729G86/TDP030 Language Technology

Sequence Labelling

Part-of-Speech Tagging & Named Entity Recognition

Marcel Bollmann

Department of Computer Science (IDA)

LINKÖPING UNIVERSITY

Partly based on Jurafsky & Martin (2026).

What is sequence labelling?

Definition

Sequence labelling comprises all tasks that annotate **each item in a sequence** (e.g. *each token in a sentence*) with predefined, discrete classes.

- The **input** is a **text document** written in natural language.
  - just as before
- The **output** is a **label** for **each token** in the document.

Reminder: BERT is a sequence model

- BERT produces one **output vector** for each **input vector**.
  - We can **fine-tune** it on any sequence labelling task!

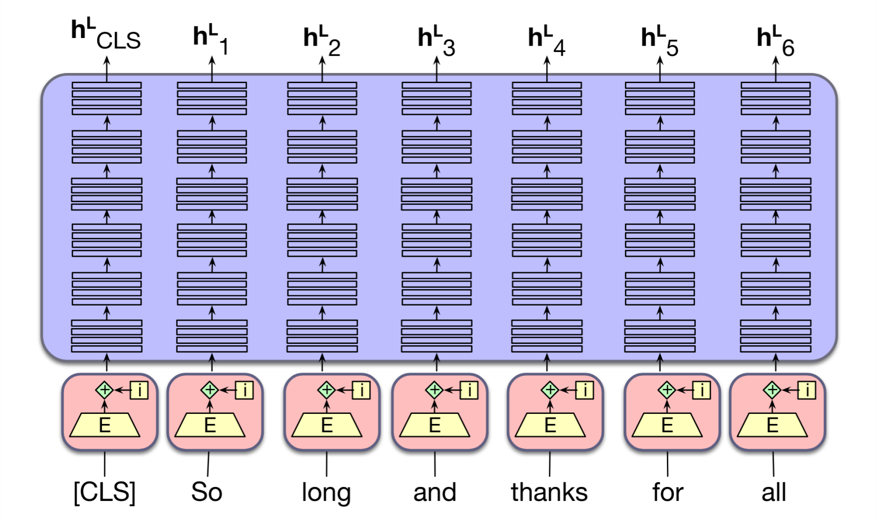


Figure 10.5 from Jurafsky & Martin (2026)

Outline

Part-of-Speech Tagging

- Parts of Speech
- Challenges
- Methods
- Evaluation

Named Entity Recognition

- Named Entities
- Challenges
- BIO Scheme
- Evaluation

Part-of-Speech Tagging

Parts of speech

- A **part of speech** is a category of words that have similar grammatical properties.

Squirrels	*are*	*adorable*	*creatures*
noun	**verb**	**adjective**	**noun**

- Dionysius Thrax of Alexandria ($\approx$ 100 BCE) described eight parts of speech.
  – noun, verb, pronoun, preposition, adverb, conjunction, participle, article

- In language technology, there are different **part-of-speech (POS) tagsets**.
  – different levels of granularity, or tailored to different languages

"Universal Dependencies" tagset

	Tag	Description	Example
Open Class	**ADJ**	Adjective: noun modifiers describing properties	*red*, *young*, *awesome*
	ADV	Adverb: verb modifiers of time, place, manner	*very*, *slowly*, *home*, *yesterday*
	NOUN	words for persons, places, things, etc.	*algorithm*, *cat*, *mango*, *beauty*
	VERB	words for actions and processes	*draw*, *provide*, *go*
	PROPN	Proper noun: name of a person, organization, place, etc..	*Regina*, *IBM*, *Colorado*
	INTJ	Interjection: exclamation, greeting, yes/no response, etc.	*oh*, *um*, *yes*, *hello*
Closed Class Words	**ADP**	Adposition (Preposition/Postposition): marks a noun's spacial, temporal, or other relation	*in, on, by, under*
	AUX	Auxiliary: helping verb marking tense, aspect, mood, etc.,	*can, may, should, are*
	CCONJ	Coordinating Conjunction: joins two phrases/clauses	*and*, *or*, *but*
	DET	Determiner: marks noun phrase properties	*a, an, the, this*
	NUM	Numeral	*one, two, 2026, 11:00, hundred*
	PART	Particle: a function word that must be associated with another word	*'s, not, (infinitive) to*
	PRON	Pronoun: a shorthand for referring to an entity or event	*she, who, I, others*
	SCONJ	Subordinating Conjunction: joins a main clause with a subordinate clause such as a sentential complement	*whether*, *because*
Other	**PUNCT**	Punctuation	; , ()
	SYM	Symbols like $ or emoji	$, %
	X	Other	asdf, qwfg

Figure 17.1 from Jurafsky & Martin (2026); more details: Universal POS tagset

Examples

1

The	*quick*	*brown*	*fox*	*jumped*	*over*	*the*	*lazy*	*dog*	.
DET	**ADJ**	**ADJ**	**NOUN**	**VERB**	**ADP**	**DET**	**ADJ**	**NOUN**	**PUNCT**

2

Preliminary	*findings*	*were*	*reported*	*in*	*today*	*’s*
ADJ	**NOUN**	**AUX**	**VERB**	**ADP**	**NOUN**	**PART**
New	*England*	*Journal*	*of*	*Medicine*		
PROPN	**PROPN**	**PROPN**	**ADP**	**PROPN**		

Part-of-speech tagging

Definition

Part-of-speech (POS) tagging is the task of tagging each word/token in a sentence with its part of speech according to some pre-defined tagset.

- Can provide **useful information** for other language technology tasks.
  - Sentiment: often expressed by adjectives, could analyse them separately
  - Text-to-speech: correct pronunciation sometimes depends on part of speech
    - *e.g.* “lead” or “object”
- Can be used for **linguistic analysis** of texts.
  - Stylometry (*e.g.* authorship attribution, forensic linguistics), linguistic change

Why do we need sequence labelling for that?

1. *a small building in the* ***back***
   NOUN
2. *earnings growth took a* ***back*** *seat*
   ADJ
3. *a majority of politicians* ***back*** *the bill*
   VERB
4. *enable the country to buy* ***back*** *the debt*
   PART
5. *I was twenty-one* ***back*** *then*
   ADV

Why do we need sequence labelling for that?

- Many word types are **unambiguous** when it comes to part of speech.
  - *'Marcel'* is always **PROPN**, *'hesitantly'* is always **ADV**
  - ≈ 85% of word types in English are unambiguous
- However, **ambiguous** word types tend to be **very common**.
  - ≈ 60% of all word *tokens* (or *instances*) that we see in English

Ambiguity causes combinatorial explosion

- Part-of-speech tags are **not independent** of each other.
  - *e.g.* predicting **DET** means that the next word is likely to be **ADJ** or **NOUN**
- Which **sequence** of part-of-speech tags has the **highest probability**?

The	*quick*	*brown*	*fox*	*jumped*	*over*	*the*	*lazy*	*dog*	*.*
DET	ADJ	ADJ	NOUN	VERB	ADP	DET	ADJ	NOUN	PUNCT
	ADV	NOUN	VERB		ADJ			VERB	
	NOUN	VERB			ADV				

- **Combinatorial explosion**: there are **108** possible sequences in this example!

Algorithms for POS tagging

- Classifiers like Naive Bayes and Logistic Regression are not really suitable here since **they don't model sequential inputs/outputs**.

- There are "traditional" machine learning algorithms that do well at this task:
  - Hidden Markov models (HMM)
  - Conditional random fields (CRF)
  - Maximum entropy Markov models (MEMM)

- Here, we'll focus on neural networks in the form of **fine-tuned BERT models**.

Reminder: The pre-train and fine-tune paradigm

1. ~~Pre-train~~ **Download a pre-trained BERT model**.
   - *e.g.* from the **HuggingFace Model Hub**
2. **Fine-tune** the model on whatever classification task we are interested in.
   - e.g part-of-speech tagging!

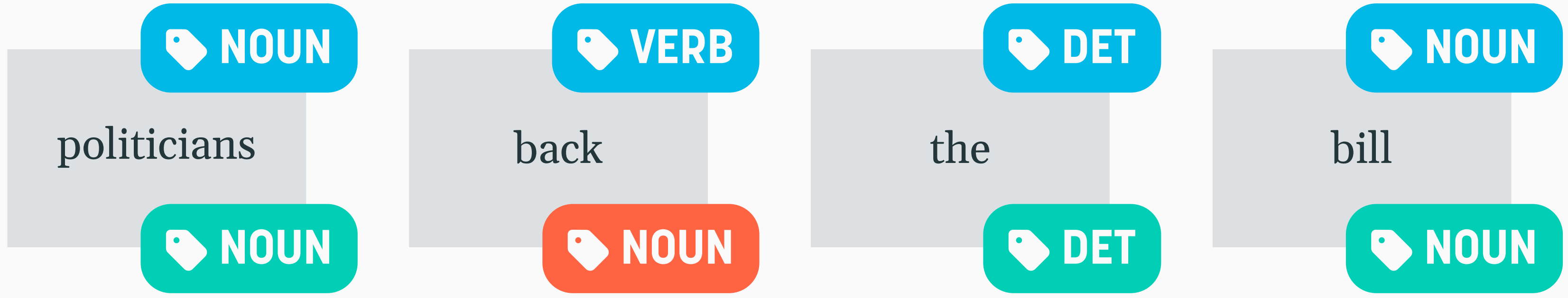
- If you simply want to **use an existing POS tagger** (that's already trained), the **spaCy** models can be a good choice. (**web demo**)

Reminder: Evaluation of text classifiers

- We need a test set with documents and **gold-standard labels**.
  - “gold-standard” = assumed to be correct; *e.g.* produced or verified manually

Evaluation of POS taggers

- We *still* need a test set with documents and **gold-standard labels**.
  - But now there is one label for each token



- We *still* evaluate our classifier by comparing them against the **predicted labels**.

We can still use confusion matrices!

predicted

true		DET	ADJ	NOUN	ADP	VERB
	DET	92	0	0	0	1
	ADJ	2	125	13	1	5
	NOUN	0	7	450	1	8
	ADP	0	0	0	233	1
	VERB	0	5	13	2	345

Important Concepts

- part-of-speech (POS) tagging
- tagsets
- *same as before:* confusion matrix, accuracy, precision, recall, …

Named Entity Recognition

What are named entities?

- A **named entity** is anything that can be referred to with a **“proper name”**.
- Most commonly used **entity tags**:
  1. **PER** – person, *e.g.* ‘Marie Curie’ or ‘Sir Elton John’
  2. **LOC** – location, *e.g.* ‘Lake Vättern’ or ‘Mount Fuji’
  3. **ORG** – organization, *e.g.* ‘Burger King’ or ‘Linköping University’
  4. **GPE** – geo-political entity, *e.g.* ‘Linköping’ or ‘Commonwealth of Australia’
- As with POS, which entities we distinguish depends entirely on the **tagset**!
  - *e.g.* spaCy’s models also include **DATE**, **TIME**, **MONEY**, **WORK_OF_ART**, …

Named entity recognition

Definition

Named entity recognition (NER) is the task of identifying named entities and labelling them with their type.

1. **Finding spans** of text that constitute named entities.
   - can be single words (‘Sweden’) or multi-word phrases (‘Republic of Ireland’)
2. **Tagging the type** of the entity.
   - *e.g.* person, location, organization, etc.

Example

Taco Bell ORG *is an* *American* LOC *- based chain of fast food restaurants founded in* *1962* DATE *by* *Glen Bell* PER *in* *Irvine , California* GPE *.*

What is NER useful for?

Taco Bell ORG *is an* *American* LOC *- based chain of fast food restaurants founded in* *1962* DATE *by* *Glen Bell* PER *in* *Irvine , California* GPE *.*

- **Sentiment analysis** with respect to a particular company or person
  - *e.g.* What do consumers think of Taco Bell?
- **Extracting facts** or **answering questions** about entities
  - *e.g.* When was Taco Bell founded? Who founded it?
- **Linking entities** to knowledge bases that contain structured information
  - *e.g.* 'Taco Bell' can be found in Wikidata as **Q752941**

What makes NER difficult?

1. We need to do **segmentation** in addition to tagging!
   - In POS tagging, each word gets one tag.
   - In NER, not everything is an entity.

2. Just as in POS tagging, there is **ambiguity**…

- *Washington* PER *was born into slavery on the farm of James Burroughs.*
- *Washington* ORG *went up 2 games to 1 in the four-game series.*
- *Blair arrived in* *Washington* LOC *for what may well be his last state visit.*
- *In June,* *Washington* GPE *passed a primary seatbelt law.*

What makes NER difficult?

- In some languages (here: Polish), even proper names can get **inflected**!

case	inflected form
nominative	Muammar Kaddafi
genitive	Muammara Kaddafiego
dative	Muammarowi Kaddafiemu
accusative	Muammara Kaddafiego
instrumental	Muammarem Kaddafim
locative	Muammarze Kaddafim
vocative	Muammarze Kaddafi

Source: Piskorski & Yangarber (2012)

But here’s the good news…

We can use the **exact same methods** as for POS tagging!

The BIO tagging scheme

Idea

Word-level tags can encode both the **boundaries** and **types** of named entities.

- One way of doing this is the **BIO scheme**, which uses three kinds of tags:
  - **B (beginning)** for tokens that begin a span
  - **I (inside)** for tokens that continue a span
  - **O (outside)** for tokens outside of any span

The BIO tagging scheme

Taco Bell ORG *is an* *American* LOC *- based chain of fast food restaurants founded in* *1962* DATE *by* *Glen Bell* PER *in* *Irvine , California* GPE *.*

Taco	Bell	is	an	American	-	based	chain	of	fast	food	restaurants
B-ORG	I-ORG	O	O	B-LOC	O	O	O	O	O	O	O

founded	in	1962	by	Glen	Bell	in	Irvine	,	California	.
O	O	B-DATE	O	B-PER	I-PER	O	B-GPE	I-GPE	I-GPE	O

Alternative tagging schemes: IO and BIOES

Words	IO Label	BIO Label	BIOES Label
Jane	I-PER	B-PER	B-PER
Villanueva	I-PER	I-PER	E-PER
of	O	O	O
United	I-ORG	B-ORG	B-ORG
Airlines	I-ORG	I-ORG	I-ORG
Holding	I-ORG	I-ORG	E-ORG
discussed	O	O	O
the	O	O	O
Chicago	I-LOC	B-LOC	S-LOC
route	O	O	O
.	O	O	O

Figure 17.7 from Jurafsky & Martin (2026)

Side note: Chinese word segmentation

- The same idea can be used for **any kind of span tagging**.
- Example: **Chinese word segmentation**!

He briefed reporters on the main contents

他向记者介绍了主要内容

他	向	记者	介绍了	主要	内容
S	S	B E	B M E	B E	B E

  – **S**ingle-character word, **B**eginning/**M**iddle/**E**nd of a word

But how does the **evaluation** work?

Evaluation of named entity recognition

- We want to evaluate NER on the **entire spans**, not individual tags.
  - It doesn't really make sense to speak of *e.g.* "precision on **I-PER**"...
  - The **O** tag is the most common, but the one we care least about.

- We can **convert tags to tuples** containing:

  1. the **start position** of the span *(e.g., index of the token)*
  2. the **end position** of the span
  3. the **entity type**

- We can then compute **span-level precision, recall, F1-score**.

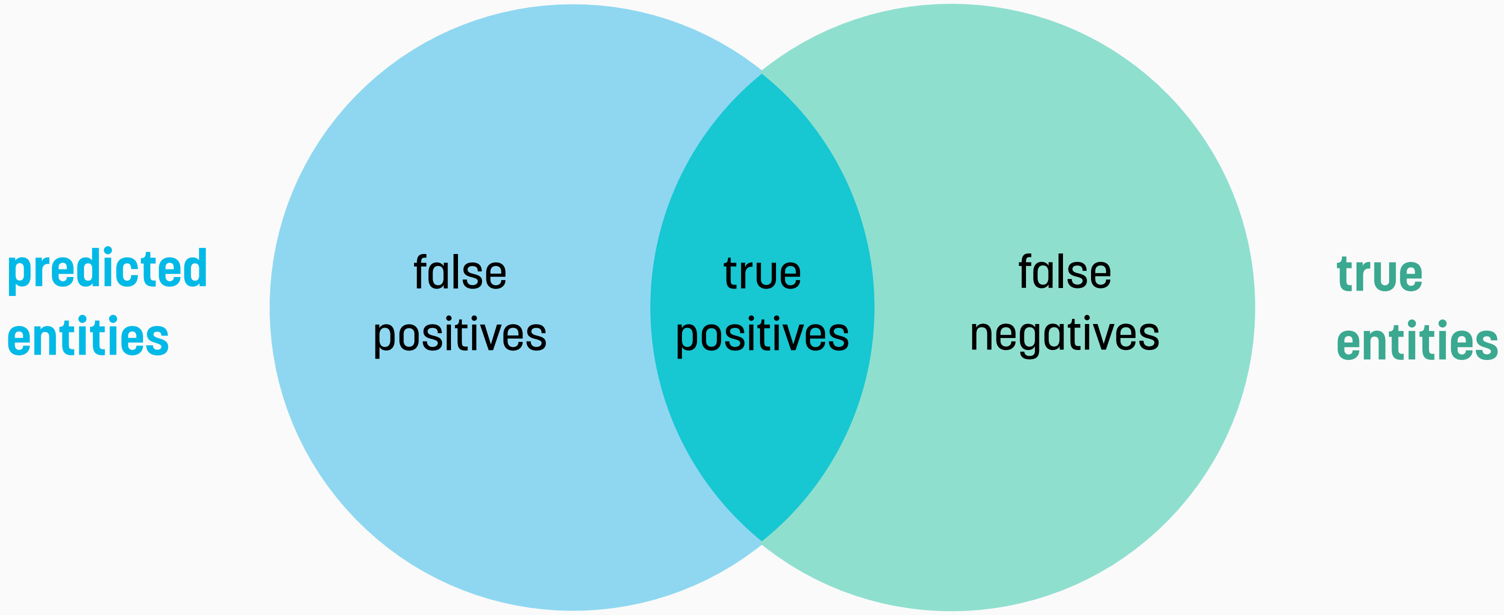
From tags to spans

1	2	3	4	5	6	7	8	9	10	11
Taco	*Bell*	*was*	*founded*	*in*	*1962*	*in*	*Irvine*	*,*	*California*	*.*
B-ORG	I-ORG	O	O	O	B-DATE	O	B-GPE	I-GPE	I-GPE	O

Three entity spans:

1. (1, 2, ORG)
2. (6, 6, DATE)
3. (8, 10, GPE)

Precision and recall, again



$$P = \frac{|\text{true positives}|}{|\text{true positives}| + |\text{false positives}|} \qquad R = \frac{|\text{true positives}|}{|\text{true positives}| + |\text{false negatives}|}$$

Span-level precision and recall

	1	2	3	4	5	6	7	8	9	10	11
	Taco	*Bell*	*was*	*founded*	*in*	*1962*	*in*	*Irvine*	*,*	*California*	*.*
true:	B-ORG	I-ORG	O	O	O	B-DATE	O	B-GPE	I-GPE	I-GPE	O
predicted:	B-ORG	I-ORG	O	O	O	B-TIME	O	B-GPE	O	B-LOC	O

Predicted entity spans:

- **(1, 2, ORG)**
- (6, 6, TIME)
- (8, 8, GPE)
- (10, 10, LOC)

True entity spans:

- **(1, 2, ORG)**
- (6, 6, DATE)
- (8, 10, GPE)

precision = $\frac{1}{4}$, or 25%

recall = $\frac{1}{3}$, or 33.3...%

Challenges for evaluation

	1	2	3	4	5	6	
	First	*Bank*	*of*	*Chicago*	*announced*	*earnings*	…
true:	B-ORG	I-ORG	I-ORG	I-ORG	O	O	
predicted:	O	B-ORG	I-ORG	I-ORG	O	O	

- There is **some overlap** between the true & predicted entity, but they don't match!
  - In tuple notation: (1, 4, ORG) vs. (2, 4, ORG)
- With our evaluation method, both precision and recall **will be zero**.
  - This example creates both a false negative *and* a false positive!
- More advanced metrics could account for **partial overlap** as well.

Important Concepts

- named entity recognition (NER)
- common named entity types
- BIO tagging scheme
- span-level precision & recall