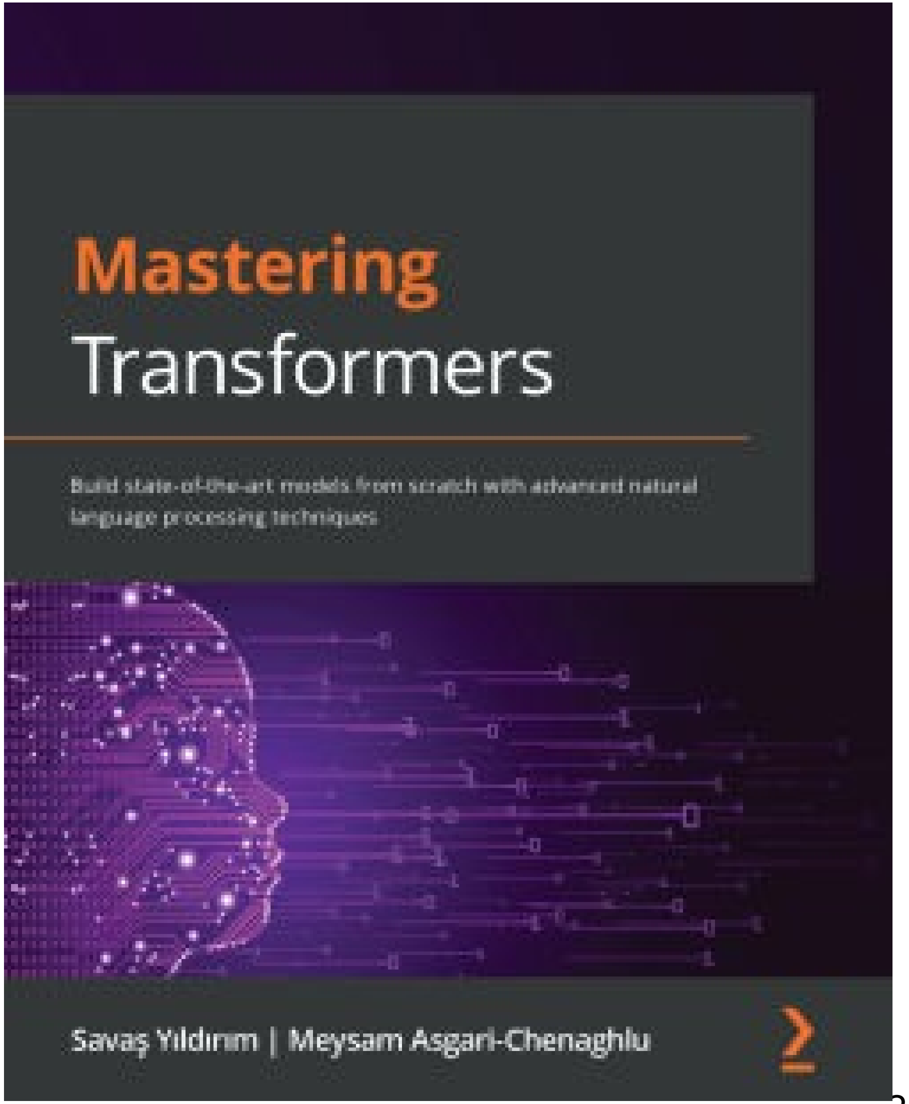
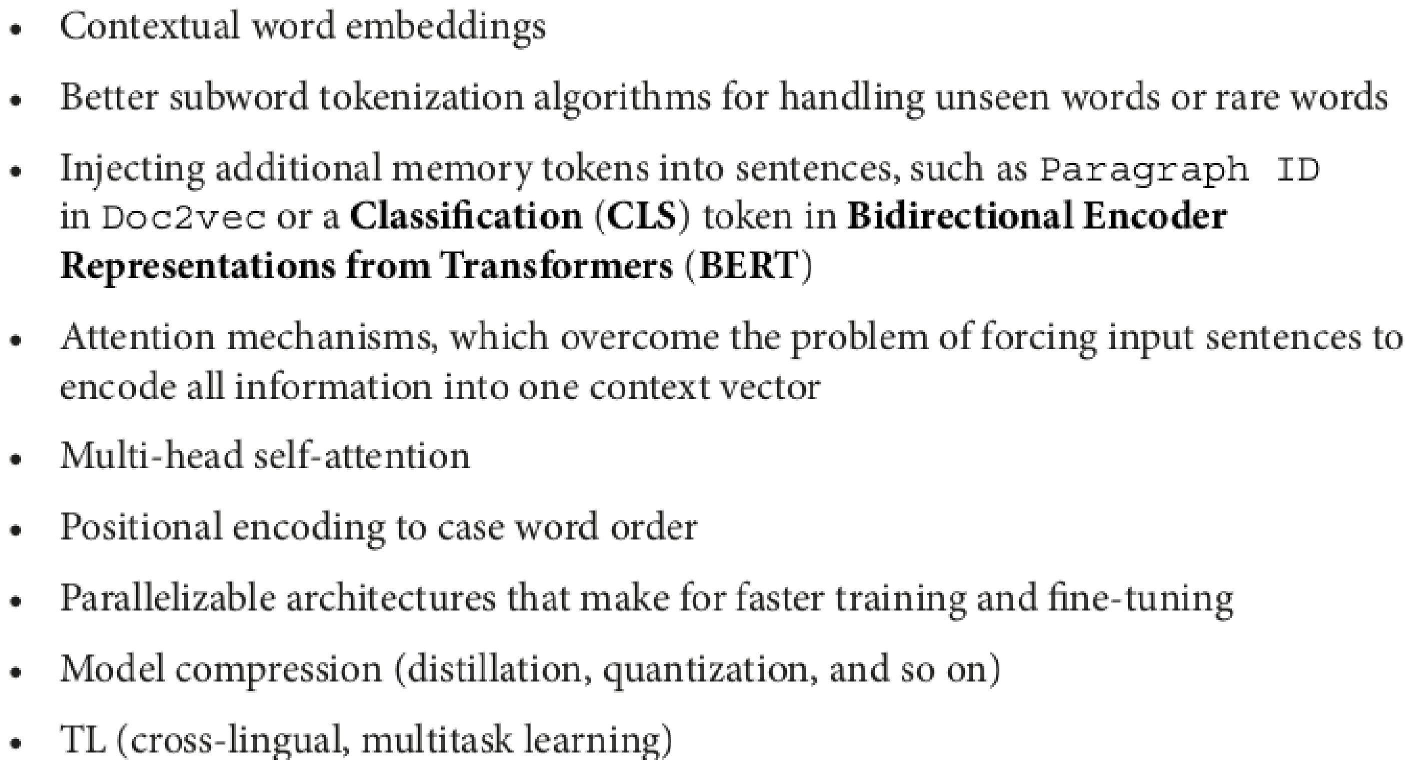
Natural Language Processing CMPE 353 AI

By Savaş Yıldırım

- Reference Text Book

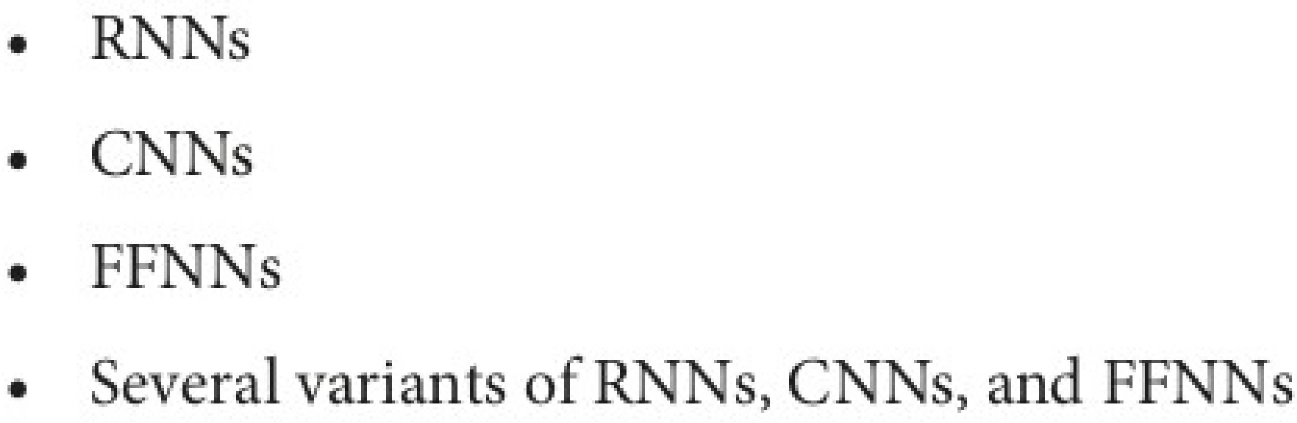

Mastering
Transformers
Build state-of-the-art models from scratch with advanced natural language processing techniques
Savaş Yıldırım | Meysam Asgari-Chenaghlu

Advances in NLP



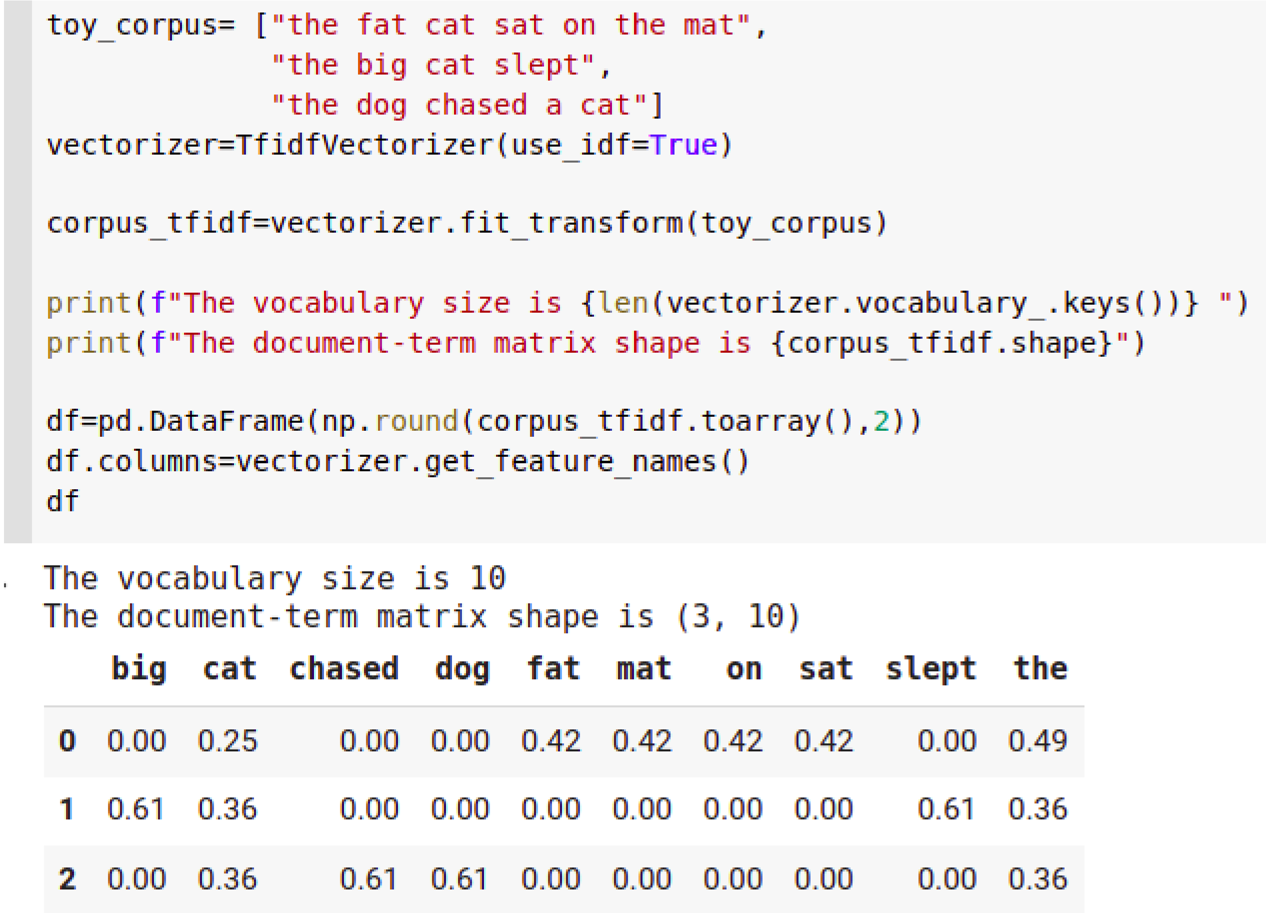
- Contextual word embeddings
- Better subword tokenization algorithms for handling unseen words or rare words
- Injecting additional memory tokens into sentences, such as `Paragraph ID` in `Doc2vec` or a **Classification** (**CLS**) token in **Bidirectional Encoder Representations from Transformers** (**BERT**)
- Attention mechanisms, which overcome the problem of forcing input sentences to encode all information into one context vector
- Multi-head self-attention
- Positional encoding to case word order
- Parallelizable architectures that make for faster training and fine-tuning
- Model compression (distillation, quantization, and so on)
- TL (cross-lingual, multitask learning)

Deep Learning Models



- RNNs
- CNNs
- FFNNs
- Several variants of RNNs, CNNs, and FFNNs

… And transformers

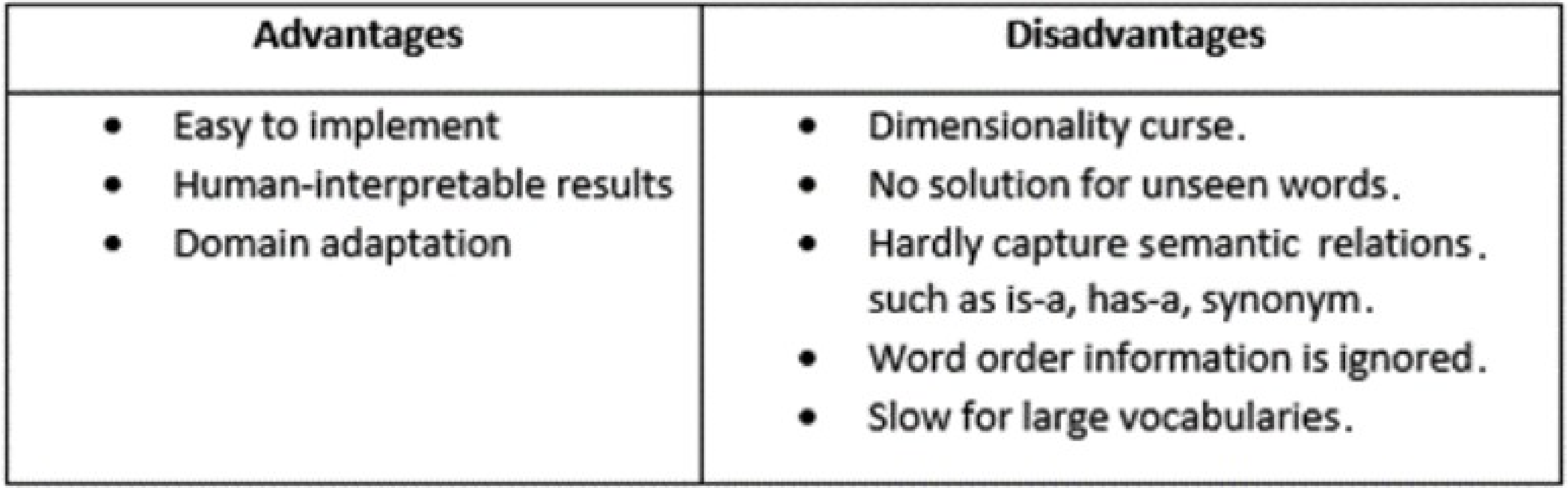


```
toy_corpus= ["the fat cat sat on the mat",
             "the big cat slept",
             "the dog chased a cat"]
vectorizer=TfidfVectorizer(use_idf=True)

corpus_tfidf=vectorizer.fit_transform(toy_corpus)

print(f"The vocabulary size is {len(vectorizer.vocabulary_.keys())} ")
print(f"The document-term matrix shape is {corpus_tfidf.shape}")

df=pd.DataFrame(np.round(corpus_tfidf.toarray(),2))
df.columns=vectorizer.get_feature_names()
df
```

```
The vocabulary size is 10
The document-term matrix shape is (3, 10)
```

	big	cat	chased	dog	fat	mat	on	sat	slept	the
0	0.00	0.25	0.00	0.00	0.42	0.42	0.42	0.42	0.00	0.49
1	0.61	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.36
2	0.00	0.36	0.61	0.61	0.00	0.00	0.00	0.00	0.00	0.36

Document x Word Matrix



Advantages	Disadvantages
• Easy to implement • Human-interpretable results • Domain adaptation	• Dimensionality curse. • No solution for unseen words. • Hardly capture semantic relations. such as is-a, has-a, synonym. • Word order information is ignored. • Slow for large vocabularies.

Table 1 – Advantages and disadvantages of a TF-IDF BoW model

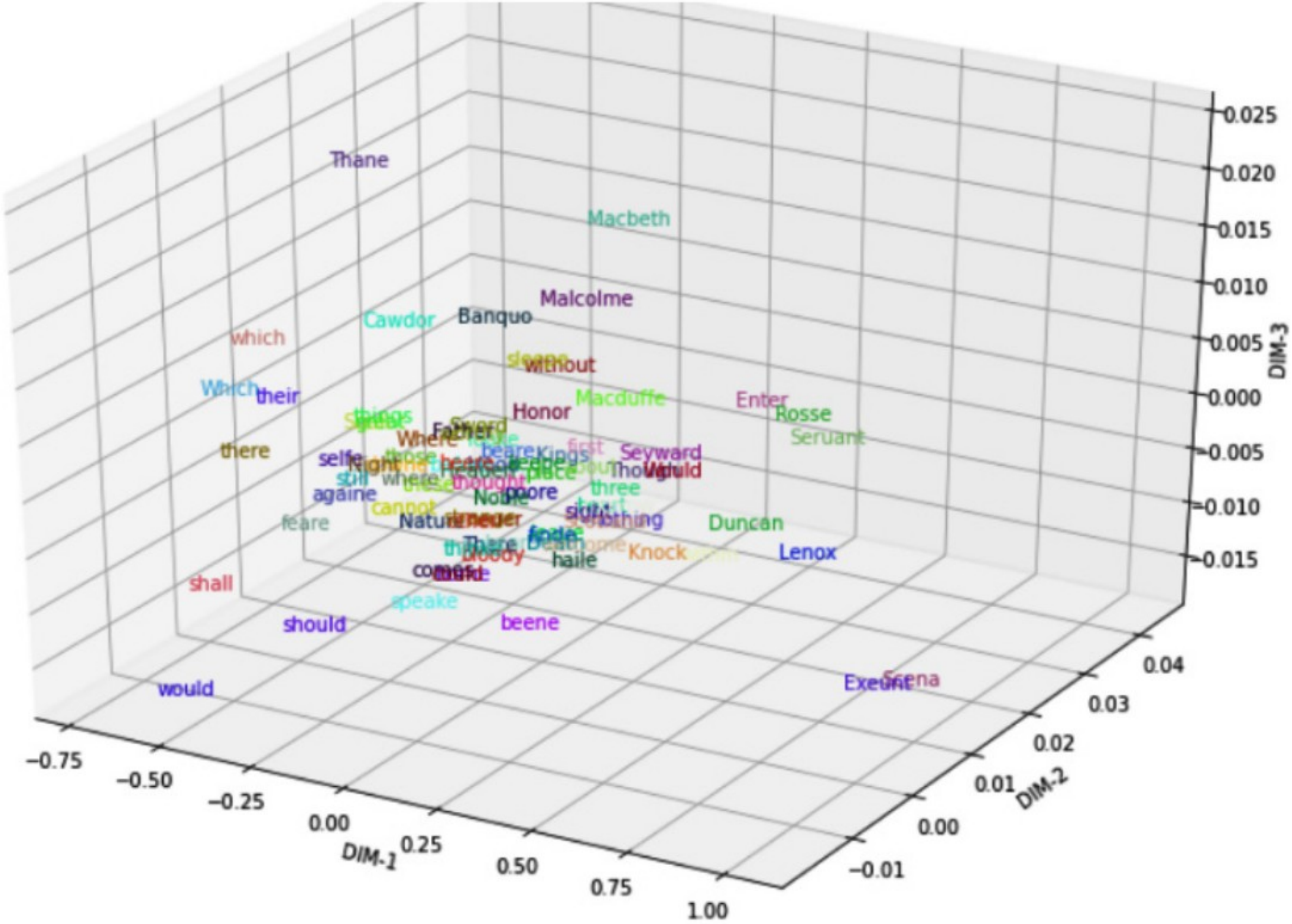


Figure 1.4 – Visualizing word embeddings with PCA

Word Embeddings

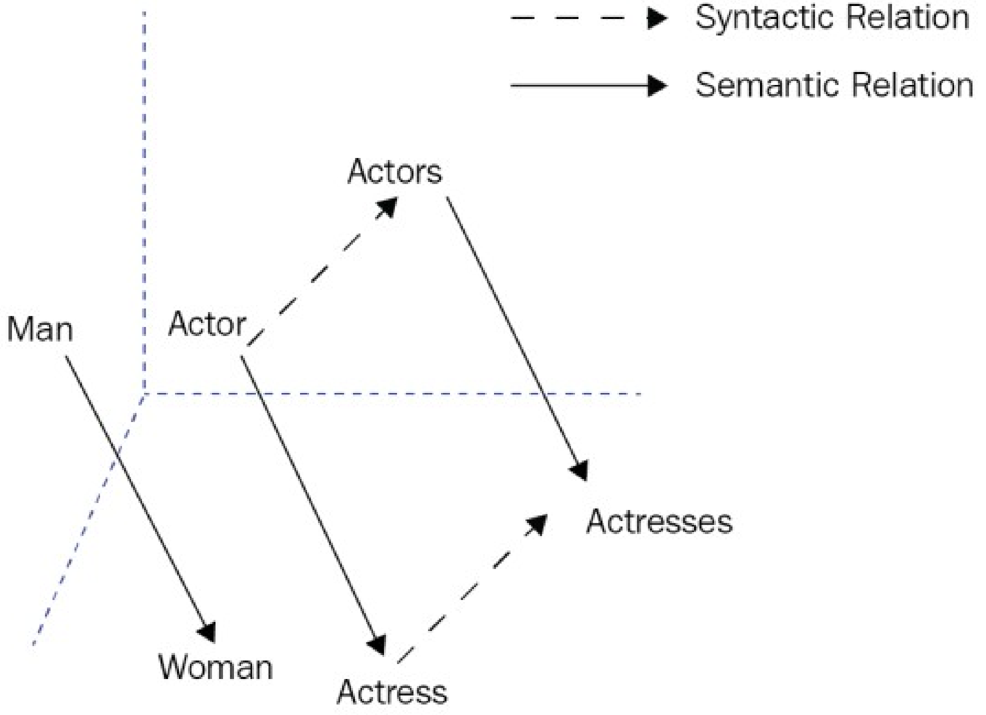


Figure 1.1 – Word embeddings offset for relation extraction

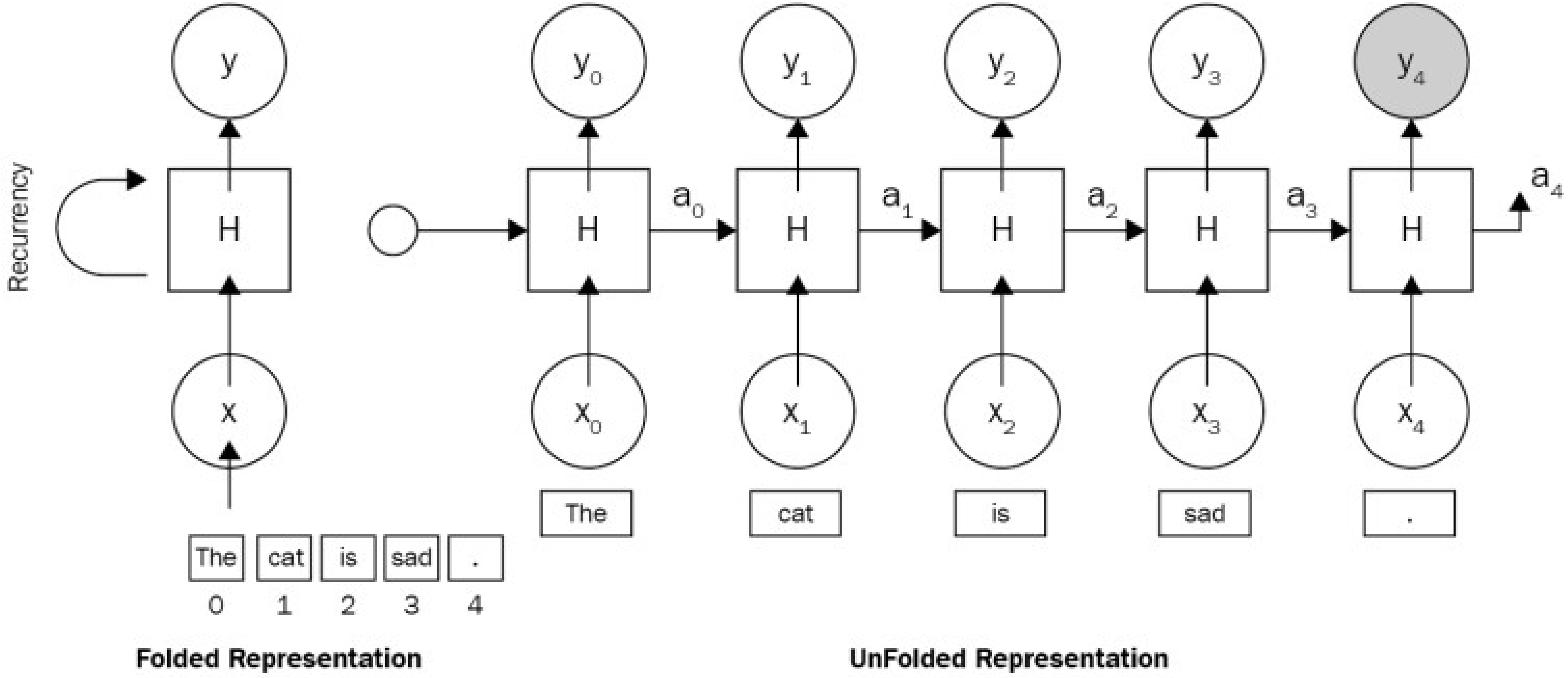


Figure 1.5 – An RNN architecture

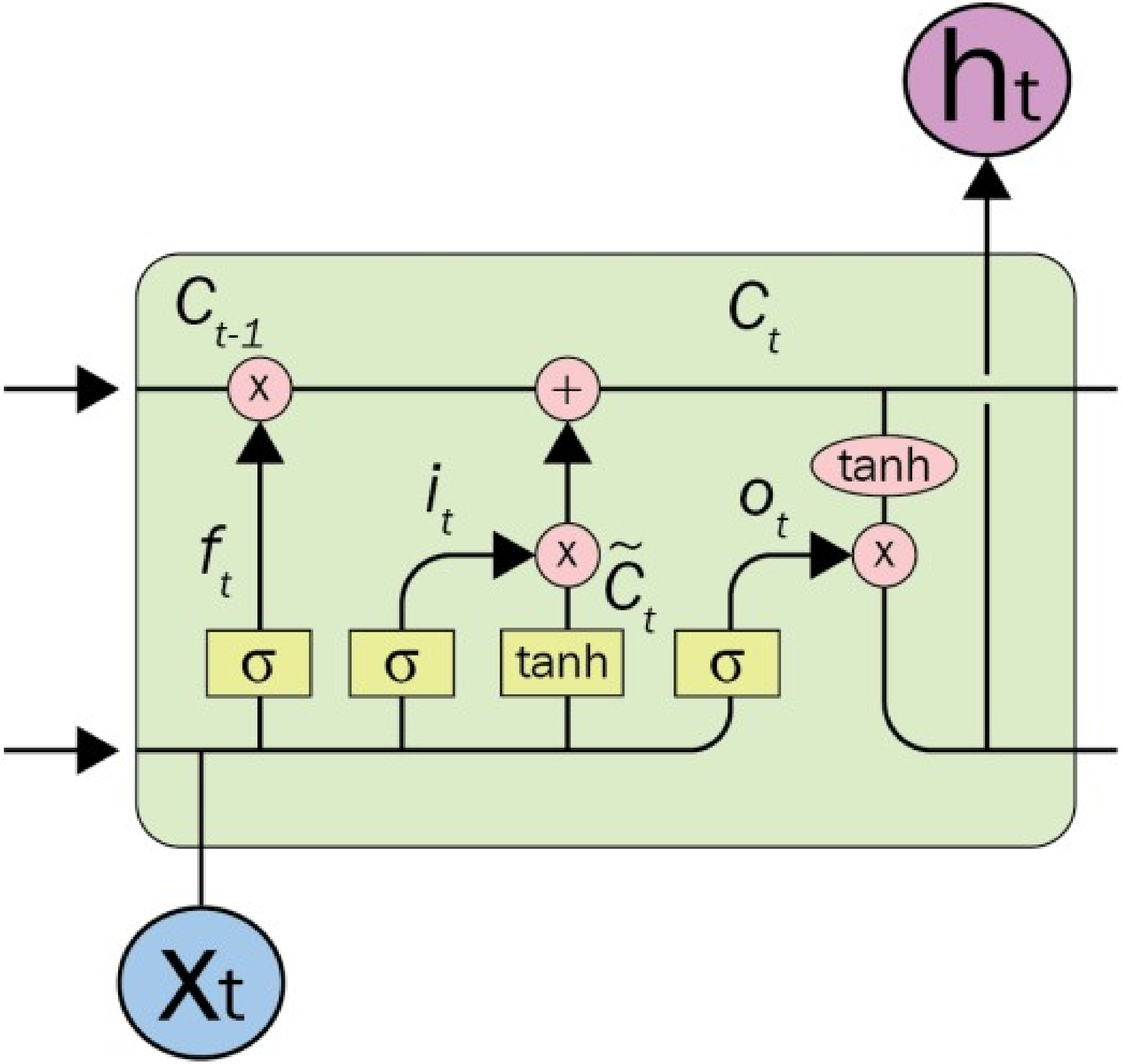


Figure 1.6 – An LSTM unit

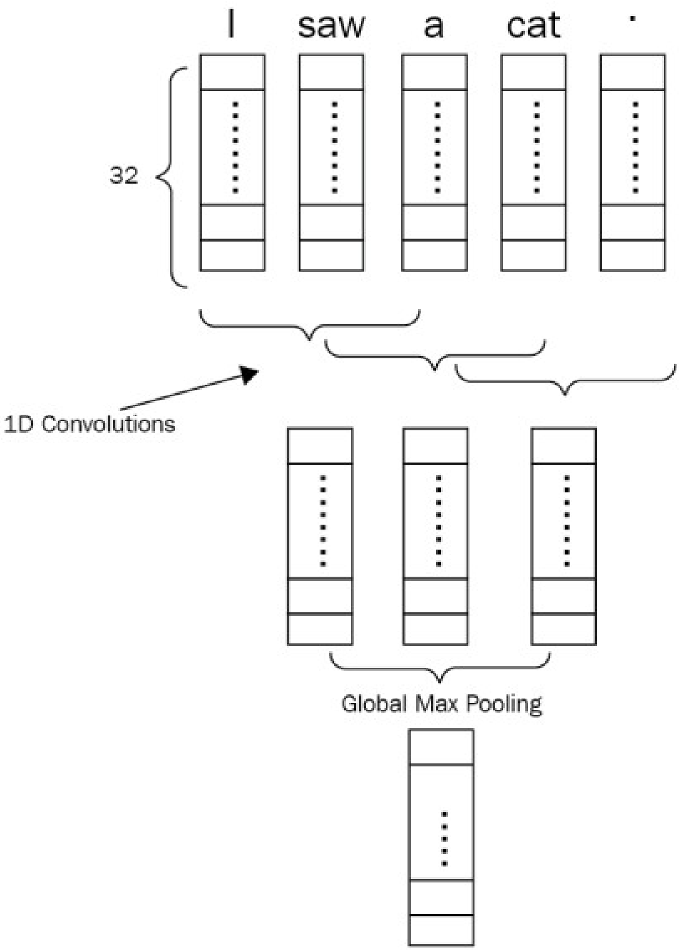


Figure 1.9 – 1D CNN network for a sentence of five tokens

Attention Mechanism

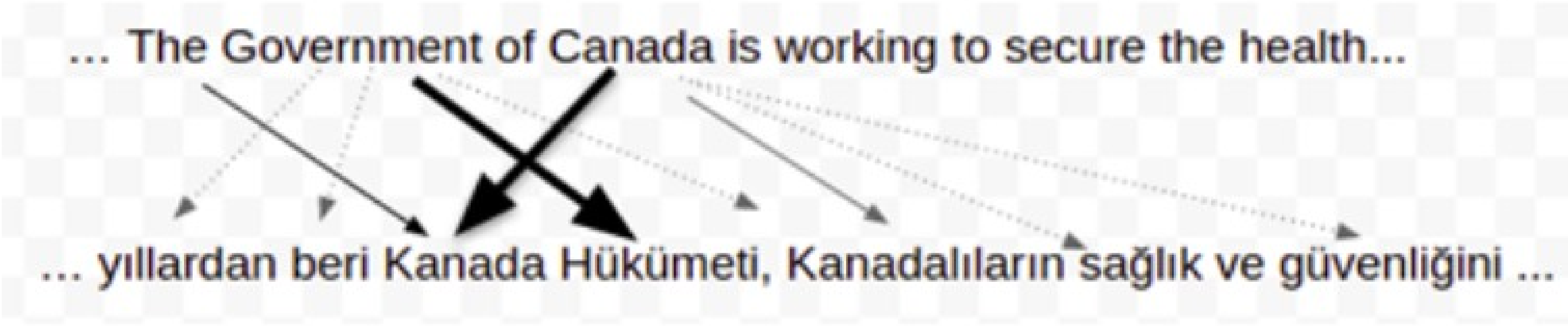

… The Government of Canada is working to secure the health...
… yıllardan beri Kanada Hükümeti, Kanadalıların sağlık ve güvenliğini ...

Figure 1.2 – Sketchy visualization of an attention mechanism

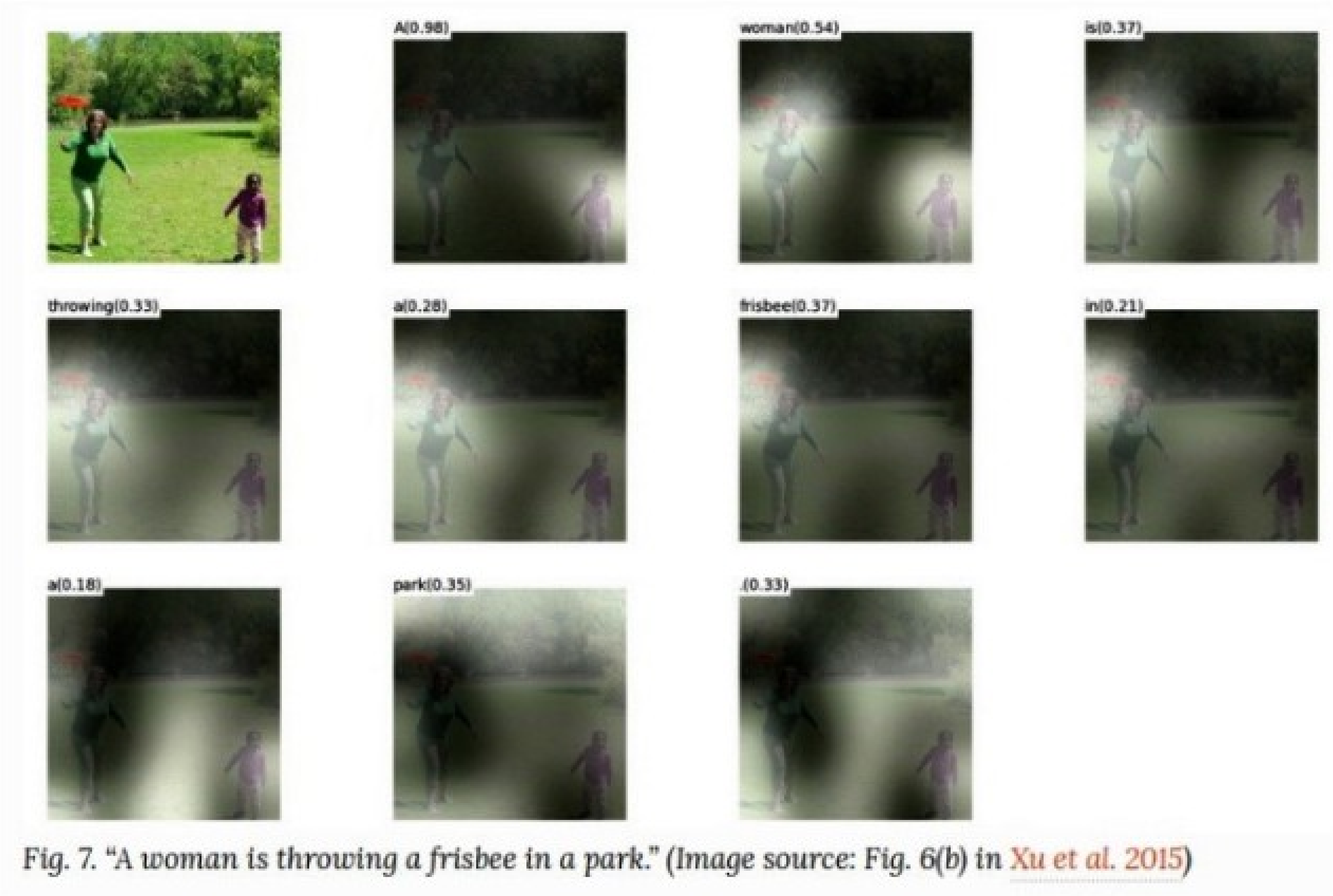


Fig. 7. "A woman is throwing a frisbee in a park." (Image source: Fig. 6(b) in Xu et al. 2015)

Figure 1.13 – Attention mechanism in computer vision

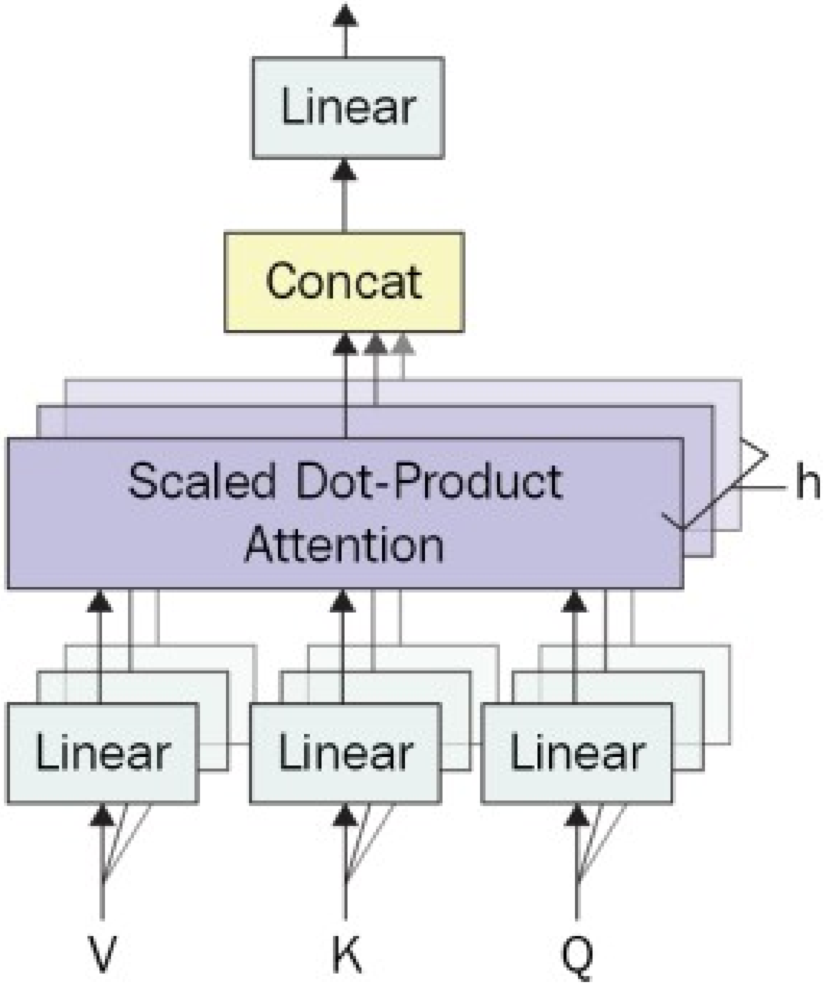


Figure 1.14 – Multi-head attention mechanism

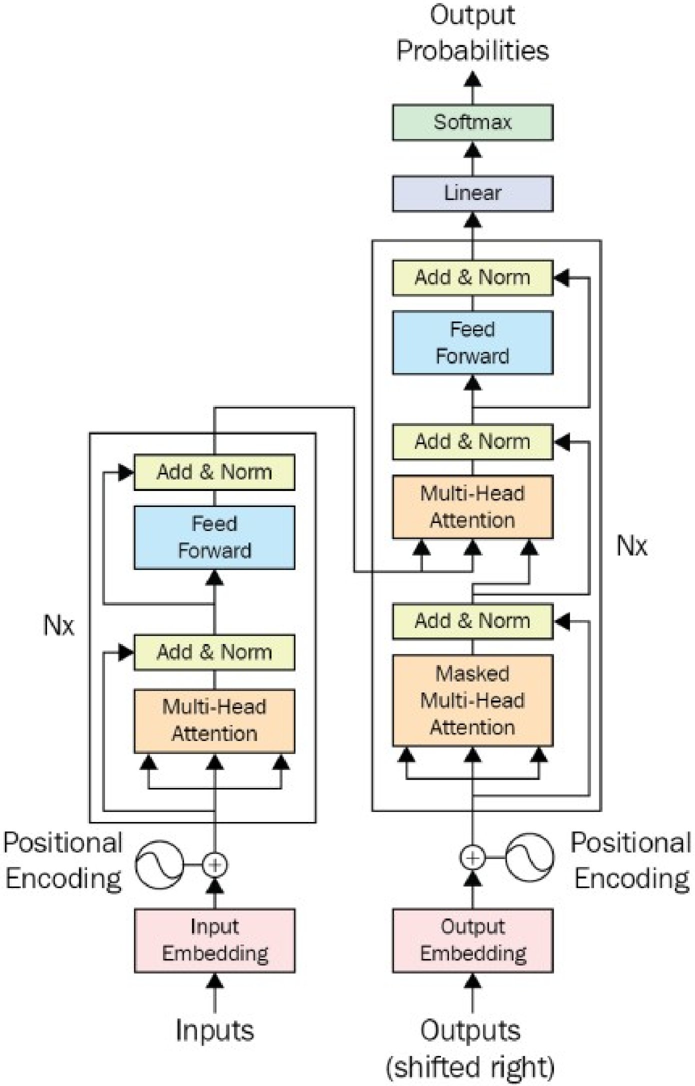
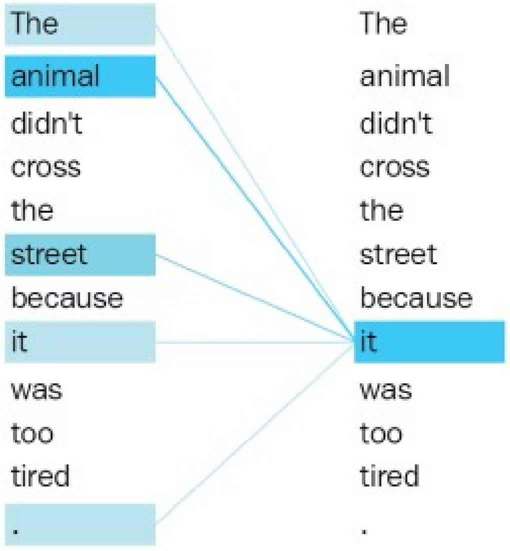
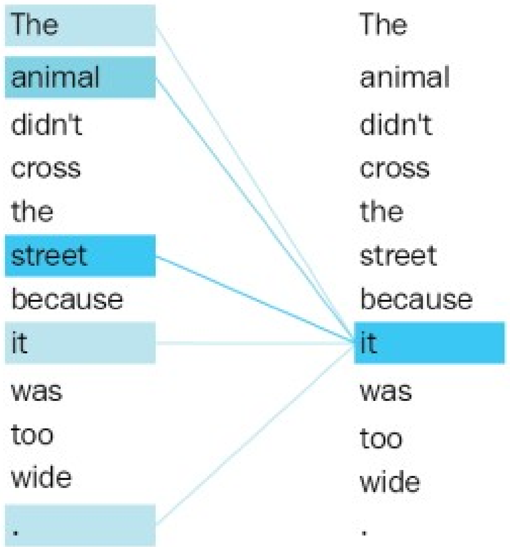


Figure 1.16 – A Transformer


The
animal
didn't
cross
the
street
because
it
was
too
tired
.
The
animal
didn't
cross
the
street
because
it
was
too
tired
.


The
animal
didn't
cross
the
street
because
it
was
too
wide
.
The
animal
didn't
cross
the
street
because
it
was
too
wide
.

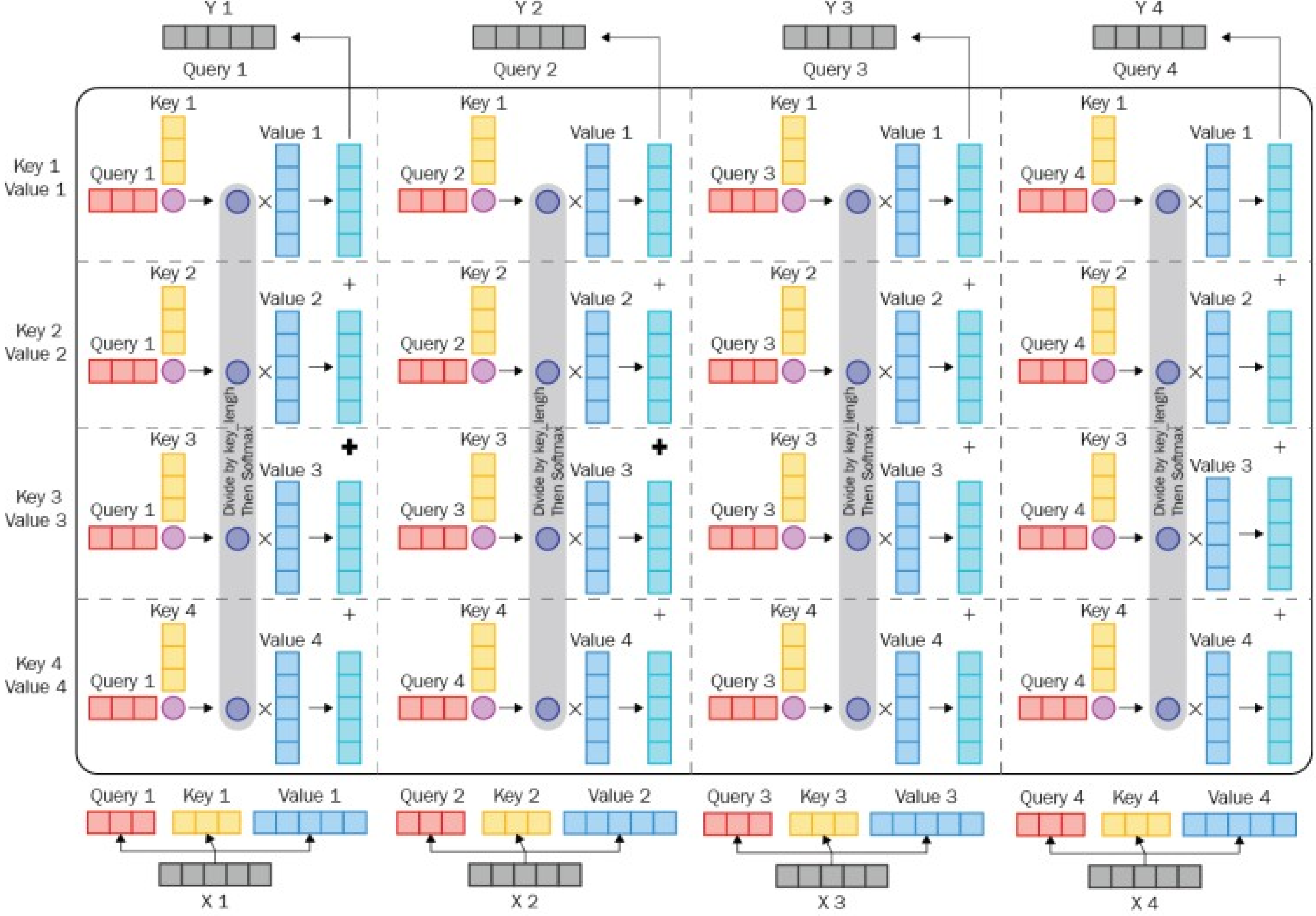


Figure 1.19 – Multi-head attention mechanism (Image inspired from https://imgur.com/gallery/FBQqrxw)

Transfer Learning in NLP

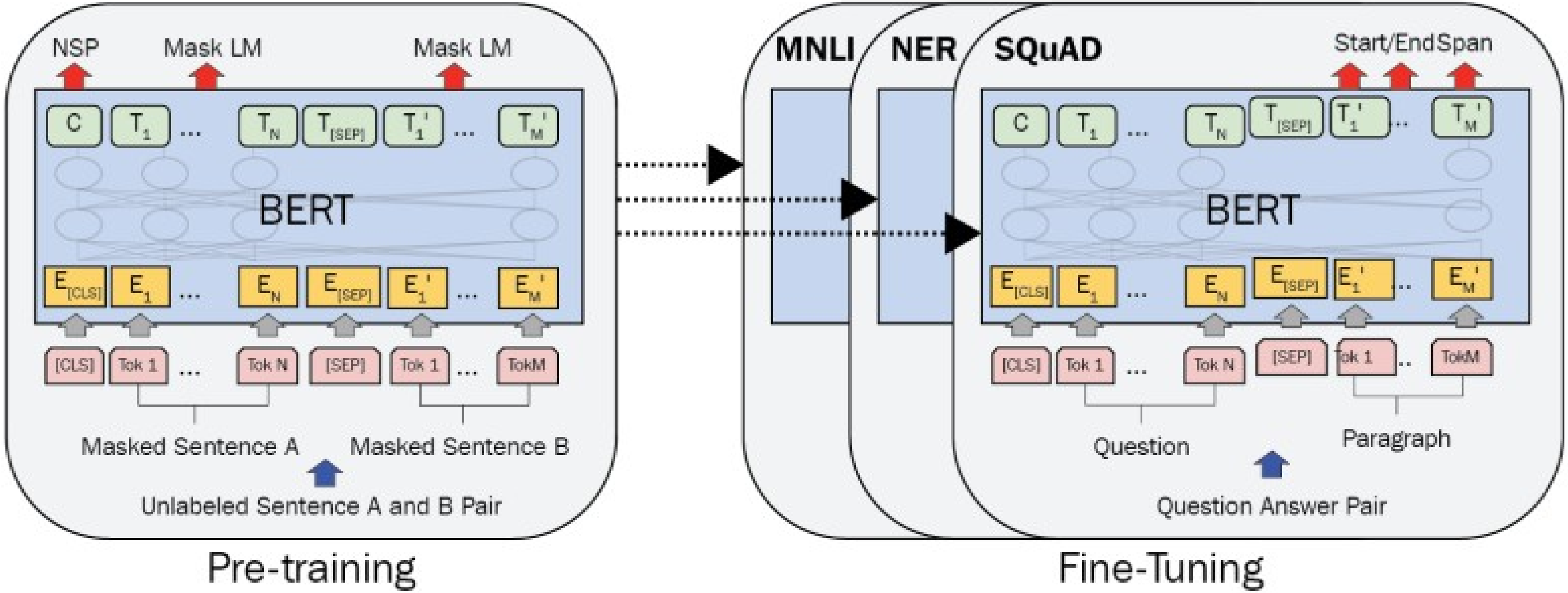


Figure 1.21 – Pre-training and fine-tuning procedures for BERT (Image inspired from J. Devlin et al., Bert: Pre-training of deep bidirectional Transformers for language understanding, 2018)

Pre-training Strategy in NLP

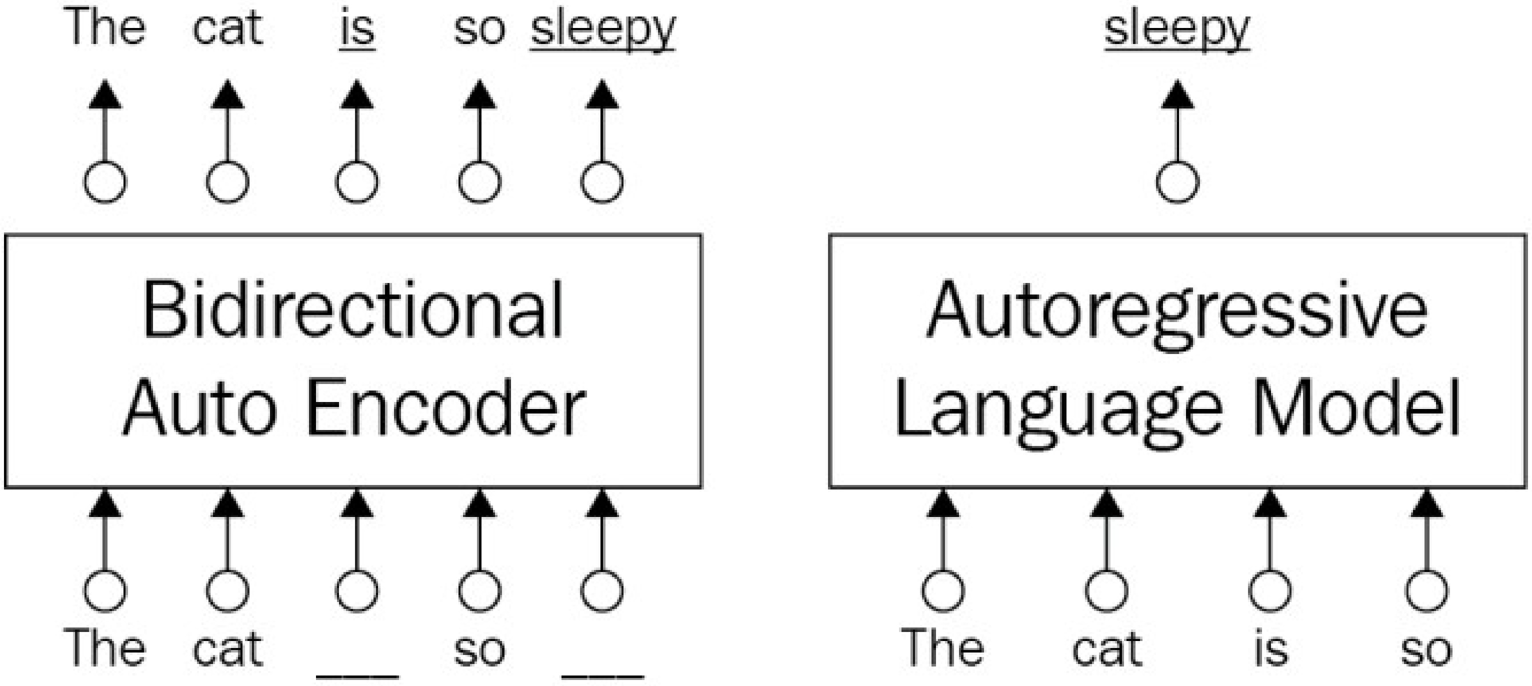


Figure 4.1 – AE versus AR language model

Advance denoising objectives

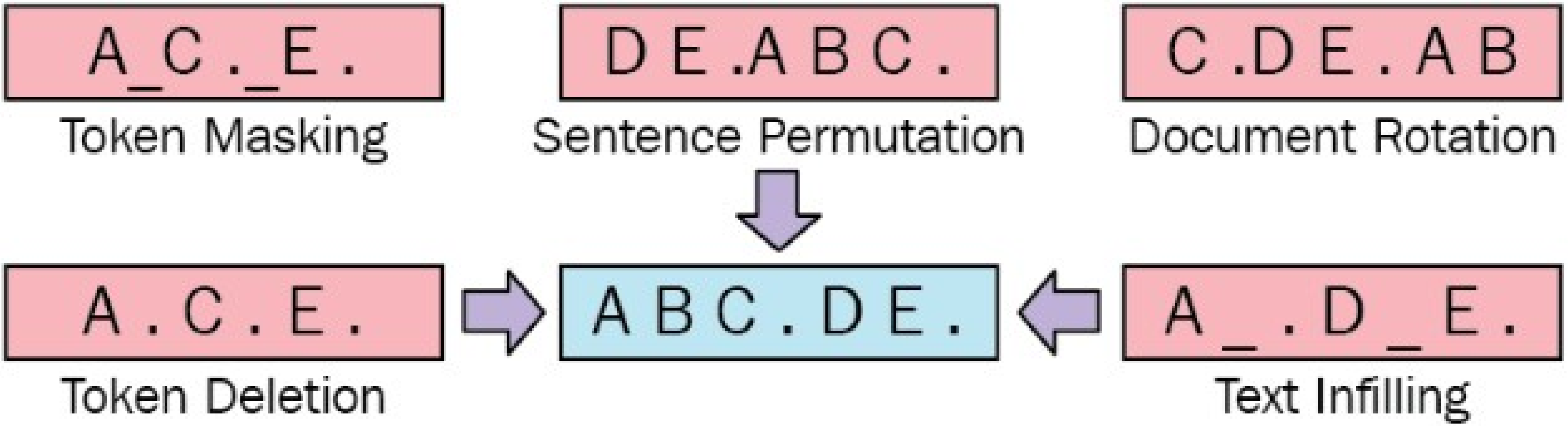


Figure 4.4 – Diagram inspired by the original BART paper

Classification Problem

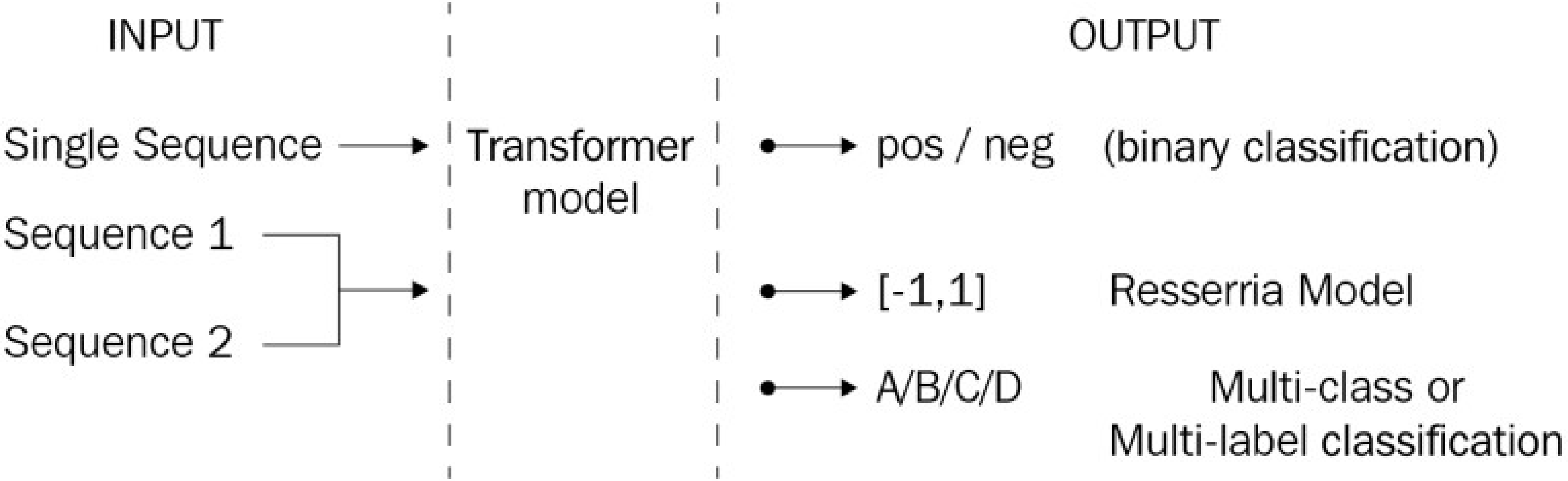
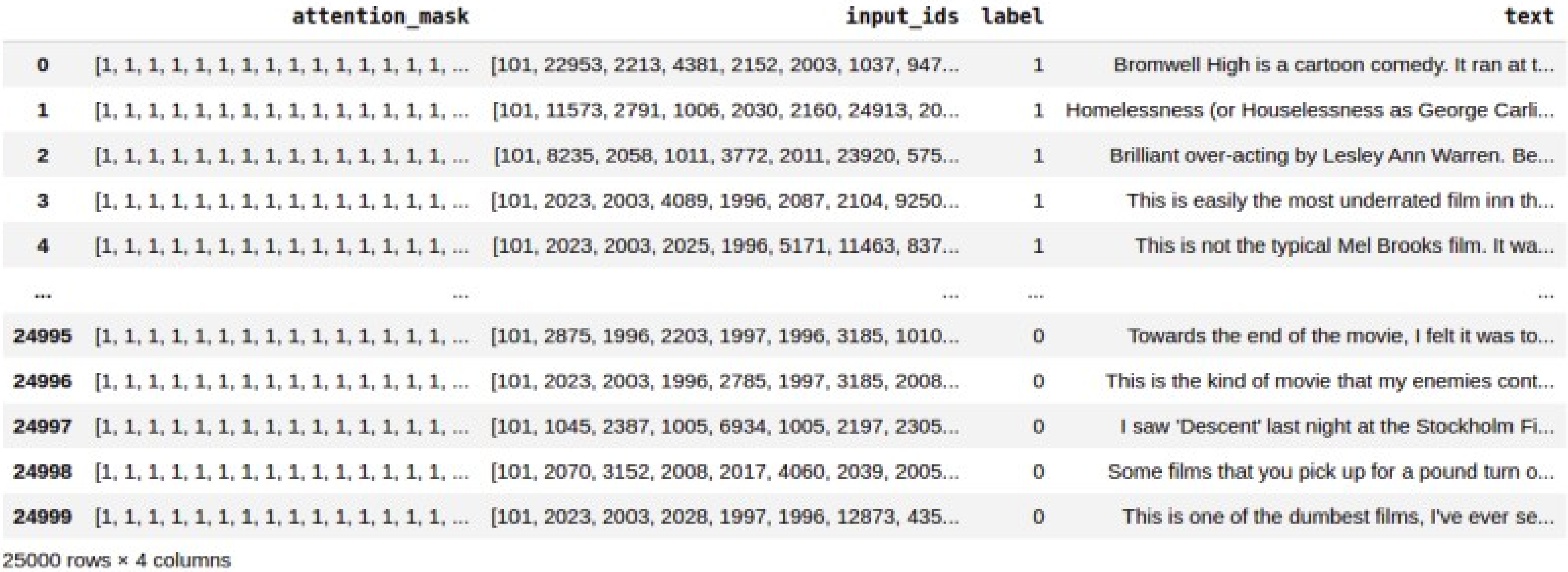


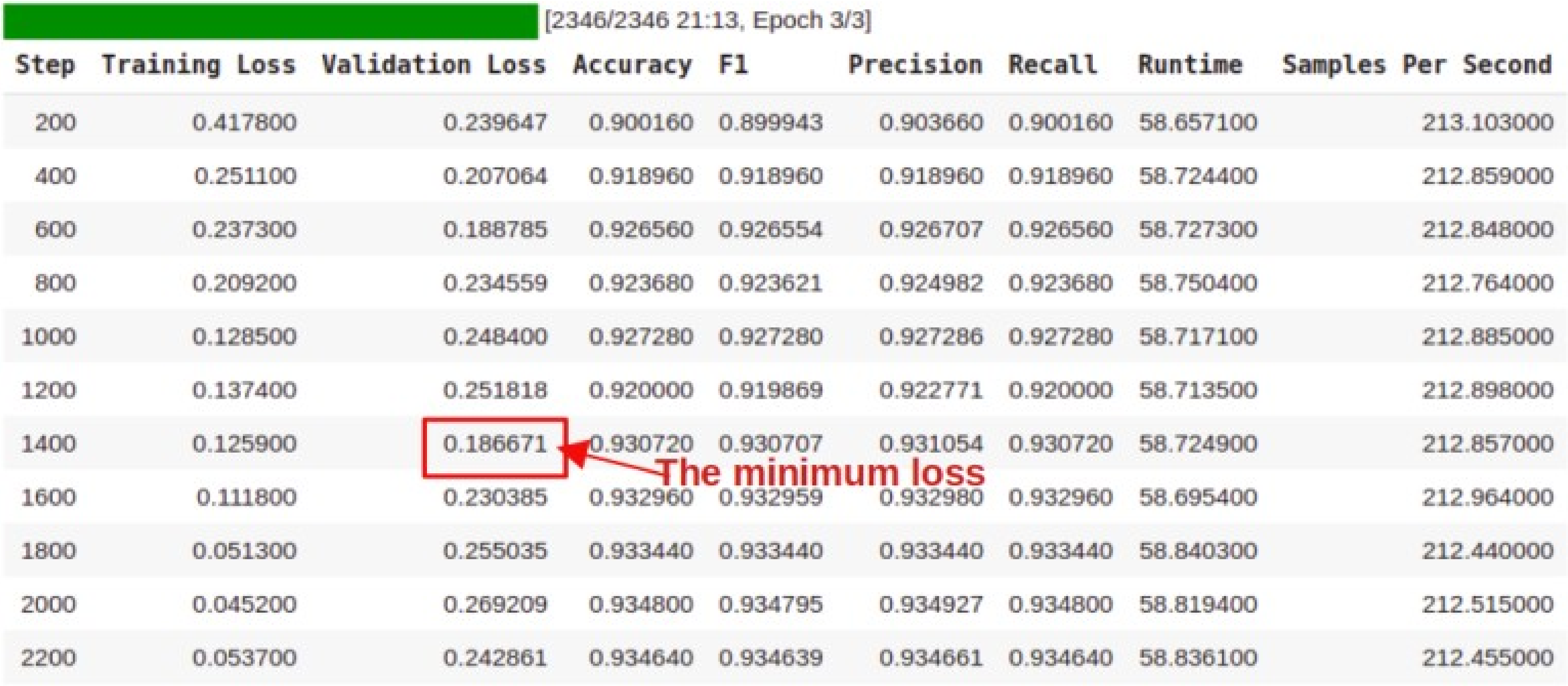
Figure 5.1 – Text classification scheme

What is happening inside



	attention_mask	input_ids	label	text
0	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 22953, 2213, 4381, 2152, 2003, 1037, 947...	1	Bromwell High is a cartoon comedy. It ran at t...
1	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 11573, 2791, 1006, 2030, 2160, 24913, 20...	1	Homelessness (or Houselessness as George Carli...
2	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 8235, 2058, 1011, 3772, 2011, 23920, 575...	1	Brilliant over-acting by Lesley Ann Warren. Be...
3	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 2023, 2003, 4089, 1996, 2087, 2104, 9250...	1	This is easily the most underrated film inn th...
4	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 2023, 2003, 2025, 1996, 5171, 11463, 837...	1	This is not the typical Mel Brooks film. It wa...
...	...	...	...	...
24995	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 2875, 1996, 2203, 1997, 1996, 3185, 1010...	0	Towards the end of the movie, I felt it was to...
24996	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 2023, 2003, 1996, 2785, 1997, 3185, 2008...	0	This is the kind of movie that my enemies cont...
24997	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 1045, 2387, 1005, 6934, 1005, 2197, 2305...	0	I saw 'Descent' last night at the Stockholm Fi...
24998	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 2070, 3152, 2008, 2017, 4060, 2039, 2005...	0	Some films that you pick up for a pound turn o...
24999	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...	[101, 2023, 2003, 2028, 1997, 1996, 12873, 435...	0	This is one of the dumbest films, I've ever se...

25000 rows × 4 columns

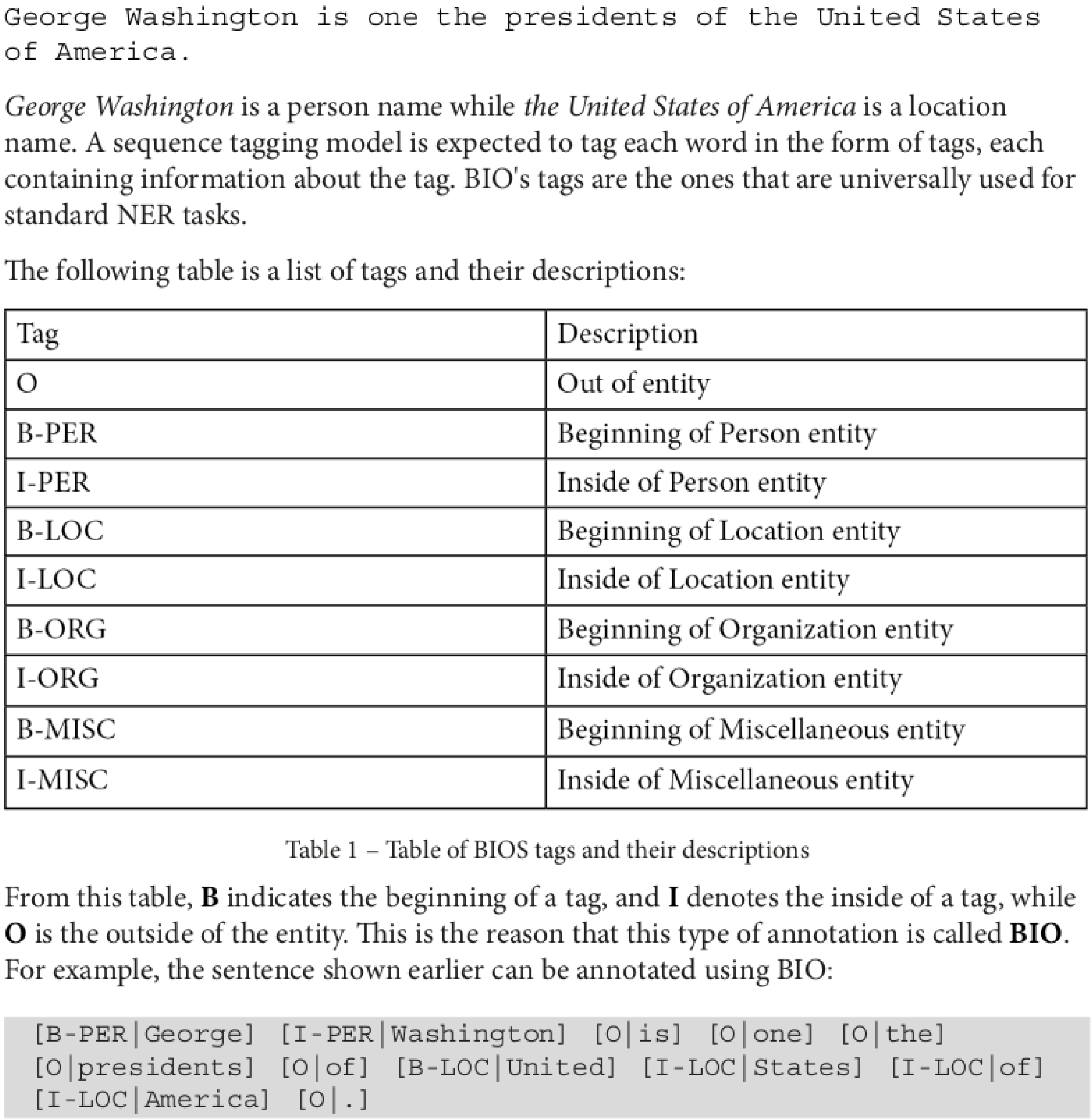


[2346/2346 21:13, Epoch 3/3]

Step	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall	Runtime	Samples Per Second
200	0.417800	0.239647	0.900160	0.899943	0.903660	0.900160	58.657100	213.103000
400	0.251100	0.207064	0.918960	0.918960	0.918960	0.918960	58.724400	212.859000
600	0.237300	0.188785	0.926560	0.926554	0.926707	0.926560	58.727300	212.848000
800	0.209200	0.234559	0.923680	0.923621	0.924982	0.923680	58.750400	212.764000
1000	0.128500	0.248400	0.927280	0.927280	0.927286	0.927280	58.717100	212.885000
1200	0.137400	0.251818	0.920000	0.919869	0.922771	0.920000	58.713500	212.898000
1400	0.125900	0.186671	0.930720	0.930707	0.931054	0.930720	58.724900	212.857000
1600	0.111800	0.230385	0.932960	0.932959	0.932980	0.932960	58.695400	212.964000
1800	0.051300	0.255035	0.933440	0.933440	0.933440	0.933440	58.840300	212.440000
2000	0.045200	0.269209	0.934800	0.934795	0.934927	0.934800	58.819400	212.515000
2200	0.053700	0.242861	0.934640	0.934639	0.934661	0.934640	58.836100	212.455000

The minimum loss

NER



```
George Washington is one the presidents of the United States
of America.
```

George Washington is a person name while *the United States of America* is a location name. A sequence tagging model is expected to tag each word in the form of tags, each containing information about the tag. BIO's tags are the ones that are universally used for standard NER tasks.

The following table is a list of tags and their descriptions:

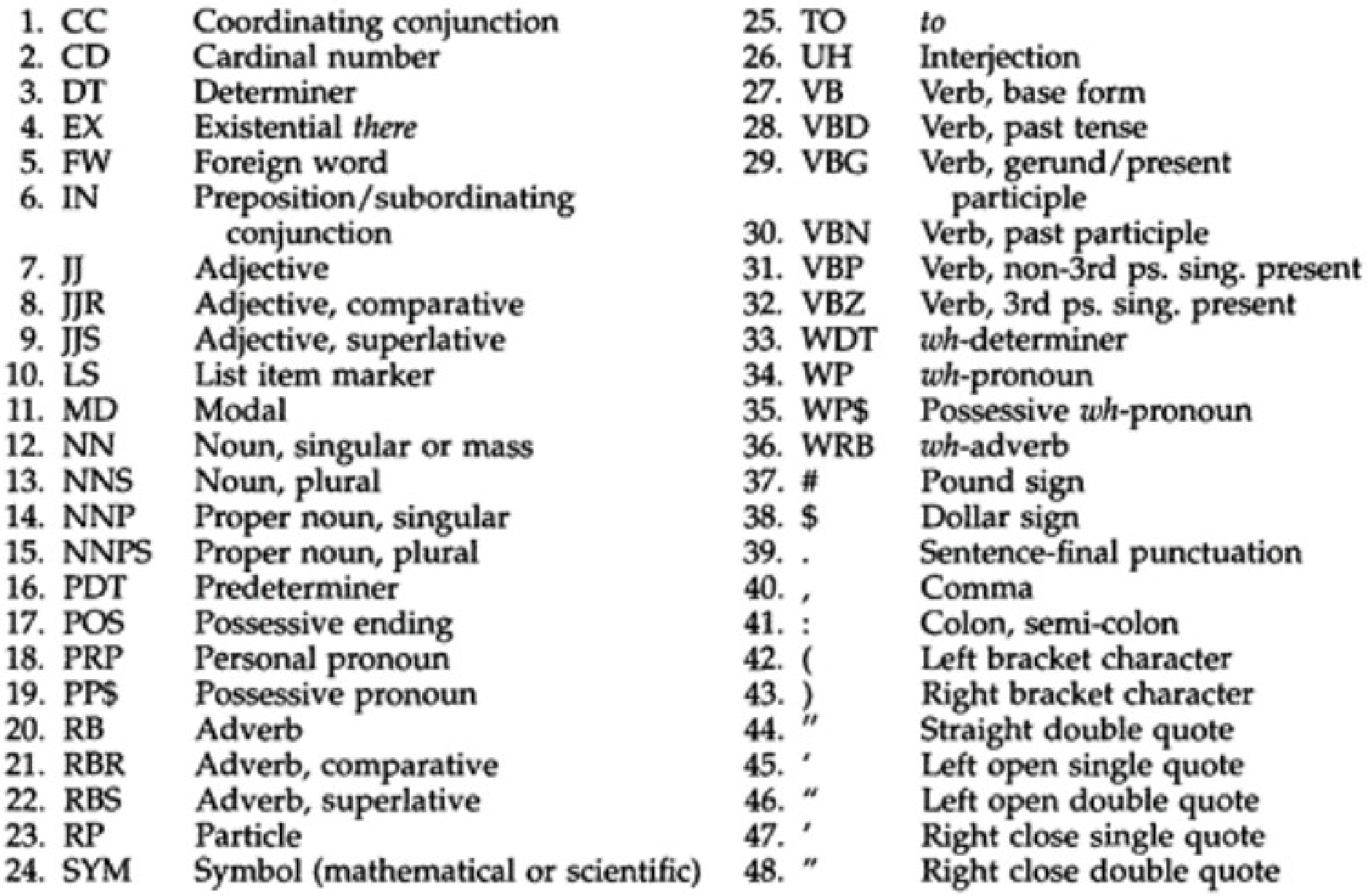
Tag	Description
O	Out of entity
B-PER	Beginning of Person entity
I-PER	Inside of Person entity
B-LOC	Beginning of Location entity
I-LOC	Inside of Location entity
B-ORG	Beginning of Organization entity
I-ORG	Inside of Organization entity
B-MISC	Beginning of Miscellaneous entity
I-MISC	Inside of Miscellaneous entity

Table 1 – Table of BIOS tags and their descriptions

From this table, **B** indicates the beginning of a tag, and **I** denotes the inside of a tag, while **O** is the outside of the entity. This is the reason that this type of annotation is called **BIO**. For example, the sentence shown earlier can be annotated using BIO:

```
[B-PER|George] [I-PER|Washington] [O|is] [O|one] [O|the]
[O|presidents] [O|of] [B-LOC|United] [I-LOC|States] [I-LOC|of]
[I-LOC|America] [O|.]
```

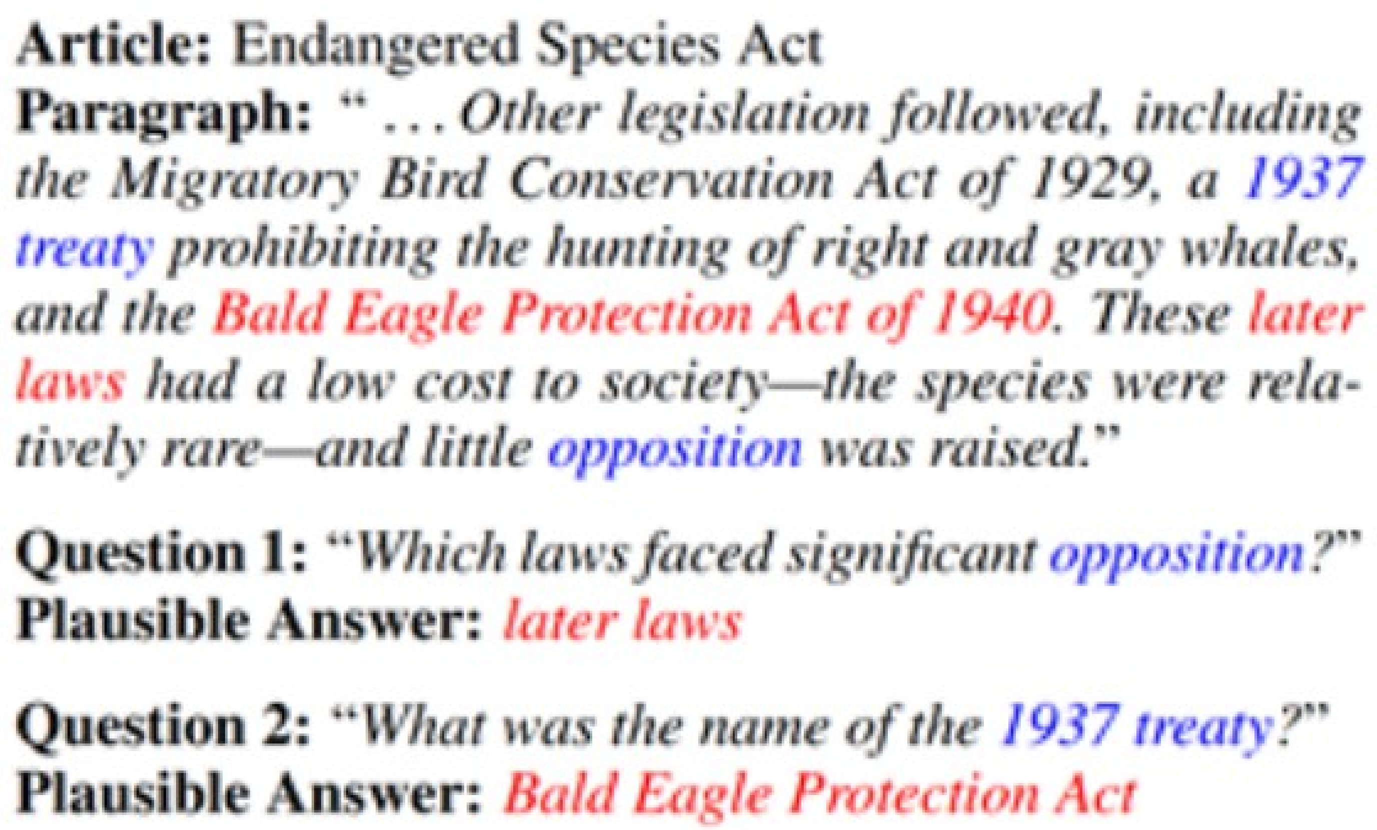
POS
Part-of-Speech



1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition/subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol (mathematical or scientific)
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund/present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd ps. sing. present
32.	VBZ	Verb, 3rd ps. sing. present
33.	WDT	*wh*-determiner
34.	WP	*wh*-pronoun
35.	WP$	Possessive *wh*-pronoun
36.	WRB	*wh*-adverb
37.	#	Pound sign
38.	$	Dollar sign
39.	.	Sentence-final punctuation
40.	,	Comma
41.	:	Colon, semi-colon
42.	(	Left bracket character
43.	)	Right bracket character
44.	"	Straight double quote
45.	‘	Left open single quote
46.	“	Left open double quote
47.	’	Right close single quote
48.	”	Right close double quote

Figure 6.2 – Penn Treebank POS tags

QA



Article: Endangered Species Act

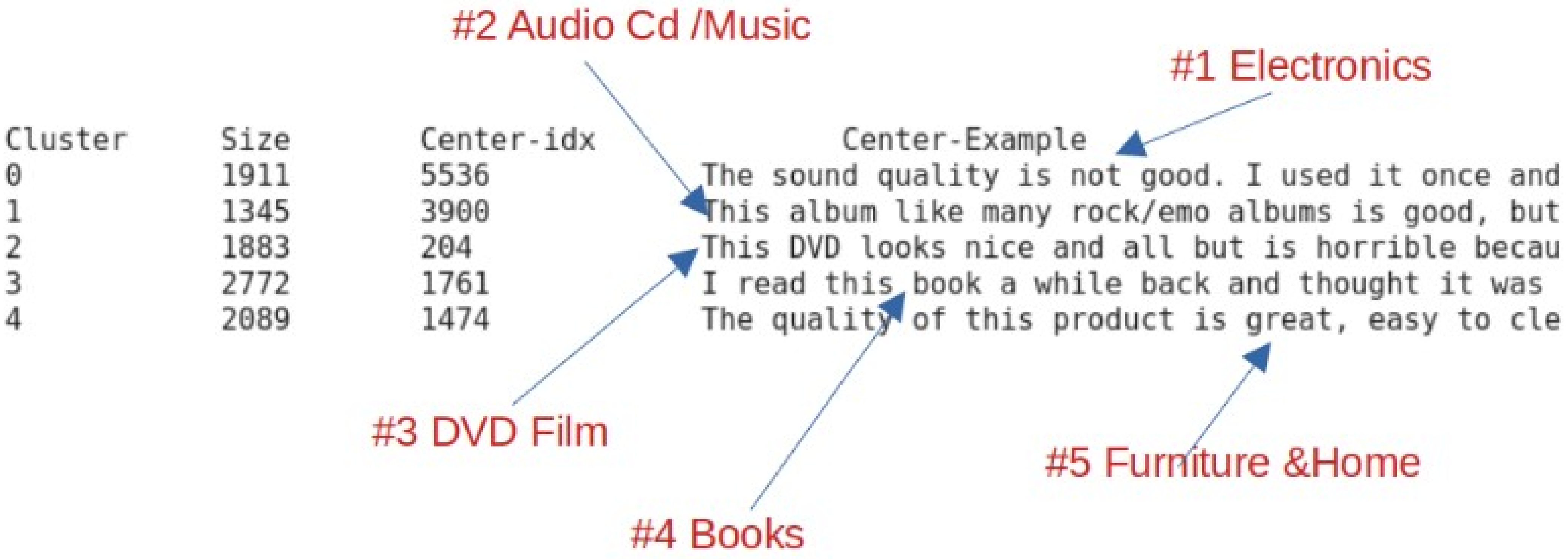
Paragraph: " *…Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*"

Question 1: "*Which laws faced significant opposition?*"
Plausible Answer: *later laws*

Question 2: "*What was the name of the 1937 treaty?*"
Plausible Answer: *Bald Eagle Protection Act*

Figure 6.3 – SQUAD dataset example

Clustering



```
Cluster     Size      Center-idx          Center-Example
0           1911      5536          The sound quality is not good. I used it once and
1           1345      3900          This album like many rock/emo albums is good, but
2           1883      204           This DVD looks nice and all but is horrible becau
3           2772      1761          I read this book a while back and thought it was
4           2089      1474          The quality of this product is great, easy to cle
```

#2 Audio Cd /Music

#1 Electronics

#3 DVD Film

#5 Furniture &Home

#4 Books

Figure 7.7 – Centroids of the cluster

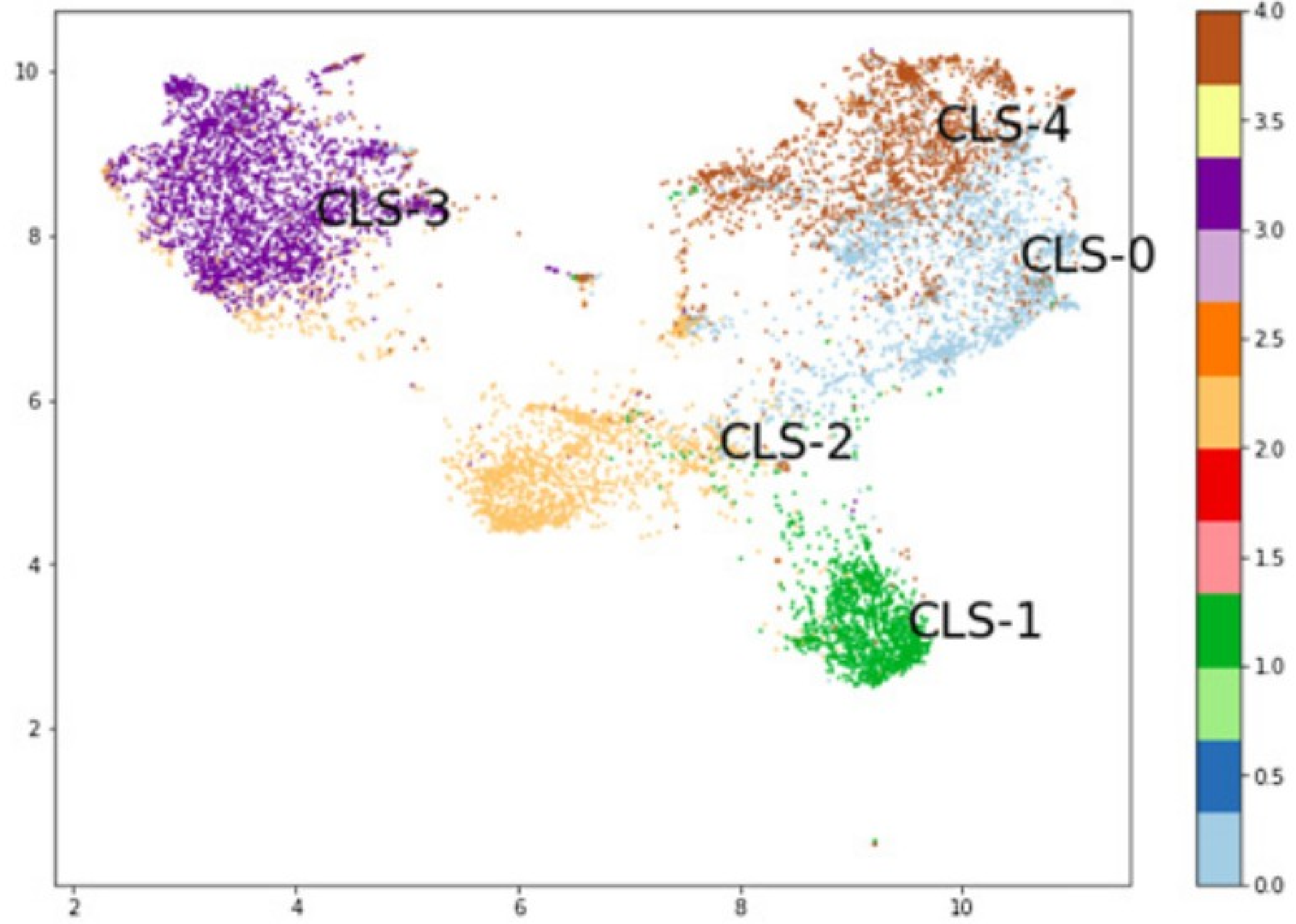
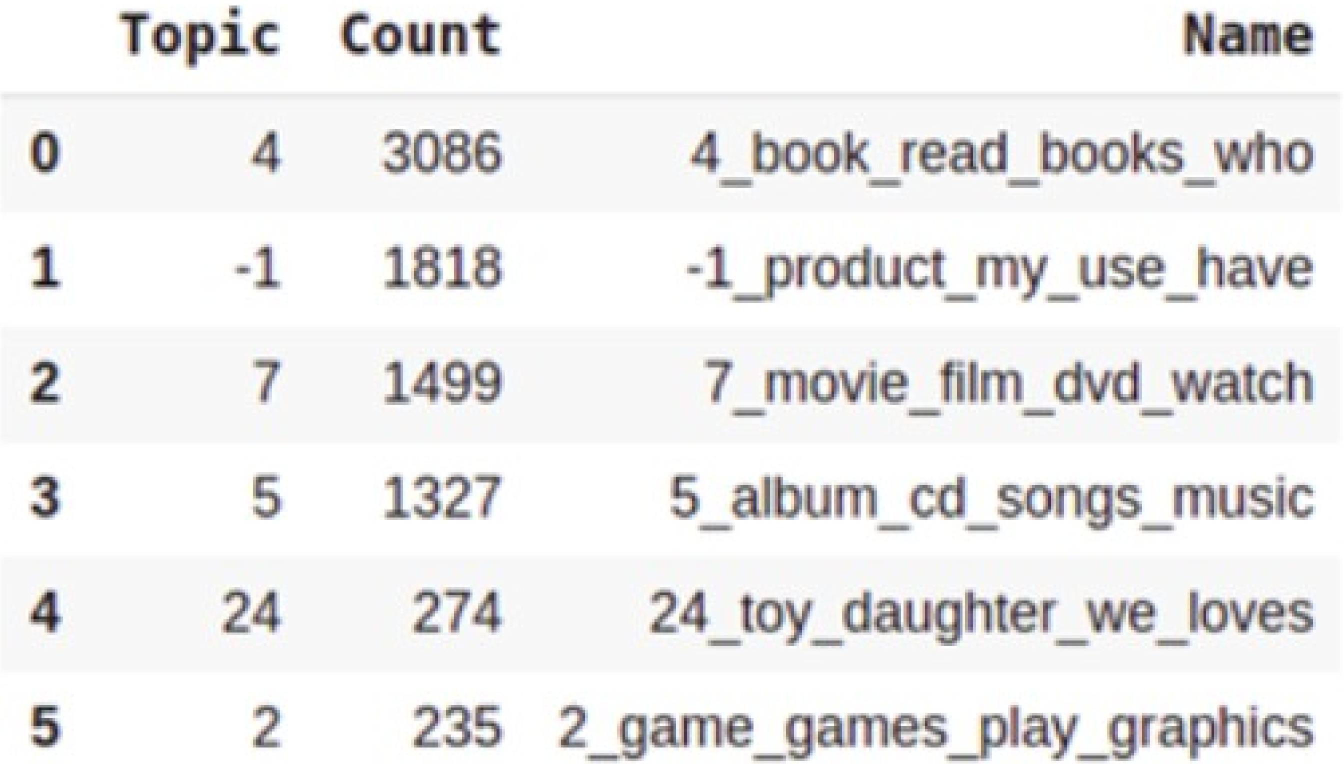


Figure 7.8 – Cluster points visualization

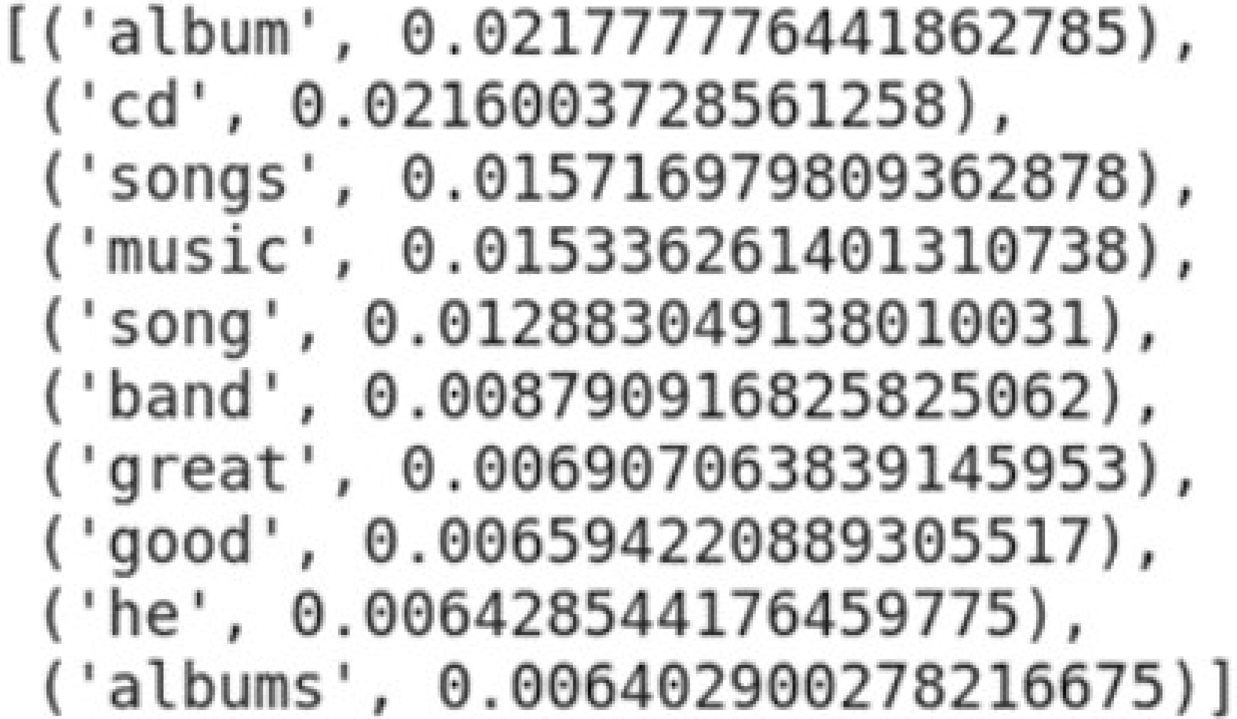


	Topic	Count	Name
0	4	3086	4_book_read_books_who
1	-1	1818	-1_product_my_use_have
2	7	1499	7_movie_film_dvd_watch
3	5	1327	5_album_cd_songs_music
4	24	274	24_toy_daughter_we_loves
5	2	235	2_game_games_play_graphics

Figure 7.9 – BERTopic results

```
topic_model.get_topic(5)
```

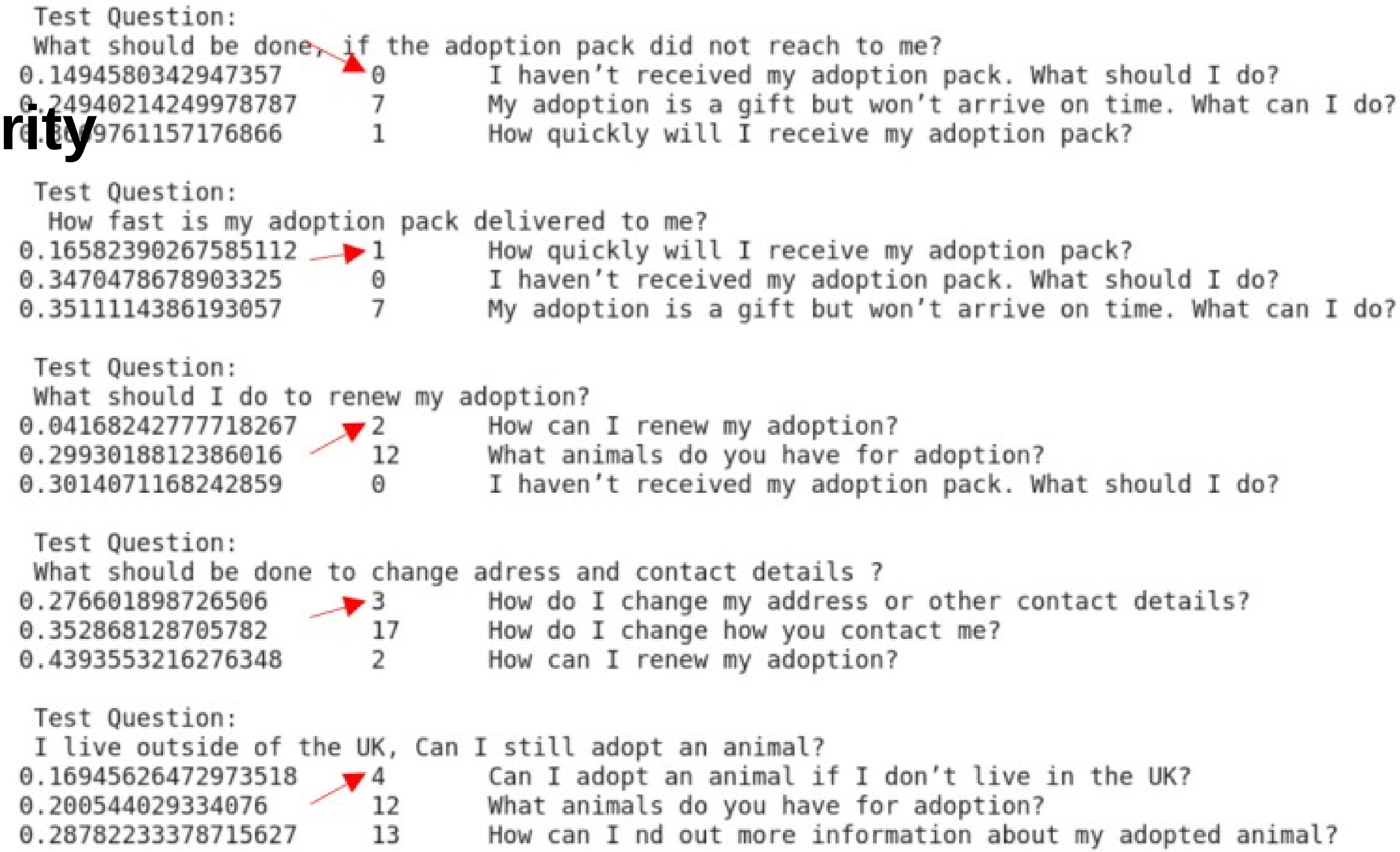
The output is as follows:



```
[('album', 0.021777776441862785),
 ('cd', 0.021600372856125 8),
 ('songs', 0.015716979809362878),
 ('music', 0.015336261401310738),
 ('song', 0.012883049138010031),
 ('band', 0.008790916825825062),
 ('great', 0.006907063839145953),
 ('good', 0.006594220889305517),
 ('he', 0.006428544176459775),
 ('albums', 0.006402900278216675)]
```

Figure 7.10 – The fifth topic words of the topic model

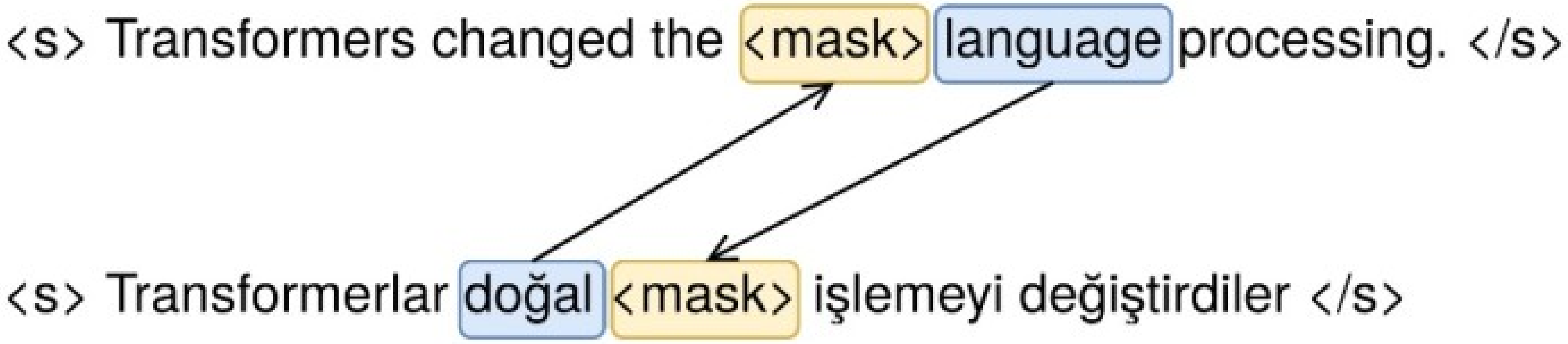
Similarity



```
 Test Question:
 What should be done, if the adoption pack did not reach to me?
0.1494580342947357      0       I haven’t received my adoption pack. What should I do?
0.24940214249978787     7       My adoption is a gift but won’t arrive on time. What can I do?
0.30[illegible]9761157176866      1       How quickly will I receive my adoption pack?

 Test Question:
  How fast is my adoption pack delivered to me?
0.16582390267585112     1       How quickly will I receive my adoption pack?
0.3470478678903325      0       I haven’t received my adoption pack. What should I do?
0.3511114386193057      7       My adoption is a gift but won’t arrive on time. What can I do?

 Test Question:
 What should I do to renew my adoption?
0.04168242777718267     2       How can I renew my adoption?
0.2993018812386016      12      What animals do you have for adoption?
0.3014071168242859      0       I haven’t received my adoption pack. What should I do?

 Test Question:
 What should be done to change adress and contact details ?
0.276601898726506       3       How do I change my address or other contact details?
0.352868128705782       17      How do I change how you contact me?
0.4393553216276348      2       How can I renew my adoption?

 Test Question:
 I live outside of the UK, Can I still adopt an animal?
0.16945626472973518     4       Can I adopt an animal if I don’t live in the UK?
0.200544029334076       12      What animals do you have for adoption?
0.28782233378715627     13      How can I nd out more information about my adopted animal?
```

Figure 7.11 – Question-question similarity

Cross Lingual Models in NLP



<s> Transformers changed the <mask> language processing. </s>

<s> Transformerlar doğal <mask> işlemeyi değiştirdiler </s>

Figure 9.1 – Cross-lingual relation example between Turkish and English

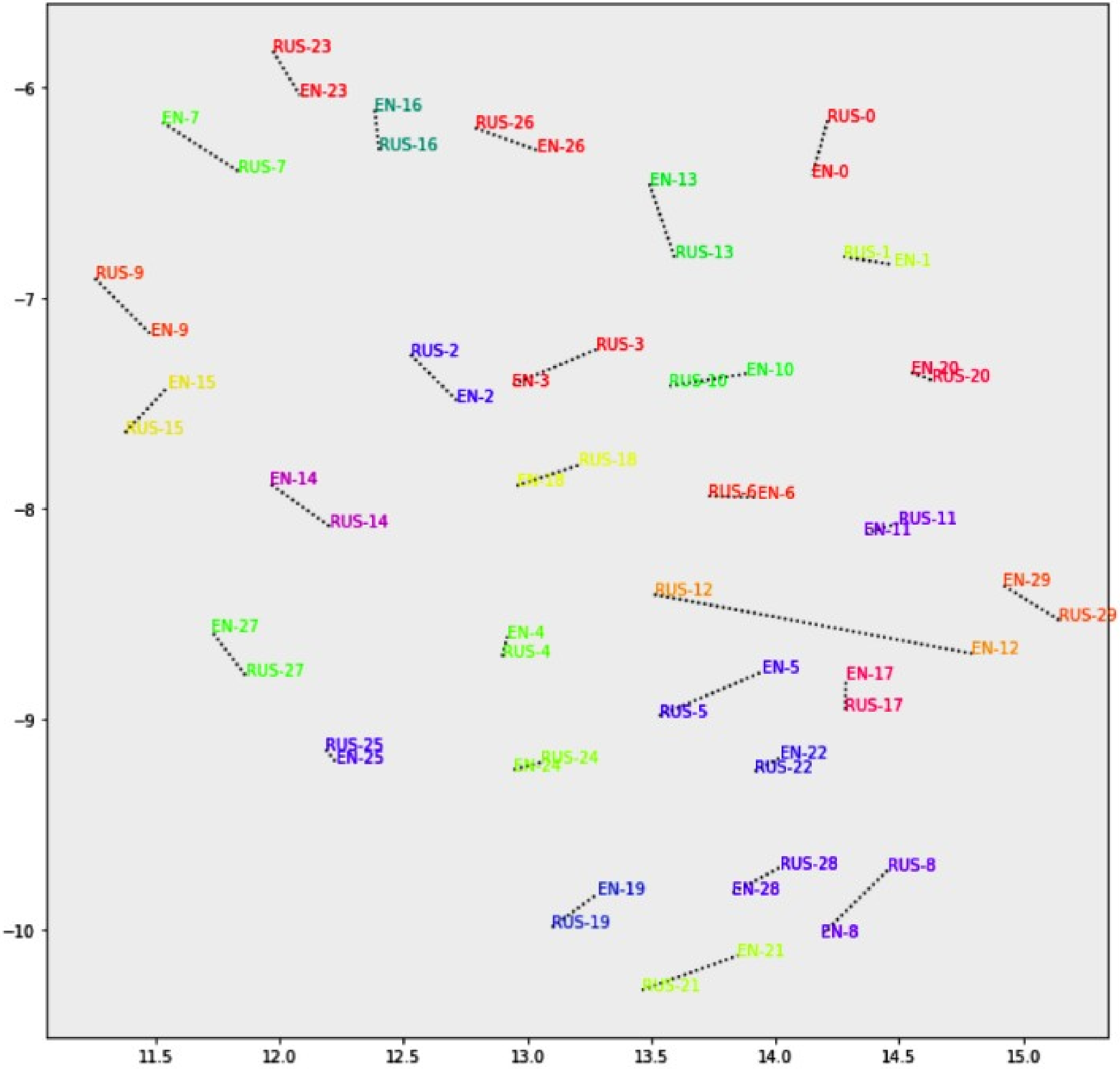


Figure 9.9 – Russian-English sentence similarity visualization

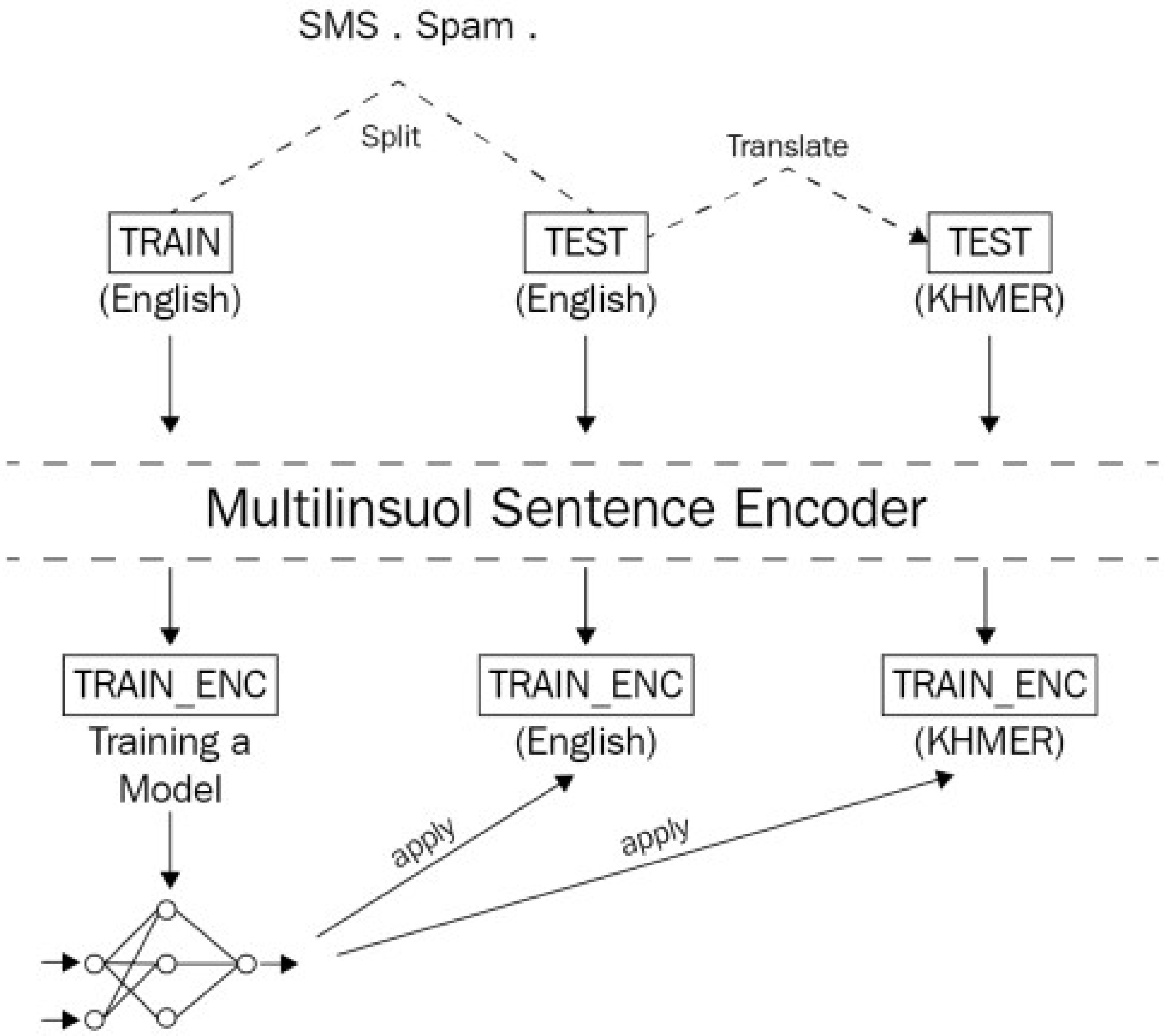


Figure 9.11 – Flow of cross-lingual classification

Interpretability

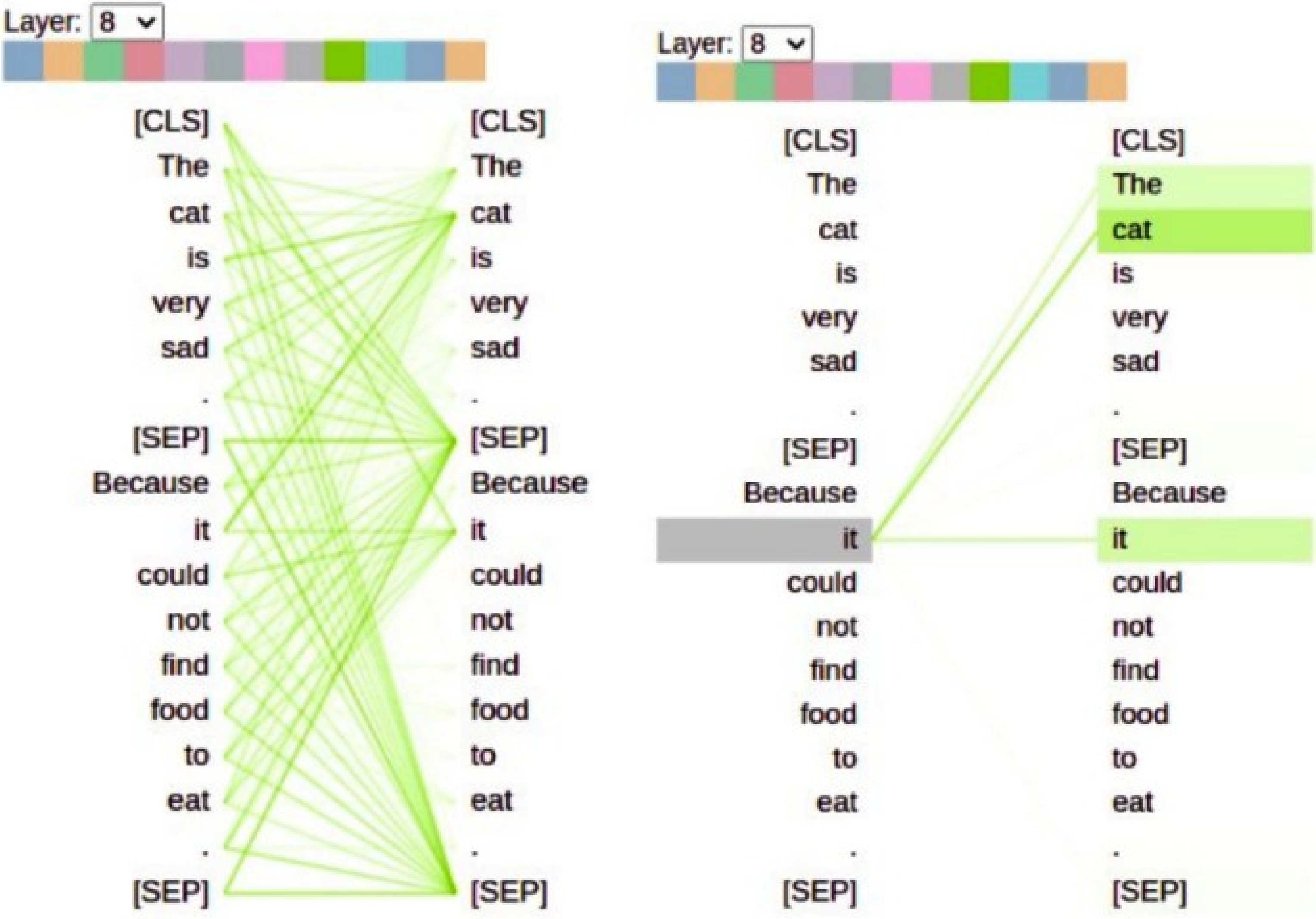
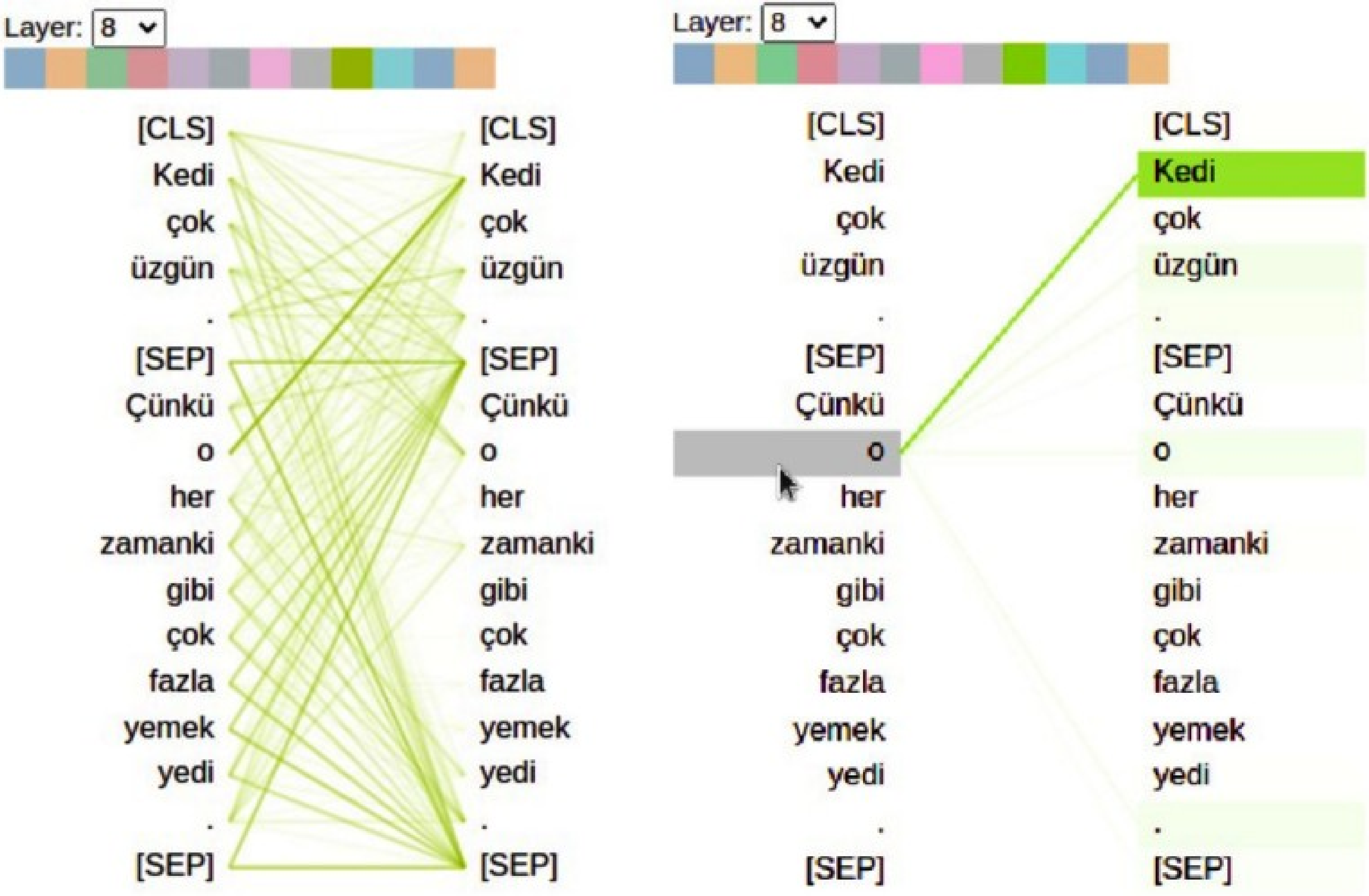


Figure 11.9 – Head-view output of BertViz



Layer: 8

Layer: 8

Figure 11.10 – Coreference pattern in the Turkish language model

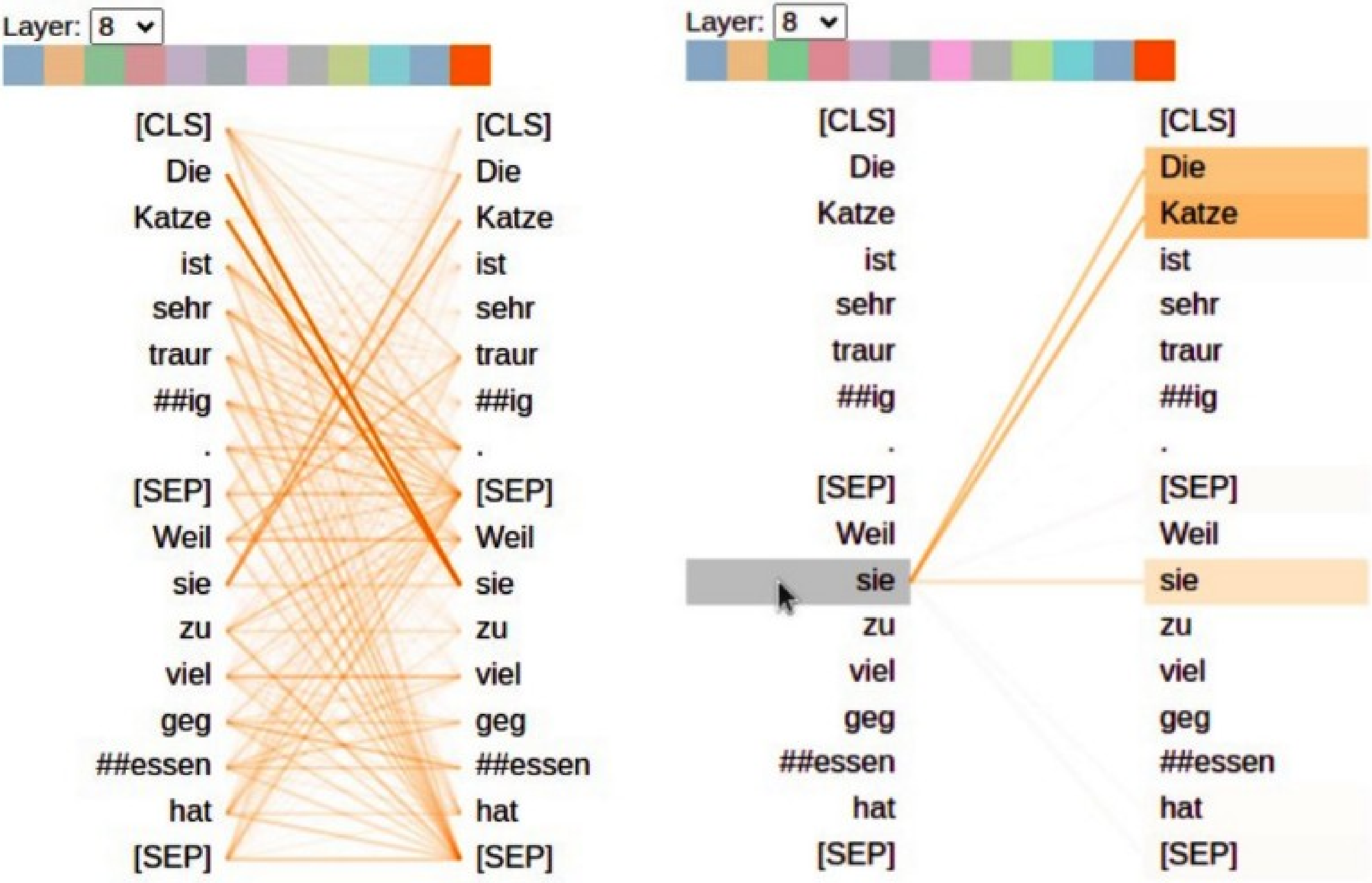
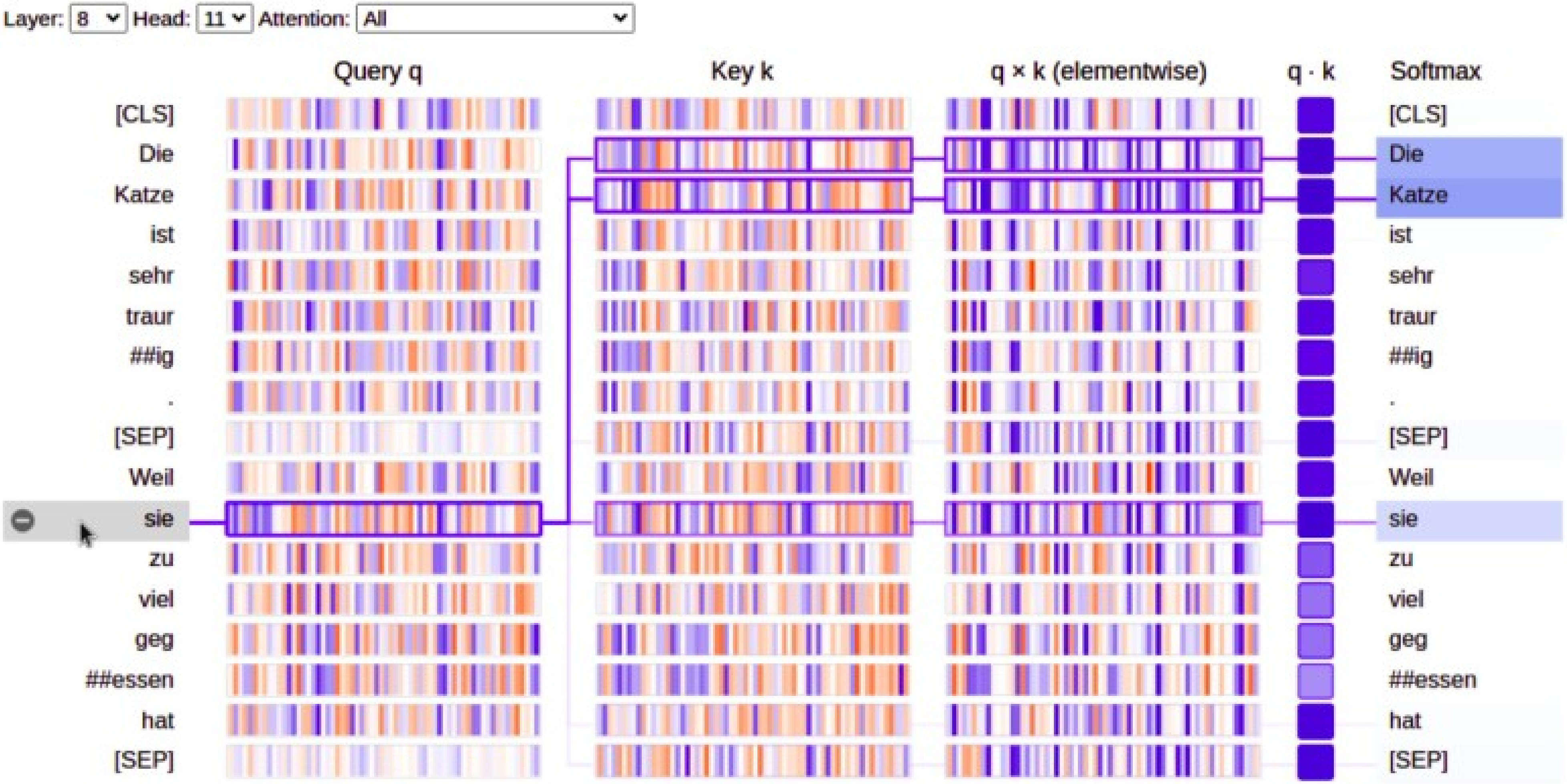


Figure 11.11 – Coreference relation pattern in the German language model

Figure 11.12 – The model view of the German language model



Layer: 8 Head: 11 Attention: All

Figure 11.14 – Neuron view of the coreference relation pattern (head <8,11>)