Deep Learning for Natural Language Processing

NLP Tasks with Sequence Outputs

UNIVERSITY OF
GOTHENBURG

CHALMERS

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

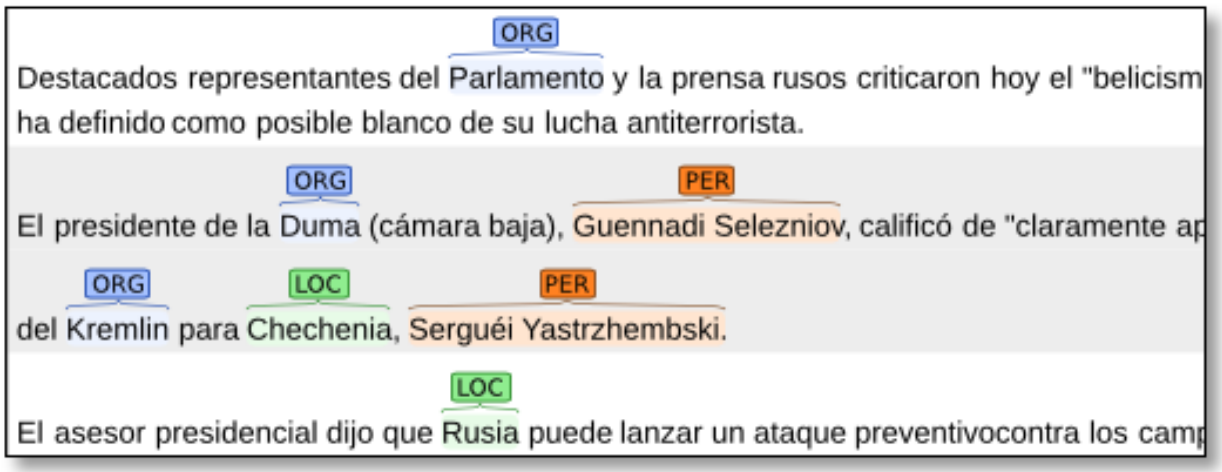
Richard Johansson

richard.johansson@gu.se

this lecture block

- first instance of **structured outputs**: sequences
- now: quick overview with applications for some different sequence-related NLP tasks:
  - **sequence labeling** tasks
  - **segmentation** tasks
  - **bracketing** tasks
- later: the implementation of models for these tasks

Astronomers	**announce**	**the**	**detection**	**of**	**water**	**in**	**the**	**atmosphere**
NNS	**VBP**	**DT**	**NN**	**IN**	**NN**	**IN**	**DT**	**NN**



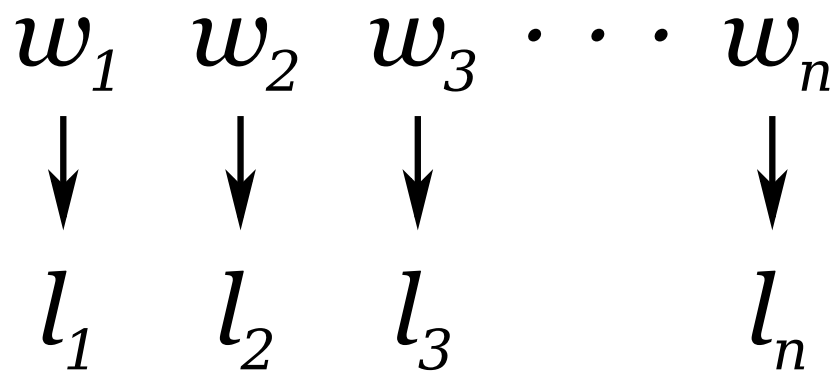
sequence labeling tasks

- in **sequence labeling** or **sequence tagging**, we are given a sequential input and our goal is to predict an output sequence of the same length
- in NLP, the input is most often a sequence of words

$$w_1 \quad w_2 \quad w_3 \cdots w_n$$

sequence labeling tasks

- in **sequence labeling** or **sequence tagging**, we are given a sequential input and our goal is to predict an output sequence of the same length
- in NLP, the input is most often a sequence of words



application example: part-of-speech tagging

- in part-of-speech tagging, we are given a sequence of words and want to assign **part-of-speech tags** (grammatical categories) to each word

She plays in many plays .

application example: part-of-speech tagging

- in part-of-speech tagging, we are given a sequence of words and want to assign **part-of-speech tags** (grammatical categories) to each word

She	**plays**	**in**	**many**	**plays**	**.**
pronoun	verb	preposition	adverb	noun	punctuation

application example: part-of-speech tagging

- in part-of-speech tagging, we are given a sequence of words and want to assign **part-of-speech tags** (grammatical categories) to each word

She	**plays**	**in**	**many**	**plays**	**.**
pronoun	verb	preposition	adverb	noun	punctuation

- the most widely studied sequence labeling task
- used to be very important in applications

part-of-speech tagging: datasets

- there are **many** dataset for PoS tagging!
- PoS datasets for many languages, using a standardized set of tags: https://universaldependencies.org/

Current UD Languages

Information about language families (and genera for families with multiple branches) is mostly taken from WALS Online (IE = Indo-European).

Afrikaans	1	49K		IE, Germanic
Akkadian	1	1K		Afro-Asiatic, Semitic
Albanian	1	<1K	W	IE, Albanian
Amharic	1	10K		Afro-Asiatic, Semitic
Ancient Greek	2	416K		IE, Greek
Arabic	3	1,042K	W	Afro-Asiatic, Semitic
Armenian	1	52K		IE, Armenian
Assyrian	1	<1K		Afro-Asiatic, Semitic
Bambara	1	13K		Mande
Basque	1	121K		Basque
Belarusian	1	13K		IE, Slavic
Bhojpuri	2	4K		IE, Indic
Breton	1	10K	W	IE, Celtic
Bulgarian	1	156K		IE, Slavic
Buryat	1	10K		Mongolic
Cantonese	1	13K		Sino-Tibetan
Catalan	1	531K		IE, Romance
Chinese	5	285K	W	Sino-Tibetan
Classical Chinese	1	74K		Sino-Tibetan
Coptic	1	40K		Afro-Asiatic, Egyptian
Croatian	1	199K	W	IE, Slavic

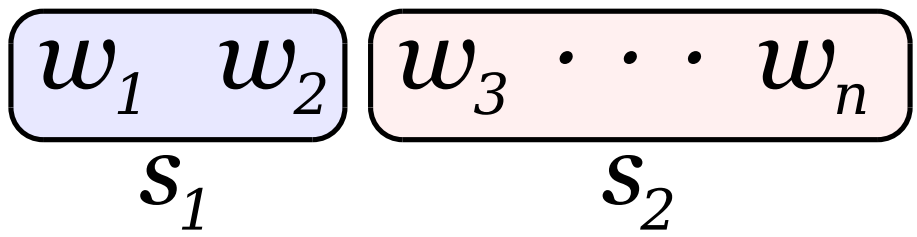
segmentation tasks

- in **segmentation** tasks, our goal is to split the text into pieces
- and optionally also assign labels to the pieces

$$w_1 \;\; w_2 \;\; w_3 \cdots w_n$$

segmentation tasks

- in **segmentation** tasks, our goal is to split the text into pieces
- and optionally also assign labels to the pieces



application example: word segmentation

- in scripts where whitespace is not used between words, **word segmentation** is nontrivial
- example by Wang and Xu (2017):

我	有	一	台	计	算	机	。
(I)	(have)	(a)		(computer)			(.)

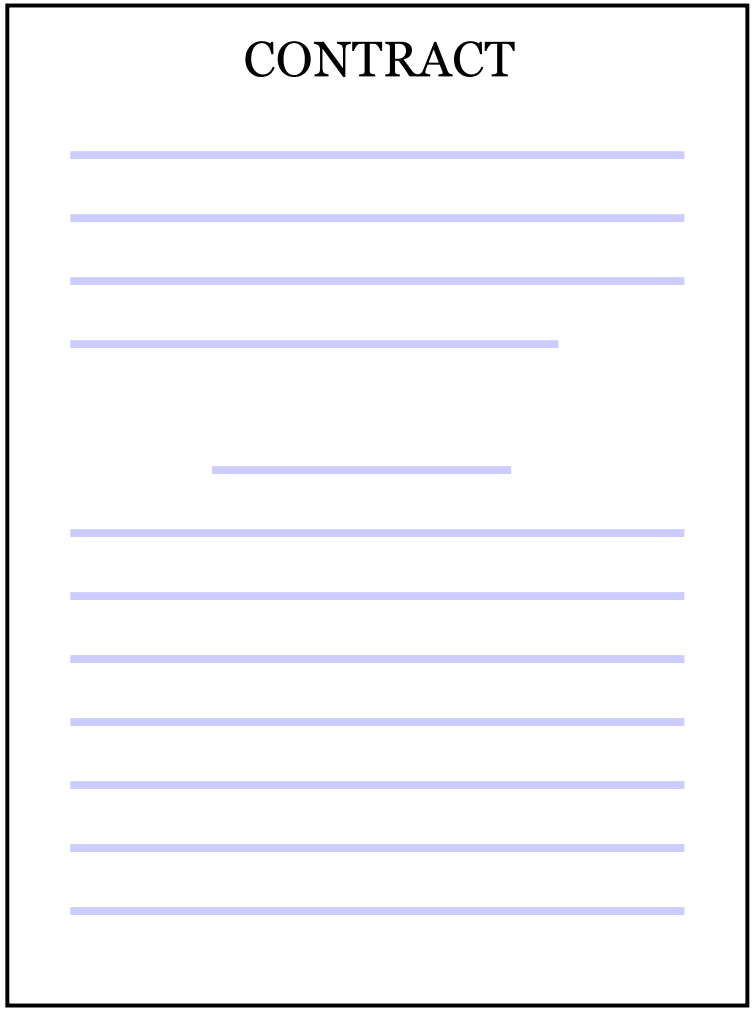
application example: word segmentation

- in scripts where whitespace is not used between words, **word segmentation** is nontrivial
- example by Wang and Xu (2017):

我	有	一 台	计 算 机	。
(I)	(have)	(a)	(computer)	(.)

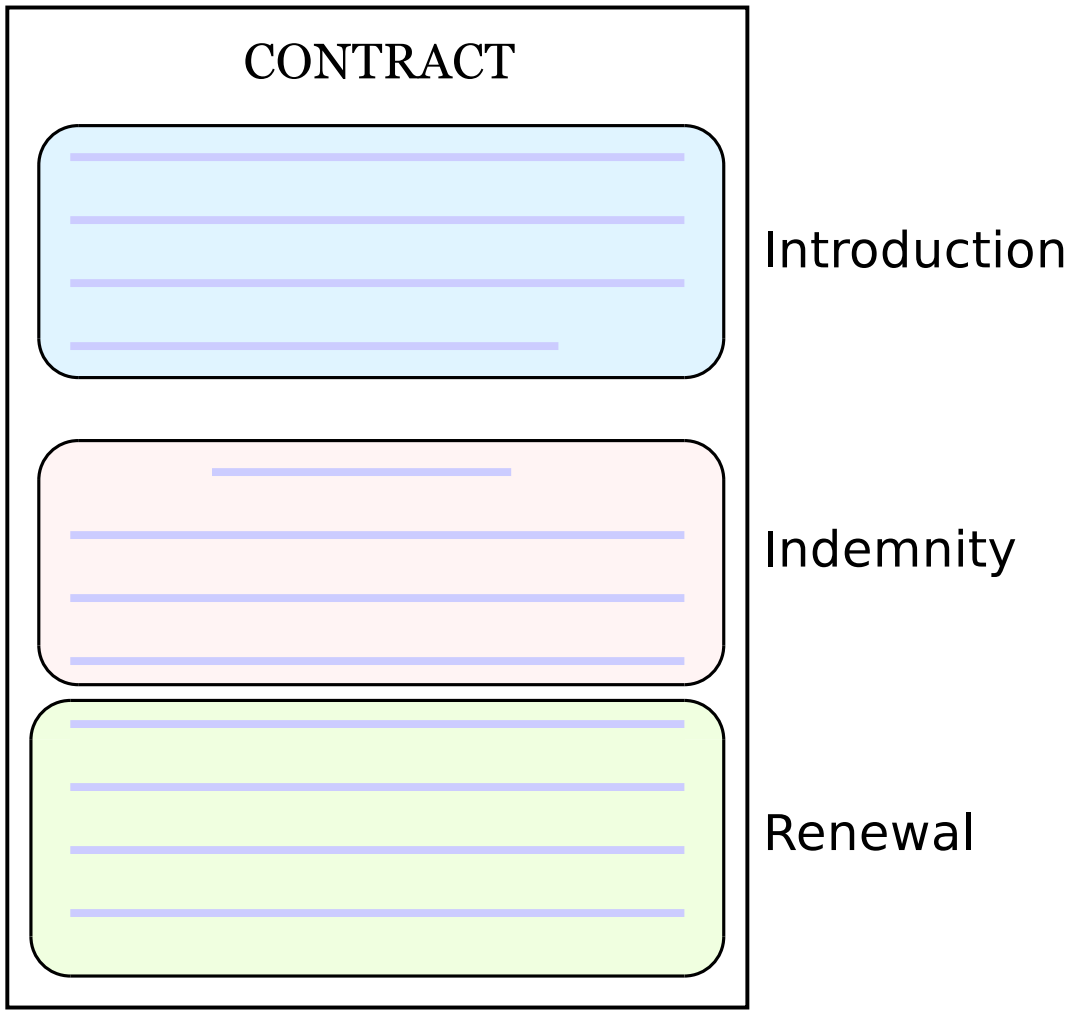
application example: contract clauses

- inspired by Jansson (2019)



application example: contract clauses

- inspired by Jansson (2019)



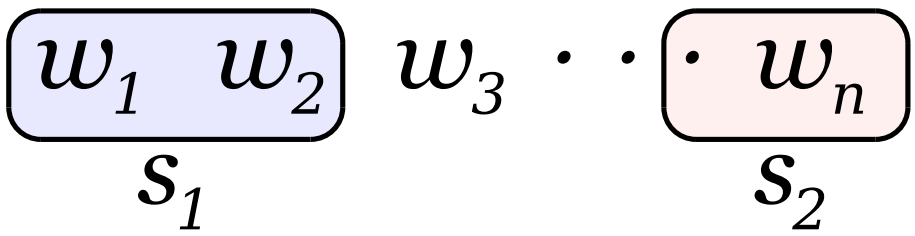
bracketing or markup tasks

- in **bracketing** or **markup** tasks, the goal is to select and categorize some pieces of text (**spans**)

$$w_1 \;\; w_2 \;\; w_3 \cdots w_n$$

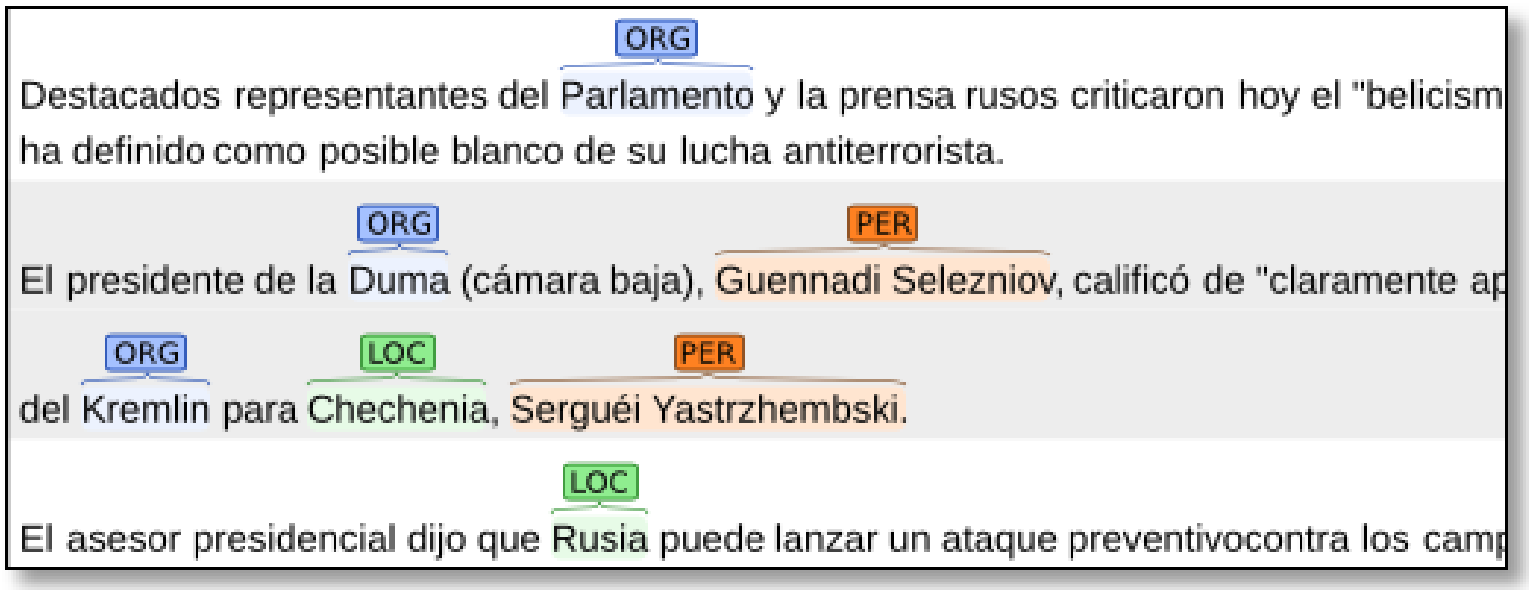
bracketing or markup tasks

- in **bracketing** or **markup** tasks, the goal is to select and categorize some pieces of text (**spans**)



named entity recognition

- the canonical example of a bracketing task is probably **named entity recognition**



other examples of bracketing tasks

- aspect-based sentiment analysis

Their spring rolls **are** **absolutely delicious**

Target Positive

semantic role labeling and "slot-filling"

- given a sentence and a target word, find the "slot-fillers"
- general tasks such as **semantic role labeling** and domain-specific **slot-filling information extraction** systems

AB Volvo	**sold**	**Volvo Cars**	**to**	**Ford**	**in**	**1999**
Seller	Event	Goods		Buyer		Time

reducing segmentation and bracketing to sequence labeling

- in practice, systems for segmentation and bracketing are often implemented as sequence labelers

The	cases	of	metastatic	cancer	of	the	gall	bladder
			Pathology				Organ	

reducing segmentation and bracketing to sequence labeling

- in practice, systems for segmentation and bracketing are often implemented as sequence labelers

The	cases	of	metastatic	cancer	of	the	gall	bladder
			Pathology				Organ	
O	O	O	B-PAT	I-PAT	O	O	B-ORG	I-ORG

reducing segmentation and bracketing to sequence labeling

- in practice, systems for segmentation and bracketing are often implemented as sequence labelers

The	cases	of	metastatic cancer	of	the	gall bladder
			Pathology			Organ
O	O	O	B-PAT	I-PAT	O	O	B-ORG	I-ORG

- **BIO coding** (a.k.a. IOB2) (Tjong Kim Sang and Veenstra, 1999) uses **B**eginning, **I**nside and **O**utside tags

Chinese word segmentation again

- example by Wang and Xu (2017):

我	有	一 台	计 算 机	。
(I)	(have)	(a)	(computer)	(.)

Chinese word segmentation again

- example by Wang and Xu (2017):

S	S	B	E	B	M	E	S
我	有	一	台	计	算	机	。
(I)	(have)	(a)		(computer)			(.)

evaluating sequence labelers

- sequence labeling tasks: usually **accuracy** at the token level

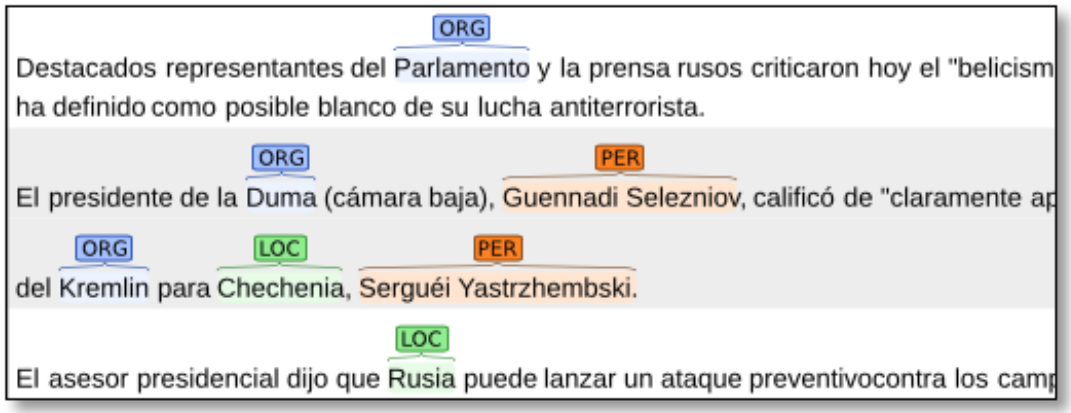
$$\text{accuracy} = \frac{\#\text{ correct tags}}{\#\text{ tokens}}$$

Astronomers	**announce**	**the**	**detection**	**of**	**water**	**in**	**the**	**atmosphere**
NNS	**VBP**	**DT**	**NN**	**IN**	**NN**	**IN**	**DT**	**NN**

evaluating bracketing and segmentation systems

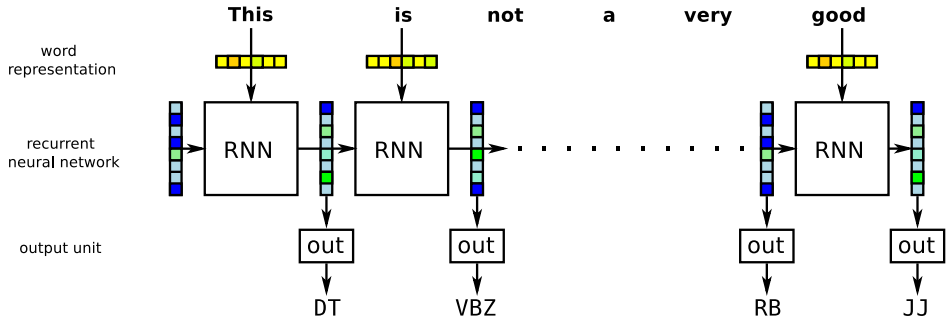
- segmentation and bracketing: usually **precision** and **recall**

$$P = \frac{\#\text{ correct spans}}{\#\text{ proposed spans}} \qquad R = \frac{\#\text{ correct spans}}{\#\text{ reference spans}}$$




```
Overall: P = 0.8755, R = 0.8837, F1 = 0.8796
LOC:     P = 0.9161, R = 0.9336, F1 = 0.9248
PER:     P = 0.9216, R = 0.9387, F1 = 0.9301
```

plan: sequence labeling and friends



- **this block**:
  - simple neural models for sequence labeling
  - improving representations
- **next block**: more sophisticated learning approaches

references

E. Jansson. 2019. Domain adapted language models. Master's Thesis, Chalmers University of Technology.

E. Tjong Kim Sang and J. Veenstra. 1999. Representing text chunks. In *EACL*.

C. Wang and B. Xu. 2017. Convolutional neural network with word embeddings for Chinese word segmentation. In *IJCNLP*. pages 163–172.