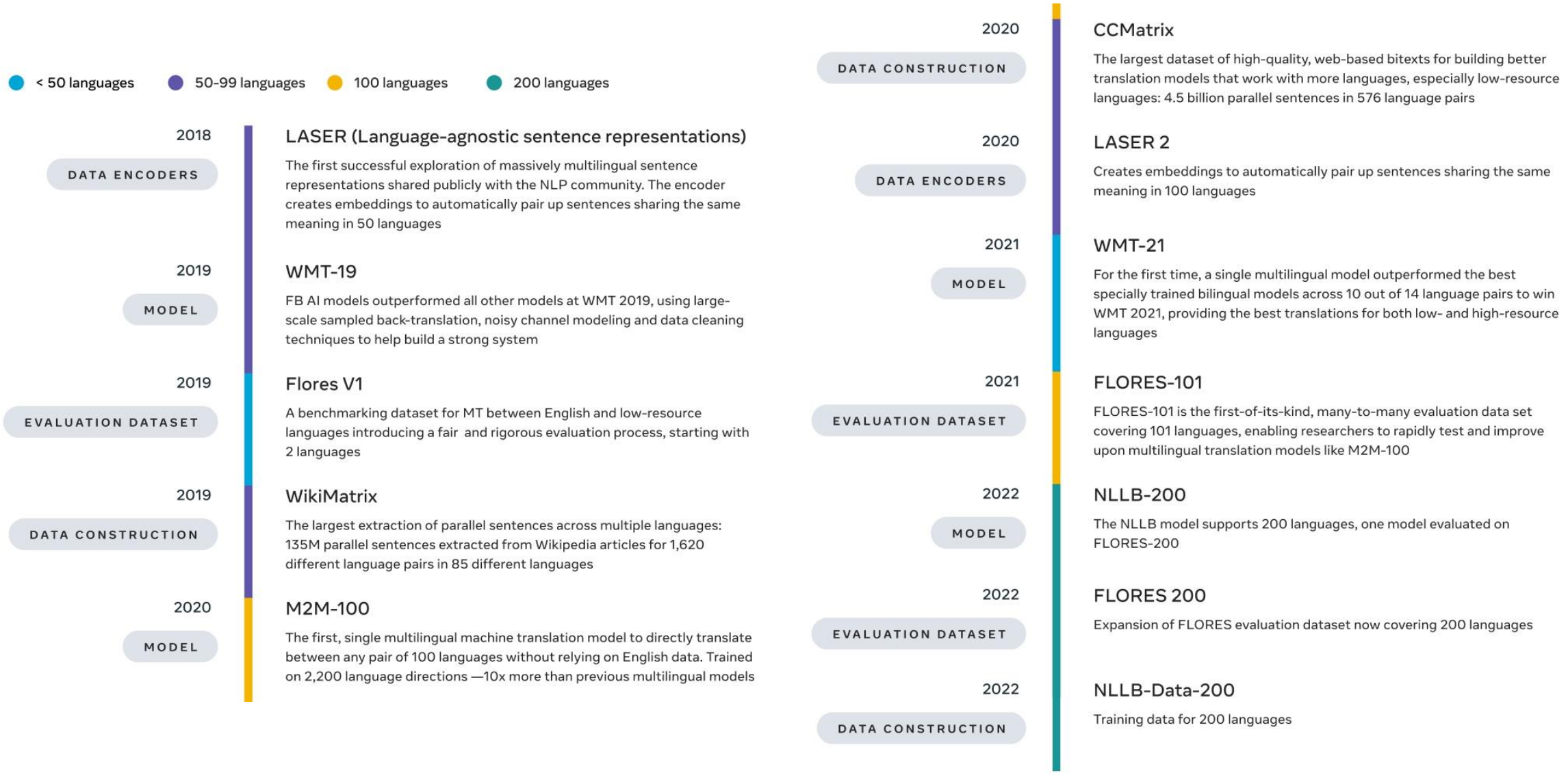
Data size	Languages
10M+	German, Czech, French, Japanese, Spanish, Russian, Polish, Chinese
1M - 10M	Finnish, Latvian, Lithuanian, Hindi, Estonian
100k to 1M	Tamil, Romanian, Pashto, Sinhala, Malayalam, Dutch, Nepali, Italian, Arabic, Korean, Hebrew, Turkish, Khmer, Farsi, Vietnamese, Croatian, Ukrainian
10K to 100K	Thai, Indonesian, Swedish, Portuguese, Xhosa, Afrikaans, Kazakh, Urdu, Macedonian, Telugu, Slovenian, Burmese, Georgia
10K-	Marathi, Gujarati, Mongolian, Azerbaijani, Bengali

Multilingual T5 (mT5)





Translation: No Language Left Behind (NLLB)



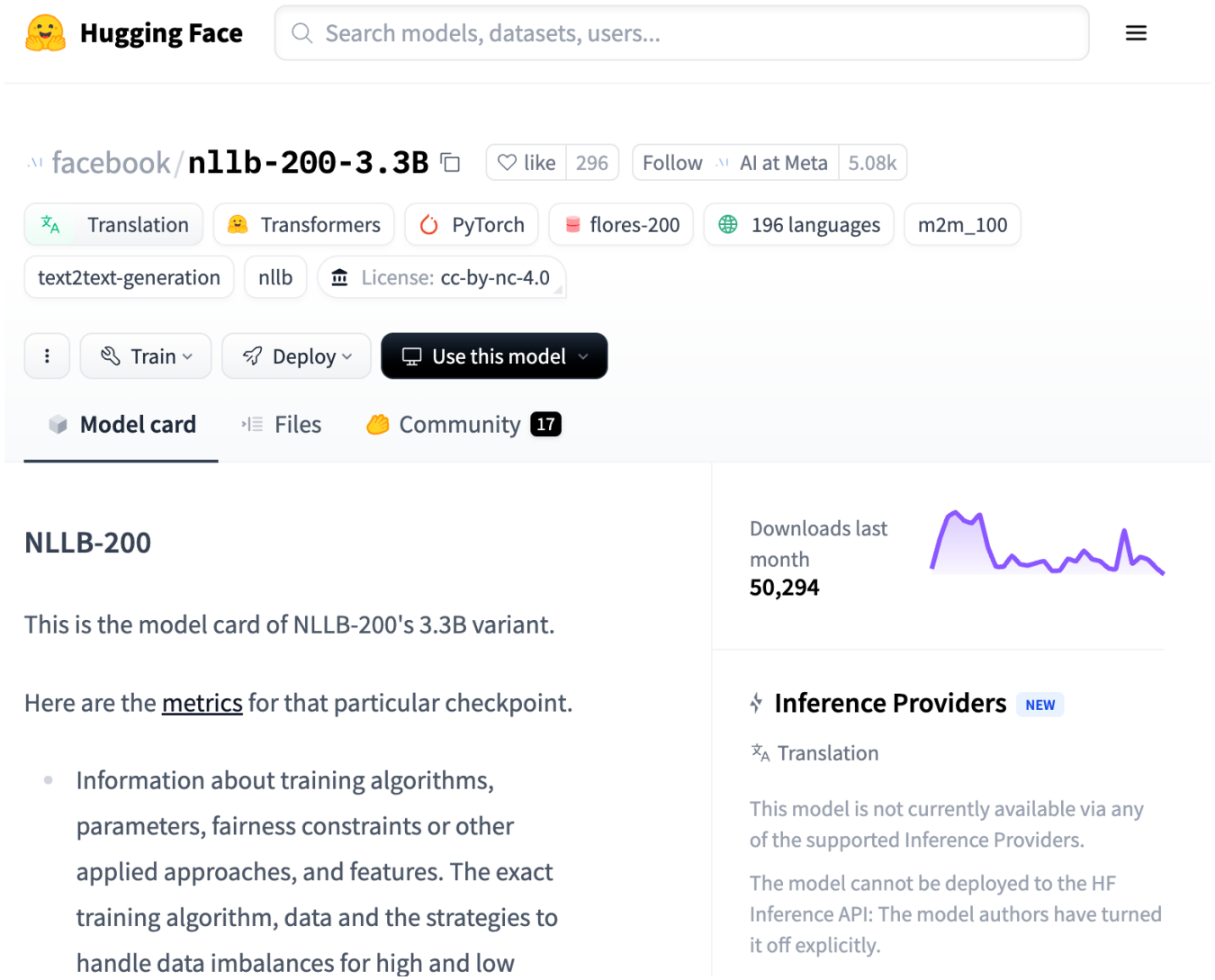
https://ai.meta.com/blog/nllb-200-high-quality-machine-translation/

Translation: No Language Left Behind (NLLB)

Acehnese (Latin script)
Arabic (Iraqi/Mesopotamian)
Arabic (Yemen)
Arabic (Tunisia)
Afrikaans
Arabic (Jordan)
Akan
Amharic
Arabic (Lebanon)
Arabic (MSA)
Arabic (Modern Standard Arabic)
Arabic (Saudi Arabia)
Arabic (Morocco)
Arabic (Egypt)
Assamese
Asturian
Awadhi
Aymara
Crimean Tatar
Welsh
Danish
German
French
Friulian
Fulfulde
Dinka(Rek)
Dyula
Dzongkha
Greek
English
Esperanto
Estonian
Basque
Ewe
Faroese
Iranian Persian
Icelandic
Italian
Javanese
Japanese
Kabyle
Kachin | Jinghpo
Kamba
Kannada
Kashmiri (Arabic script)
Kashmiri (Devanagari script)
Georgian
Kanuri (Arabic script)
Kanuri (Latin script)
Kazakh
Kabiye
Thai
Khmer
Kikuyu
South Azerbaijani
North Azerbaijani
Bashkir
Bambara
Balinese
Belarusian
Bemba
Bengali
Bhojpuri
Banjar (Latin script)
Tibetan
Bosnian
Buginese
Bulgarian
Catalan
Cebuano
Czech
Chokwe
Central Kurdish
Fijian
Finnish
Fon
Scottish Gaelic
Irish
Galician
Guarani
Gujarati
Haitian Creole
Hausa
Hebrew
Hindi
Chhattisgarhi
Croatian
Hugarian
Armenian
Igobo
Ilocano
Indonesian
Kinyarwanda
Kyrgyz
Kimbundu
Konga
Korean
Kurdish (Kurmanji)
Lao
Latvian (Standard)
Ligurian
Limburgish
Lingala
Lithuanian
Lombard
Latgalian
Luxembourgish
Luba-Kasai
Ganda
Dholuo
Mizo

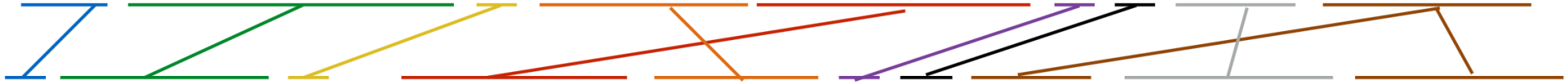
https://ai.meta.com/research/no-language-left-behind/

Translation: No Language Left Behind (NLLB)



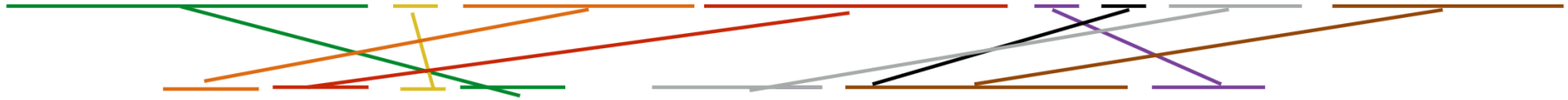
Word Alignment

The development of artificial intelligence is a really big deal.



El desarrollo de la inteligencia artificial es un asunto realmente importante.

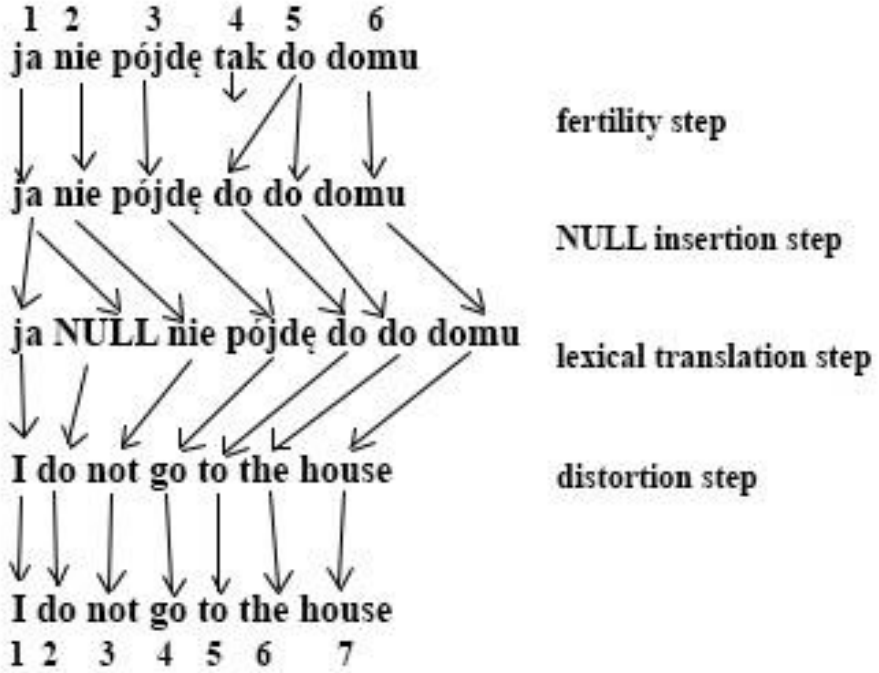
The development of artificial intelligence is a really big deal.



人工知能の発展は本当にすごいことです。

IBM Alignment Models

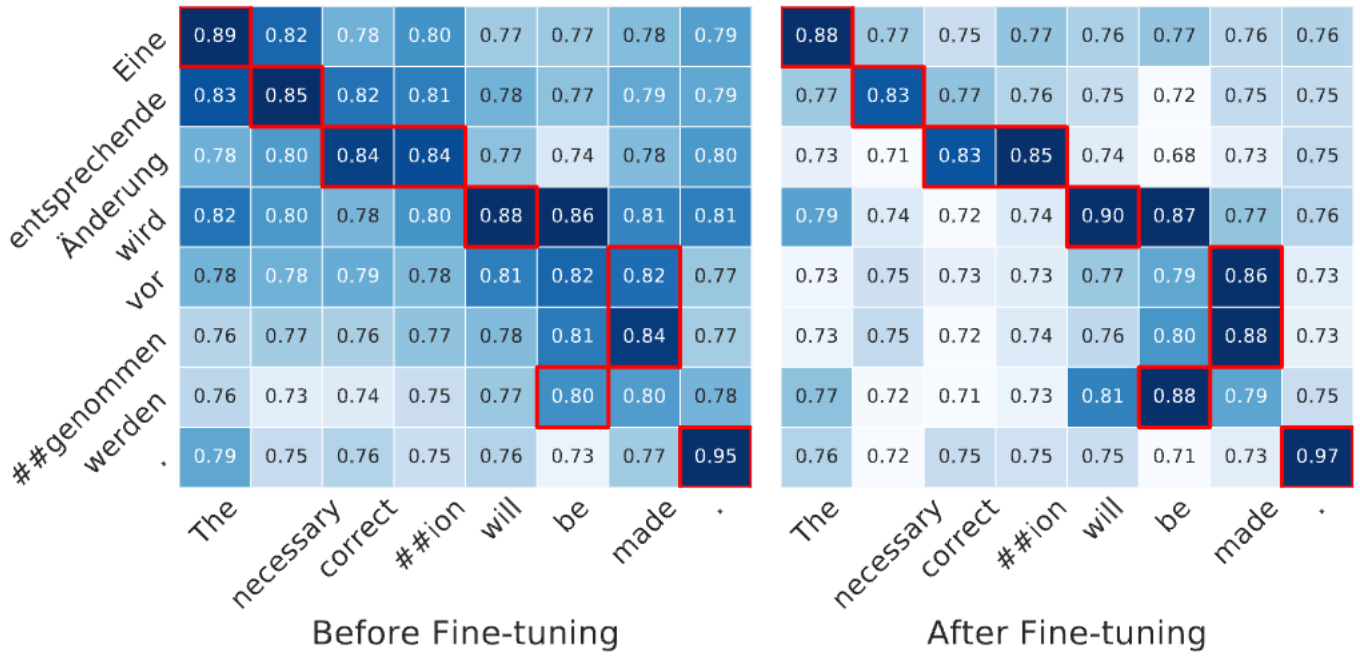
- Based on statistics
  - Model 1: lexical translation
  - Model 2: additional absolute alignment model
  - Model 3: extra fertility model
  - Model 4: added relative alignment model
  - Model 5: fixed deficiency problem.
  - Model 6: Model 4 combined with
    a HMM alignment model in a log linear way



https://en.wikipedia.org/wiki/IBM_alignment_models

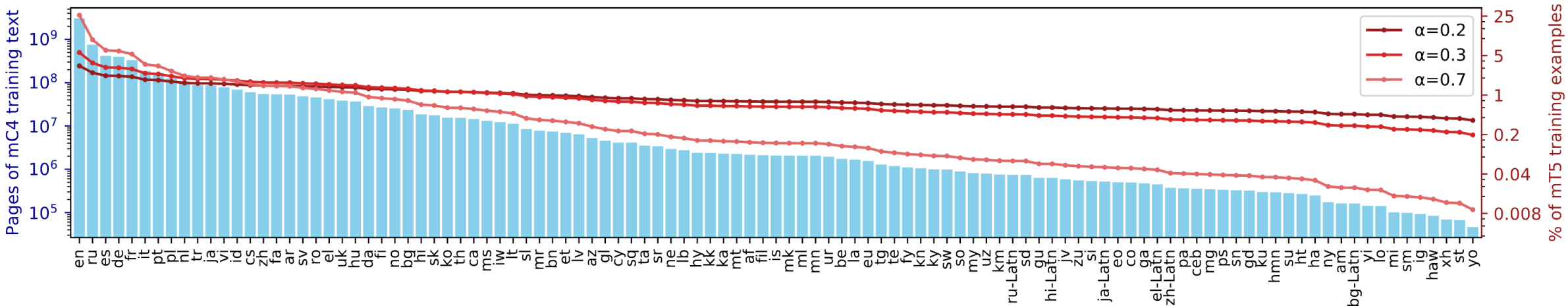
Learning-Based Alignment Models

- Awesome-Align
  - https://github.com/neulab/awesome-align



Cross-Lingual Transfer Learning

- Language resources are not evenly distributed



Can we transfer the learned knowledge from languages to languages?

Cross-Lingual Transfer Learning

- Training examples in a source language
- Testing examples in a target language

I like this restaurant because its food is good. ☺

我喜欢这家餐厅，因为它的食物很好。 ?

J'aime ce restaurant car sa cuisine est bonne. ?

Ich mag dieses Restaurant, weil das Essen gut ist. ?

मुझे यह रेस्टोरेंट पसंद है क्योंकि इसका खाना अच्छा है। ?

A Simple Baseline: Translation-Train

- Training examples in a source language
  - Translate training examples to the target language
  - Train the model f with translated training examples
- Testing examples in a target language
  - Test the model f

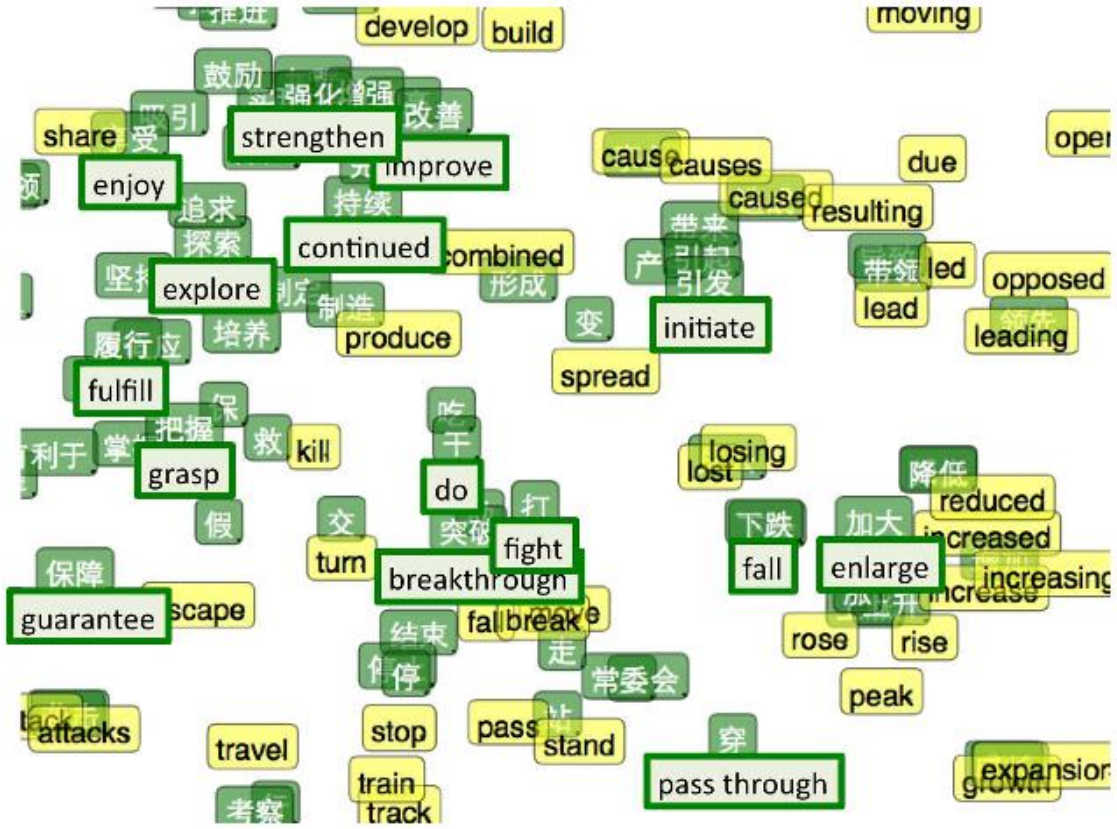
A Simple Baseline: Translation-Test

- Training examples in a source language
  - Train the model f with training examples
- Testing examples in a target language
  - Translate testing examples to the source language
  - Test the model f on translated examples

What if we don't have translator?

Multilingual Embedding

- Train the model f in multilingual embedding

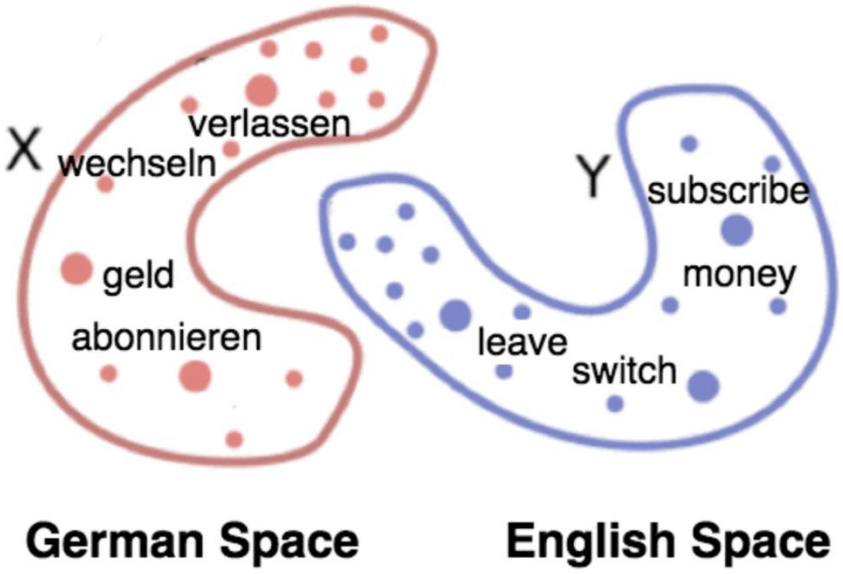


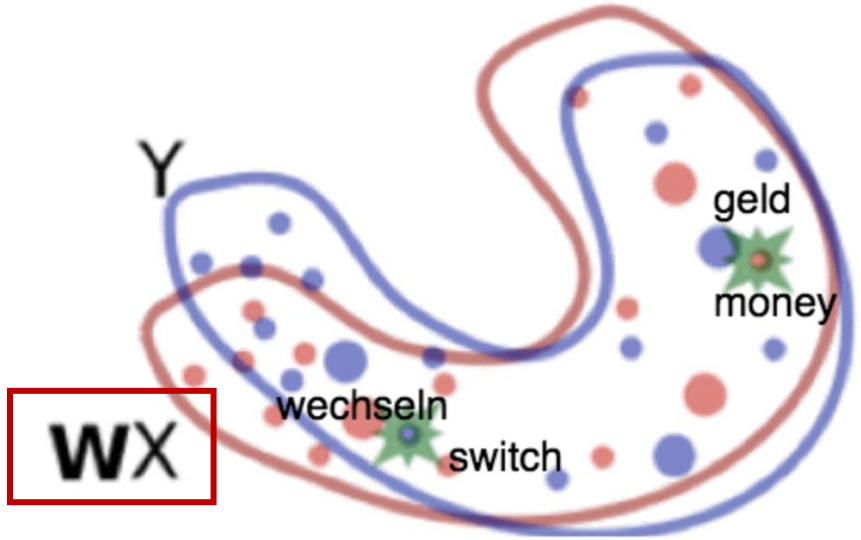
How to learn multilingual embedding?

https://www.ruder.io/cross-lingual-embeddings/

Multilingual Embedding Alignment

- Train word embeddings for languages separately
- Alignment different embedding spaces to one universal space





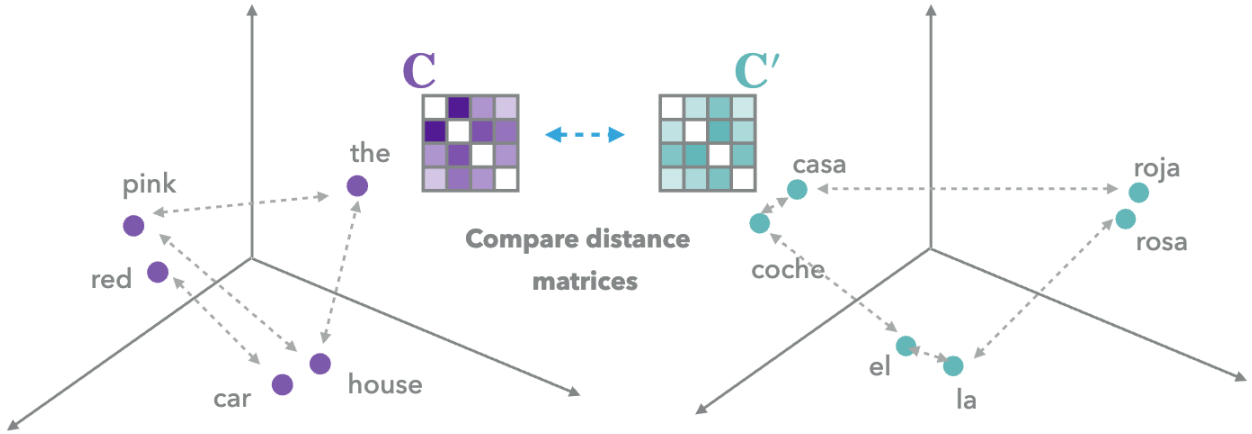
Linear projection!

https://meryemmhamdi1.github.io/transfer-learning-multilingual-emb

Multilingual Embedding Alignment

The Gromov-Wasserstein Distance

- Generalizes OT to the non-registered case
- Main idea: compare **distances** instead of absolute **positions**

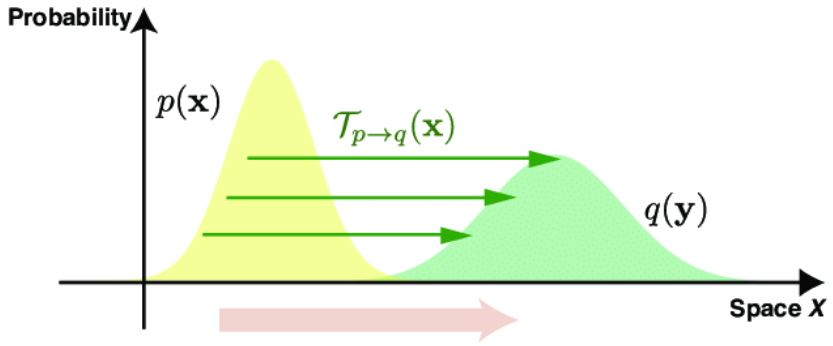


cost of matching $\mathbf{x}^{(i)}$ to $\mathbf{y}^{(j)}$ *and* $\mathbf{x}^{(k)}$ to $\mathbf{y}^{(l)}$ =

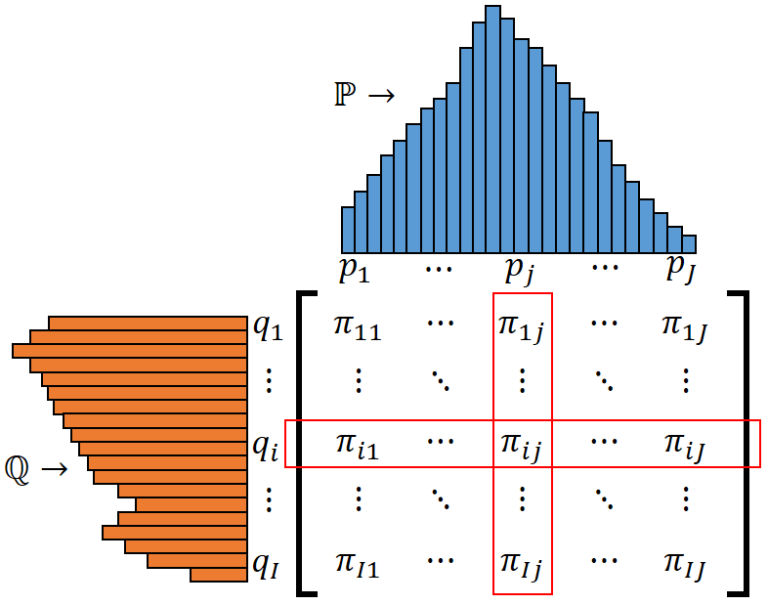
$$\mathcal{L}\big(d(\mathbf{x}^{(i)}, \mathbf{x}^{(i)})\big), d(\mathbf{x}^{(i)}, \mathbf{x}^{(i)})\big)$$

- The objective:

$$\mathrm{GW}(\mathbf{C}, \mathbf{C}', \mathbf{p}, \mathbf{q}) = \min_{\Gamma \in \Pi(\mathbf{p},\mathbf{q})} \sum_{i,j,k,l} \mathcal{L}(\mathbf{C}_{ik}, \mathbf{C}'_{jl}) \Gamma_{ij} \Gamma_{kl}$$



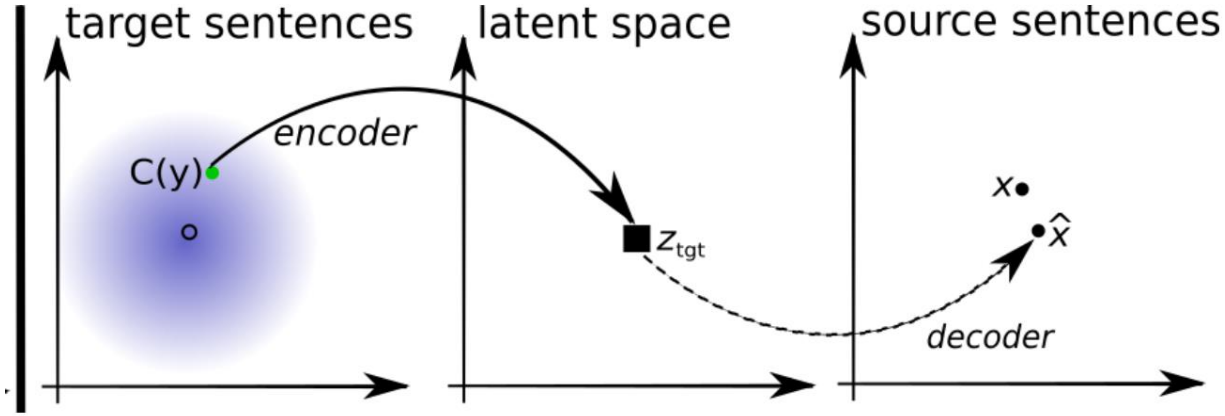
$\mathcal{W}(p,q)^2 = \int d\mathbf{x} \|\mathbf{x} - \mathcal{T}_{p\to q}(\mathbf{x})\|^2 p(\mathbf{x})$: **L^2-Wasserstein distance**



Gromov-Wasserstein Alignment of Word Embedding Spaces

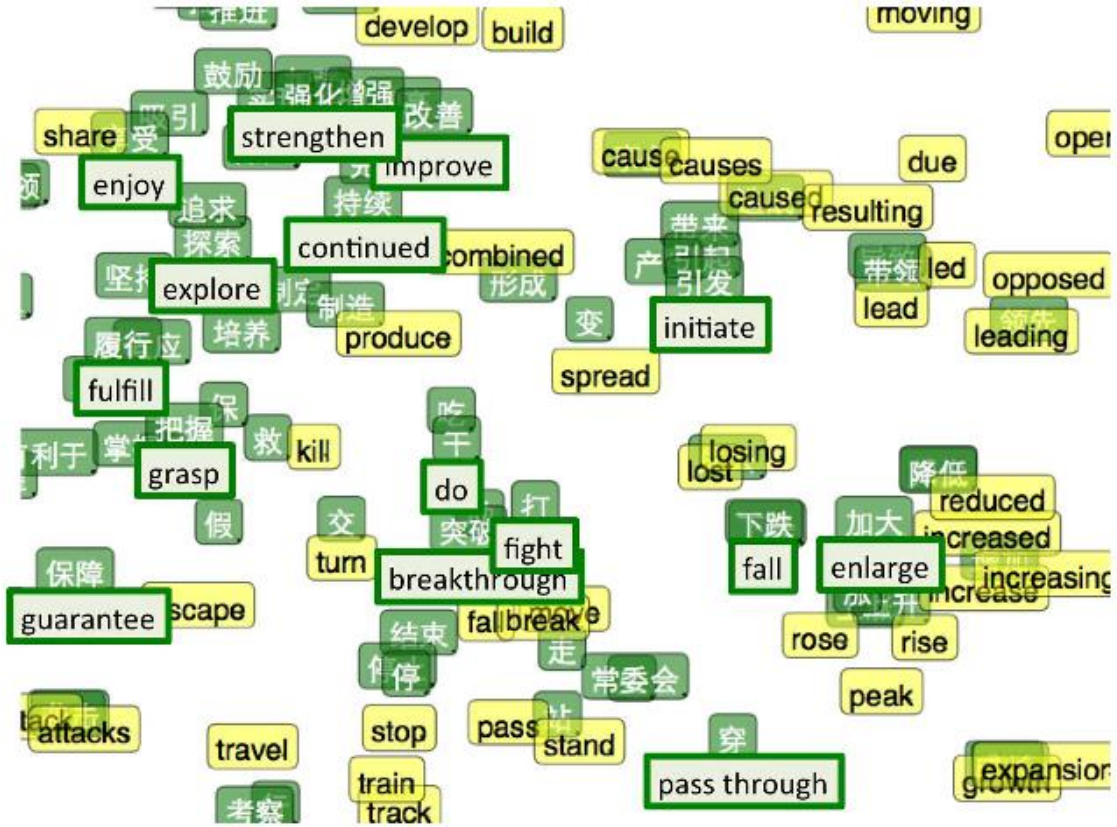
Unsupervised Machine Translation

- Unparallel corpus
  - Corpus in language 1
  - Corpus in language 2
- High-level idea
  - Unsupervised multilingual word alignment
  - Word-level translation
  - Learn syntax from corpus



Multilingual Embedding

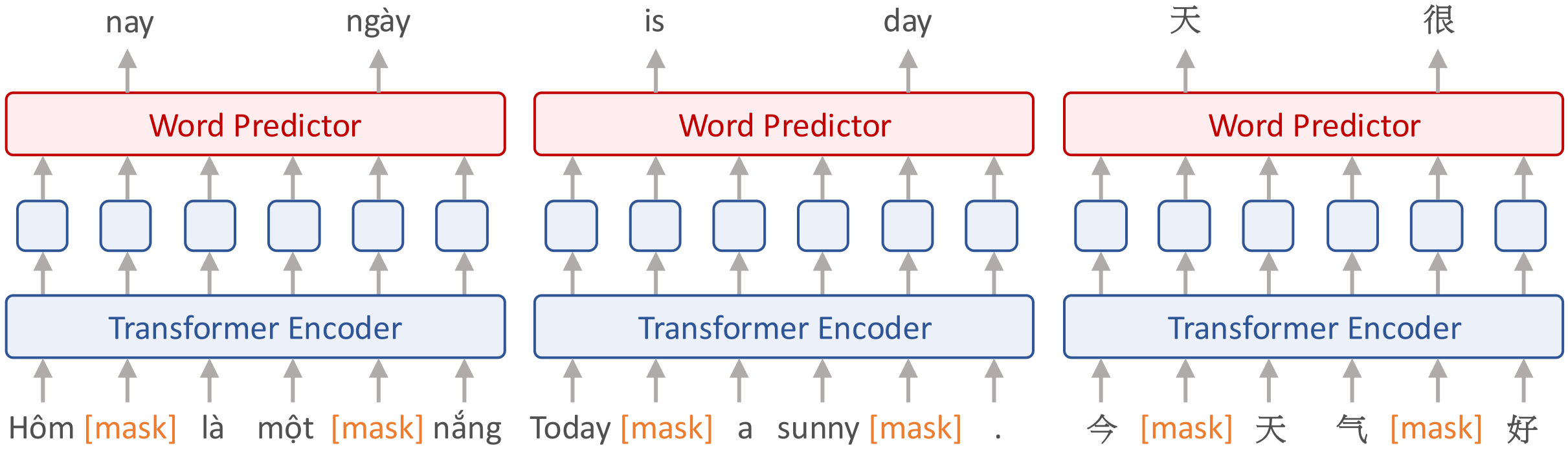
- Train the model f in multilingual embedding



https://www.ruder.io/cross-lingual-embeddings/

Joint Learning for Multilingual Embedding

- BERT/RoBERTa
  - Masked language modeling



Multilingual version: mBERT / XLM-R

Impressive Cross-Lingual Transfer Performance

Model	D	#M	#lg	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	Avg
Fine-tune multilingual model on English training set (Cross-lingual Transfer)																			
Lample and Conneau (2019)	Wiki+MT	N	15	85.0	78.7	78.9	77.8	76.6	77.4	75.3	72.5	73.1	76.1	73.2	76.5	69.6	68.4	67.3	75.1
Huang et al. (2019)	Wiki+MT	N	15	85.1	79.0	79.4	77.8	77.2	77.2	76.3	72.8	73.5	76.4	73.6	76.2	69.4	69.7	66.7	75.4
Devlin et al. (2018)	Wiki	N	102	82.1	73.8	74.3	71.1	66.4	68.9	69.0	61.6	64.9	69.5	55.8	69.3	60.0	50.4	58.0	66.3
Lample and Conneau (2019)	Wiki	N	100	83.7	76.2	76.6	73.7	72.4	73.0	72.1	68.1	68.4	72.0	68.2	71.5	64.5	58.0	62.4	71.3
Lample and Conneau (2019)	Wiki	1	100	83.2	76.7	77.7	74.0	72.7	74.1	72.7	68.7	68.6	72.9	68.9	72.5	65.6	58.2	62.4	70.7
XLM-R$_{\text{Base}}$	CC	1	100	85.8	79.7	80.7	78.7	77.5	79.6	78.1	74.2	73.8	76.5	74.6	76.7	72.4	66.5	68.3	76.2
XLM-R	CC	1	100	**89.1**	**84.1**	**85.1**	**83.9**	**82.9**	**84.0**	**81.2**	**79.6**	**79.8**	**80.8**	**78.1**	**80.2**	**76.9**	**73.9**	**73.8**	**80.9**
Translate everything to English and use English-only model (TRANSLATE-TEST)																			
BERT-en	Wiki	1	1	88.8	81.4	82.3	80.1	80.3	80.9	76.2	76.0	75.4	72.0	71.9	75.6	70.0	65.8	65.8	76.2
RoBERTa	Wiki+CC	1	1	**91.3**	82.9	84.3	81.2	81.7	83.1	78.3	76.8	76.6	74.2	74.1	77.5	70.9	66.7	66.8	77.8
Fine-tune multilingual model on each training set (TRANSLATE-TRAIN)																			
Lample and Conneau (2019)	Wiki	N	100	82.9	77.6	77.9	77.9	77.1	75.7	75.5	72.6	71.2	75.8	73.1	76.2	70.4	66.5	62.4	74.2

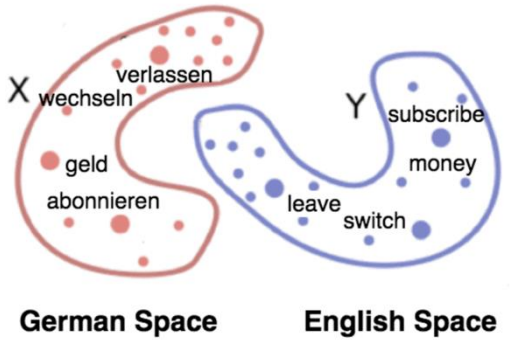
Why Work?

- Anchor word (word-piece overlap)?
  - I like this restaurant
  - J'aime ce restaurant
  - Me gusta este restaurante

Fake-English Test

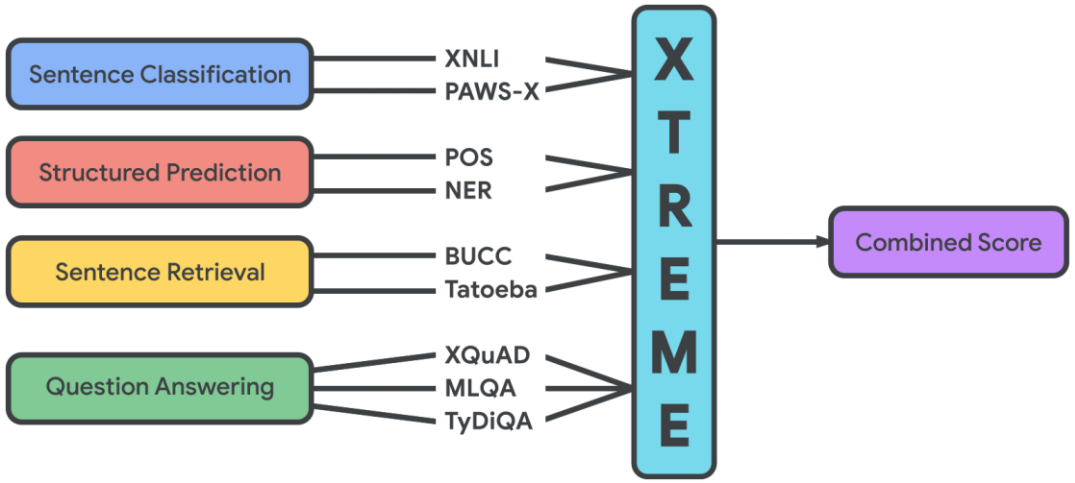
- Fake-English (enfake)
  - A different language than English, but having the exact same properties except word surface forms

B-BERT	Train	Test	XNLI Accuracy	XNLI Word-piece Contribution	NER F_1-Score
en-es	en	es	72.3	1.4	61.9 (±0.8)
enfake-es	enfake		70.9		62.6 (±1.6)
en-hi	en	hi	60.1	0.5	61.6 (±0.7)
enfake-hi	enfake		59.6		62.9 (±0.7)
en-ru	en	ru	66.4	0.7	57.1* (±0.9)
enfake-ru	enfake		65.7		54.2 (±0.7)
en-enfake	enfake	enfake	78.0	0.5	78.9* (±0.7)
en-enfake	enfake	en	77.5		76.6(±0.8)



The contribution of word-piece overlap is small!

XTREME: Benchmark for Cross-Lingual Transfer

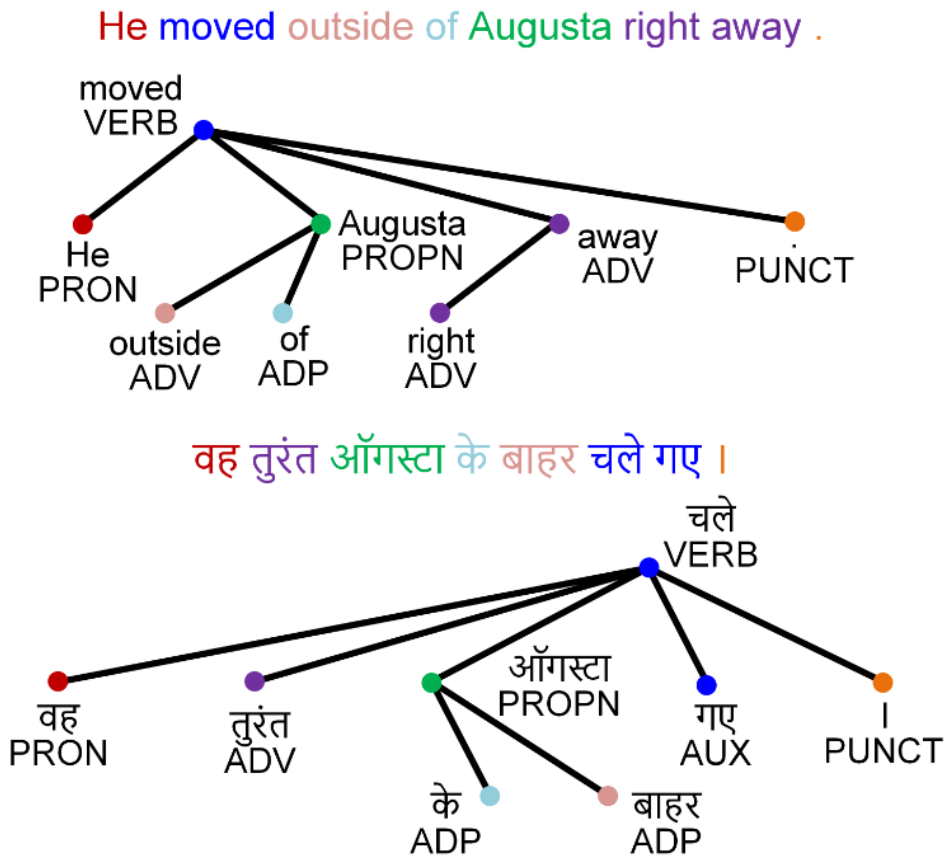


Task	Corpus	|Train|	|Dev|	|Test|	Test sets	|Lang.|	Task	Metric	Domain
Classification	XNLI	392,702	2,490	5,010	translations	15	NLI	Acc.	Misc.
	PAWS-X	49,401	2,000	2,000	translations	7	Paraphrase	Acc.	Wiki / Quora
Struct. pred.	POS	21,253	3,974	47-20,436	ind. annot.	33 (90)	POS	F1	Misc.
	NER	20,000	10,000	1,000-10,000	ind. annot.	40 (176)	NER	F1	Wikipedia
QA	XQuAD	87,599	34,726	1,190	translations	11	Span extraction	F1 / EM	Wikipedia
	MLQA			4,517–11,590	translations	7	Span extraction	F1 / EM	Wikipedia
	TyDiQA-GoldP	3,696	634	323–2,719	ind. annot.	9	Span extraction	F1 / EM	Wikipedia
Retrieval	BUCC	-	-	1,896–14,330	-	5	Sent. retrieval	F1	Wiki / news
	Tatoeba	-	-	1,000	-	33 (122)	Sent. retrieval	Acc.	misc.

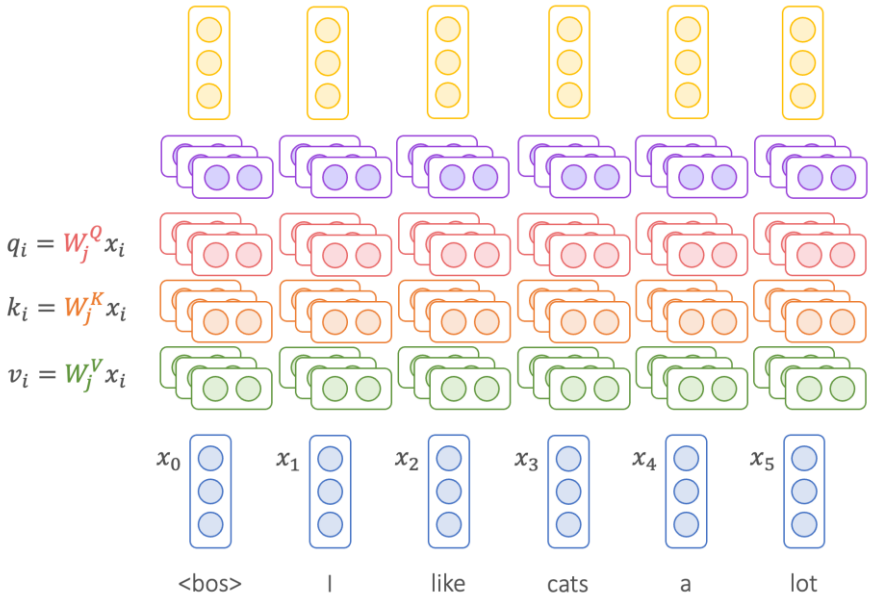
XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalization

Parse Trees Help Cross-Lingual Transfer

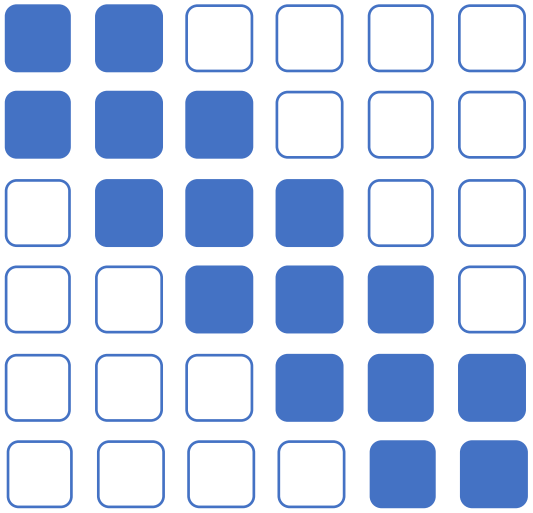
- Translation sentences share similar dependency structures

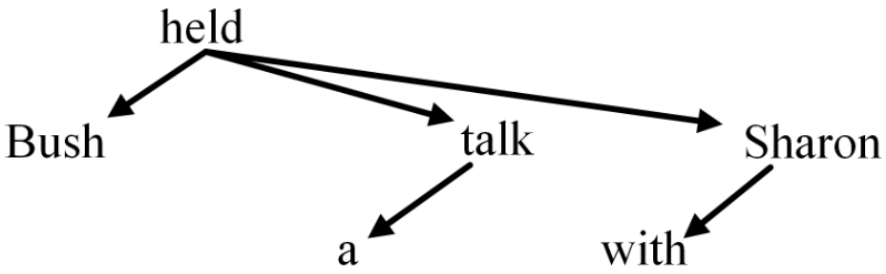


Parse Trees Help Cross-Lingual Transfer



Attention Mask





	Bush	held	a	talk	with	Sharon
Bush	1	1	3	2	3	2
held	1	1	2	1	2	1
a	3	2	1	1	4	3
talk	2	1	1	1	3	2
with	3	2	4	3	1	1
Sharon	2	1	3	2	1	1

GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction

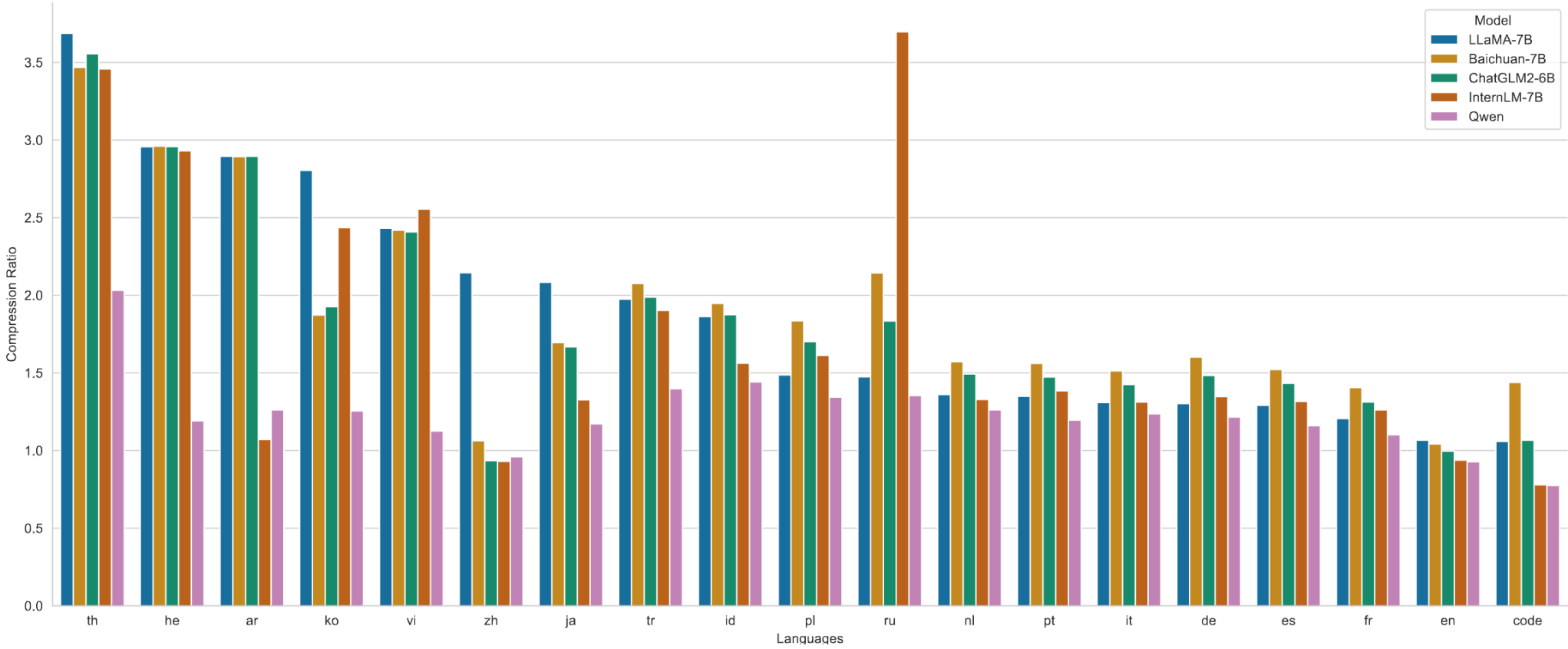
Parse Trees Help Cross-Lingual Transfer

Model	Event Argument Role Labeling						Relation Extraction					
	En ⇓ Zh	En ⇓ Ar	Zh ⇓ En	Zh ⇓ Ar	Ar ⇓ En	Ar ⇓ Zh	En ⇓ Zh	En ⇓ Ar	Zh ⇓ En	Zh ⇓ Ar	Ar ⇓ En	Ar ⇓ Zh
CL_Trans_GCN	41.8	55.6	41.2	52.9	39.6	40.8	56.7	65.3	65.9	59.7	59.6	46.3
CL_GCN	51.9	50.4	53.7	51.5	50.3	51.9	49.4	58.3	65.0	55.0	56.7	42.4
CL_RNN	60.4	53.9	55.7	52.5	50.7	50.9	53.7	63.9	70.9	57.6	67.1	55.7
Transformer	61.5	55.0	58.0	57.7	54.3	57.0	57.1	63.4	69.6	60.6	67.0	52.6
Transformer_RPR	62.3	60.8	57.3	66.3	**57.5**	**59.8**	**58.0**	59.9	70.0	55.6	66.5	**56.5**
GATE (this work)	**63.2**	**68.5**	**59.3**	**69.2**	53.9	57.8	55.1	**66.8**	**71.5**	**61.2**	**69.0**	54.3

GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction

Multilingual Large Language Models

- Llama series / Qwen series



Large Language Models Are Multilingual

If I have 50 dollars and I buy a boba tea that costs 7 dollars, how much do I have left?

You start with **$50** and spend **$7** on boba tea.

Calculation:

50 – 7 = **43**

So you have **$43 left**. 🧋💵

如果我有 50 美元，買了一杯要 7 美元的珍珠奶茶，我還剩多少錢？

你原本有 **50 美元**，買珍珠奶茶花了 **7 美元**。

計算：

50 – 7 = **43**

所以你還剩 **43 美元**。 💰

Do LLMs reason in the same way across different input languages?

Cross-Lingual In-Context Learning

English **ICL Q**	Michael had 58 golf balls. On Tuesday, he lost 23 golf balls. On Wednesday, he lost 2 more. How many golf balls did he have at the end of Wednesday?
ICL A	Michael started with 58 golf balls and lost 23, so he has 58 - 23 = 35. After he lost 2 more, he has 35 - 2 = 33 balls.
English **ICL Q**	Olivia has $23. ⋯ How much does she have left?
ICL A	5 bagels for $3 ⋯ The answer is 8.
English **ICL Q**	⋯ How many lollipops did Jason give to Denny?
ICL A	⋯ The answer is 8.
Thai **Test Q**	⋯ รายได้ของเธอในสัปดาห์นี้จะเท่ากับกี่ดอลลาร์ (Eliza's rate per hour for the first 40 hours she works each week is $10. She also receives an overtime pay of 1.2 times her regular hourly rate. If Eliza worked for 45 hours this week, how much are her earnings this week?)
Model Answer ✗	⋯ Elisa's total earnings for the week are $400 (from the first 40 hours) + $48 (from that overtime hour) = $448.

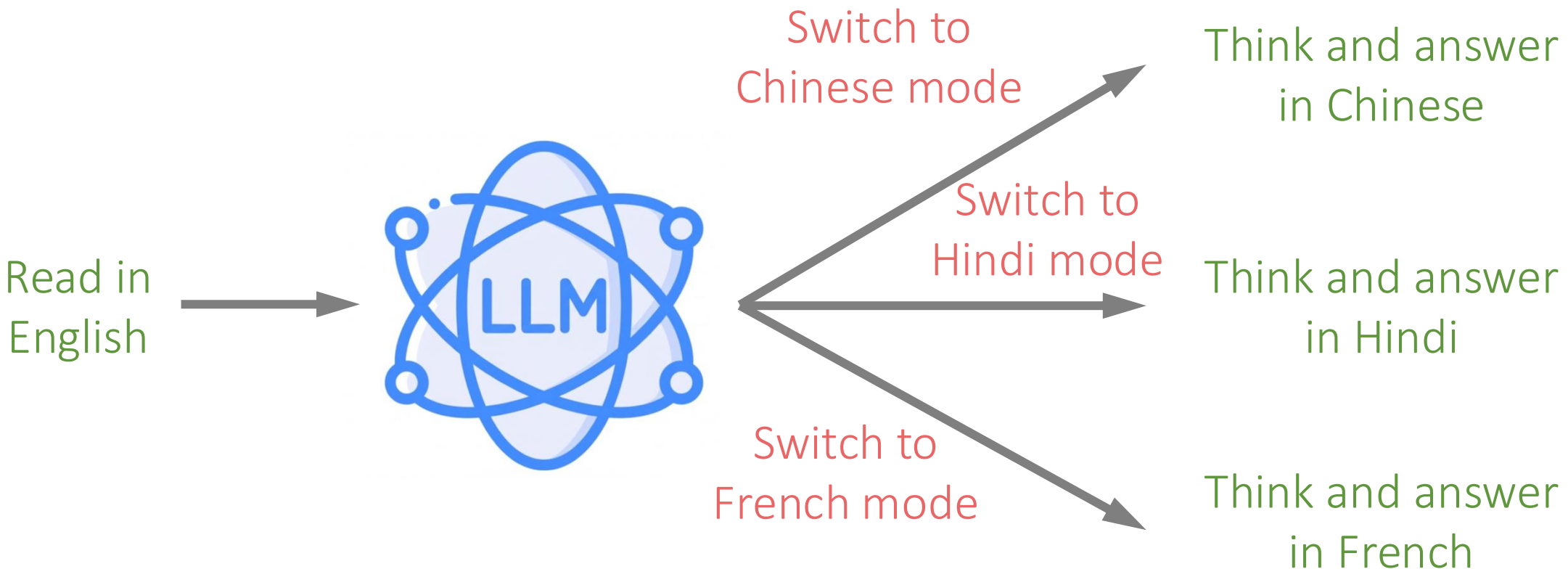
Spanish **ICL Q**	Michael tenía 58 pelotas de golf. El martes, perdió 23 pelotas de golf. El miércoles, perdió 2 más. ¿Cuántas pelotas de golf tenía al final del miércoles?
ICL A	Michael started with 58 golf balls and lost 23, so he has 58 - 23 = 35. After he lost 2 more, he has 35 - 2 = 33 balls.
German **ICL Q**	Olivia hat $23. ⋯ Wie viel Geld hat sie übrig?
ICL A	5 bagels for $3 ⋯ The answer is 8.
Chinese **ICL Q**	⋯ 杰森给了丹尼多少根棒棒糖？
ICL A	⋯ The answer is 8.
Thai **Test Q**	⋯ รายได้ของเธอในสัปดาห์นี้จะเท่ากับกี่ดอลลาร์
Model Answer ✓	Calculate the earnings for the first 40 hours at the regular rate: 40 hours * $10/hour = $400 ⋯ earnings for this week will be $460.

Cross-Lingual In-Context Learning

- Multilingual context-irrelevant sentences also help

Llama3.1-8B-Instruct			
English	57.10	44.42	55.91
English + CIS-En	55.90	47.88	55.46
English + CIS-Fr	$52.10^{*}_{3.80\downarrow}$	$52.82^{***}_{4.94\uparrow}$	$59.40^{***}_{3.94\uparrow}$
English + CIS-Ja	$58.80_{2.90\uparrow}$	$55.96^{***}_{8.08\uparrow}$	$59.86^{***}_{4.40\uparrow}$
English + CIS-Zh	$55.00_{0.90\downarrow}$	$54.68^{***}_{6.80\uparrow}$	$64.66^{***}_{9.20\uparrow}$
English + CIS-Multi	$62.50^{***}_{6.60\uparrow}$	$56.03^{***}_{8.15\uparrow}$	$64.74^{***}_{9.28\uparrow}$
Multilingual + CIS-Multi	$68.60^{***}_{6.10\uparrow}$	$57.24_{1.21\uparrow}$	$66.20^{*}_{1.46\uparrow}$
Qwen2-7B-Instruct			
English	43.70	48.46	62.29
English + CIS-En	43.00	50.90	62.31
English + CIS-Fr	$43.50_{0.50\uparrow}$	$53.65^{***}_{2.75\uparrow}$	$62.31_{0.00-}$
English + CIS-Ja	$43.80_{0.80\uparrow}$	$56.22^{**}_{5.32\uparrow}$	$63.09_{0.78\uparrow}$
English + CIS-Zh	$42.60_{0.40\downarrow}$	$56.79^{***}_{5.89\uparrow}$	$62.49_{0.18\uparrow}$
English + CIS-Multi	$42.70_{0.30\downarrow}$	$54.94^{***}_{4.04\uparrow}$	$62.51_{0.20\uparrow}$
Multilingual + CIS-Multi	$47.30^{***}_{4.70\uparrow}$	$55.83_{0.89\uparrow}$	$63.51_{1.00\uparrow}$

Can We Switch Language Mode for LLMs?

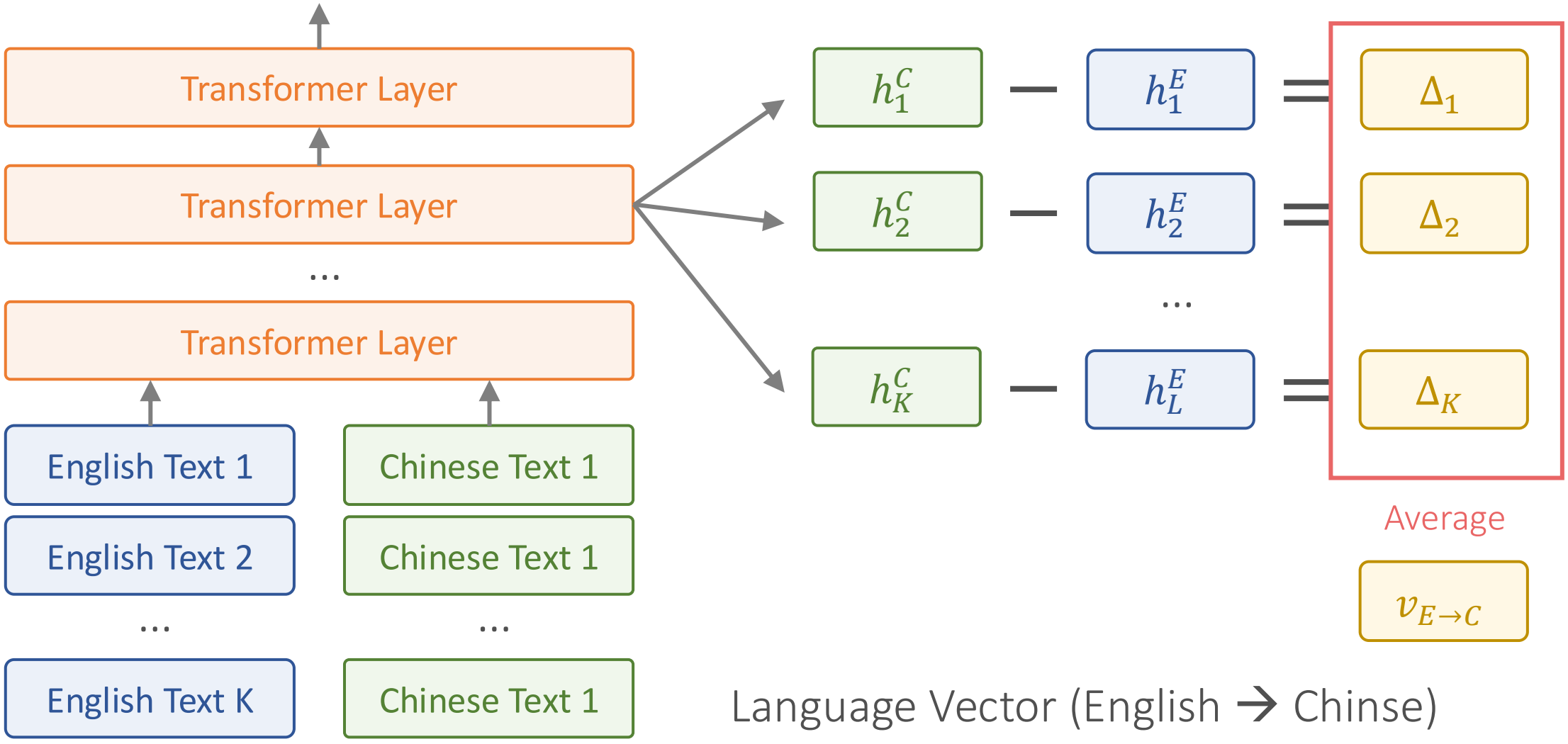


It’s possible to switch language mode for LLMs → Reasoning is language-agnostic

Multilingual In-Context Learning

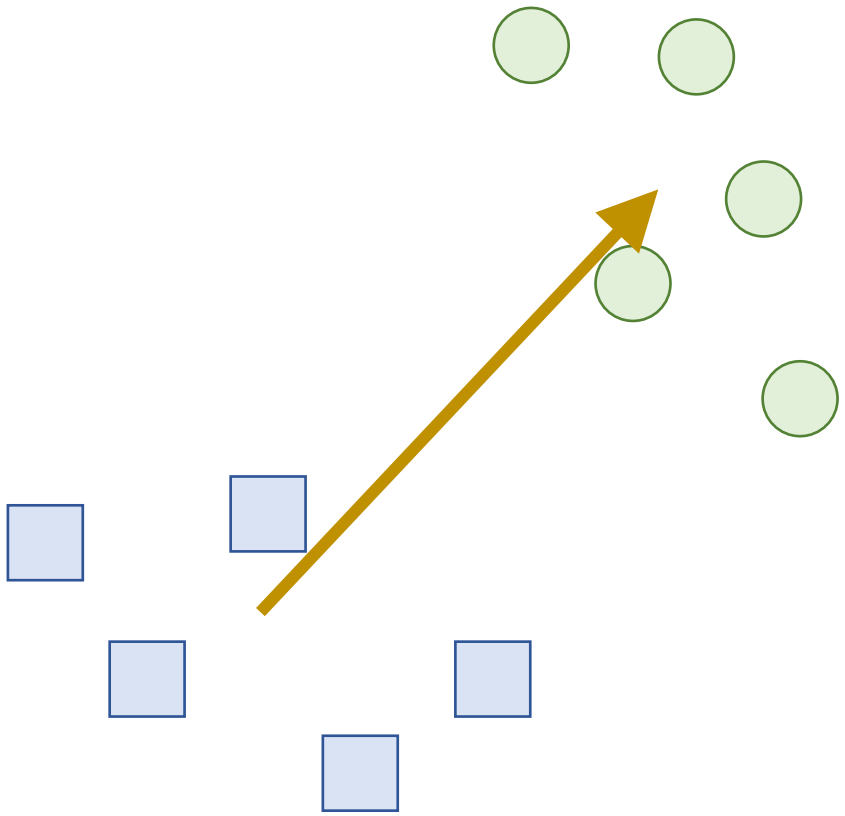
English **ICL Q**	Michael had 58 golf balls. On Tuesday, he lost 23 golf balls. On Wednesday, he lost 2 more. How many golf balls did he have at the end of Wednesday?
ICL A	Michael started with 58 golf balls and lost 23, so he has 58 - 23 = 35. After he lost 2 more, he has 35 - 2 = 33 balls.
English **ICL Q**	Olivia has $23. ⋯ How much does she have left?
ICL A	5 bagels for $3 ⋯ The answer is 8.
English **ICL Q**	⋯ How many lollipops did Jason give to Denny?
ICL A	⋯ The answer is 8.
Thai **Test Q**	⋯ รายได้ของเธอในสัปดาห์นี้จะเท่ากับกี่ดอลลาร์? (Eliza‘s rate per hour for the first 40 hours she works each week is $10. She also receives an overtime pay of 1.2 times her regular hourly rate. If Eliza worked for 45 hours this week, how much are her earnings this week?)

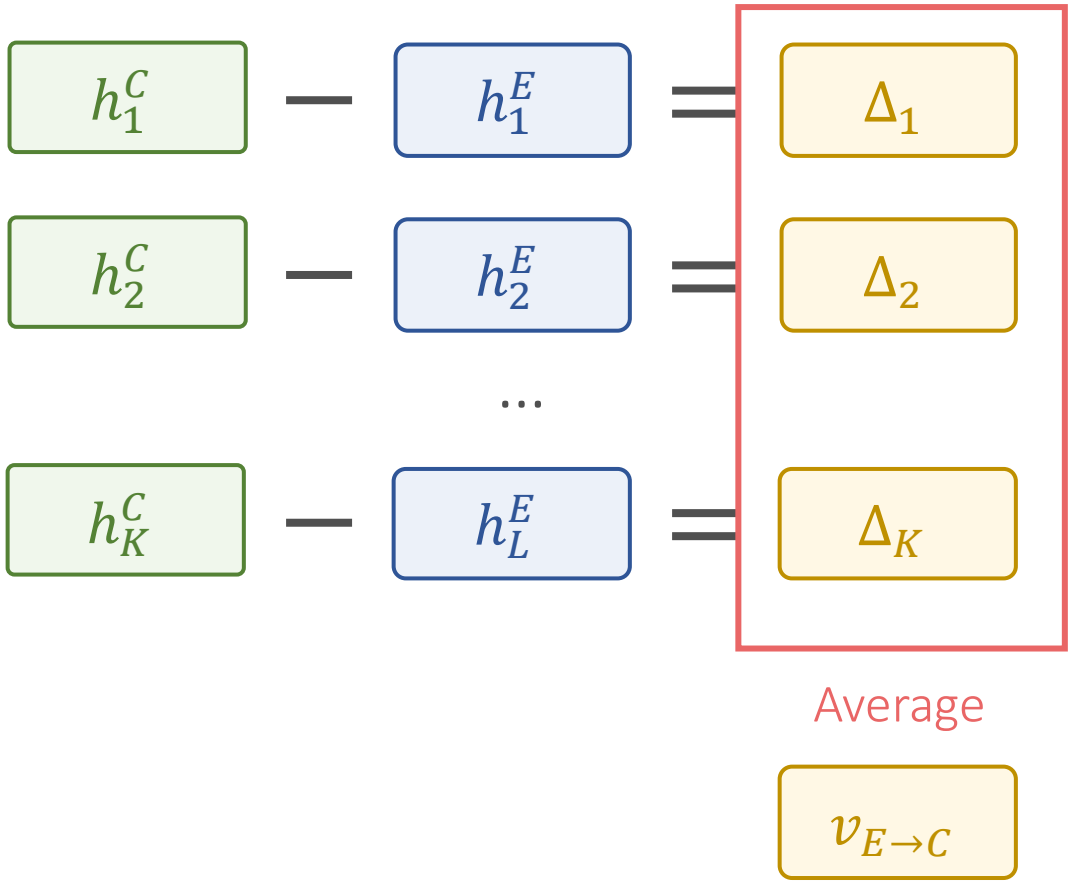
Language Steering Vectors



Language Vector (English → Chinse)

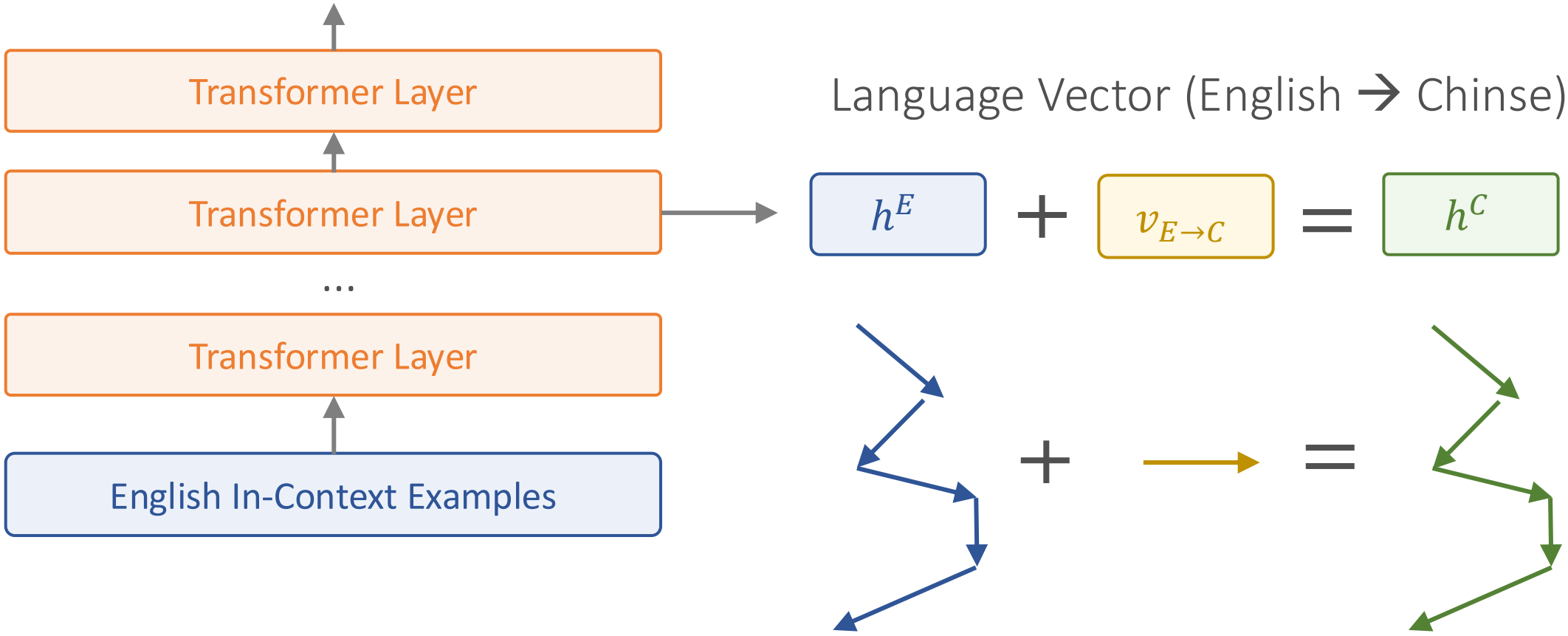
Language Steering Vectors





Language Vector (English → Chinse)

Applying Language Steering Vectors



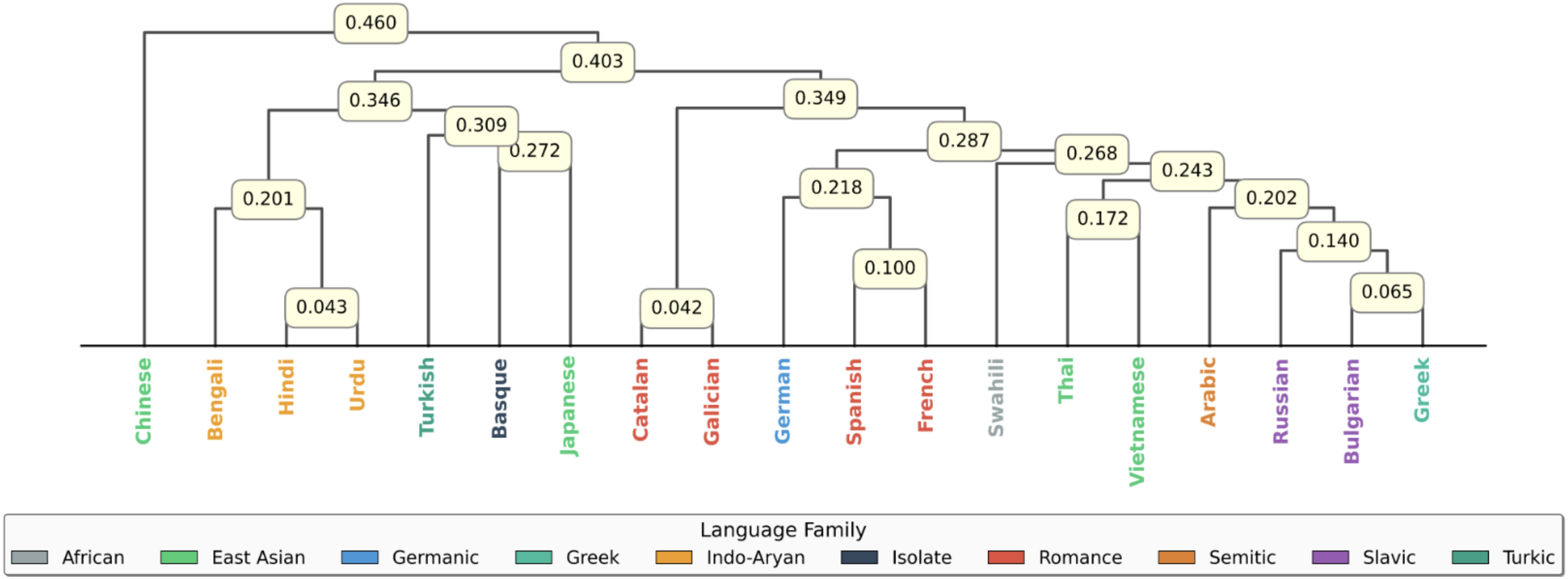
Language Vectors Improve Multilingual In-Context Learning

Language	MGSM				XNLI				MSVAMP			
	B	MFS	Ours	OR	B	MFS	Ours	OR	B	MFS	Ours	OR
Arabic	–	–	–	–	62.57	60.78	**62.87**	63.77	–	–	–	–
Basque	32.14	35.70	**36.90**	52.38	–	–	–	–	–	–	–	–
Bengali	57.14	55.95	**61.90**	58.33	–	–	–	–	57.49	**62.87**	59.58	61.38
Bulgarian	–	–	–	–	56.29	**61.98**	61.68	66.17	–	–	–	–
Catalan	64.29	64.28	**69.05**	76.19	–	–	–	–	–	–	–	–
Chinese	67.86	67.86	**71.43**	72.62	59.88	**61.38**	59.28	61.98	69.76	**73.05**	71.26	73.35
French	61.90	64.29	**70.24**	65.48	67.37	68.26	**72.75**	71.26	71.56	73.05	**74.55**	73.05
Galician	64.29	69.04	**73.81**	77.38	–	–	–	–	–	–	–	–
German	66.67	66.67	**75.00**	71.43	64.07	65.27	**66.47**	65.27	71.26	70.06	**71.26**	76.95
Greek	–	–	–	–	68.26	66.17	**70.06**	72.75	–	–	–	–
Hindi	–	–	–	–	61.38	57.78	**64.97**	61.98	–	–	–	–
Japanese	55.95	**61.90**	55.95	63.10	–	–	–	–	63.17	**68.26**	67.96	70.06
Russian	71.43	71.43	**72.62**	76.19	58.98	**64.37**	63.77	60.78	68.86	**72.46**	72.16	71.56
Spanish	77.38	70.24	**76.19**	78.57	66.77	66.17	**70.66**	67.37	74.55	73.95	**75.75**	74.55
Swahili	55.95	63.10	**65.48**	66.67	52.40	51.20	**55.99**	56.89	56.29	59.58	**60.78**	62.87
Thai	57.14	**67.86**	61.90	59.52	59.88	**64.37**	60.78	70.66	59.58	**64.67**	64.07	66.77
Turkish	–	–	–	–	62.28	57.19	**65.87**	65.57	–	–	–	–
Urdu	–	–	–	–	55.09	55.39	**56.29**	57.19	–	–	–	–
Vietnamese	–	–	–	–	64.07	63.17	**68.86**	68.86	–	–	–	–
Average	61.01	63.19	**65.87**	68.16	61.38	61.68	**64.31**	65.04	65.84	**68.66**	68.60	70.06

Language Vectors are Task-Agnostic

Language	MGSM (vector) → XNLI (eval)					XNLI (vector) → MGSM (eval)				
	B	MFS	Ours	CT	OR	B	MFS	Ours	CT	OR
Chinese	78.14	77.84	77.84	**78.14**	79.34	84.52	88.10	86.90	**90.48**	79.76
Thai	70.96	**73.05**	70.06	70.06	72.46	79.76	**82.14**	79.76	**82.14**	85.71
Swahili	53.29	47.31	**55.09**	54.79	54.79	19.05	16.67	**26.19**	25.00	33.33
Russian	78.14	79.04	**79.64**	**79.94**	78.74	85.71	85.71	**88.10**	**88.10**	85.71
French	81.44	82.34	82.34	**82.63**	82.63	78.57	80.95	**79.76**	**79.76**	84.52
Spanish	82.63	82.63	**83.53**	83.23	81.74	84.52	83.33	**89.29**	88.10	86.90
German	81.74	77.54	82.93	**82.98**	79.34	83.33	83.33	**85.71**	**85.71**	86.90
Average	75.19	74.22	75.92	**75.96**	75.58	73.64	74.32	**76.53**	76.36	77.55

Language Vectors Capture Linguistic Features



How Do LLMs Understand Multilingual Inputs?

- Still an open research problem

Do Multilingual LLMs Think In English?

Lisa Schut [1] **Yarin Gal** [1] **Sebastian Farquhar** [2]

Code-Switching Inputs

- Another open research problem!

这个 weekend 我要去 shopping。

Party बहुत amazing थी, food भी tasty था

Voy a comprar un laptop porque mi old one ya no funciona

أنا عندي meeting بعدين نروح، مهم بعد ساعة cafe