Named Entity Recognition

Omar Bayoumi 1747042

What is NER?

- **The Named Entity Recognition (NER) problem belongs to the category of tagging problems.**

- **I used the IOB scheme for NER:**
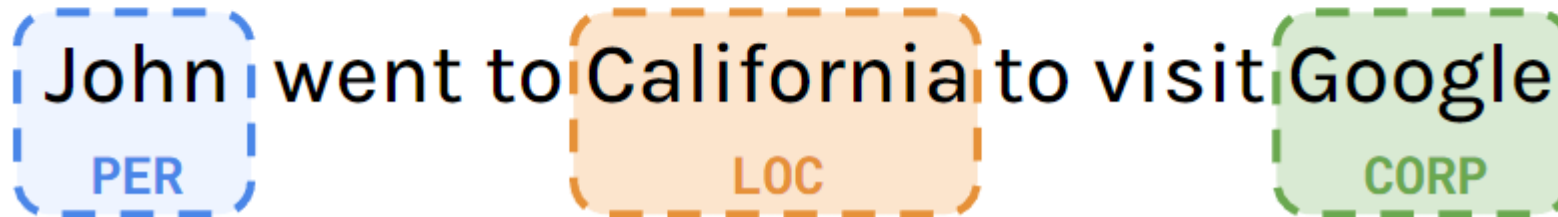 - **I: inside**
  - **O: outside**
  - **B: beginning**

Categories

PER: Person   **CORP**: Corporation

LOC: Location   **PROD**: Product

GRP: Group   **CW**: Creative Work

John went to California to visit Google
PER   LOC   CORP

Categories: IOB

My	name	is	Robin	Hood

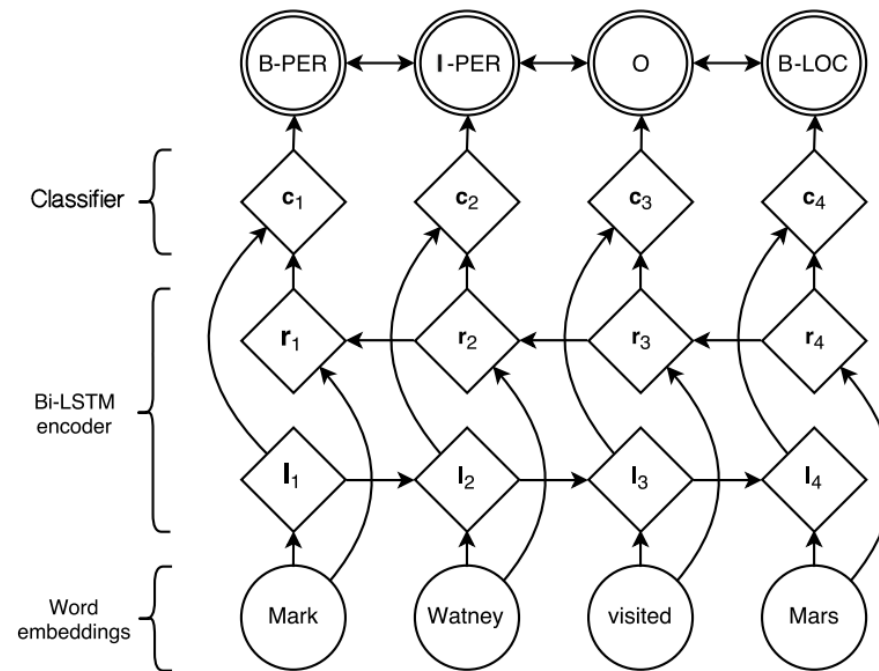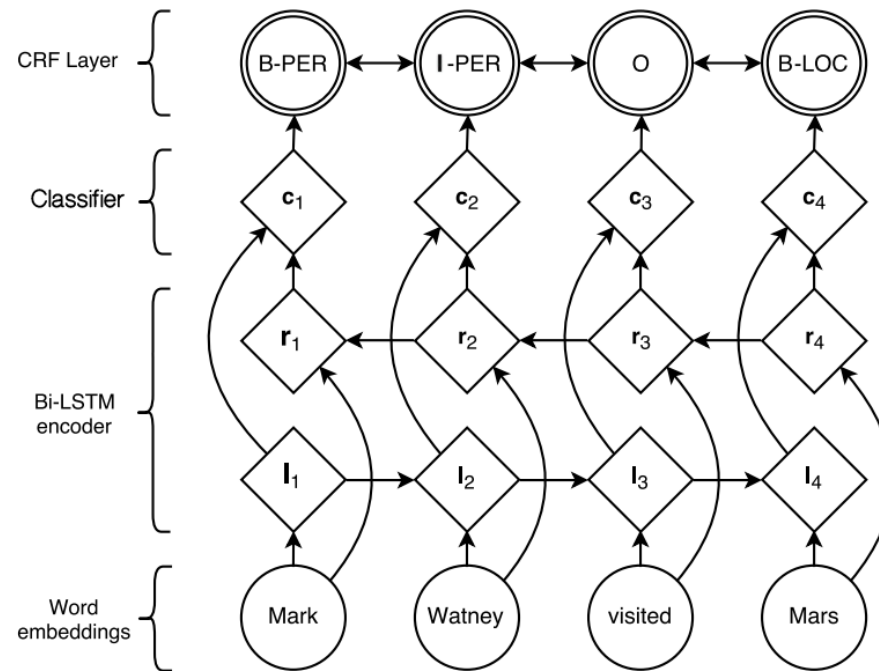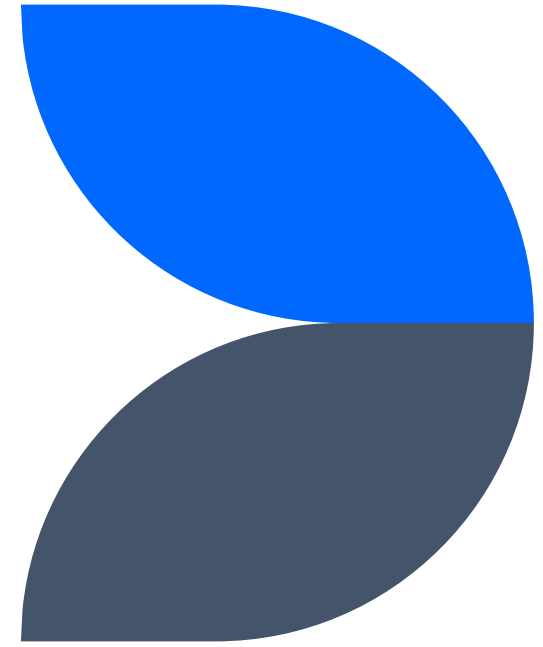| O	O	O	**B-Per**	**I-Per**
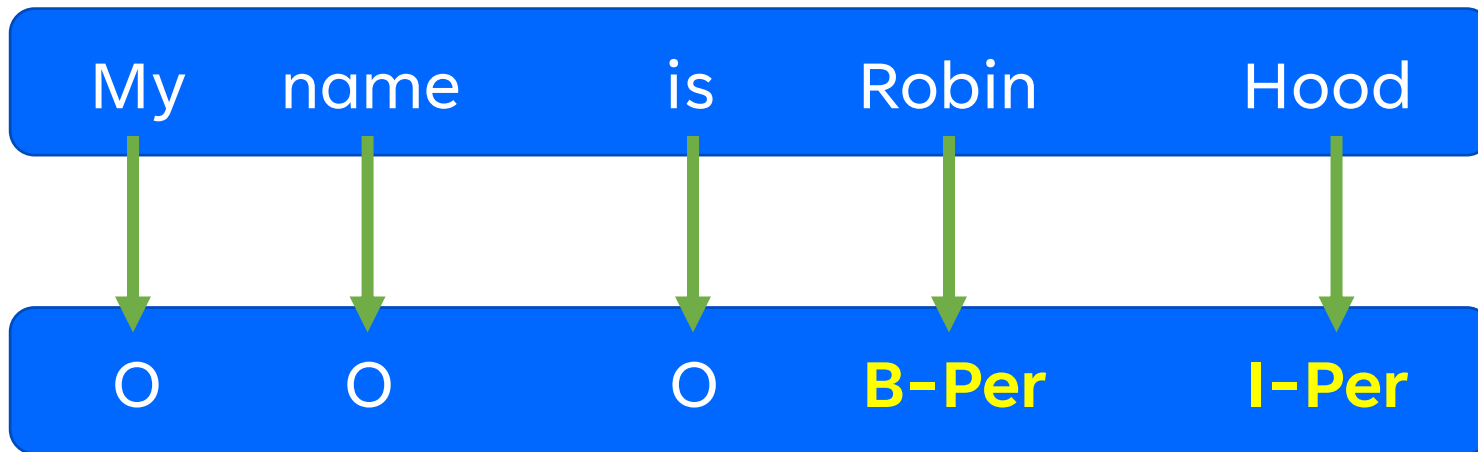
Architectures

Model 1: Base

Model 2: Base + CRF

Conditional Random Field (CRF)

A CRF is an output layer that encourages the neural network to produce a valid sequence of output labels, through **log-likelihood maximization** of the correct label sequence
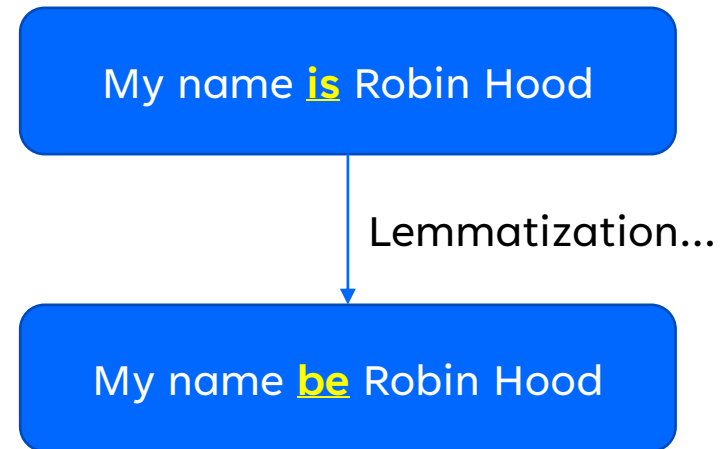
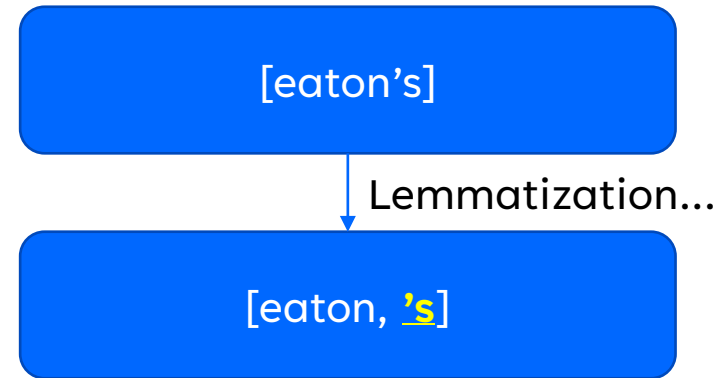Preprocess

Dataset Augmentation

Spacy lemmatizer

- Component for assigning base forms to words using rules based on part-of-speech tags, or lookup tables.

My name **is** Robin Hood

Lemmatization...

My name **be** Robin Hood

Dataset Augmentation
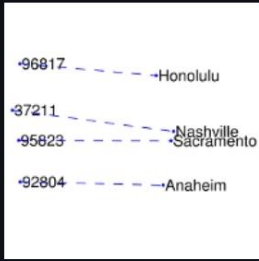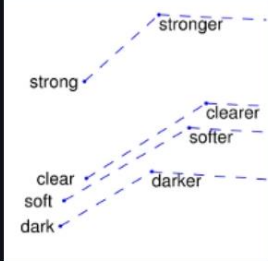
Spacy lemmatizer

- Component for assigning base forms to words using rules based on part-of-speech tags, or lookup tables.

- Some word is splitted

[eaton's]

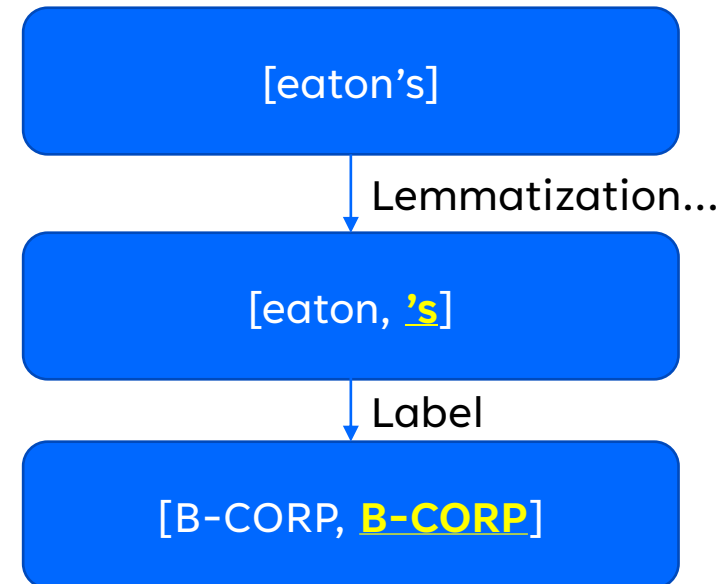Lemmatization...

[eaton, **'s**]

Dataset Augmentation

Spacy lemmatizer

- Component for assigning base forms to words using rules based on part-of-speech tags, or lookup tables.

- Some word is splitted

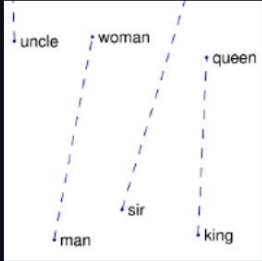- The label is B-CLASS for each split

[eaton's]

Lemmatization...

[eaton, **'s**]

Label

[B-CORP, **B-CORP**]

GloVe

- Different implementations

GloVe

- Different implementations
- I used the one based on twitter
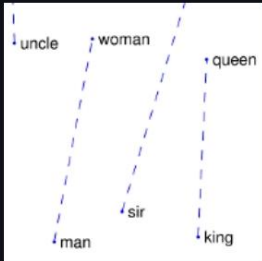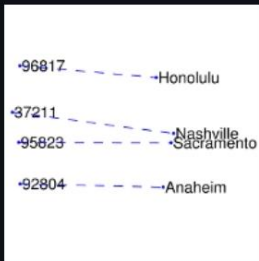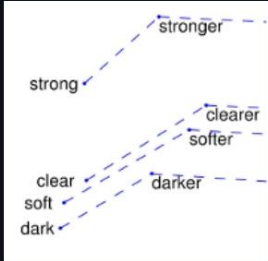
nearest neighbors of *frog*	Litoria	Leptodactylidae	Rana	Eleutherodactylus
Pictures				

Comparisons	man -> woman	city -> zip	comparative -> superlative
GloVe Geometry			

GloVe

- Different implementations
- I used the one based on twitter
- Embedding size: 200

Vocabulary

- I calculated the frequencies

Token index in
the vocabulary

39 1592 10 2548 5

I saw a cat .

Embedding dimension

39
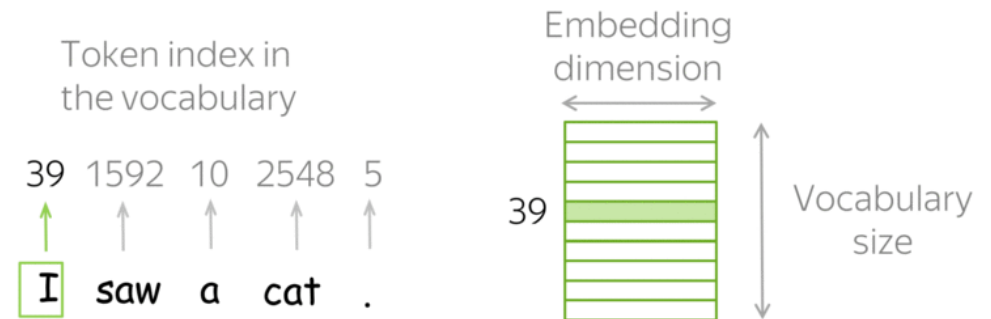
Vocabulary size

Vocabulary

- I calculated the frequencies
- I assigned an id to each word
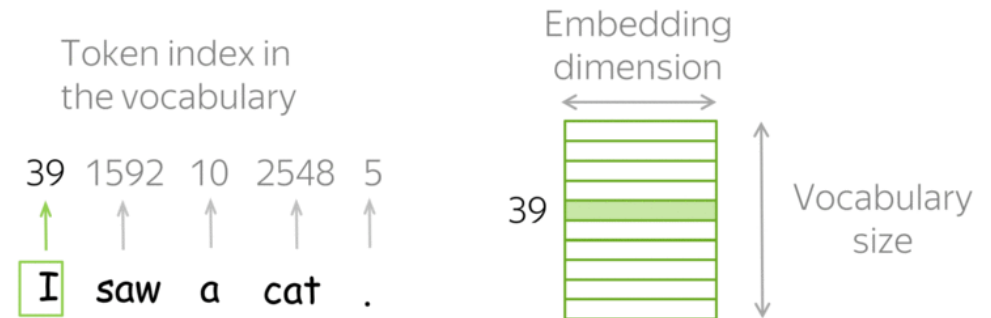
Vocabulary

- I calculated the frequencies

- I assigned an id to each word

- If I use GloVe, in case the word doesn't exist in the vocabulary I assign <unk> id.

- If I don't use GloVe, if the word appear less than a min frequency, I assign <unk> id
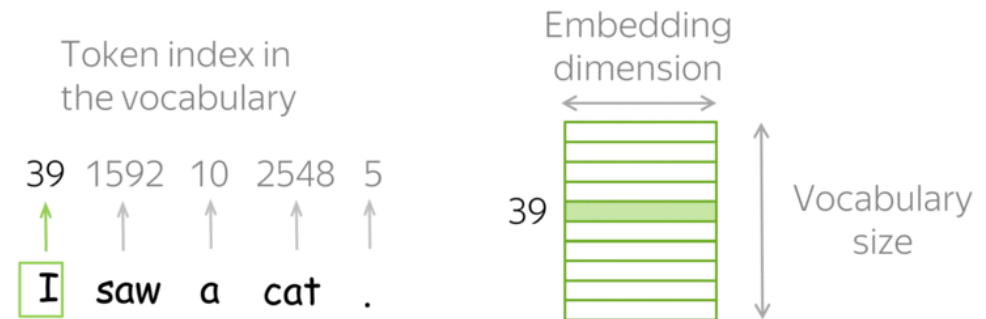
Input of the model
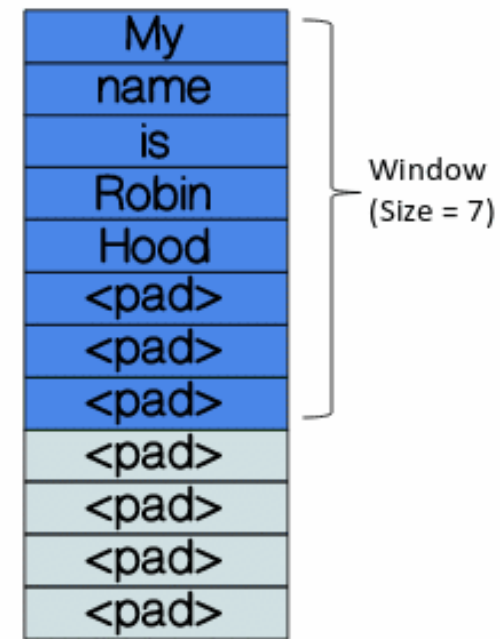
Windows

- Finding the length of the input
  - Small number
  - Big number

Input of the model

Windows

- Finding the length of the input
  - Small number
  - Big number
- The fixed size window moves to the sentence and each frame is an input
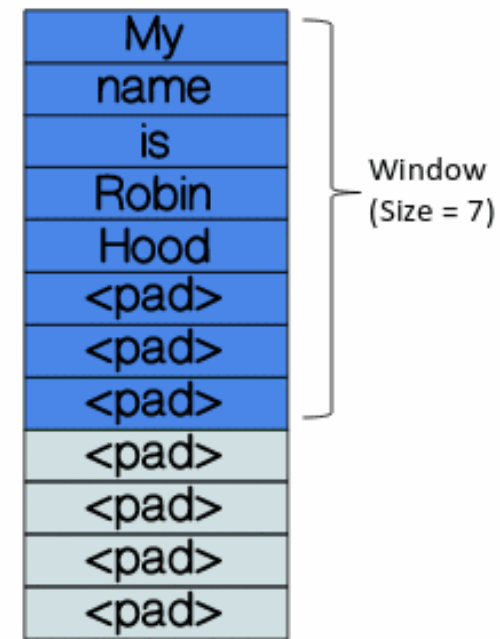
Input of the model

Windows

- Finding the length of the input
  - Small number
  - Big number

- The fixed size window moves to the sentence and each frame is an input.

- Padding is added at the end for small sentences
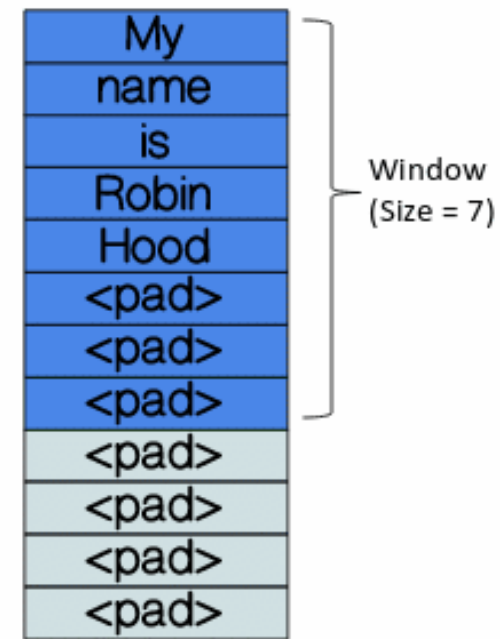
Training
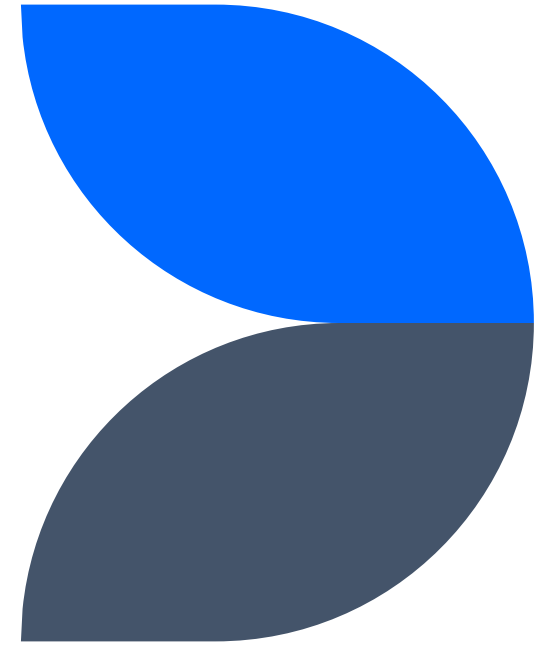
Hyperparameters
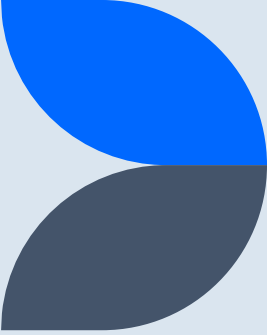
Hyperparameter	Base model	Base model + CRF
Min frequency	3	0
Glove Embedding	No	Yes
Window size/shift	100	100
LSTM hidden dimention	128	128
Dropout	0.5	0.5
Learning rate	0.001	0.001

Loss function

**Model 1:
Cross Entropy loss
Best = 0.252**
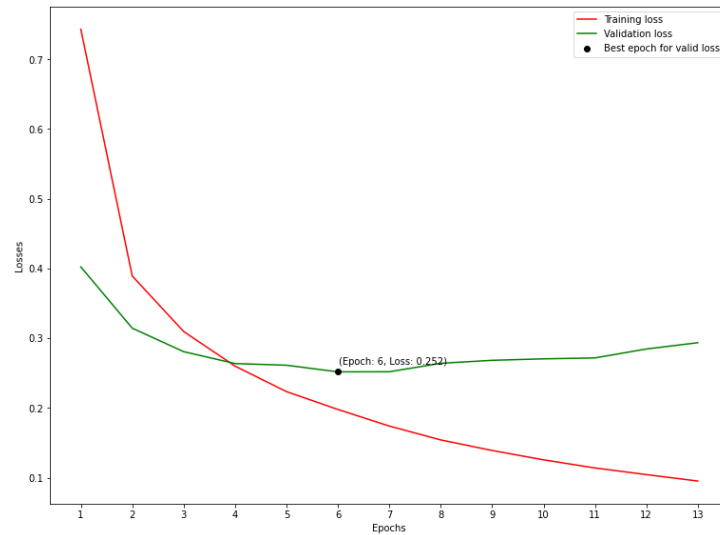
**Model 2:
Negative Loglikelihood of the CRF layer
Best = 0.178**
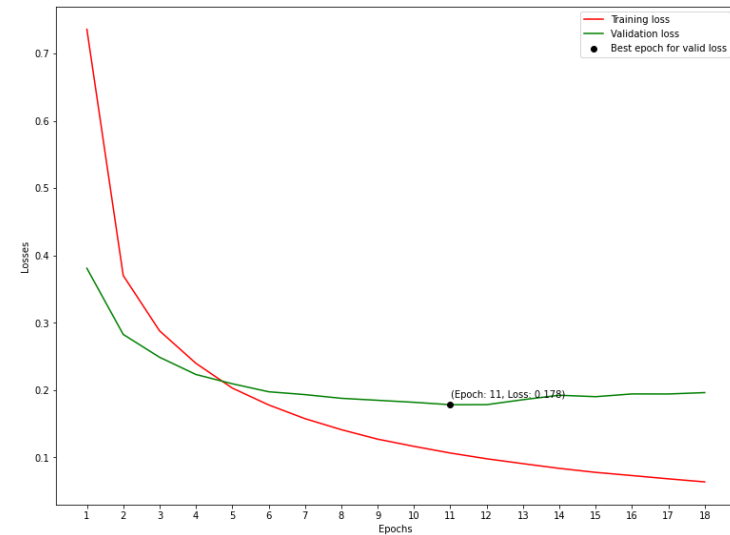
Results

Scores

Model 1

```
Micro Precision: 0.9279272213944004
Macro Precision: 0.7640977764243746
F1 score: 0.7574234829133377
Seqeval Accuracy: 0.9279272213944004
Seqeval F1 score: 0.6611607177439702
Per class Precision:
        O 0.9659632402995235
        I-PER 0.8875
        B-PER 0.8707482993197279
        I-GRP 0.842391304347826
        I-LOC 0.8289473684210527
        B-LOC 0.8073770491803278
        I-CORP 0.78
        B-GRP 0.7421052631578947
        B-CORP 0.7107438016528925
        I-PROD 0.6875
        I-CW 0.6653061224489796
        B-PROD 0.6282051282051282
        B-CW 0.5164835164835165
        <pad> 0.0
```

Model 2

```
Micro Precision: 0.9309073798133479
Macro Precision: 0.7809130948896775
F1 score: 0.760941086542142
Seqeval Accuracy: 0.9309073798133479
Seqeval F1 score: 0.6729829048219201
Per class Precision:
        O 0.9669092673459486
        I-GRP 0.8645533141210374
        B-PER 0.8407643312101911
        I-PER 0.8379888268156425
        I-CORP 0.8130841121495327
        I-PROD 0.7971014492753623
        B-GRP 0.7701149425287356
        I-LOC 0.7654320987654321
        B-LOC 0.7326007326007326
        B-CORP 0.7272727272727273
        B-PROD 0.7272727272727273
        I-CW 0.7075098814229249
        B-CW 0.6012658227848101
        <pad> 0.0
```

Confusion Matrices

Model 1
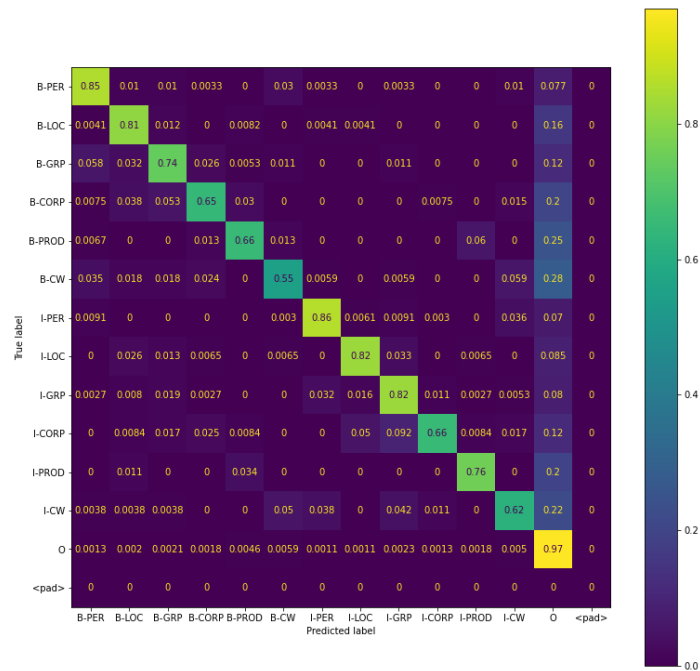
Model 2
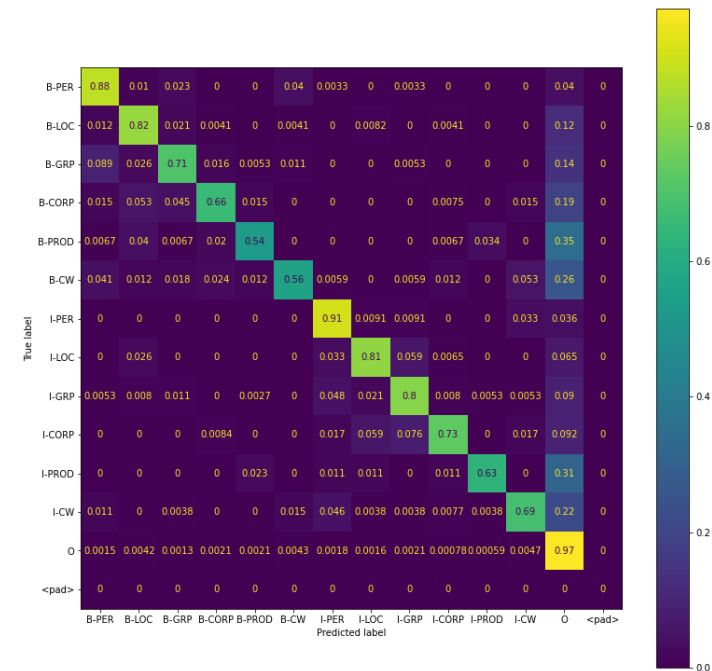
Semantic Role Labeling

Omar Bayoumi 1747042

Who did What to Whom, how, Where and When?

The **cat** **ate** the **fish**

Who?   Did what?   To whom?

Algorithm

Step 1

Predicate Identification

Step 3

Argument Identification

Predicate Disambiguation

Step 2

Argument Classification

Step 4
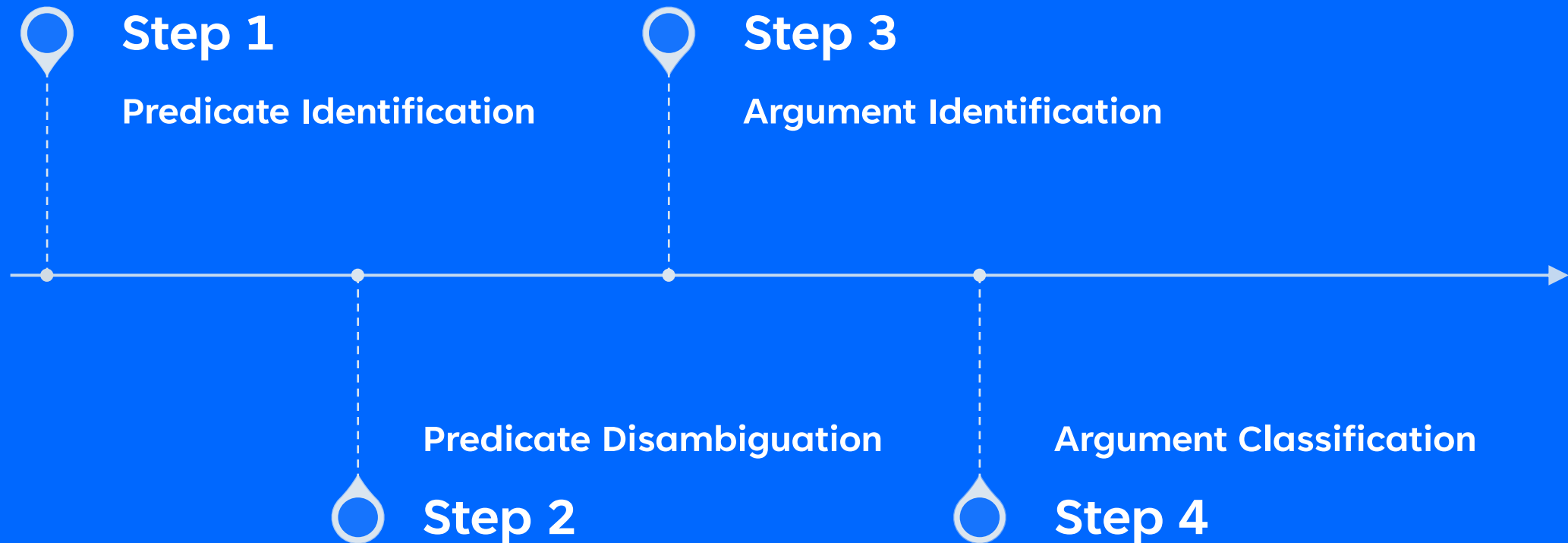
Architectures

Step 3-4: Base model

Transformer input

Predicate input

Transformer

Predicate Embedding

Concatenation

B-LSTM

Classifier

SEMANTIC ROLE LABELING

Step 3-4:
Base model + POS

Transformer input   Predicate input   POS input

```
┌──────────────┐    ┌──────────────┐    ┌──────────────┐
│ Transformer  │    │  Predicate   │    │     POS      │
│              │    │  Embedding   │    │  Embedding   │
└──────────────┘    └──────────────┘    └──────────────┘
                           │
                    ┌──────────────┐
            ──────▶ │Concatenation │ ◀──────
                    └──────────────┘
                           │
┌──────────────┐    ┌──────────────┐
│  Classifier  │ ◀──│   B-LSTM     │
└──────────────┘    └──────────────┘
```
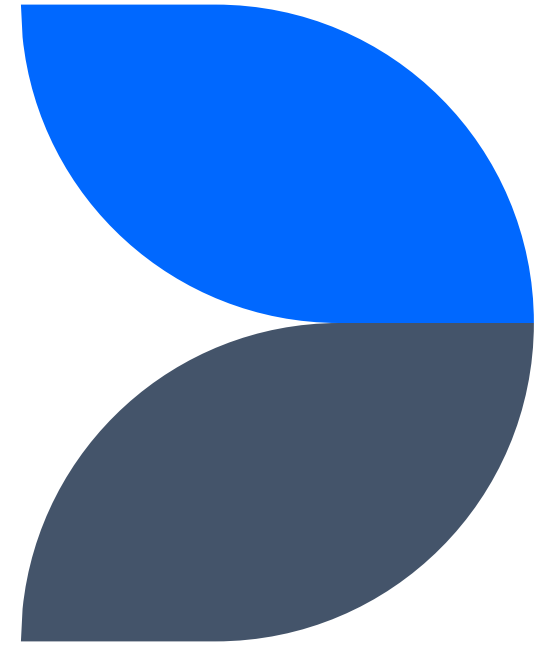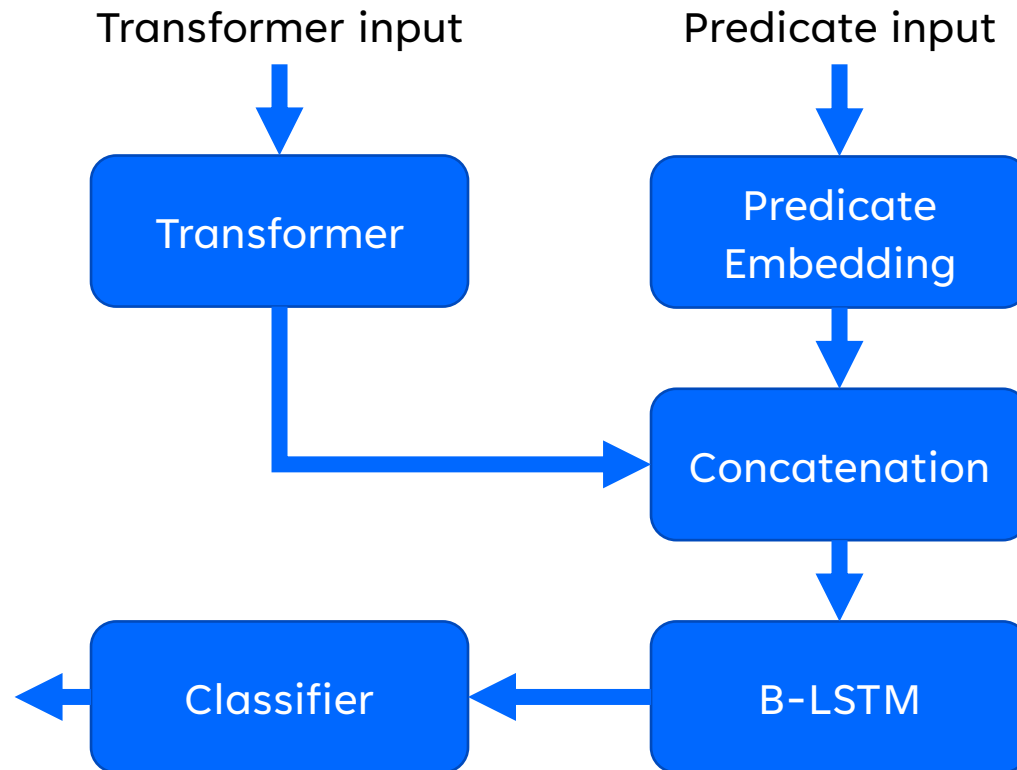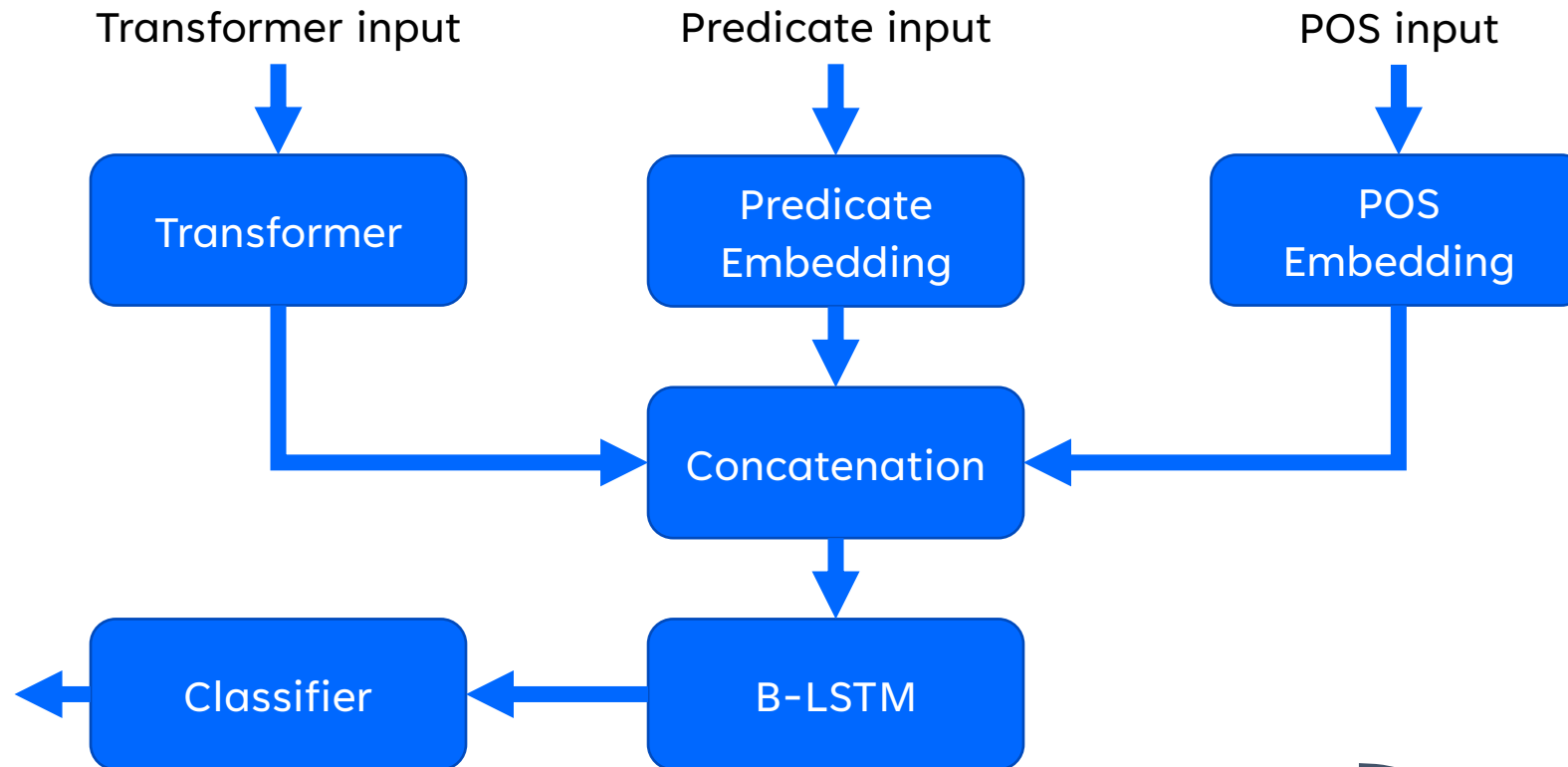
Step 1:
Base model – Predicate Embedding

Transformer input

Transformer

Concatenation

Classifier ← B-LSTM

Step 2: Base model

Transformer input

Predicate input

```
Transformer
```

```
Predicate
Embedding
```

```
Concatenation
```

```
Classifier
```

```
B-LSTM
```
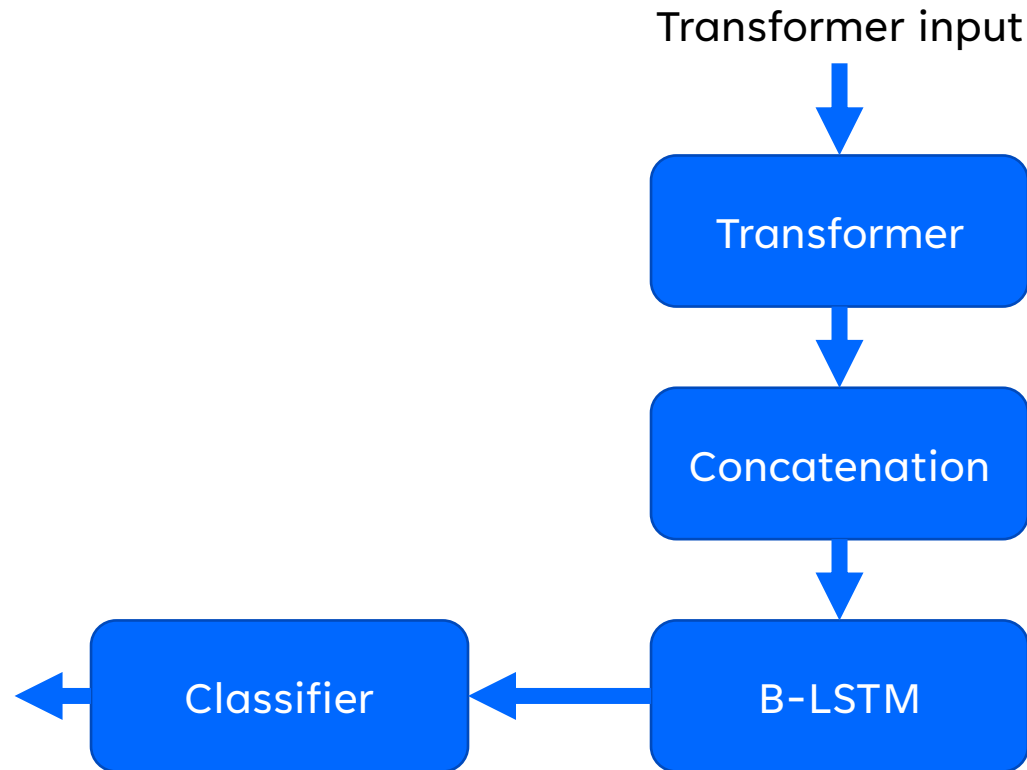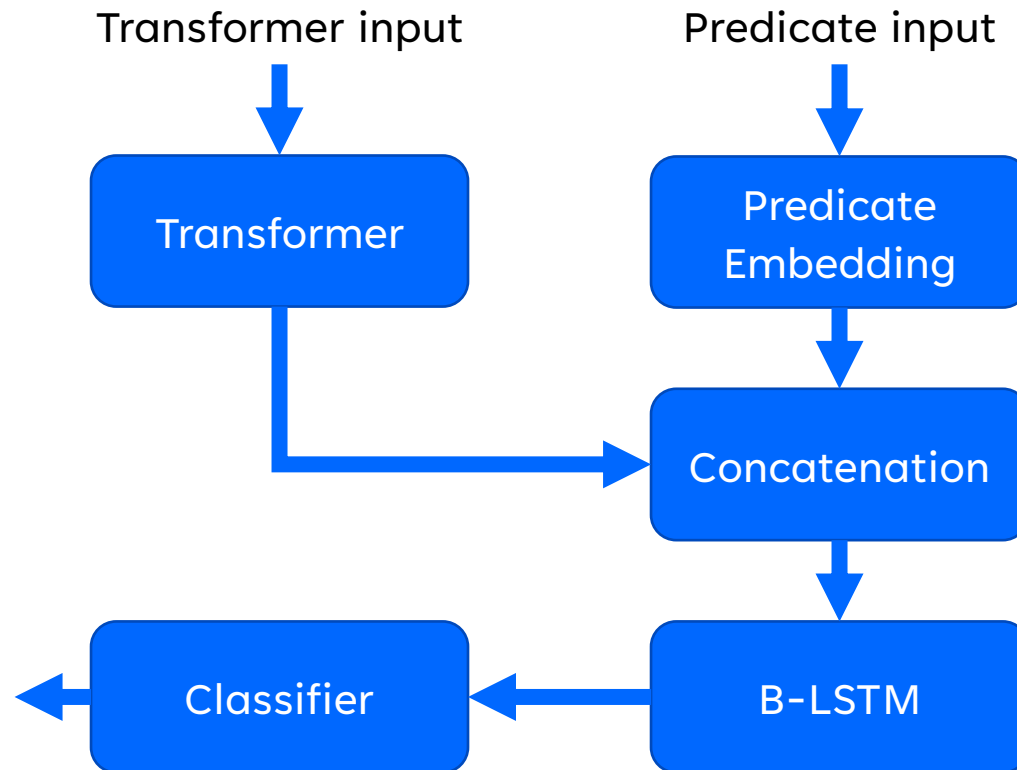
Preprocess

Preprocess: Base

"I used a Transformer"

[[CLS], I, used, a, Trans, **##former**, [SEP]]

[101, 146, 1215, 170, 13809, 23763, 102]
Transformer input

Input

[0,   1,   2,   1,   1,   0,   0]
Predicate Embedding input

Preprocess: Base + POS

"I used a Transformer"

[[CLS], I, used, a, Trans, ##former, [SEP]]

[101, 146, 1215, 170, 13809, 23763, 102]
Transformer input

Input

[0, 1, 2, 1, 1, 0, 0]
Predicate Embedding input

[0, 11, 16, 6, 8, 0, 0]
POS Embedding input
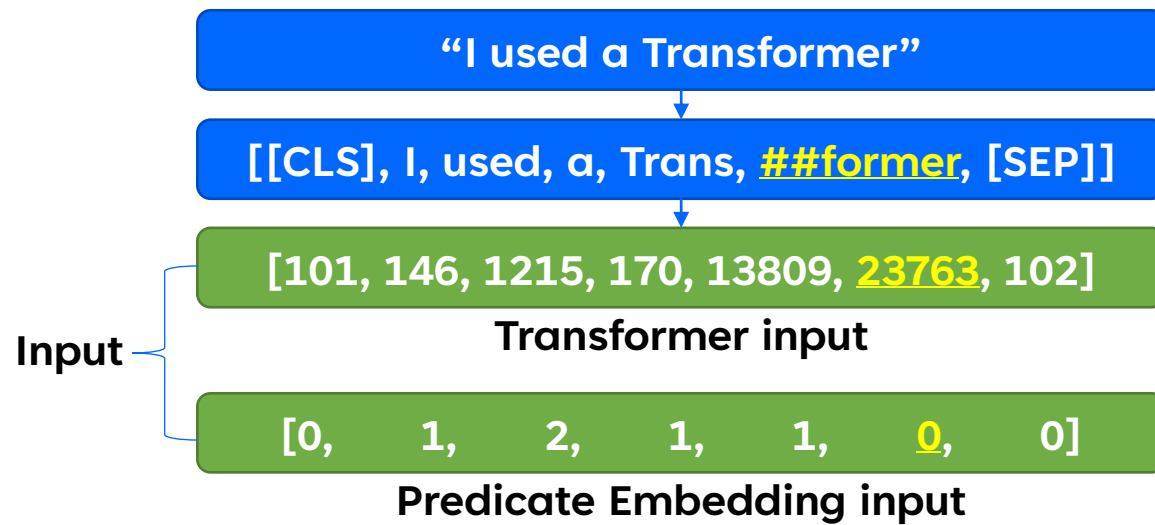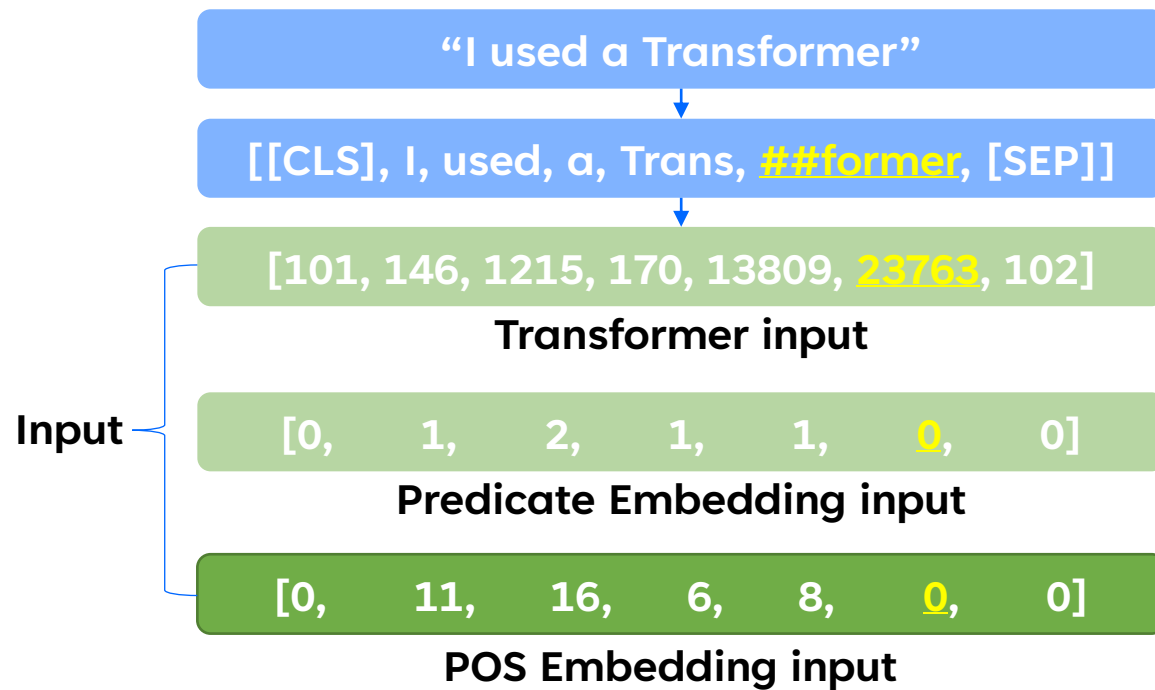
Preprocess: Base + POS + VerbAtlas

VerbAtlas Manipulation

"I used a Transformer" + "[SEP] use agent patient"

[[CLS], I, used, a, Trans, ##former, [SEP]] + [use, agent, patient, [SEP]]

[101, 146, 1215, 170, 13809, 23763, 102] + [1329, 3677, 5351, 102]

Transformer input

Input

[0, 1, 2, 1, 1, 0, 0] + [0, 0, 0, 0]

Predicate Embedding input

[0, 11, 16, 6, 8, 0, 0] + [0, 0, 0, 0]

POS Embedding input
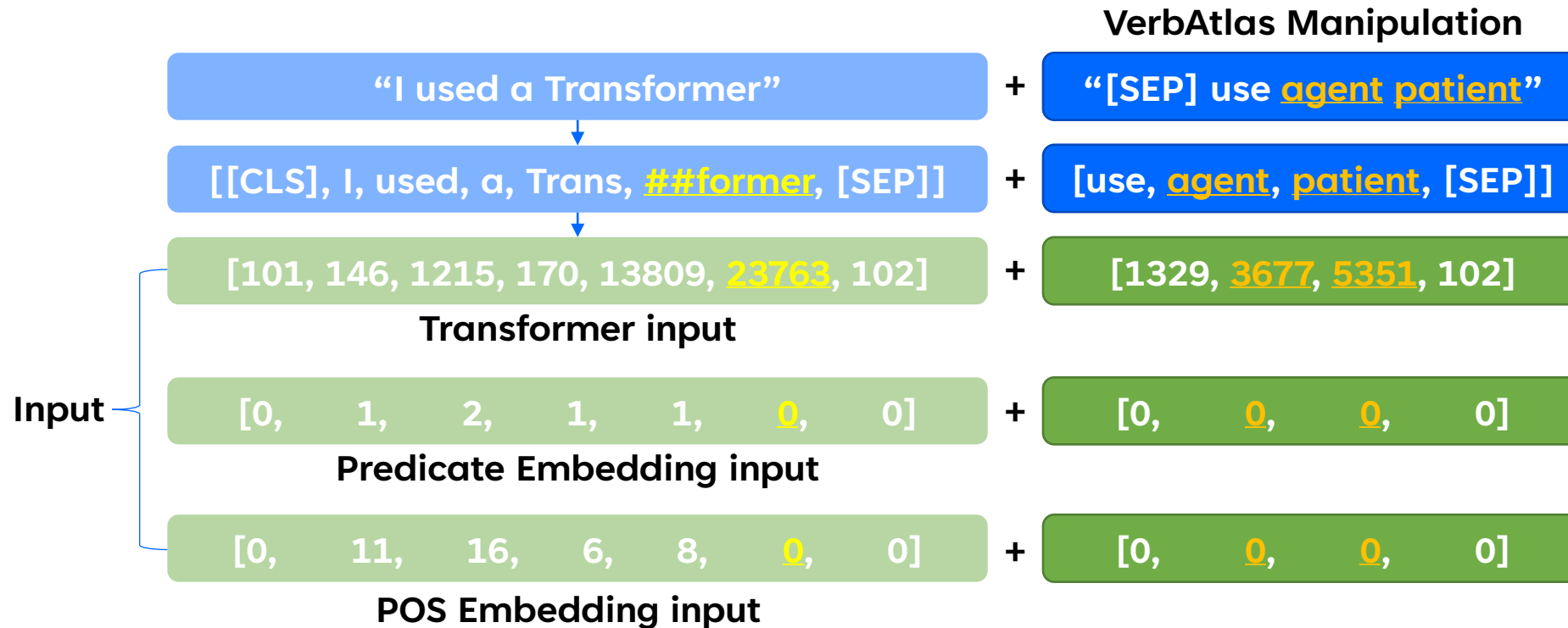
Training
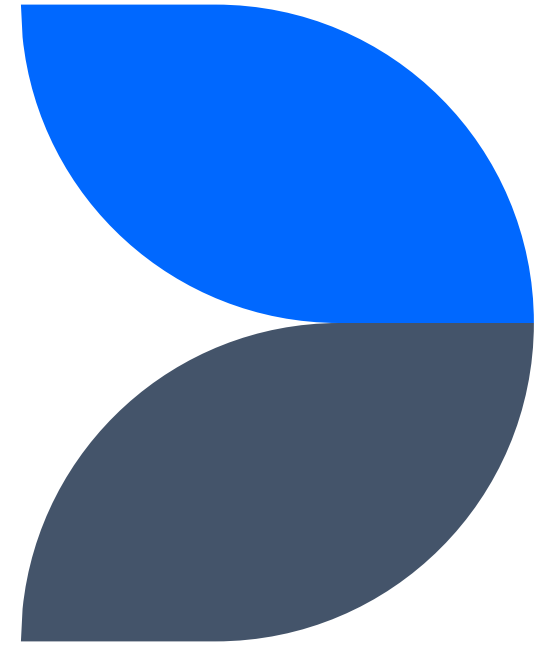
Hyperparameters

Model name	lr	Transformer lr	Batch	Hidden dim BLSTM	Predicate Embedding dim	POS Embedding dim	Transformer name
Step 1							
base model without predicate embedding	4e-4	4e-5	32	200	-	-	bert-base-uncased
Step 2							
base model	1e-3	1e-4	32	200	200	-	bert-base-uncased
Step 3-4							
base model hparams_1	1e-3	1e-4	80	200	128	-	bert-base-cased
base model hparams_2	1e-3	1e-4	80	200	128	-	bert-base-uncased
base model + POS hparams_1	1e-3	1e-4	32	200	128	128	bert-base-uncased
base model + POS hparams_2	1e-3	1e-4	80	200	128	128	bert-base-uncased
base model + POS hparams_3	1e-3	1e-4	32	200	200	200	bert-base-uncased
base model + POS hparams_4	8e-4	1e-4	32	1600	200	200	bert-base-uncased
base model + POS + VerbAtlas	1e-3	1e-4	32	200	128	128	bert-base-uncased

Loss Function

Step 1
Best = 0.023
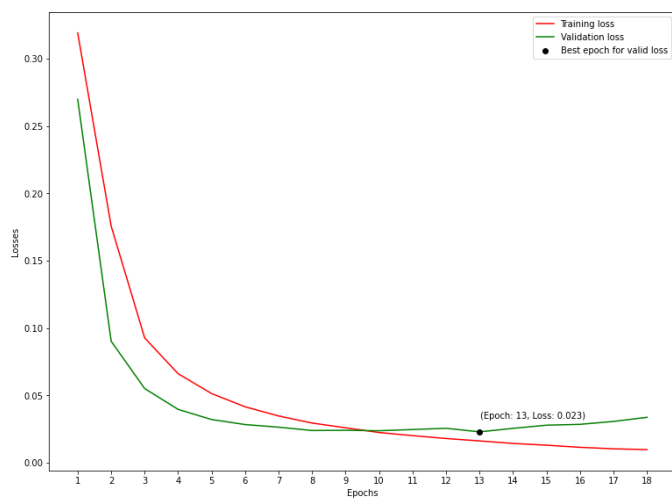
Step 2
Best = 0.098

Step 3-4
Best = 0.072
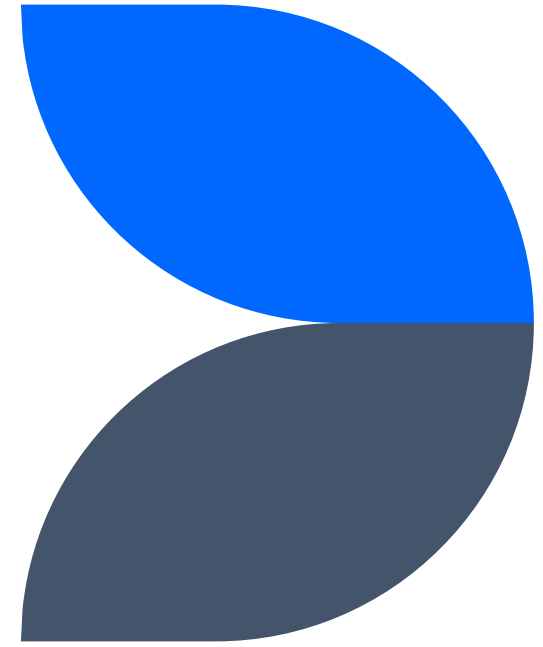
SEMANTIC ROLE LABELING

Results

Scores

Name	F1 score Predicate Identification	F1 score Predicate Disambiguation	F1 score Argument Identification	F1 score Argument Classification
Results 1-2-3-4	0.9482	0.7943	0.8481	0.7808
Results 2-3-4	-	0.8309	0.8779	0.8072
Results 3-4	-	-	0.882	0.8468

Other Languages

Hyperparameters

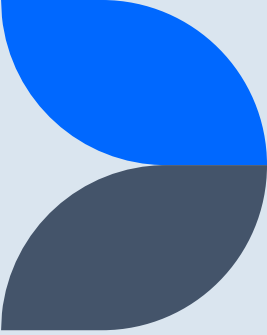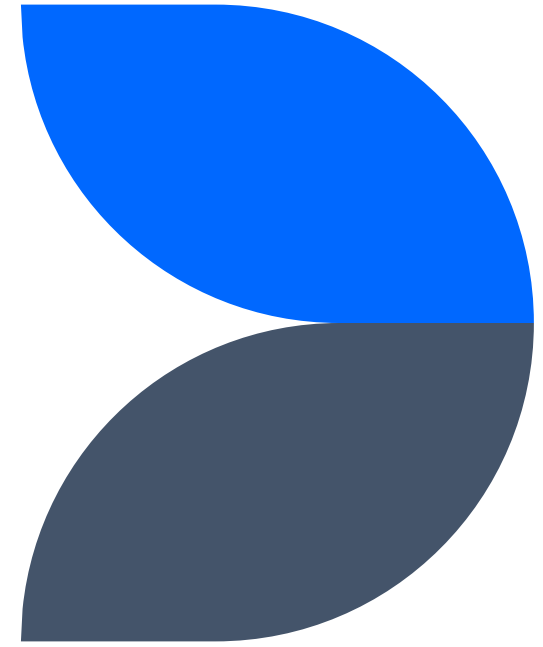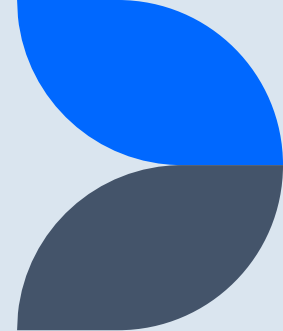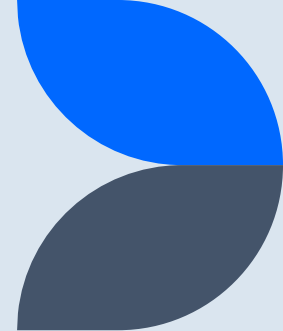| Language	Model name	lr	Transformer lr	Batch	Hidden dim BLSTM	Predicate Embedding dim	POS Embedding dim	Transformer name
Step 1								
Spanish/French	base model without predicate embedding	4e-4	4e-5	32	200	-	-	bert-base-uncased
Step 2								
Spanish/French	fine tune from english checkpoint	1e-3	1e-4	32	200	200	-	bert-base-uncased
Step 3-4								
Spanish/French	fine tune from base model	5e-4	5e-5	32	200	128	-	bert-base-uncased
Spanish/French	fine tune from base model + POS hparams_3	5e-4	5e-5	32	200	200	200	bert-base-uncased
Spanish	base model + POS + VerbAtlas hparams_1	1e-3	1e-4	32	200	128	128	bert-base-uncased
Spanish	base model + POS + VerbAtlas hparams_2	1e-3	1e-4	32	200	128	128	bert-base-multilingual-uncased
Spanish/French	fine tune from base model + POS + VerbAtlas	5e-4	5e-5	32	200	128	128	bert-base-uncased

Hyperparameters

Language	Model name	lr	Transformer lr	Batch	Hidden dim BLSTM	Predicate Embedding dim	POS Embedding dim	Transformer name
				Step 1				
Spanish/French	base model without predicate embedding	4e-4	4e-5	32	200	-	-	bert-base-uncased
				Step 2				
Spanish/French	fine tune from english checkpoint	1e-3	1e-4	32	200	200	-	bert-base-uncased
				Step 3-4				
Spanish/French	fine tune from base model	5e-4	5e-5	32	200	128	-	bert-base-uncased
Spanish/French	fine tune from base model + POS hparams_3	5e-4	5e-5	32	200	200	200	bert-base-uncased
Spanish	base model + POS + VerbAtlas hparams_1	1e-3	1e-4	32	200	128	128	bert-base-uncased
Spanish	base model + POS + VerbAtlas hparams_2	1e-3	1e-4	32	200	128	128	bert-base-multilingual-uncased
Spanish/French	fine tune from base model + POS + VerbAtlas	5e-4	5e-5	32	200	128	128	bert-base-uncased

Scores

Name	F1 score Predicate Identification	F1 score Predicate Disambiguation	F1 score Argument Identification	F1 score Argument Classification
Spanish				
Results 1-2-3-4	0.9176	0.4465	0.6773	0.5078
Results 2-3-4	-	0.4661	0.7256	0.5422
Results 3-4	-	-	0.7360	0.6600
French				
Results 1-2-3-4	0.8821	0.4179	0.6586	0.4889
Results 2-3-4	-	0.4357	0.7099	0.5177
Results 3-4	-	-	0.7230	0.6455

Thanks 🤭