Полиязычные модели

Булатова Екатерина

План презентации

- Полиязычные модели
- Модели:
    - mBERT
    - XLM
    - XLM-R
    - MMTE
- Сравнение точности моделей: XTREME
- Анализ моделей

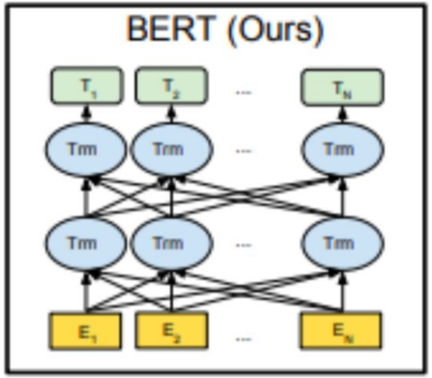
Полиязычные модели

- Проблема: доминирование английского языка
- Решение:
  - Переобучение моделей для новых языков
  - Создание полиязычных моделей
- Второй вариант дает точность выше

mBERT

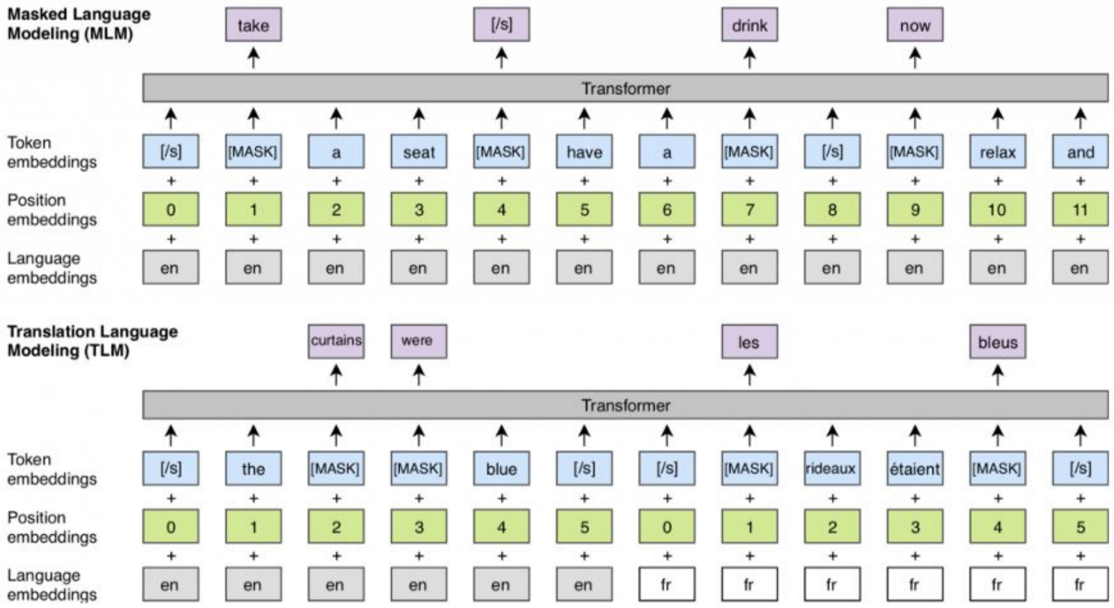
- == BERT, предобученный на Википедии
  - Unsupervised
  - Трансформеры
  - (M)MLM
  - NSP
  - Fine-tuning
- Не ищет специально зависимости

System	English	Chinese	Spanish	German	Arabic	Urdu
XNLI Baseline - Translate Train	73.7	67.0	68.8	66.5	65.8	56.6
XNLI Baseline - Translate Test	73.7	68.3	70.7	68.7	66.8	59.3
BERT - Translate Train Cased	**81.9**	**76.6**	**77.8**	**75.9**	**70.7**	61.6
BERT - Translate Train Uncased	81.4	74.2	77.3	75.2	70.5	61.7
BERT - Translate Test Uncased	81.4	70.1	74.9	74.4	70.4	**62.1**
BERT - Zero Shot Uncased	81.4	63.8	74.3	70.5	62.1	58.3



XLM

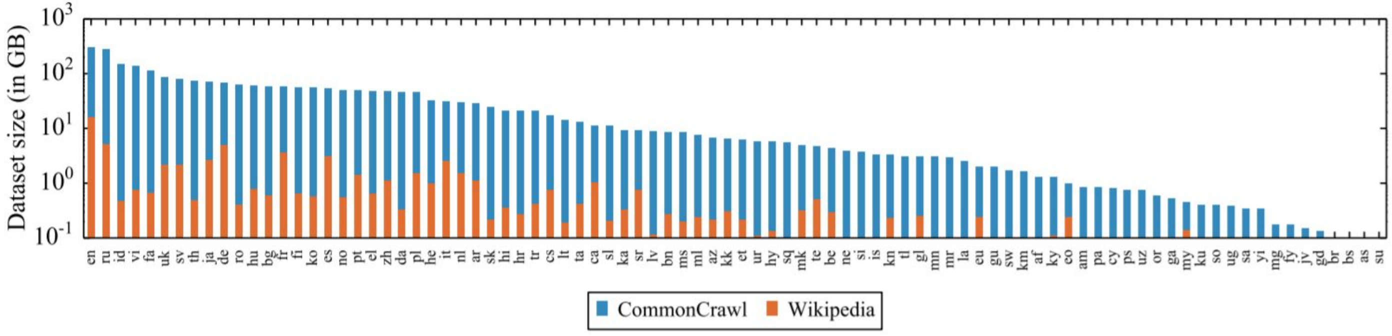
- Новый способ обучения BERT для классификации
- Особенности:
  - BPE
  - Трейн -- пары из одинакового текста на двух языках
  - ID языка + PE
  - Embeddings
- **TLM**

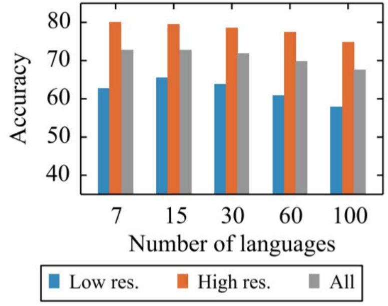


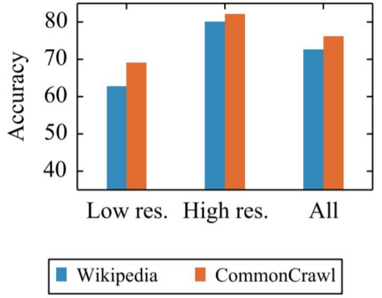
	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	Δ
Machine translation baselines (TRANSLATE-TRAIN)																
Devlin et al. (2018)	81.9	-	77.8	75.9	-	-	-	-	70.7	-	-	76.6	-	-	61.6	-
XLM (MLM+TLM)	85.0	80.2	80.8	80.3	78.1	79.3	78.1	74.7	76.5	76.6	75.5	78.6	72.3	70.9	63.2	76.7
Machine translation baselines (TRANSLATE-TEST)																
Devlin et al. (2018)	81.4	-	74.9	74.4	-	-	-	-	70.4	-	-	70.1	-	-	62.1	-
XLM (MLM+TLM)	85.0	79.0	79.5	78.1	77.8	77.6	75.5	73.7	73.7	70.8	70.4	73.6	69.0	64.7	65.1	74.2
Evaluation of cross-lingual sentence encoders																
Conneau et al. (2018b)	73.7	67.7	68.7	67.7	68.9	67.9	65.4	64.2	64.8	66.4	64.1	65.8	64.1	55.7	58.4	65.6
Devlin et al. (2018)	81.4	-	74.3	70.5	-	-	-	-	62.1	-	-	63.8	-	-	58.3	-
Artetxe and Schwenk (2018)	73.9	71.9	72.9	72.6	73.1	74.2	71.5	69.7	71.4	72.0	69.2	71.4	65.5	62.2	61.0	70.2
XLM (MLM)	83.2	76.5	76.3	74.2	73.1	74.0	73.1	67.8	68.5	71.2	69.2	71.9	65.7	64.6	63.4	71.5
XLM (MLM+TLM)	**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**

XLM-R

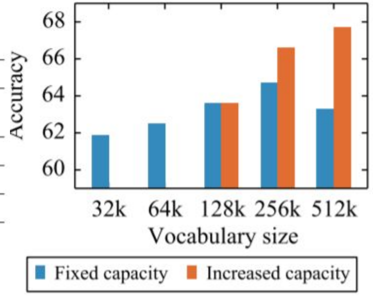
- “Проклятие полиязычности” (Curse of Multilinguality)
- Common Crawl

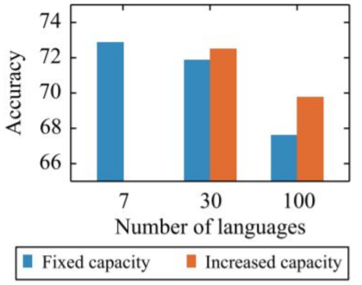






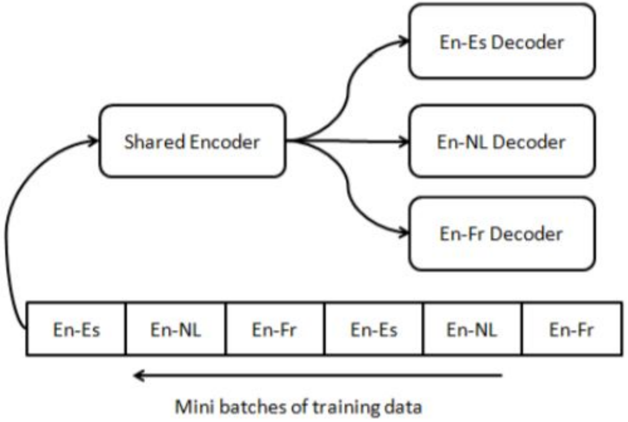
Model	D	#M	#lg	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	Avg
Fine-tune multilingual model on English training set (Cross-lingual Transfer)																			
Lample and Conneau (2019)	Wiki+MT	N	15	85.0	78.7	78.9	77.8	76.6	77.4	75.3	72.5	73.1	76.1	73.2	76.5	69.6	68.4	67.3	75.1
Huang et al. (2019)	Wiki+MT	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$_{Base}$	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
Fine-tune multilingual model on all training sets (TRANSLATE-TRAIN-ALL)																			
Lample and Conneau (2019)[†]	Wiki+MT	1	15	85.0	80.8	81.3	80.3	79.1	80.9	78.3	75.6	77.6	78.5	76.0	79.5	72.9	72.8	68.5	77.8
Huang et al. (2019)	Wiki+MT	1	15	85.6	81.1	82.3	80.9	79.5	81.4	79.7	76.8	78.2	77.9	77.1	80.5	73.4	73.8	69.6	78.5
Lample and Conneau (2019)	Wiki	1	100	84.5	80.1	81.3	79.3	78.6	79.4	77.5	75.2	75.6	78.3	75.7	78.3	72.1	69.2	67.7	76.9
XLM-R$_{Base}$	CC	1	100	85.4	81.4	82.2	80.3	80.4	81.3	79.7	78.6	77.3	79.7	77.9	80.2	76.1	73.1	73.0	79.1
XLM-R	CC	1	100	**89.1**	**85.1**	**86.6**	**85.7**	**85.3**	**85.9**	**83.5**	**83.2**	**83.1**	**83.7**	**81.5**	**83.7**	**81.6**	**78.0**	**78.1**	**83.6**

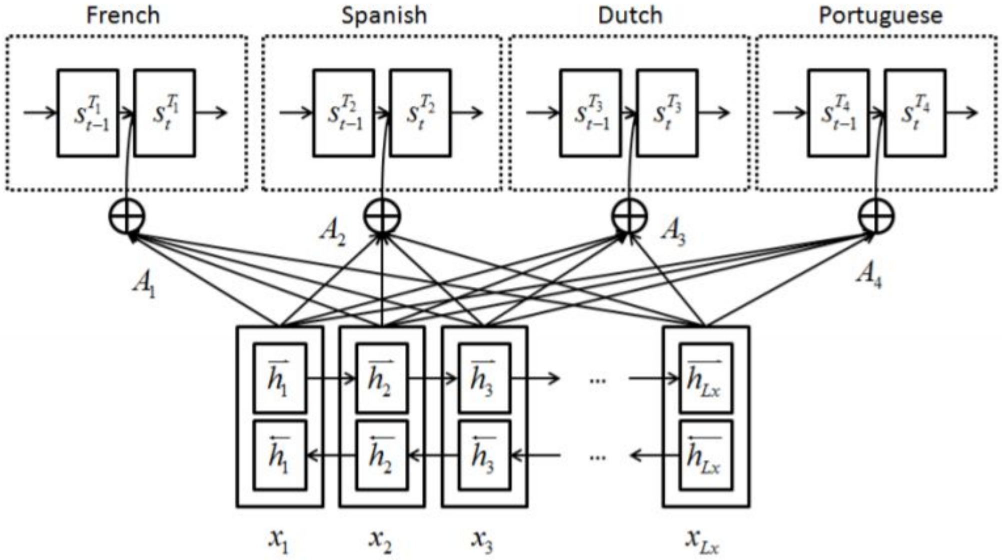




Massively Multilingual Neural Machine Translation

- RNN encoder-decoder с несколькими задачами
- Один и тот же encoder
- Центрирована к английскому





XTREME

- 9 межъязыковых бенчмарок для 40 языков

Rank	Model	Participant	Affiliation	Attempt Date	Avg	Sentence-pair Classification	Structured Prediction	Question Answering	Sentence Retrieval
0		Human	-	-	93.3	95.1	97.0	87.8	-
1	Polyglot	MLNLC	ByteDance	Feb 25, 2021	81.3	88.1	80.6	71.8	89.4
2	VECO	DAMO NLP Team	Alibaba	Feb 11, 2021	81.1	88.6	75.4	72.4	92.1
3	ERNIE-M	ERNIE Team	Baidu	Jan 1, 2021	80.9	87.9	75.6	72.3	91.9
4	T-ULRv2 + StableTune	Turing	Microsoft	Oct 7, 2020	80.7	88.8	75.4	72.9	89.3
5	Anonymous3	Anonymous3	Anonymous3	Jan 3, 2021	79.9	88.2	74.6	71.7	89.0
6	FILTER	Dynamics 365 AI Research	Microsoft	Sep 8, 2020	77.0	87.5	71.9	68.5	84.4
7	X-STILTs	Phang et al.	New York University	Jun 17, 2020	73.5	83.9	69.4	67.2	76.5
8	XLM-R (large)	XTREME Team	Alphabet, CMU	-	68.2	82.8	69.0	62.3	61.6
9	Creative	Creative	Microsoft	Feb 3, 2021	70.0	65.0	88.8	53.3	81.3
10	mBERT	XTREME Team	Alphabet, CMU	-	59.6	73.7	66.3	53.8	47.7

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

Model	Avg	Pair sentence		Structured prediction		Question answering			Sentence retrieval	
		XNLI	PAWS-X	POS	NER	XQuAD	MLQA	TyDiQA-GoldP	BUCC	Tatoeba
Metrics		Acc.	Acc.	F1	F1	F1 / EM	F1 / EM	F1 / EM	F1	Acc.
Cross-lingual zero-shot transfer (models are trained on English data)										
mBERT	59.8	65.4	81.9	71.5	62.2	64.5 / 49.4	61.4 / 44.2	59.7 / 43.9	56.7	38.7
XLM	55.7	69.1	80.9	71.3	61.2	59.8 / 44.3	48.5 / 32.6	43.6 / 29.1	56.8	32.6
XLM-R Large	68.2	79.2	86.4	73.8	65.4	76.6 / 60.8	71.6 / 53.2	65.1 / 45.0	66.0	57.3
MMTE	59.5	67.4	81.3	73.5	58.3	64.4 / 46.2	60.3 / 41.4	58.1 / 43.8	59.8	37.9
Translate-train (models are trained on English training data translated to the target language)										
mBERT	-	74.6	86.3	-	-	70.0 / 56.0	65.6 / 48.0	55.1 / 42.1	-	-
mBERT, multi-task	-	75.1	88.9	-	-	72.4 / 58.3	67.6 / 49.8	64.2 / 49.3	-	-
Translate-test (models are trained on English data and evaluated on target language data translated to English)										
BERT-large	-	76.8	84.4	-	-	76.3 / 62.1	72.9 / 55.3	72.1 / 56.0	-	-
In-language models (models are trained on the target language training data)										
mBERT, 1000 examples	-	-	-	87.6	77.9	-	-	58.7 / 46.5	-	-
mBERT	-	-	-	89.8	88.3	-	-	74.5 / 62.7	-	-
mBERT, multi-task	-	-	-	91.5	89.1	-	-	77.6 / 68.0	-	-
Human	-	92.8	97.5	97.0	-	91.2 / 82.3	91.2 / 82.3	90.1 / -	-	-

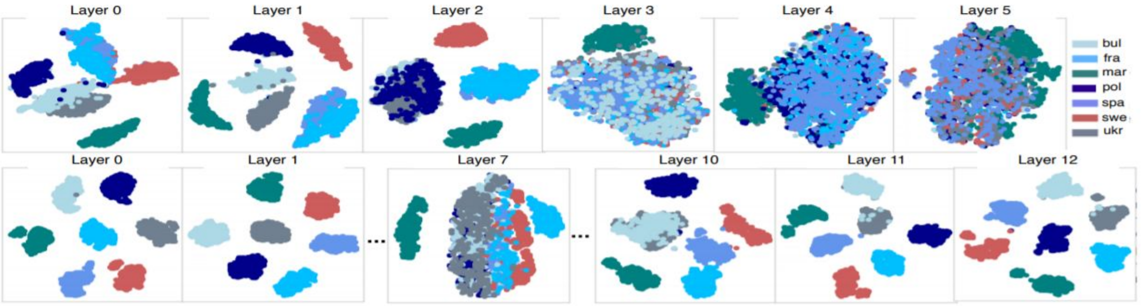
Анализ моделей

- mBERT разбивает языки на семьи
- Представления с верхнего слоя

XLM
vs.
mBERT & XLM-R

Code	Type	Feature name
37A	Nom	Definite articles
38A	Nom	Indefinite articles
45A	Nom	Politeness distinctions in pronouns
47A	Nom	Intensifiers and reflexive pronouns
51A	Nom	Position of case affixes
70A	Verb	The morphological imperative
71A	Verb	The prohibitive
72A	Verb	Imperative-hortative systems
79A	Verb	Suppletion according to tense and aspect
79B	Verb	Suppletion in impertatives and hortatives
81A	WO	Order of Subject, Object and Verb (SOV)
82A	WO	Order of Subject and Verb (SV)
83A	WO	Order of Object and Verb (OV)
85A	WO	Order of adposition and noun phrase
86A	WO	Order of genitive and noun
87A	WO	Order of adjective and noun
92A	WO	Position of polar question particles
93A	WO	Position of interrogative phrases in content questions
95A	WO	Relationship between OV and adposition and noun phrase order
97A	WO	Relationship between OV and adjective and noun order
115A	SC	Negative indefinite pronouns and predicate negation
116A	SC	Polar questions
143F	WO	Postverbal negative morphemes
144D	WO	Position of negative morphemes
144J	WO	Subject verb negative word object order

Code	Type	LASER	M-BERT	XLM	XLM-R	Baseline
37A	Nom	0.864	0.957	0.83	**0.997**	0.199
38A*	Nom	0.571	**0.597**	0.595	0.579	0.334
45A†	Nom	0.997	**1.0**	0.989	**1.0**	0.428
47A†	Nom	0.97	0.995	0.934	**0.999**	0.333
51A‡	Nom	0.682	**0.763**	0.752	0.762	0.375
70A	Verb	0.64	0.69	0.603	**0.695**	0.243
71A	Verb	0.347	0.522	0.452	**0.576**	0.243
72A	Verb	0.422	0.763	0.557	**0.769**	0.417
79A§	Verb	0.456	0.94	0.646	**0.978**	0.4
79B§	Verb	0.212	0.528	0.382	**0.544**	0.25
81A	WO	0.993	**1.0**	0.959	0.998	0.462
82A	WO	0.429	0.352	**0.449**	0.368	0.363
83A	WO	0.993	**1.0**	0.939	0.999	0.462
85A	WO	0.993	**1.0**	0.873	0.995	0.462
86A†	WO	0.763	0.811	0.757	**0.82**	0.166
87A	WO	0.976	**0.999**	0.944	0.998	0.416
92A|	WO	0.212	0.16	0.231	0.206	**0.285**
93A¶	WO	0.647	0.65	0.627	**0.665**	0.25
95A	WO	0.993	**1.0**	0.96	0.999	0.462
97A	WO	0.983	0.996	0.941	**0.998**	0.243
115A#	SC	0.998	**1.0**	0.984	0.999	0.4
116A◊	SC	0.584	0.622	0.602	**0.634**	0.4
143F	WO	0.608	0.644	0.599	**0.65**	0.364
144D↓	WO	0.978	0.998	0.979	**1.0**	0.429
144J$^{\delta}$	WO	0.983	0.996	0.954	**0.999**	0.445



Анализ моделей

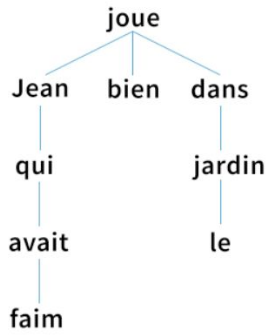
Векторное представление

The	chef	who	ran	to	the	store	was	out	of	food
.4	.1	.3	.7	.4	.1	.3	.1	.3	-.8	0
-.2	.9	-.4	-.4	0	-.6	.1	.9	.1	.3	.7
.3	-.2	.2	0	-.5	.2	-.6	-.8	.8	-.6	-.9

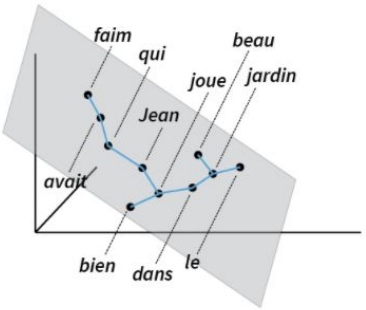
Structural probe method

Jean qui avait faim joue bien dans le jardin
(Jean, who was hungry, plays in the garden)

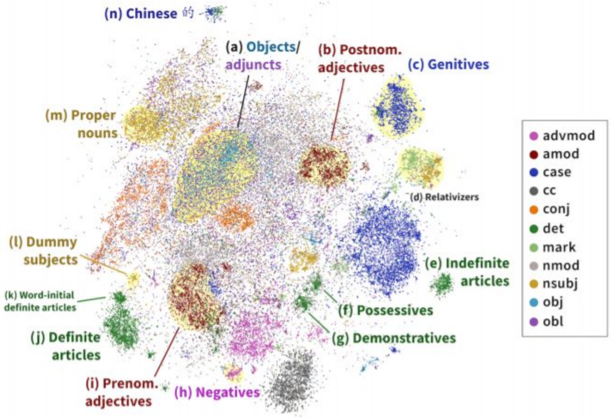
Синтаксическое дерево



Линейное преобразование

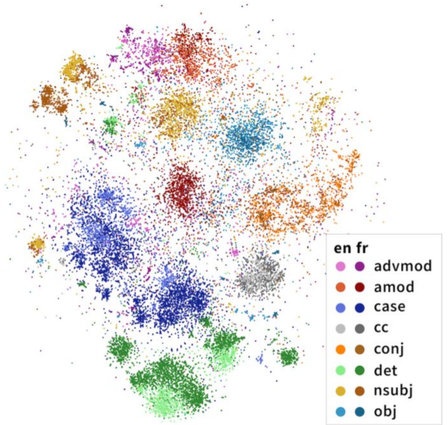


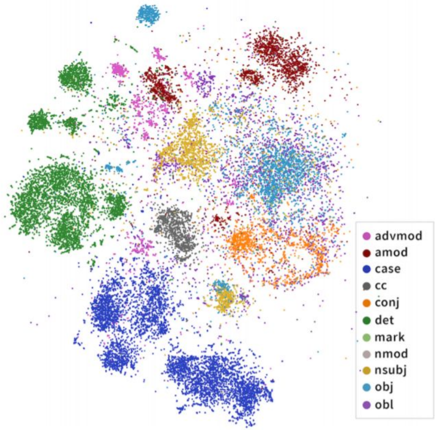
Анализ mBERT



Example sentences (trimmed for clarity). Heads in **bold**; dependents in ***bold italic***.

(b) Postnominal adjectives	fr	Le gaz développe ses **applications *domestiques***.
	id	**Film *lain*** yang menerima penghargaan istimewa.
	fa	وی تصمیمات ـ اویک در تنظیم قیمت نفت خام و ـ را ـ مؤثر ـ دانست
(c) Genitives	en	The assortment ***of*** **customers** adds entertainment.
	es	Con la recuperación ***de*** la **democracia** y las libertades
	lv	Svešiniece piecēlās, atvadījās ***no*** vecā **vīra**
(j) Definite articles	en	The value of ***the*** highest **bid**
	fr	Merak est une ville d'Indonésie sur ***la*** côte **occidentale**.
	de	Selbst mitten in ***der*** **Woche** war das Lokal gut besucht.





Ссылки

- https://arxiv.org/abs/1809.05053
- https://arxiv.org/abs/1911.02116
- https://arxiv.org/abs/2003.11080
- https://peltarion.com/blog/data-science/a-deep-dive-into-multilingual-nlp-models
- https://www.aclweb.org/anthology/N19-1388/
- https://github.com/google-research/bert/blob/master/multilingual.md
- https://huggingface.co/bert-base-multilingual-cased
- https://arxiv.org/pdf/1901.07291.pdf
- https://github.com/facebookresearch/XLM
- https://www.aclweb.org/anthology/P15-1166.pdf
- https://arxiv.org/pdf/2005.04511.pdf
- https://arxiv.org/pdf/2009.12862.pdf