Homeworks presentation

Natural Language Processing course

Gioele Migno
1795826

Homework 1

Named Entity Recognition (NER)

Task description

In Named Entity Recognition (NER) we are interested in locate and classify **named entities** mentioned in a text.

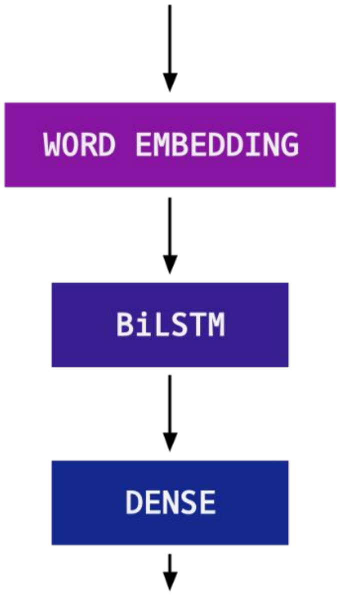
A **named entity** is a real-world object that can be denoted with a proper name. We consider six types of entities:

person, *corporation*, *location*, *product*, *group* and *creative work*

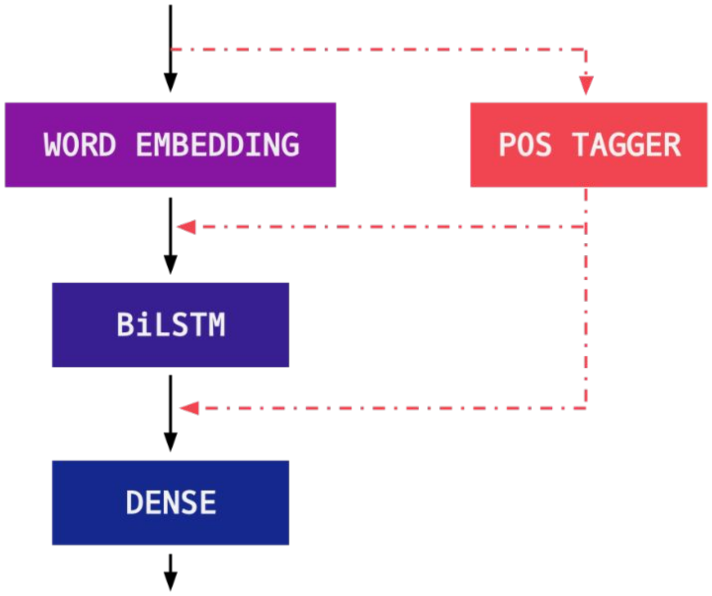
`Obama` `went to` `Italy`

PERSON LOCATION

Model architectures



BASELINE



BASELINE + POS TAGGER

Model architectures: **Baseline parameters**

Word embedding tested:

`word2vec-google-news-300` (17% unknown words)

`glove.6B.100d` (2% unknown words)

BiLSTM - Hidden size:

100 — 220 — 300

BiLSTM - Number of layers:

2	3	4

Model architectures: **Baseline + POS Tagger parameters**

POS Tagging:

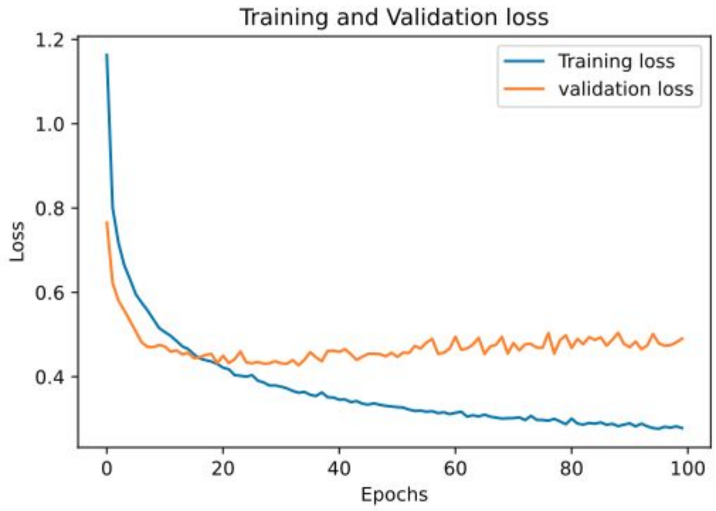
- Performed with `nltk` library
- Penn Treebank tagset
- POS Tags with one-hot encoding

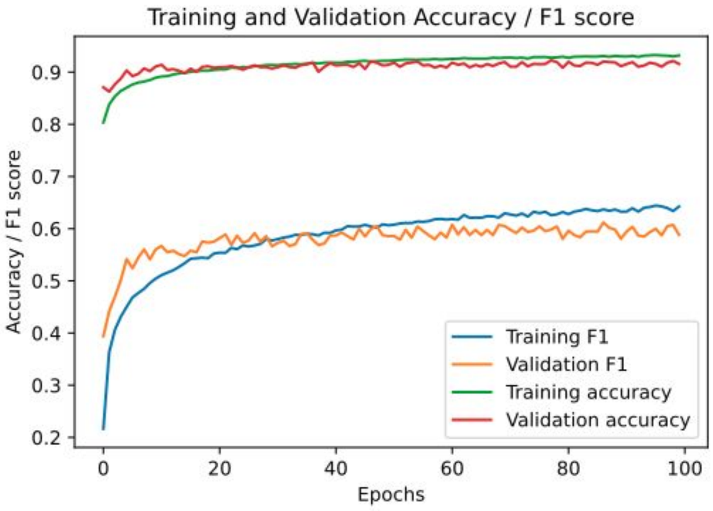
Dense layer changes w.r.t. baseline:

- BiLSTM output + POS Tags as input
- Increased the number of dense layers from 1 to 3
  - Hidden layer of size `(300 x 300)`
  - `ReLU` activation function

Training: Baseline

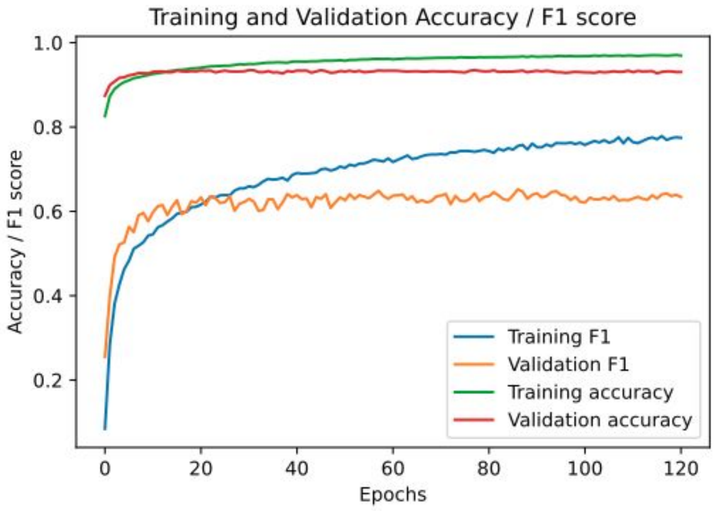
Cross entropy weighted loss, Adam optimizer lr=0.001, dropout 50%, l2=5e-05





Training: **Baseline + POS Tagger**

Cross entropy weighted loss, Adam optimizer lr=0.001, dropout 50%, l2=1e-05



Results: Baseline

Baseline

	precision	recall	f1-score	support
CORP	0.51	0.61	0.56	133
CW	0.40	0.59	0.48	170
GRP	0.64	0.67	0.65	190
LOC	0.67	0.81	0.74	243
PER	0.82	0.86	0.84	300
PROD	0.35	0.56	0.43	149
micro avg	0.58	0.71	0.64	1185
macro avg	0.57	0.68	$\boxed{0.62}$	1185
weighted avg	0.61	0.71	0.65	1185

Results: **Baseline + POS Tagger**

POS Tagger increases macro F1-score by 3%

```
--------------------------------------------------------------------
Baseline + POS Tagger

              precision    recall  f1-score   support

        CORP       0.56      0.59      0.58       133
          CW       0.55      0.49      0.52       170
         GRP       0.69      0.69      0.69       190
         LOC       0.81      0.80      0.81       243
         PER       0.78      0.87      0.82       300
        PROD       0.51      0.52      0.51       149

   micro avg       0.68      0.70      0.69      1185
   macro avg       0.65      0.66      0.65      1185
weighted avg       0.68      0.70      0.69      1185
--------------------------------------------------------------------
```

Homework 2

Semantic Role Labeling (SRL)

Task description

In Semantic Role Labeling (SRL) we are interested in analyzing the **predicate-argument structures** within a sentence.

PREDICATE: Word or a multi-word expression denoting an event or an action

ARGUMENT: Part of the text linked in some way to the predicate

GO-FORWARD

`Obama` `went` `to` `Italy`

AGENT DESTINATION

Task description: **Pipeline**

SRL can be seen as a pipeline of four subtasks:

1) **Predicate identification**

   Obama went to Italy

2) **Predicate disambiguation**

   GO-FORWARD

   Obama went to Italy

3) **Argument identification**

   GO-FORWARD

   Obama went to Italy

4) **Argument classification**

   GO-FORWARD

   Obama went to Italy

   AGENT DESTINATION

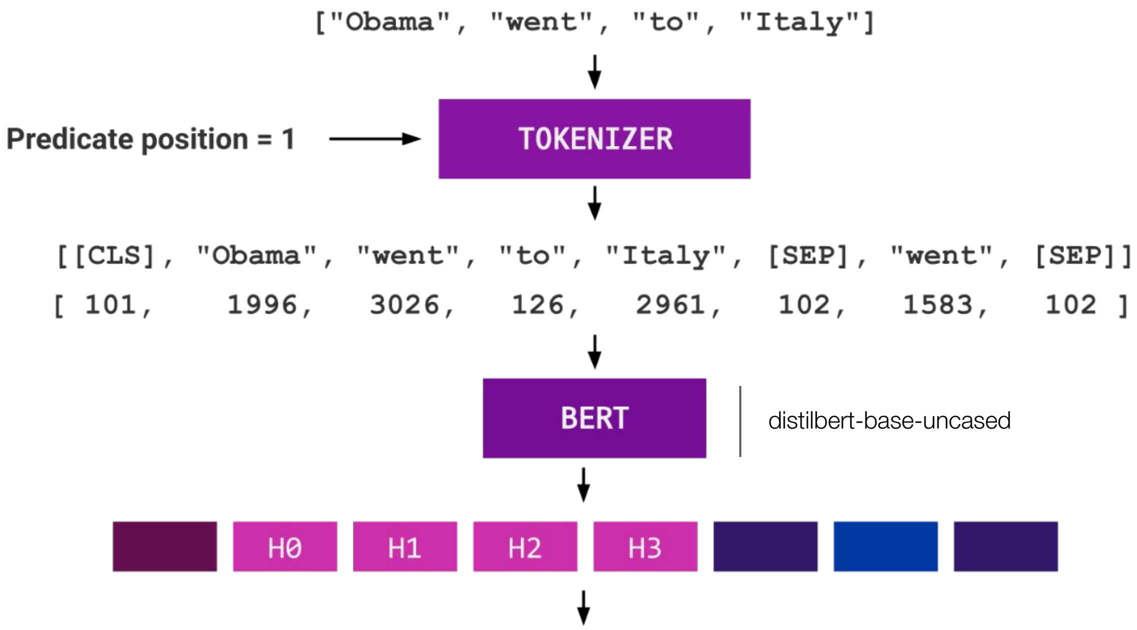
SRL scenarios

We consider two different scenarios

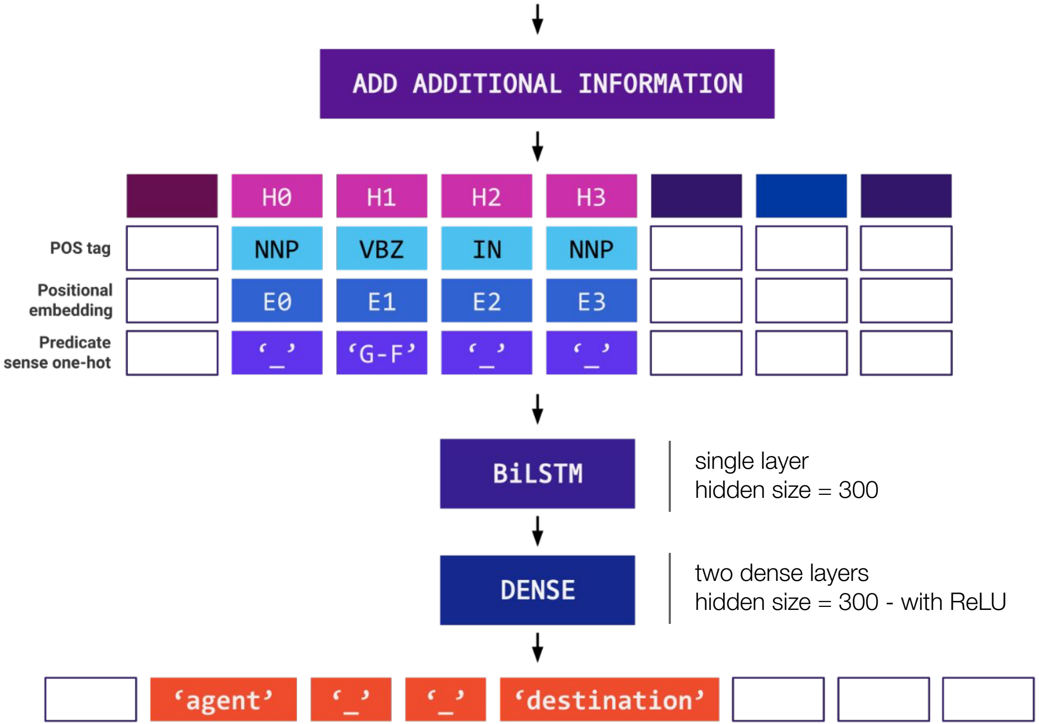
SRL_34: *Predicate identification* and *disambiguation* have already been done

SRL_234: Only *predicate identification* has already been done

SRL_34 Model: Sentence encoding



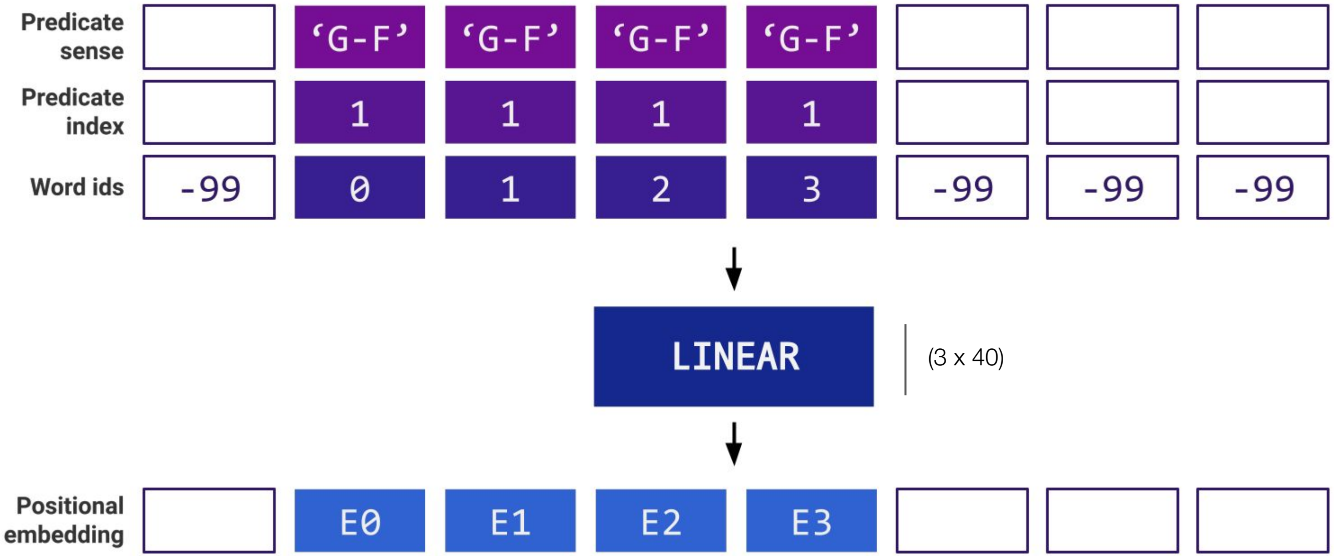
SRL_34 Model: **Add additional information**



single layer
hidden size = 300

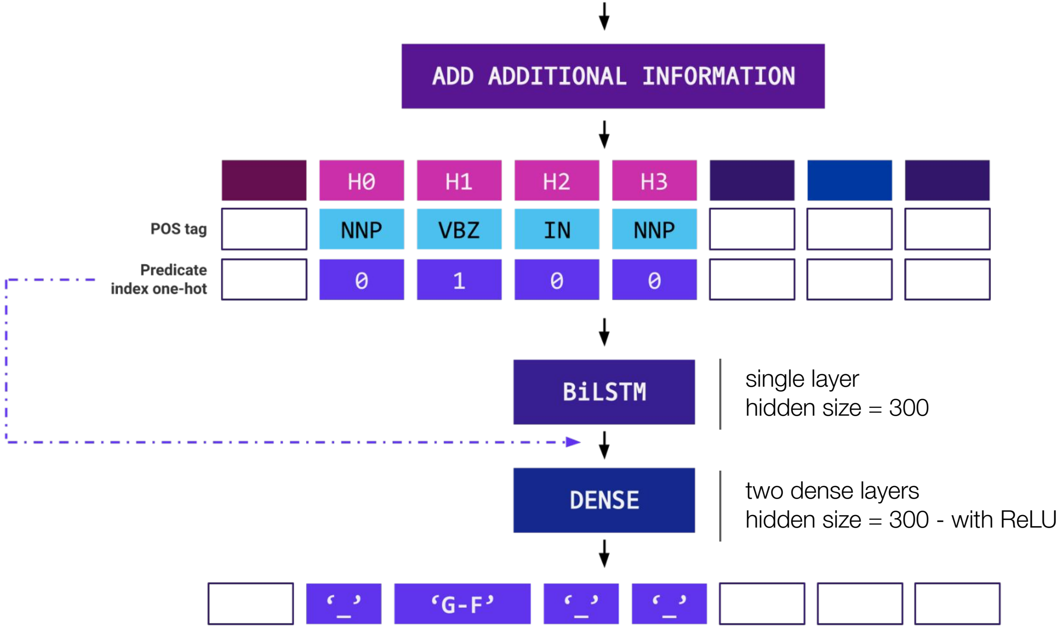
two dense layers
hidden size = 300 - with ReLU

Add additional information: **Positional embedding**

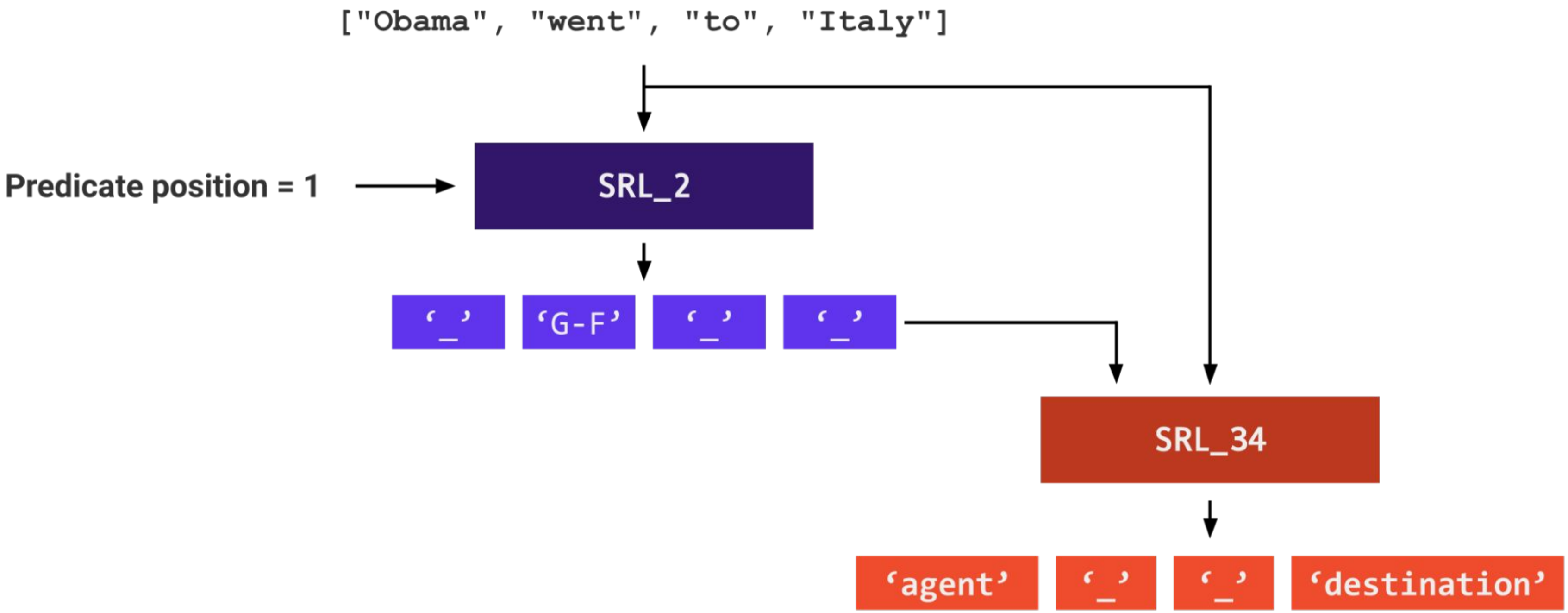


SRL_234: Predicate disambiguation

To perform step 2, SRL_234 uses a sub-model SRL_2 similar to SRL_34



SRL_234: **Overall model**

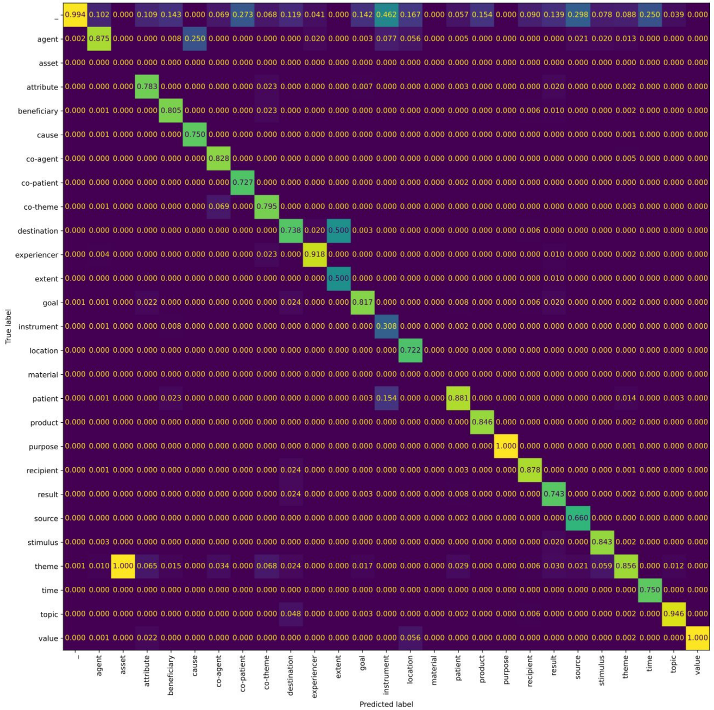


Results: Performance comparison

SRL_34 Tasks	P	R	F1
arg_ident	90.57	88.69	89.62
arg_class	85.96	84.18	**85.06**

SRL_234 Tasks	P	R	F1
pred_disamb	88.37	88.09	88.23
arg_ident	91.26	88.37	**89.79**
arg_class	85.04	82.35	83.67

Results: **Confusion matrix SRL_34**



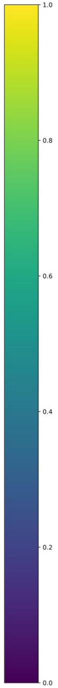
Best accuracy argument classification on:

- EXPERIENCER

Worst accuracy argument classification on:

- INSTRUMENT

  (46% of the time is a false positive)



Confusion matrix SRL_34: **True positive INSTRUMENT**

Example of true positive INSTRUMENT argument class

AGENT ✓

2004/a/59/357_210:1

INCREASE_ENLARGE_MULTIPLY

In 2003, the Office of Human Resources Management expanded its monitoring activities by including on - site monitoring missions covering a broad range of human resources management issues.

PATIENT ✓

INSTRUMENT ✓

Confusion matrix SRL_34: **False positive INSTRUMENT**

Example of false positive argument identification.

AGENT ✓

BENEFICIARY ✓

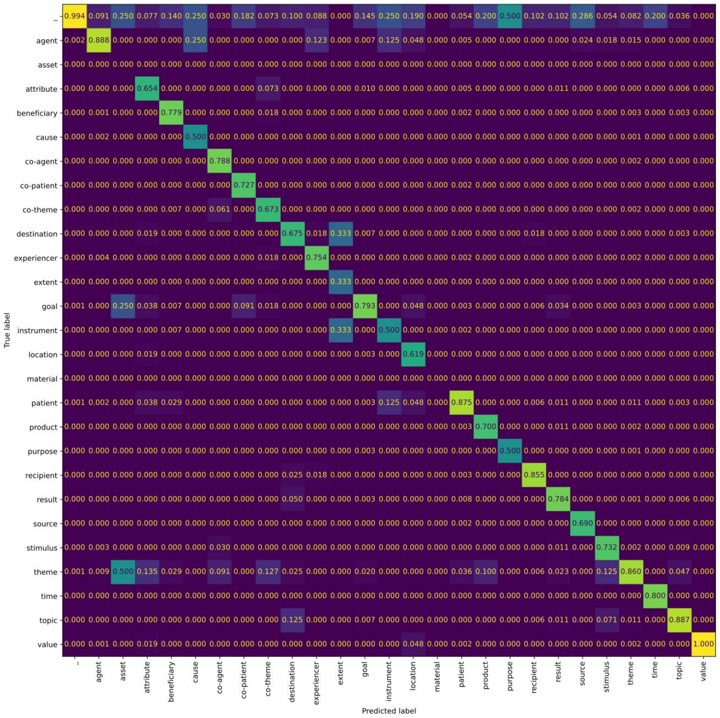
1998/a/c_2/52/sr_20_85:4

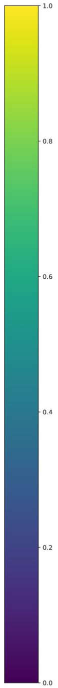
HELP_HEAL_CARE_CURE

The international community, especially the developed countries, should also assist the development of small and medium - sized enterprises in the developing countries through joint ventures, stock participation, technology transfer or licensing agreements.

INSTRUMENT ✗

Results: **Confusion matrix SRL_234 drop**





Drop accuracy argument classification on:

- EXPERIENCER

  (12% of the time assigned to AGENT)

Confusion matrix SRL_234 drop: **EXPERIENCER**

SRL_234 performs worse due to wrong predictions of SRL_2.

SRL_2: BELIEVE ✗
INTERPRET ✓

1999/a/c_2/54/sr_29_21:1

Ms. Critchlow (Guyana), introducing the draft resolution on behalf of the Group of 77 and China, said that the use of solar energy was viewed by the Group of 77 and China as another step in the implementation of Agenda 21.

ATTRIBUTE

SRL_234: STIMULUS ✗
SRL_34: THEME ✓

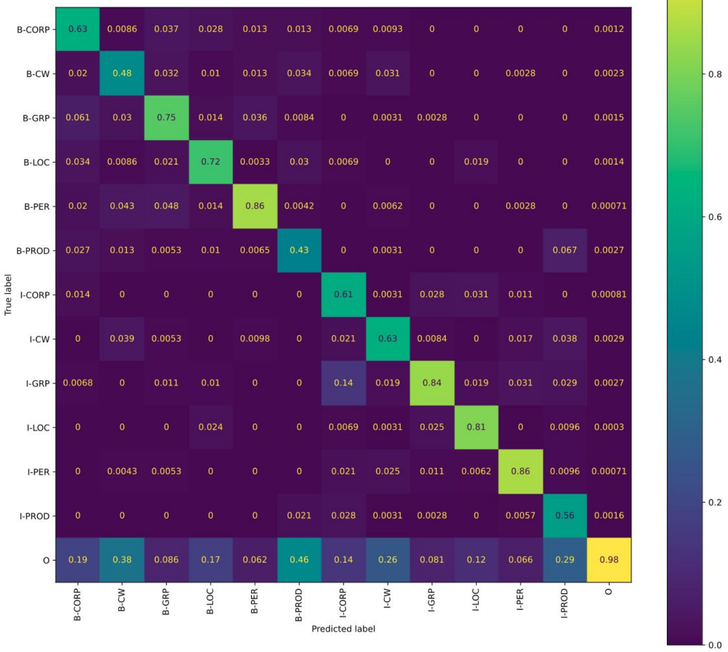
SRL_234: EXPERIENCER ✗
SRL_34: AGENT ✓

Thanks for your attention

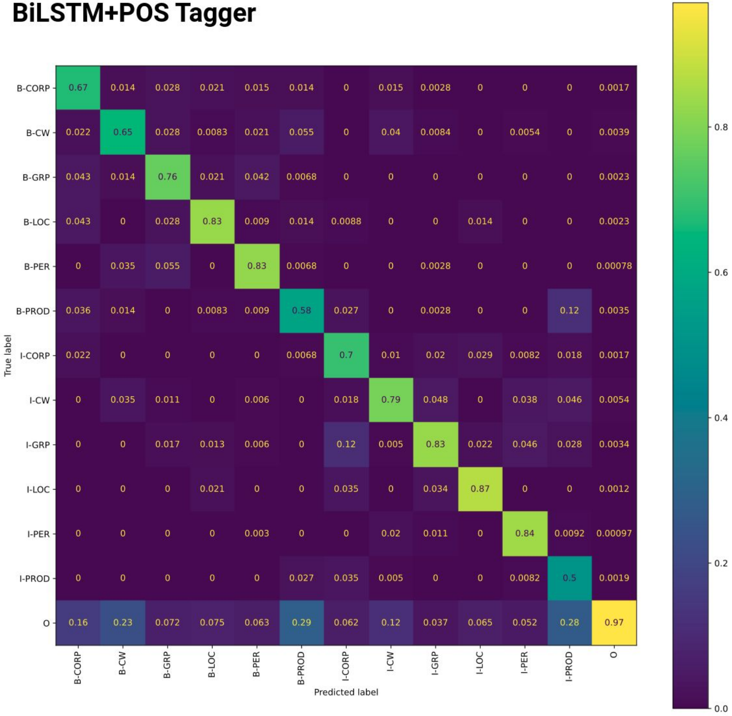
End of the presentation

Results: **Confusion Matrices**

Baseline



BiLSTM+POS Tagger



Confusion matrix SRL_34: **False positive INSTRUMENT - 2**

Second example of false positive argument identification.

AGENT ✓ PATIENT ✓

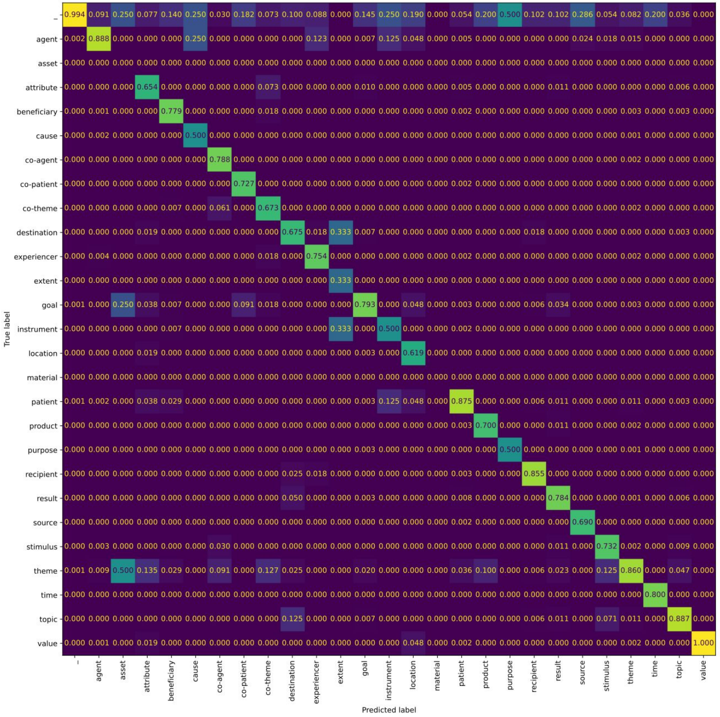
2005/e/cn_4/2005/25_101:1

AMELIORATE

The Working Group agrees that the right to development enriches such strategies through systematically incorporating human rights and the principles of transparency, equality, participation, accountability and non-discrimination into the development processes, at both national and international levels.

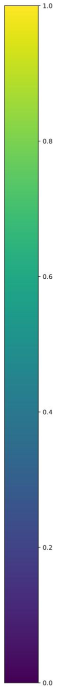
INSTRUMENT ✗ CO-PATIENT ✗

Results: **Confusion matrix SRL_234 drop**



Drop accuracy argument classification on:

- EXPERIENCER (12% of the time assigned to AGENT)
- CAUSE (25% of the time assigned to AGENT)*



*only 4 samples in validation set

Confusion matrix SRL_234 drop: **EXPERIENCER**

SRL_234 performs worse due to wrong predictions of SRL_2.

SRL_2: BELIEVE ✗
INTERPRET ✓

1999/a/c_2/54/sr_29_21:1

Ms. Critchlow (Guyana), introducing the draft resolution on behalf of the Group of 77 and China, said that the use of solar energy was viewed by the Group of 77 and China as another step in the implementation of Agenda 21.

ATTRIBUTE

SRL_34: THEME ✓
SRL_234: STIMULUS ✗

SRL_34: AGENT ✓
SRL_234: EXPERIENCER ✗

Confusion matrix SRL_234 drop: **CAUSE**

SRL_234 performs worse due to wrong predictions of SRL_2

SRL_34: INCITE_INDUCE
SRL_234: IMPLY

1998/a/cn_9/sr_620_123:1

Mr. MARKUS (Observer for Switzerland) said that, although he agreed with the substance of the text, it could lead to misunderstandings.

SRL_34: AGENT
SRL_234: CAUSE

SRL_34: RESULT
SRL_234: TOPIC