

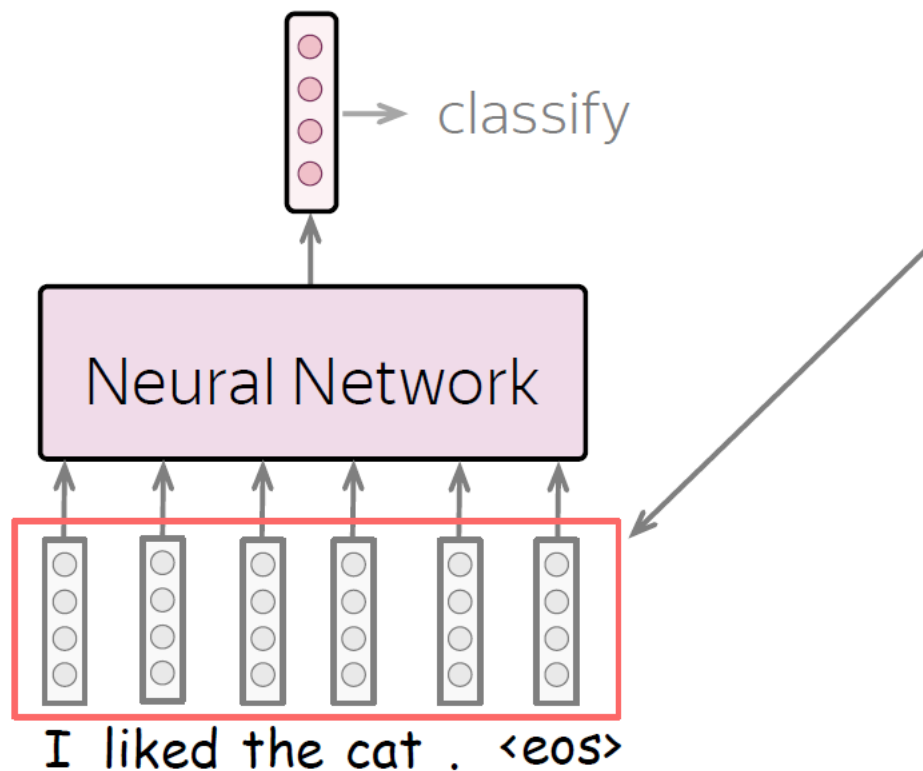
Лекция 5.

Transfer Learning (обучение с переносом, трансферное обучение)

Transfer Learning

- Идея трансферного обучения
- Две важные идеи
- ELMO
- BERT

Обзор классификации текстов: word embeddings



Входные векторные представления слов:

- Обучение с нуля
- Использование предварительно обученных данных (Word2Vec, GloVe)
- Инициализация с использованием предварительно обученных данных, затем тонкая настройка

Обзор классификации текстов: word embeddings

Обучающие данные для классификации текста (с метками):

- Небольшой объём, неразнообразные
- Тема специфичная для конкретной задачи

Обучающие данные для векторных представлений слов (без меток)

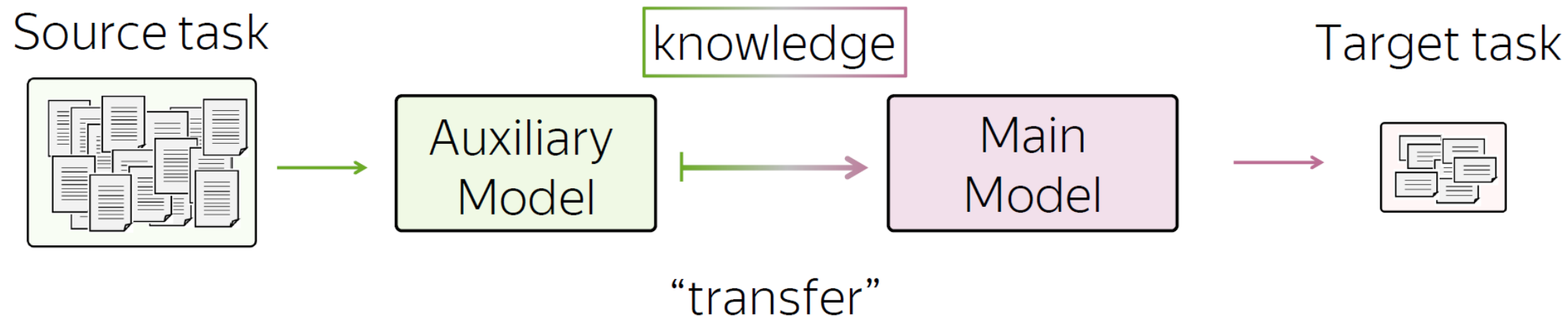
- Огромный и разнообразный корпус данных (например, Википедия)
- Тема: общая

Обзор классификации текстов: word embeddings

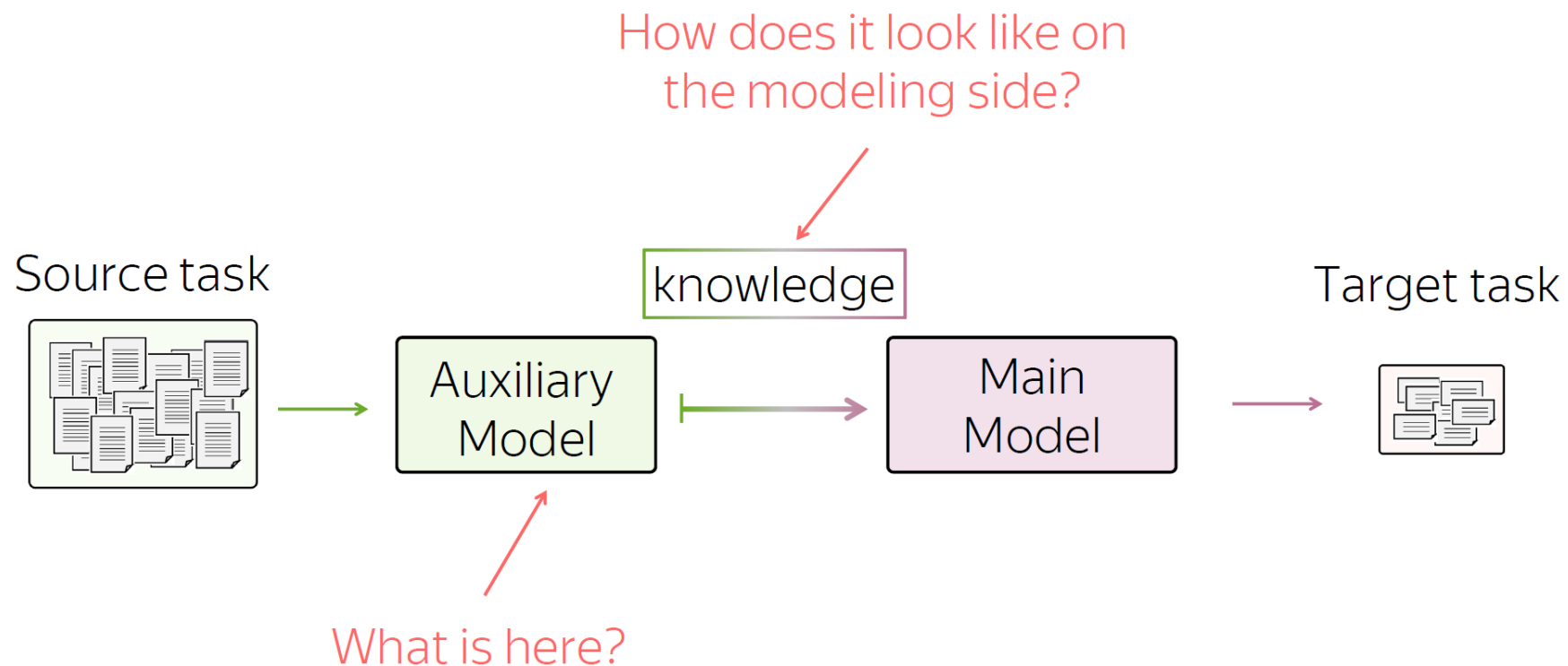
- Обучение с нуля: может быть недостаточно, чтобы изучить взаимосвязи между словами.
- Предобученные модели (Word2Vec, GloVe): известны взаимосвязи между словами, но неизвестна специфика задачи.
- Инициализация с помощью предобученных моделей данных, затем тонкая настройка: известны взаимосвязи между словами и известна специфика задачи.

Происходит «перенос» знаний из огромного немаркированного корпуса в вашу специфическую для задачи модель.

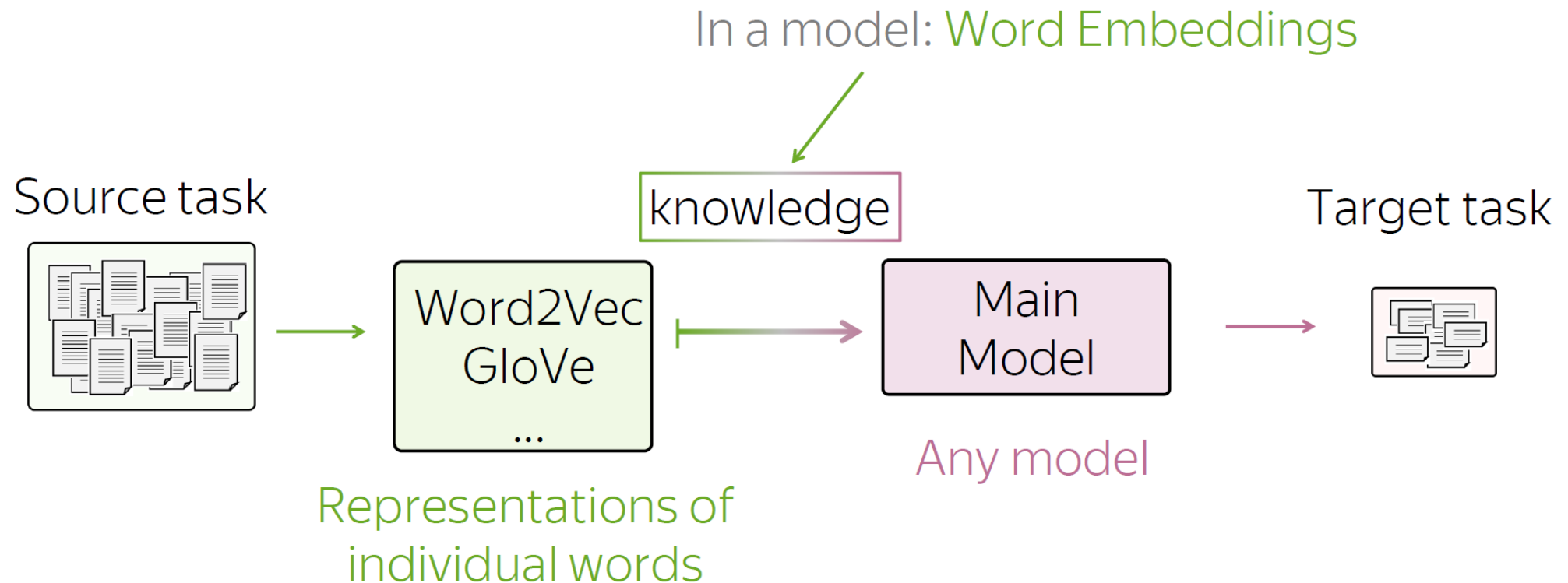
Идея трансферного обучения



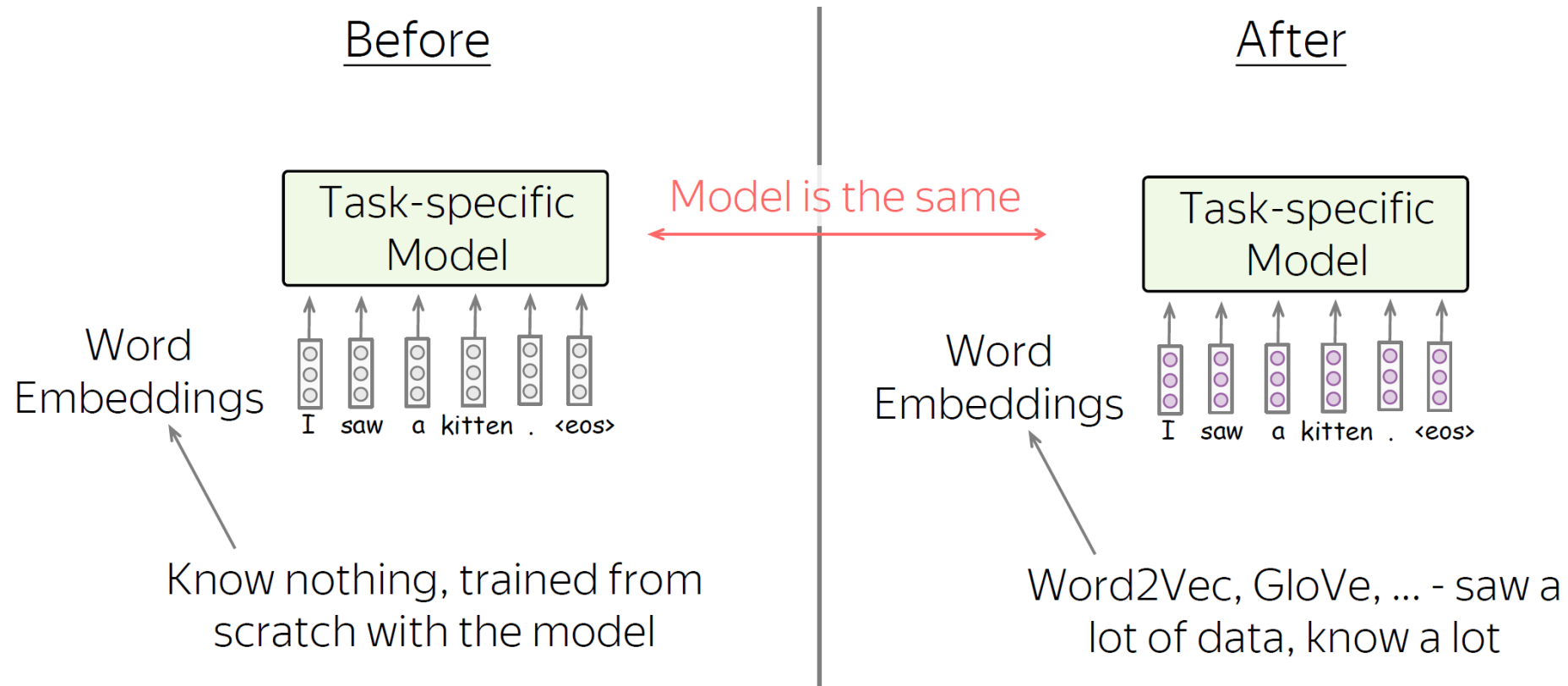
Обзор классификации текстов: word embeddings



Простейший вариант: word embeddings (Word2Vec, GloVe)

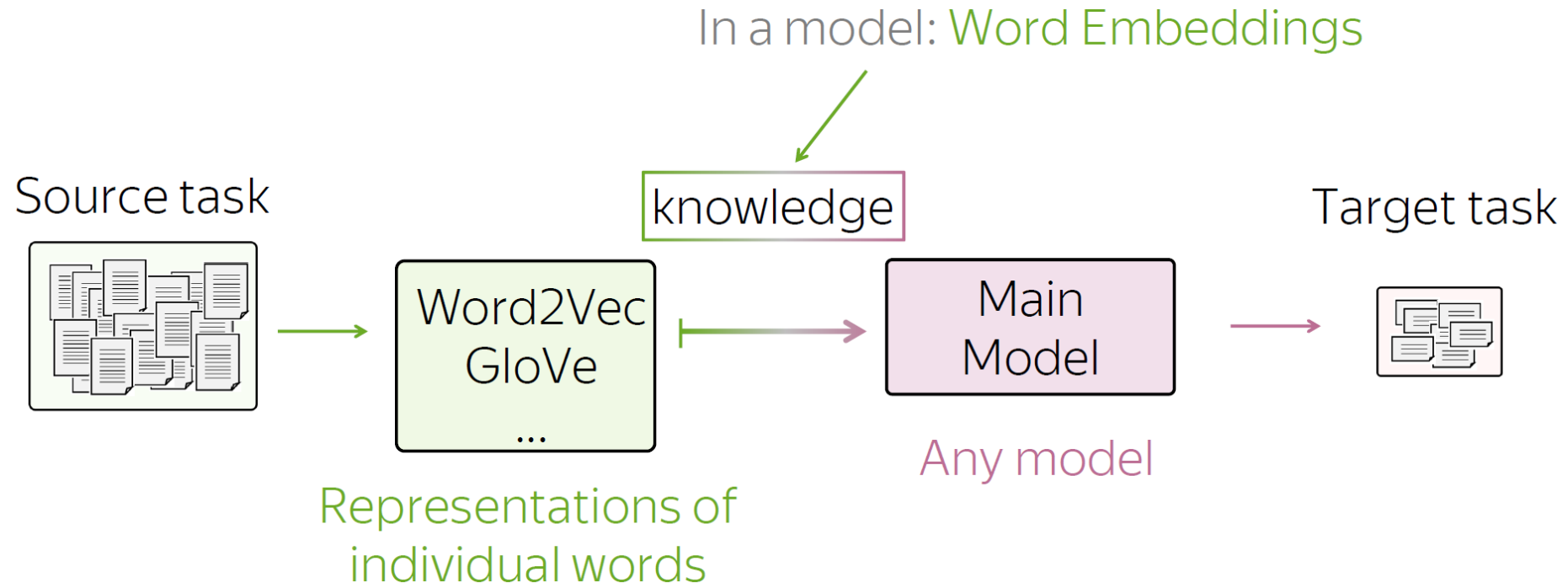


Transfer Through Word Embedding



Простейший вариант: word embeddings (Word2Vec, GloVe)

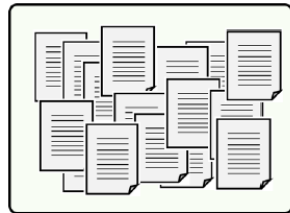
Идея: Благодаря **word embeddings** мы «переносим» знания из их обучающих данных в нашу специфическую для задачи модель.



Простейший вариант: word embeddings (Word2Vec, GloVe)

Идея: С помощью **Model** мы «переносим» знания их обучающих данных в нашу специфическую для задачи модель.

Source task



Model

knowledge



Main
Model

Any model



Target task



Две важные идеи

GloVe, Word2Vec → CoVe, ELMo → BERT, GPT

Great idea 1

What is encoded

Words → Words in context

Words in context

Great idea 2

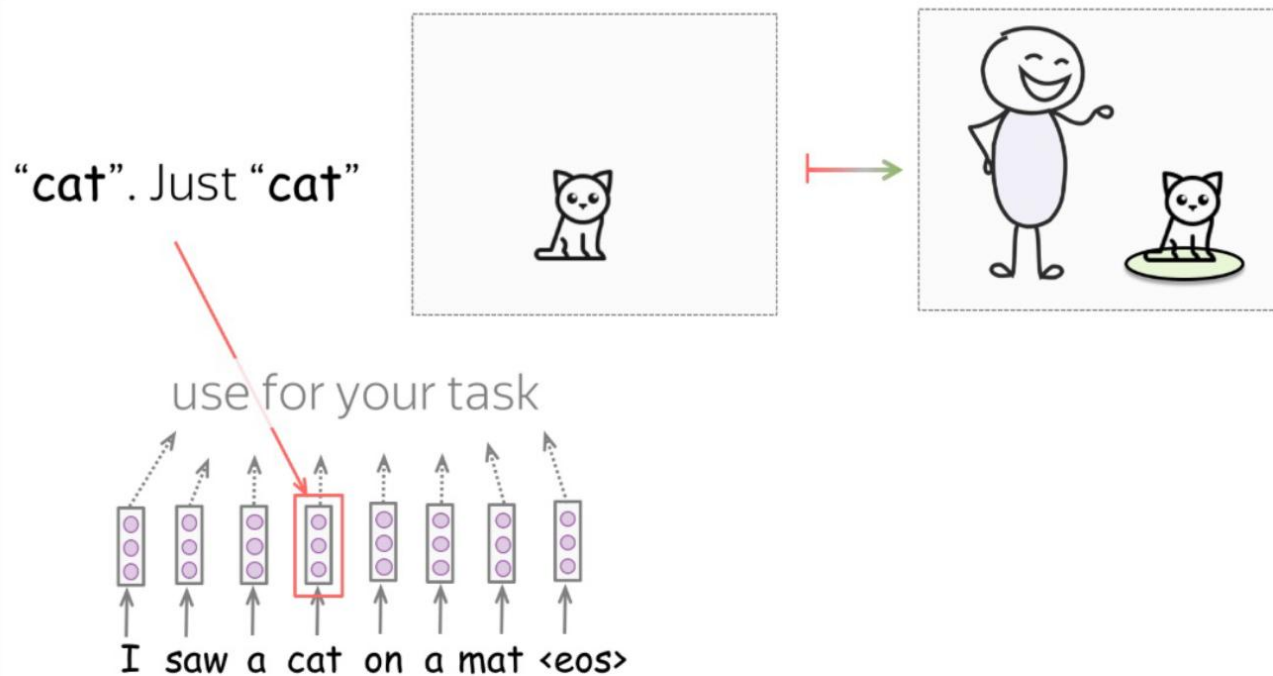
How it is used for downstream tasks

Input for task-specific models

Input for task-specific models → Instead of task-specific models

Идея 1: от слов к словам в контексте

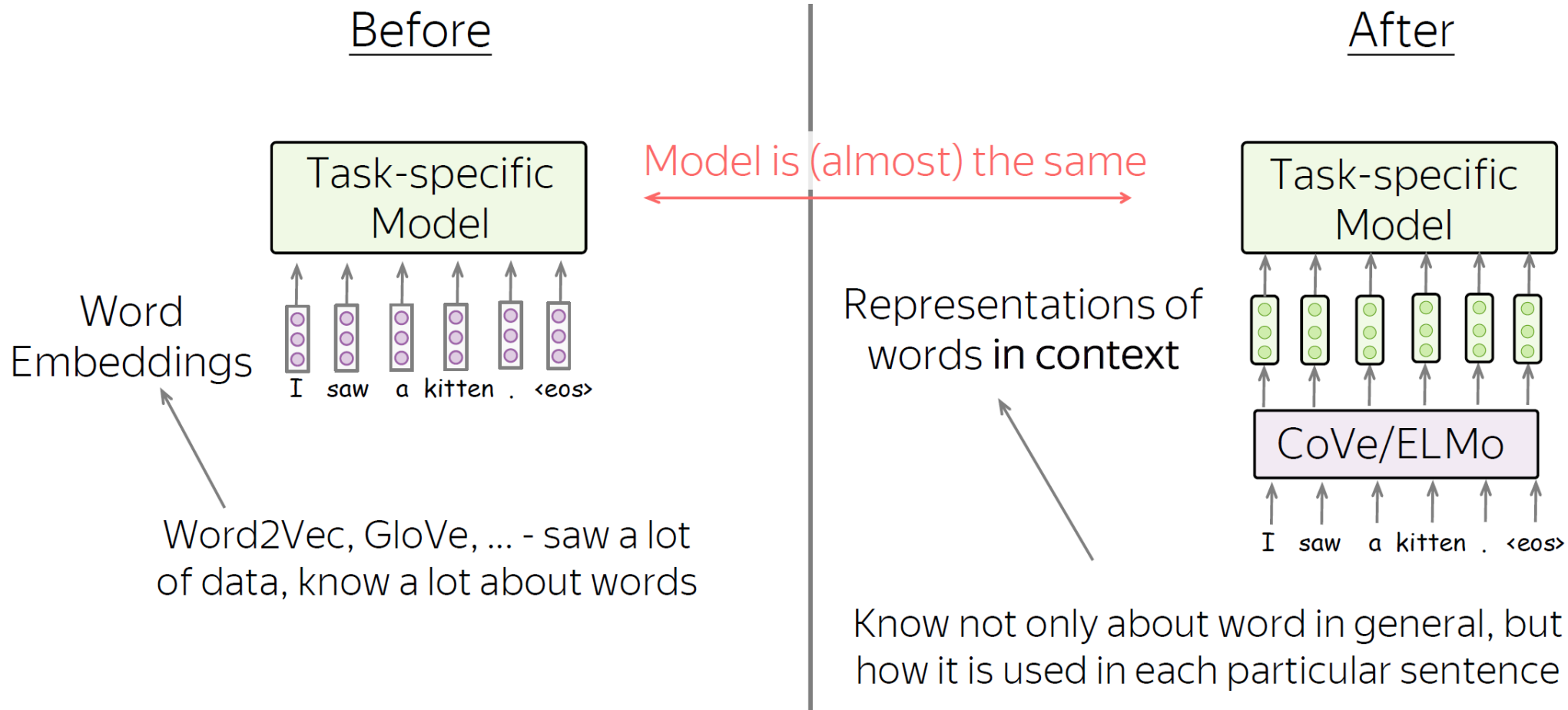
Before



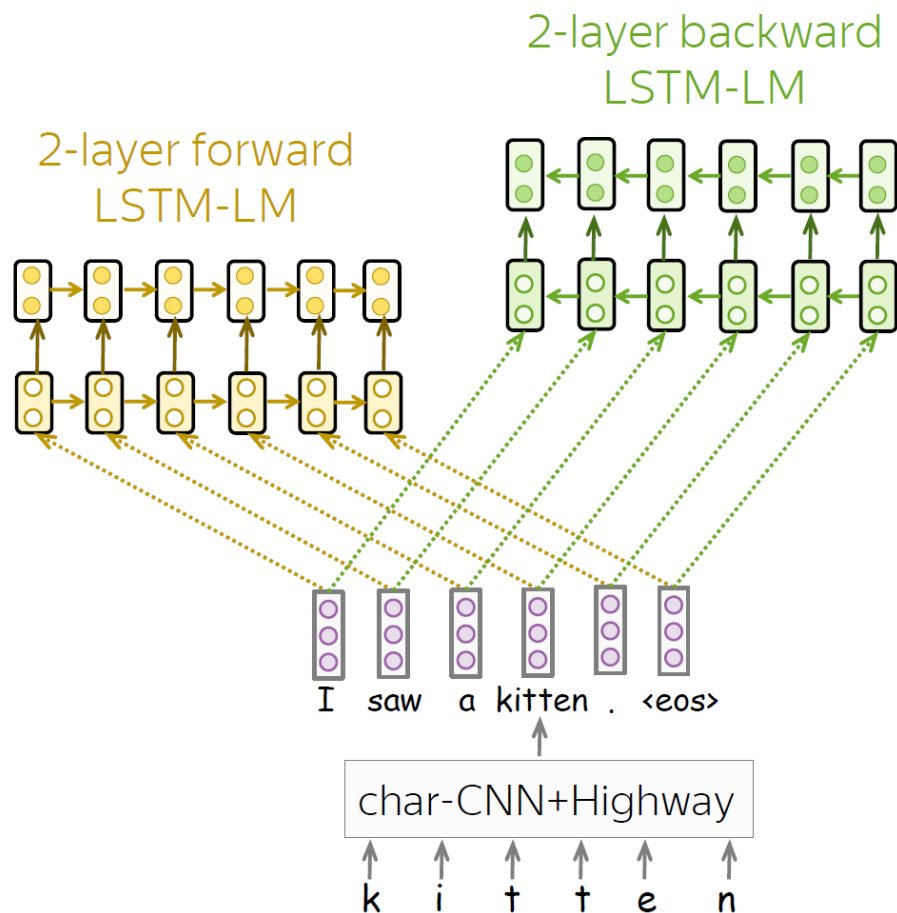
After

Not just any “cat”!
This is the **cat** that **I** saw, the
one who **sat on the mat**

Идея 1: от слов к словам в контексте

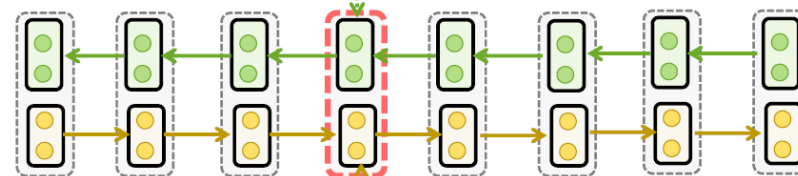


ELMo: от слов к словам в контексте



I saw a cat on a mat <eos>

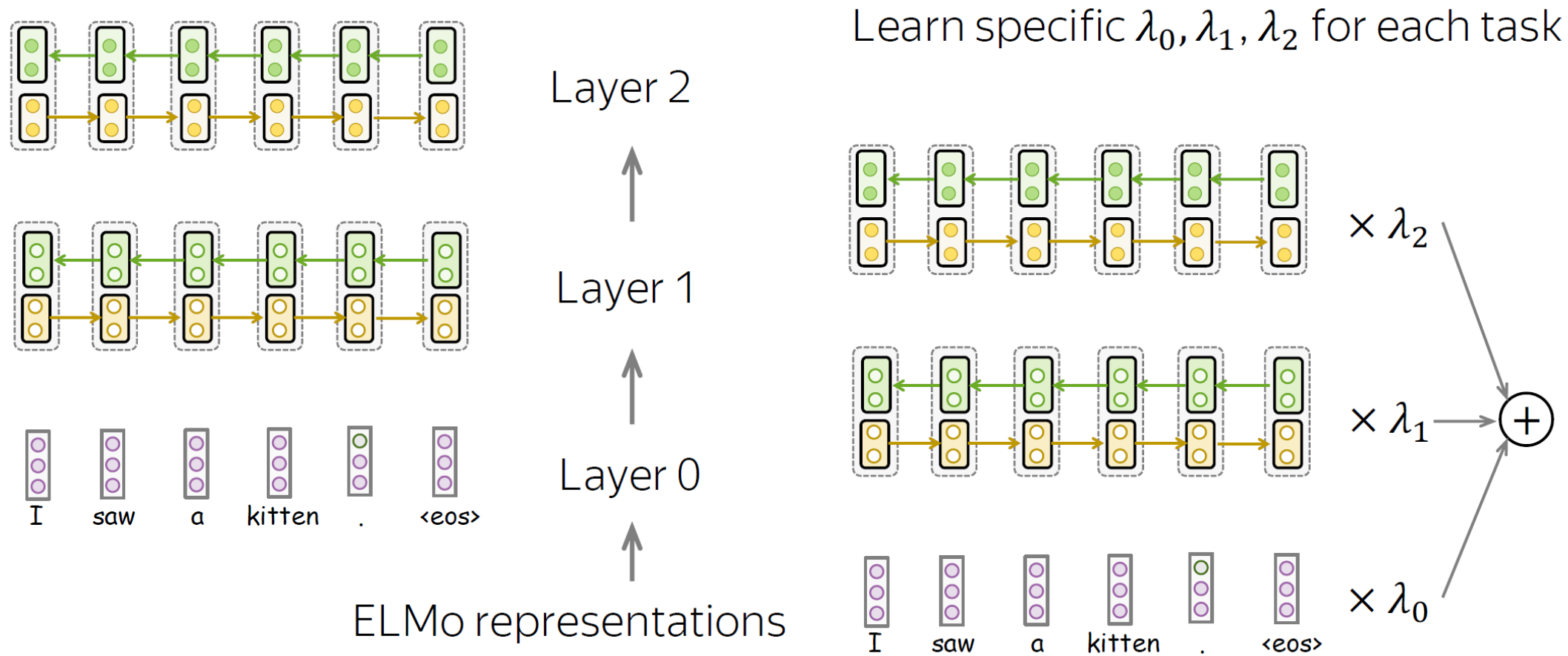
"cat on a mat <eos>"



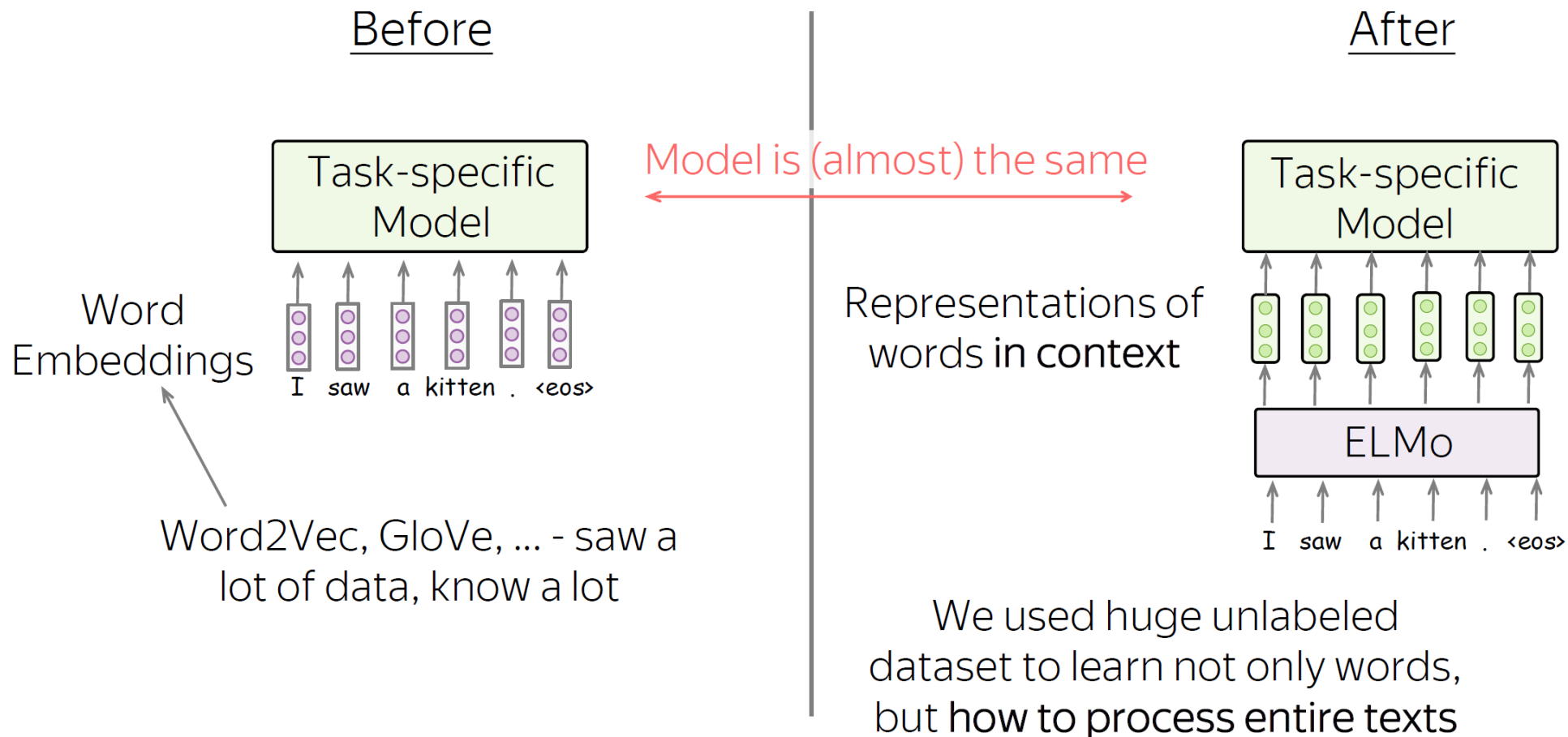
"I saw a cat"

This is the **cat** that
I saw, the one who
was **on the mat**

ELMo: от слов к словам в контексте

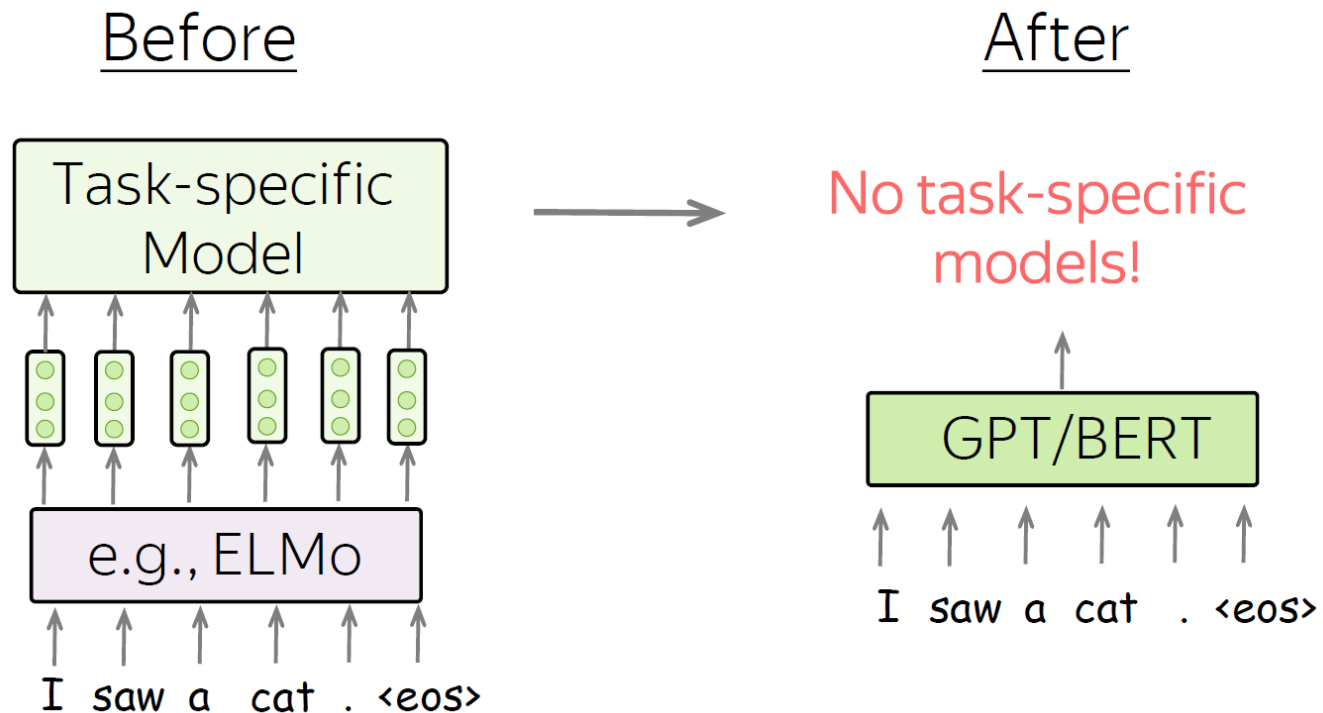


ELMo: как использовать?



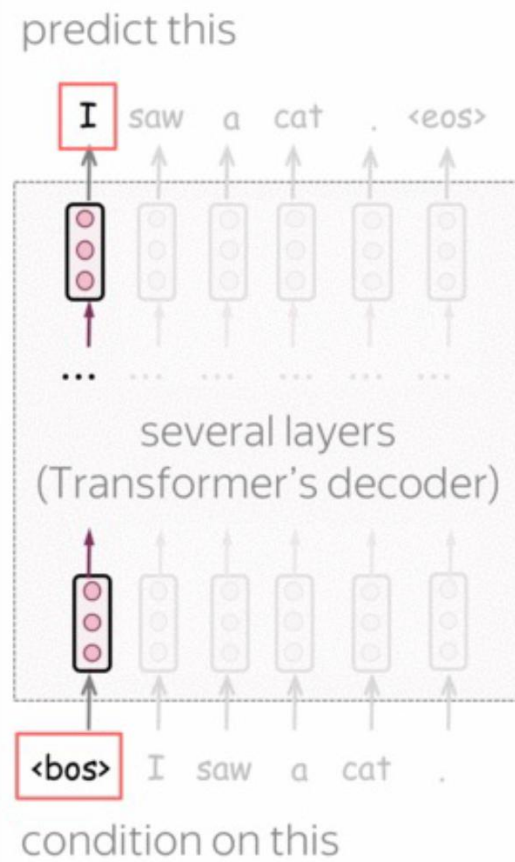
Идея 2: от специализированных моделей к унифицированным

Идея: вместо замены word embeddings, заменим входную модель.



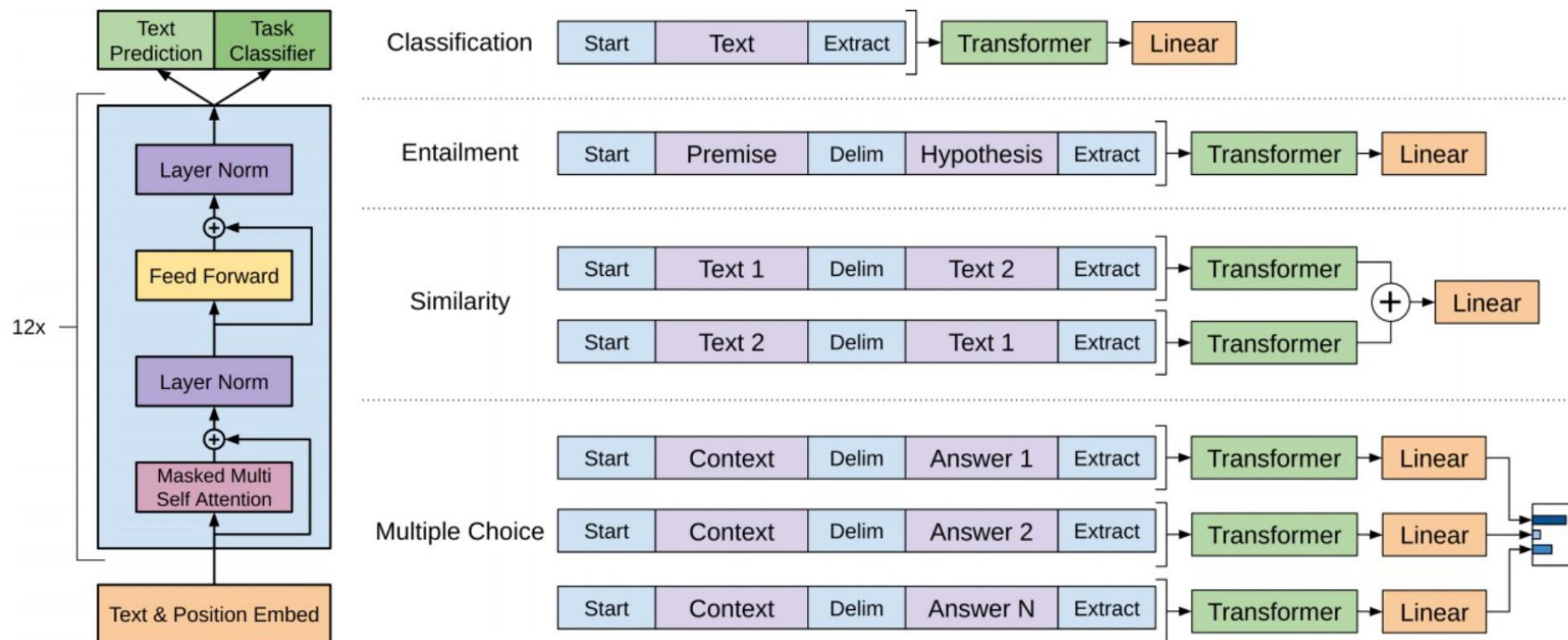
GPT(1-2-3): Декодер трансформера

- Трансформер слева направо
- GPT-1: декодер с 12 слоями



Тонкая настройка: использование GPT для нисходящих задач

Fine-tuning loss: $L = L_{xent} + \lambda \cdot L_{task}$

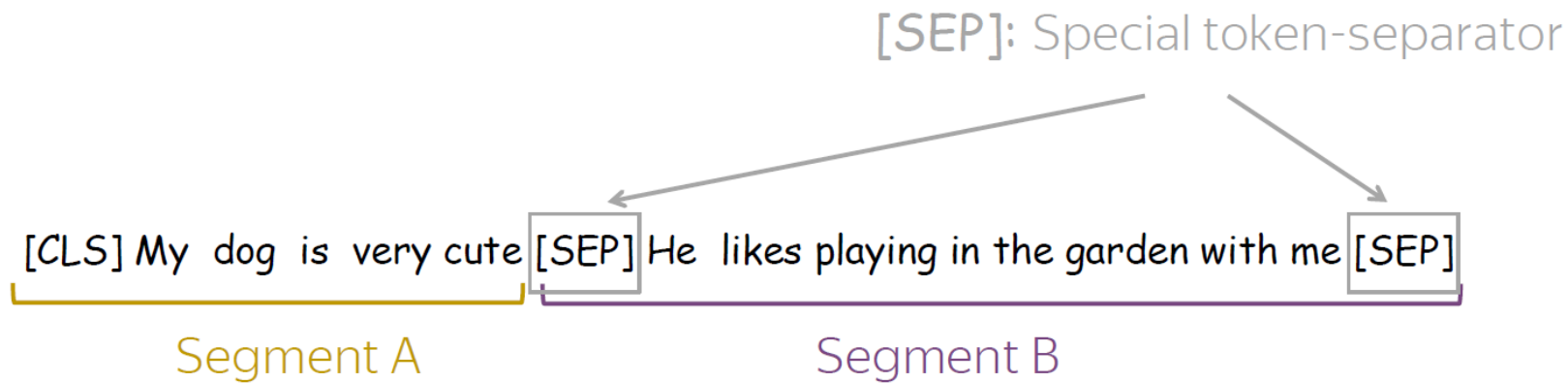


BERT: Transformer Encoder with Fancy Training

Архитектура модели:

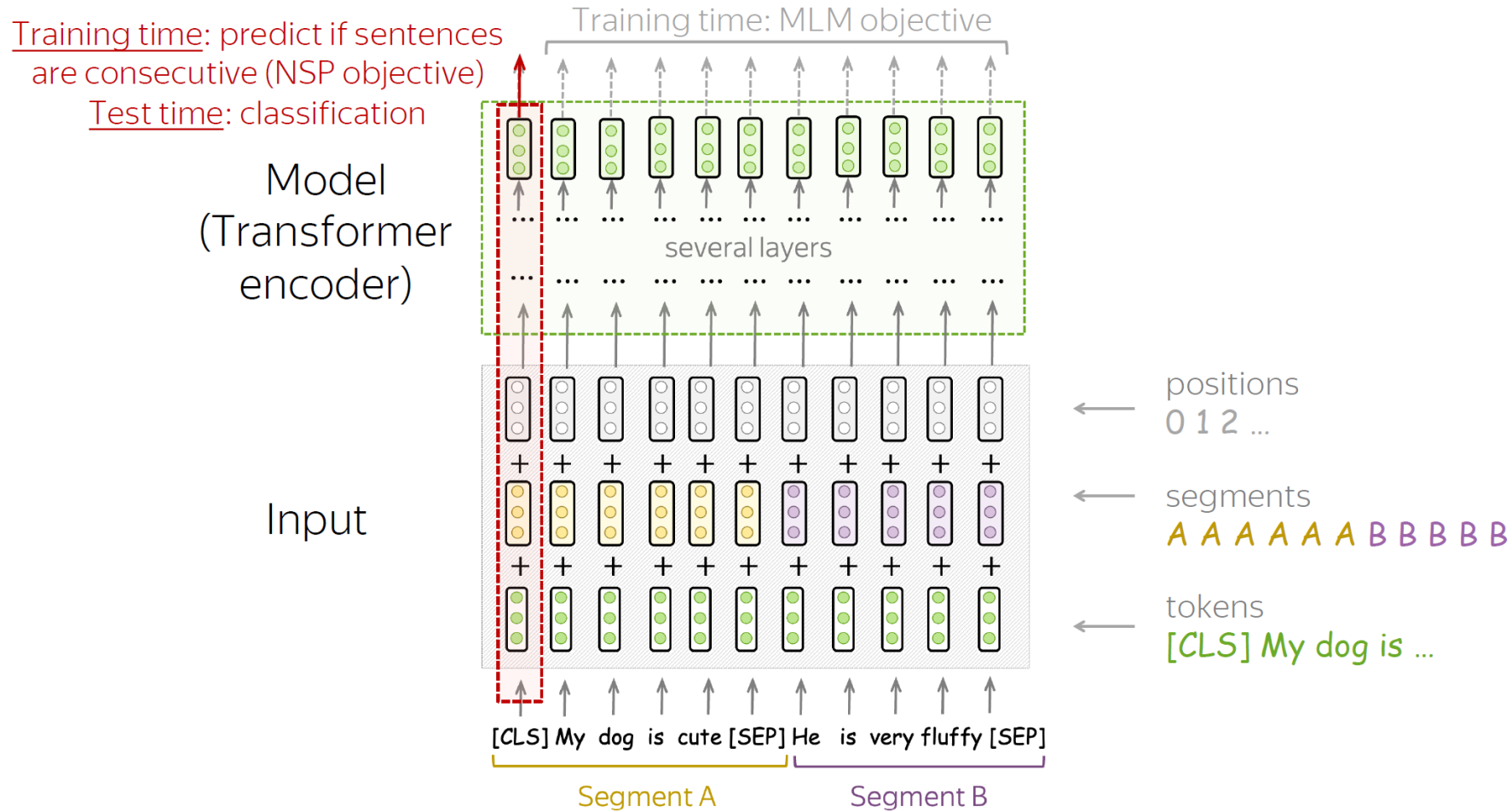
- Кодировщик Transformer
- Цели обучения (MLM: моделирование маскированного языка, NSP: прогнозирование следующего предложения)
- Большой объём данных

BERT: Transformer Encoder with Fancy Training



Training on pairs of sentences: either
consecutive or random (50%/50%)

BERT: Transformer Encoder with Fancy Training

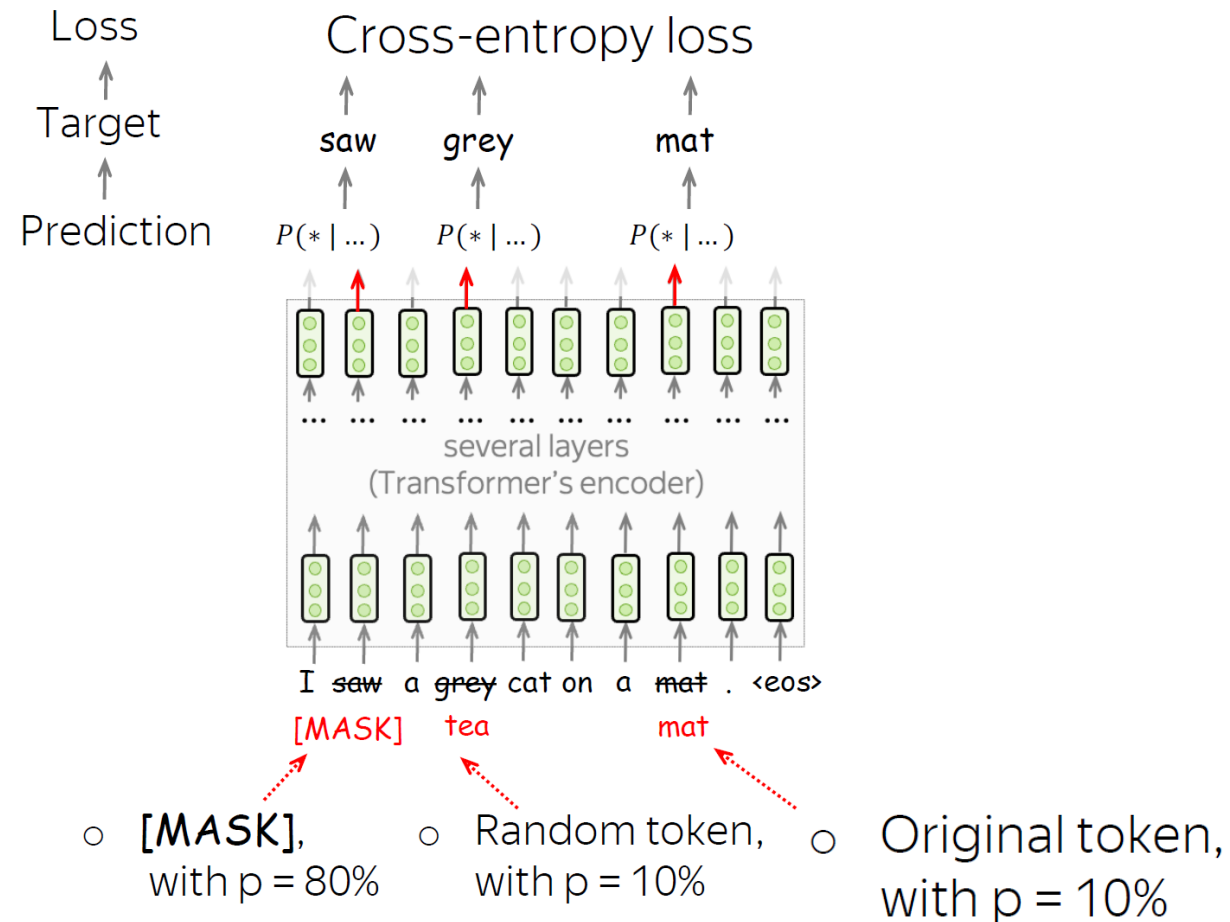


BERT: Цель моделирования маскированного языка

На каждом этапе обучения:

- случайным образом выбрать 15% токенов
- заменить каждый из выбранных токенов чем-либо
- предсказать случайно выбранные токены

BERT: Цель моделирования маскированного языка



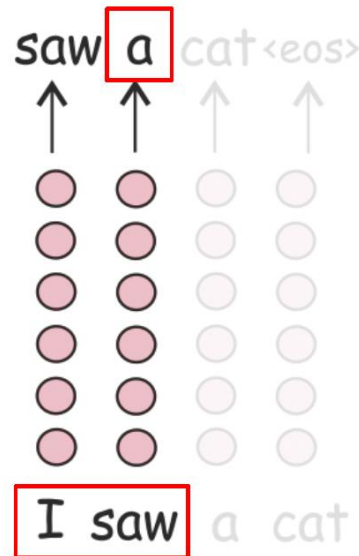
Цели обучения: LM и MLM

Language Modeling

Target: next token

Prediction: $P(* | \text{I saw})$

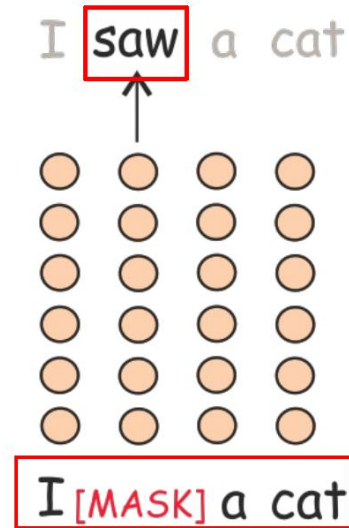
left-to-right, does
not see future



Masked Language Modeling

Target: current token (the true one)

Prediction: $P(* | \text{I [MASK] a cat})$



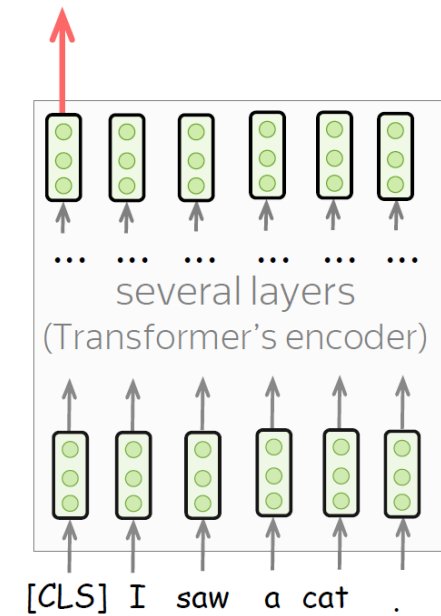
sees the whole text, but
something is corrupted

Тонкая настройка BERT: классификация отдельных предложений

Примеры заданий:

- SST-2 – бинарная классификация тональности
- CoLA (Корпус лингвистической приемлемости) – определить, является ли предложение лингвистически приемлемым

class label

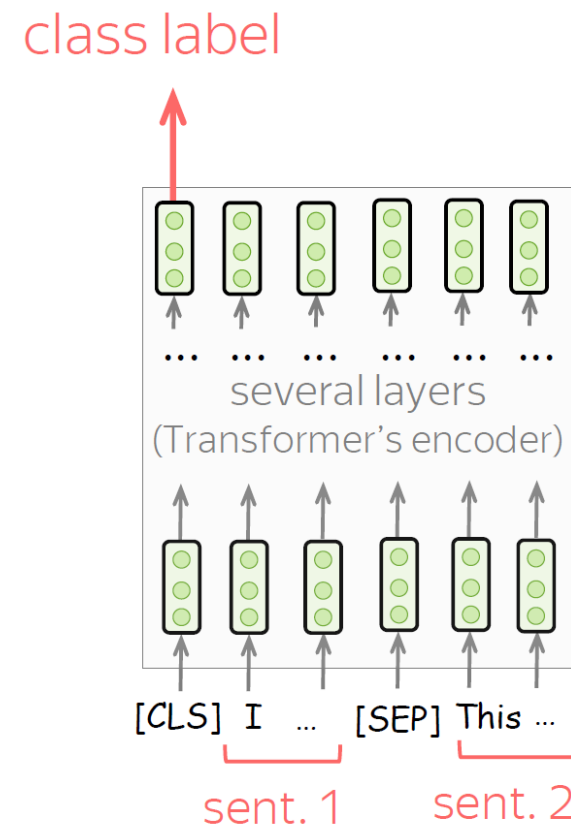


No second sentence!

Тонкая настройка BERT: классификация пар предложений

Примеры заданий:

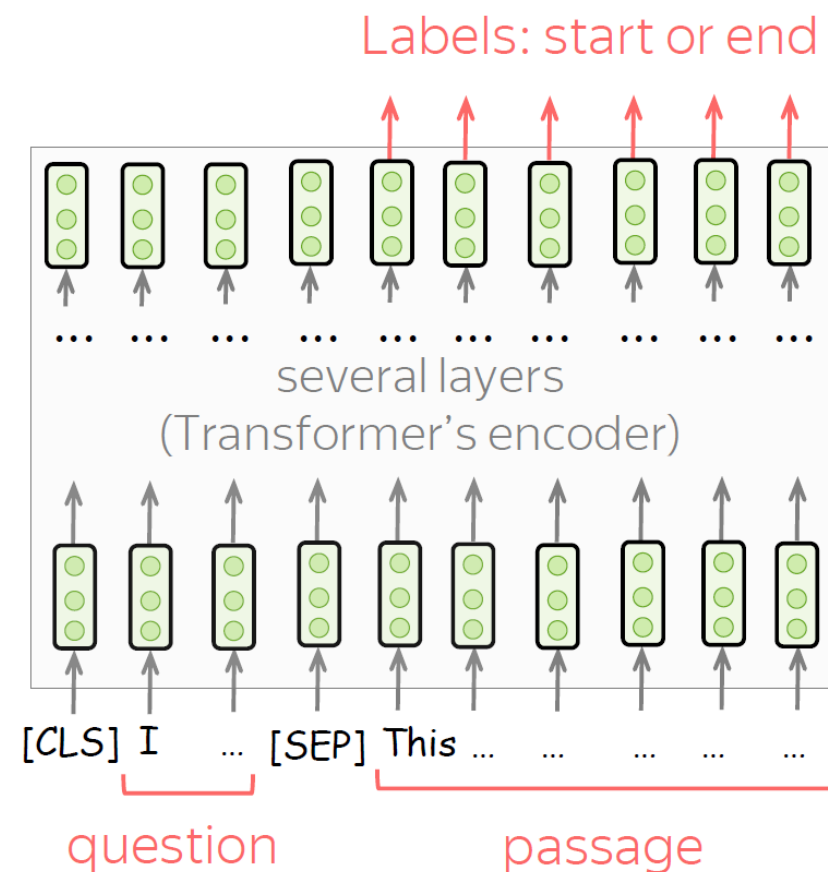
- MLNI – классификация выводов. Для пары предложений определить, является ли второе из них выводом, противоречием или нейтральным.
- QQP (Quora Question Pairs) – для двух вопросов определить, эквивалентны ли они семантически.
- STS-B – для двух предложений определите степень сходства от 1 до 5.



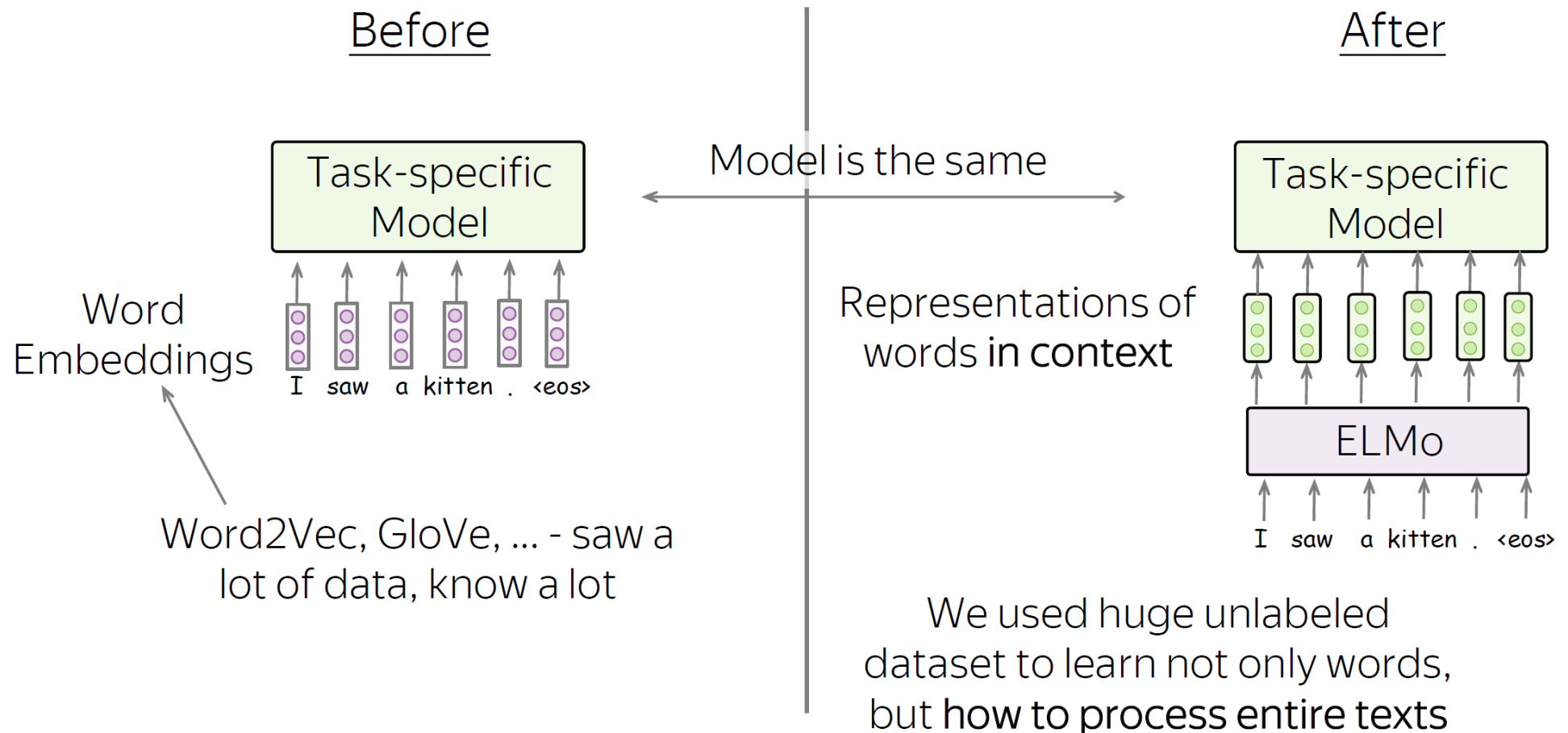
Тонкая настройка BERT: ответ на вопрос

Примеры заданий:

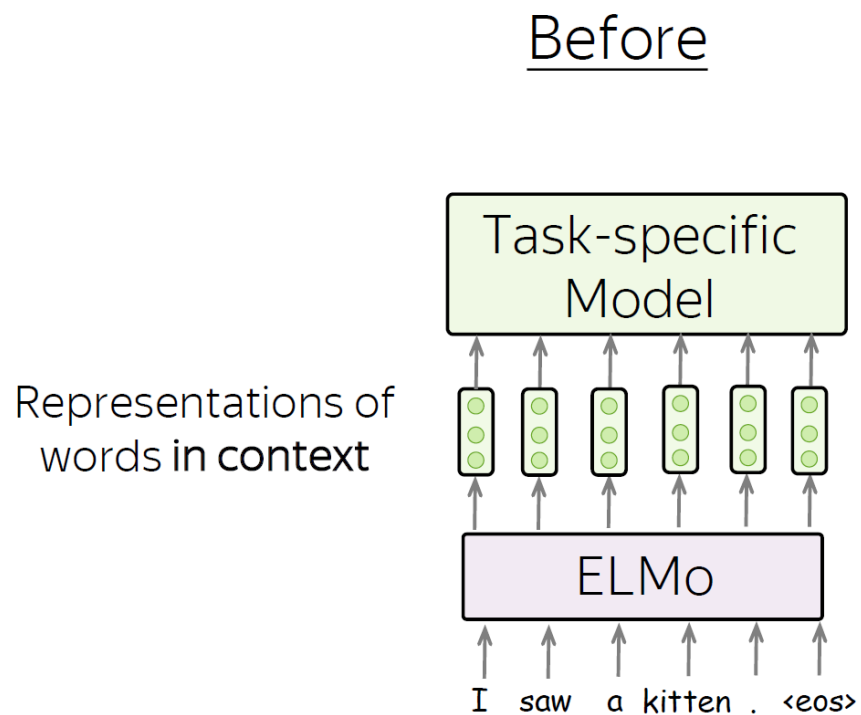
- SQUAD – набор данных с парами «вопрос-отрывок»; отрывок содержит ответ



ELMo: Что изменилось?

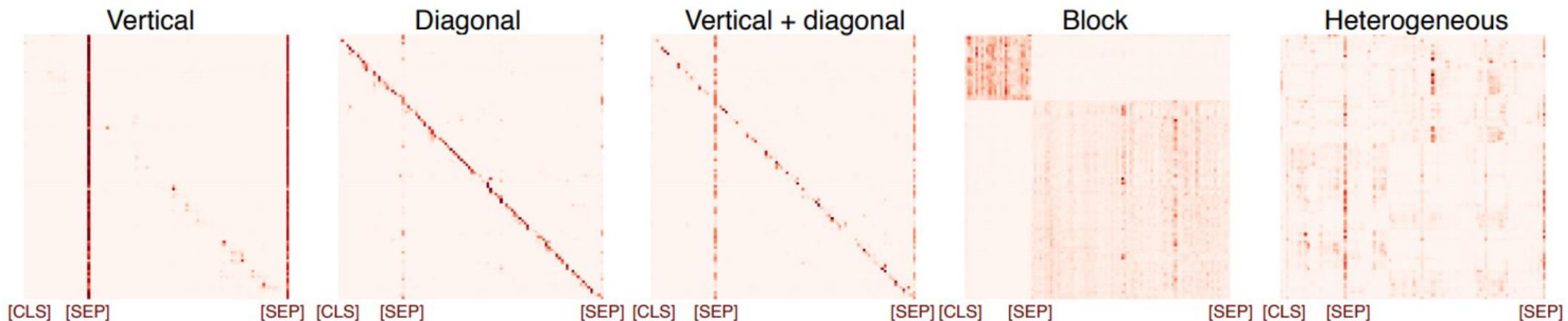


BERT: Что изменилось?



Методы анализа. BERT Self-Attention Heads

Типичные модели самовнимания:



BERT и Classical NLP Pipeline

Part of speech: I want to find more , [something] bigger or deeper . → NN (Noun)

Constituents: I want to find more , [something bigger or deeper] . → NP (Noun Phrase)

Dependencies: [I]₁ am not [sure]₂ how reliable that is , though . → nsubj (nominal subject)

Entities: The most fascinating is the maze known as [Wind Cave] . → LOC

Semantic Role Labeling: I want to [find]₁ [more , something bigger or deeper]₂ . → Agr1 (Agent)

Coreference: So [the followers]₁ waited to say anything about what [they]₂ saw . → True

BERT и Classical NLP Pipeline

Part of speech:

Constituents:

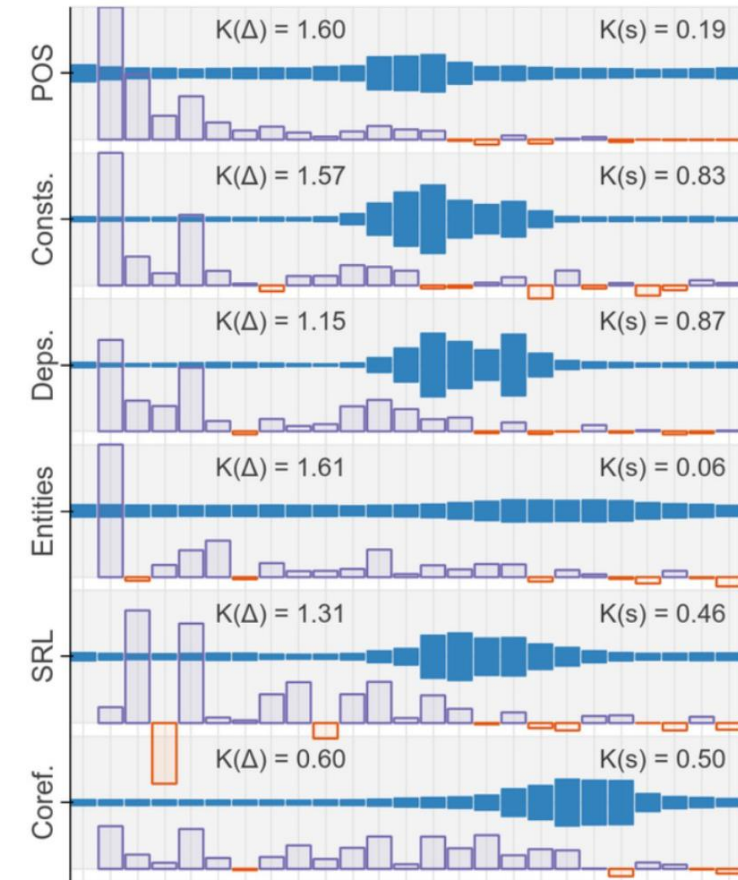
Dependencies:

Entities:

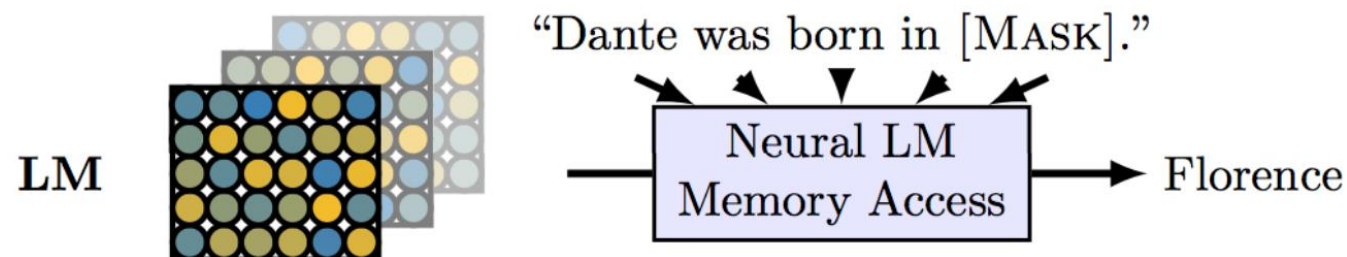
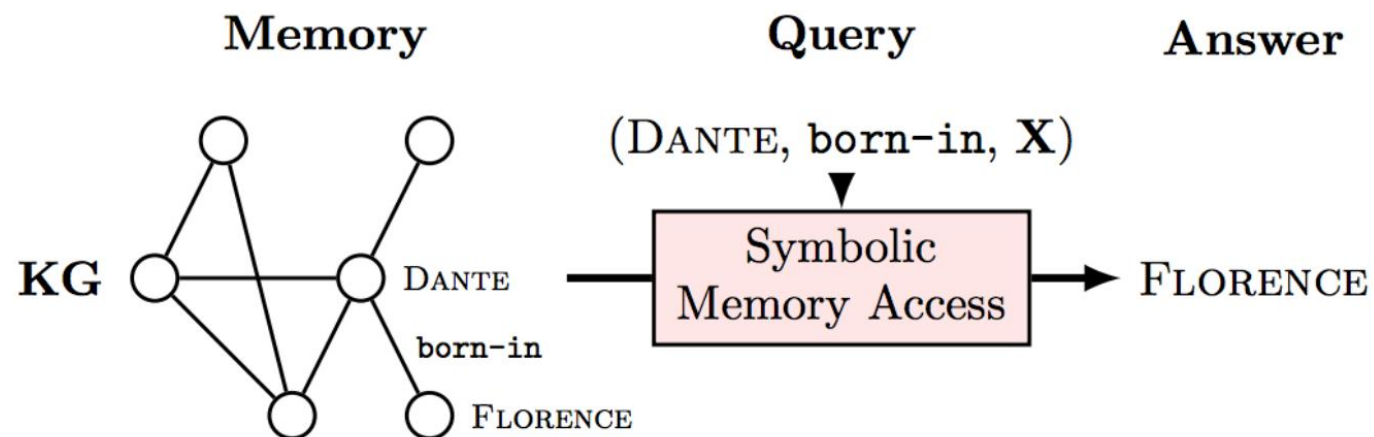
Semantic Role
Labeling:

Coreference:

In classical NLP, to
solve a subsequent
task is was required to
solve the previous one



Языковые модели как базы знаний



e.g. ELMo/BERT

Языковые модели как базы знаний

Relation	Query	Answer	Generation
P19	Francesco Bartolomeo Conti was born in ____.	Florence	Rome [-1.8], Florence [-1.8], Naples [-1.9], Milan [-2.4], Bologna [-2.5]
P20	Adolphe Adam died in ____.	Paris	Paris [-0.5], London [-3.5], Vienna [-3.6], Berlin [-3.8], Brussels [-4.0]
P279	English bulldog is a subclass of ____.	dog	dogs [-0.3], breeds [-2.2], dog [-2.4], cattle [-4.3], sheep [-4.5]
P37	The official language of Mauritius is ____.	English	English [-0.6], French [-0.9], Arabic [-6.2], Tamil [-6.7], Malayalam [-7.0]
P413	Patrick Oboya plays in ____ position.	midfielder	centre [-2.0], center [-2.2], midfielder [-2.4], forward [-2.4], midfield [-2.7]
P138	Hamburg Airport is named after ____.	Hamburg	Hess [-7.0], Hermann [-7.1], Schmidt [-7.1], Hamburg [-7.5], Ludwig [-7.5]
P364	The original language of Mon oncle Benjamin is ____.	French	French [-0.2], Breton [-3.3], English [-3.8], Dutch [-4.2], German [-4.9]
P54	Dani Alves plays with ____.	Barcelona	Santos [-2.4], Porto [-2.5], Sporting [-3.1], Brazil [-3.3], Portugal [-3.7]
P106	Paul Toungui is a ____ by profession.	politician	lawyer [-1.1], journalist [-2.4], teacher [-2.7], doctor [-3.0], physician [-3.7]
P527	Sodium sulfide consists of ____.	sodium	water [-1.2], sulfur [-1.7], sodium [-2.5], zinc [-2.8], salt [-2.9]
P102	Gordon Scholes is a member of the ____ political party.	Labor	Labour [-1.3], Conservative [-1.6], Green [-2.4], Liberal [-2.9], Labor [-2.9]
P530	Kenya maintains diplomatic relations with ____.	Uganda	India [-3.0], Uganda [-3.2], Tanzania [-3.5], China [-3.6], Pakistan [-3.6]
P176	iPod Touch is produced by ____.	Apple	Apple [-1.6], Nokia [-1.7], Sony [-2.0], Samsung [-2.6], Intel [-3.1]
P30	Bailey Peninsula is located in ____.	Antarctica	Antarctica [-1.4], Bermuda [-2.2], Newfoundland [-2.5], Alaska [-2.7], Canada [-3.1]
P178	JDK is developed by ____.	Oracle	IBM [-2.0], Intel [-2.3], Microsoft [-2.5], HP [-3.4], Nokia [-3.5]
P1412	Carl III used to communicate in ____.	Swedish	German [-1.6], Latin [-1.9], French [-2.4], English [-3.0], Spanish [-3.0]
P17	Sunshine Coast, British Columbia is located in ____.	Canada	Canada [-1.2], Alberta [-2.8], Yukon [-2.9], Labrador [-3.4], Victoria [-3.4]
P39	Pope Clement VII has the position of ____.	pope	cardinal [-2.4], Pope [-2.5], pope [-2.6], President [-3.1], Chancellor [-3.2]
P264	Joe Cocker is represented by music label ____.	Capitol	EMI [-2.6], BMG [-2.6], Universal [-2.8], Capitol [-3.2], Columbia [-3.3]
P276	London Jazz Festival is located in ____.	London	London [-0.3], Greenwich [-3.2], Chelsea [-4.0], Camden [-4.6], Stratford [-4.8]
P127	Border TV is owned by ____.	ITV	Sky [-3.1], ITV [-3.3], Global [-3.4], Frontier [-4.1], Disney [-4.3]
P103	The native language of Mammootty is ____.	Malayalam	Malayalam [-0.2], Tamil [-2.1], Telugu [-4.8], English [-5.2], Hindi [-5.6]
P495	The Sharon Cuneta Show was created in ____.	Philippines	Manila [-3.2], Philippines [-3.6], February [-3.7], December [-3.8], Argentina [-4.0]

T-Rex