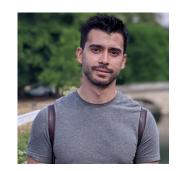


#### Ana Ezquerro Carlos Gómez-Rodríguez David Vilares









#### How do SoTA parsers work?



State-of-the-art System

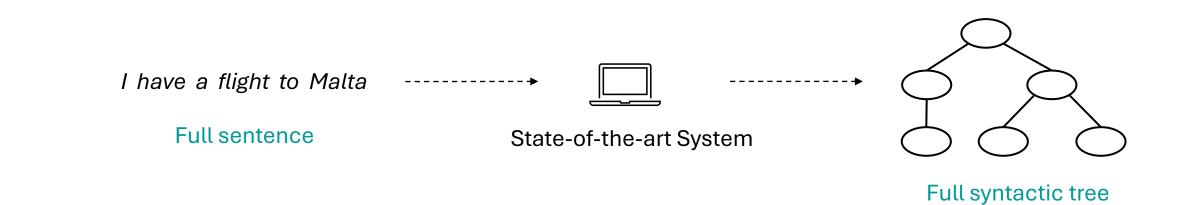
I have a flight to Malta ------

		L
ш	I	

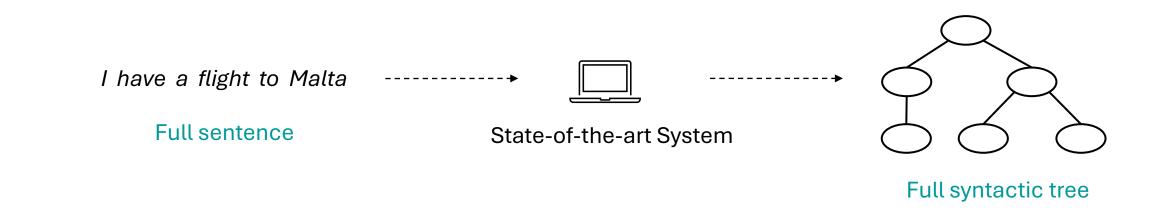
Full sentence

State-of-the-art System

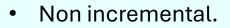
#### How do SoTA parsers work?



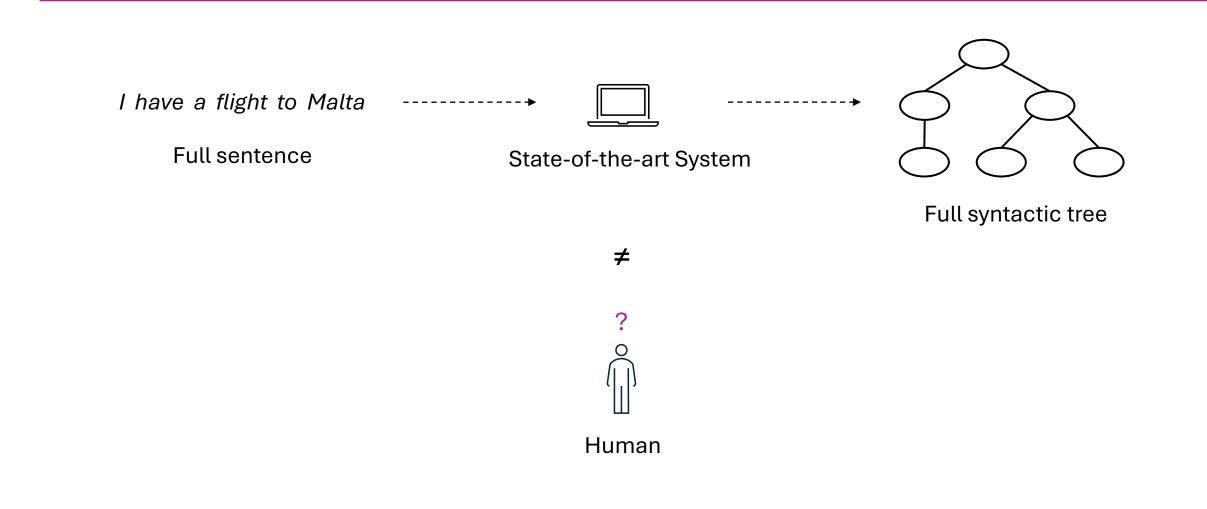
#### How do SoTA parsers work?

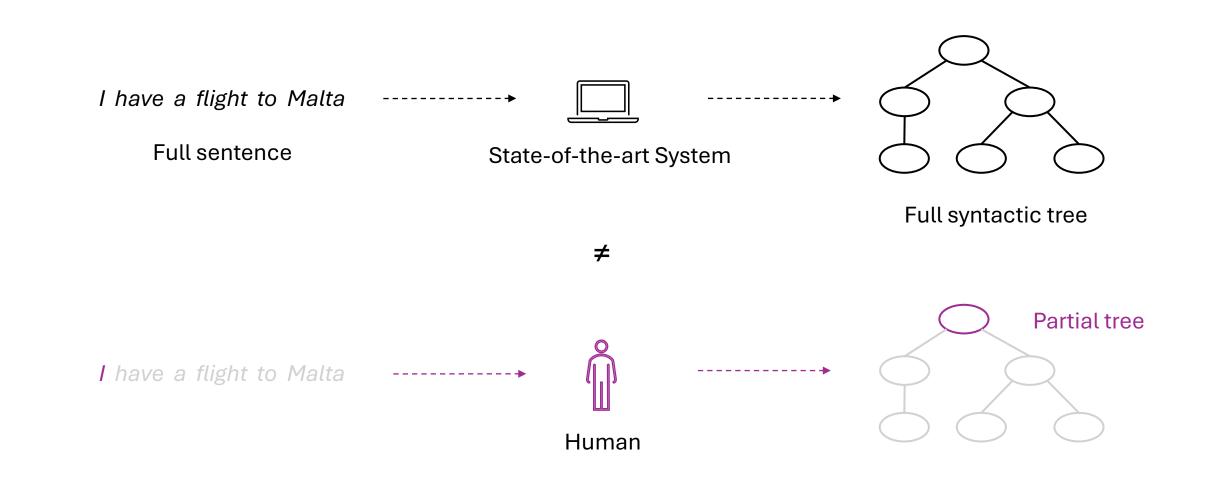


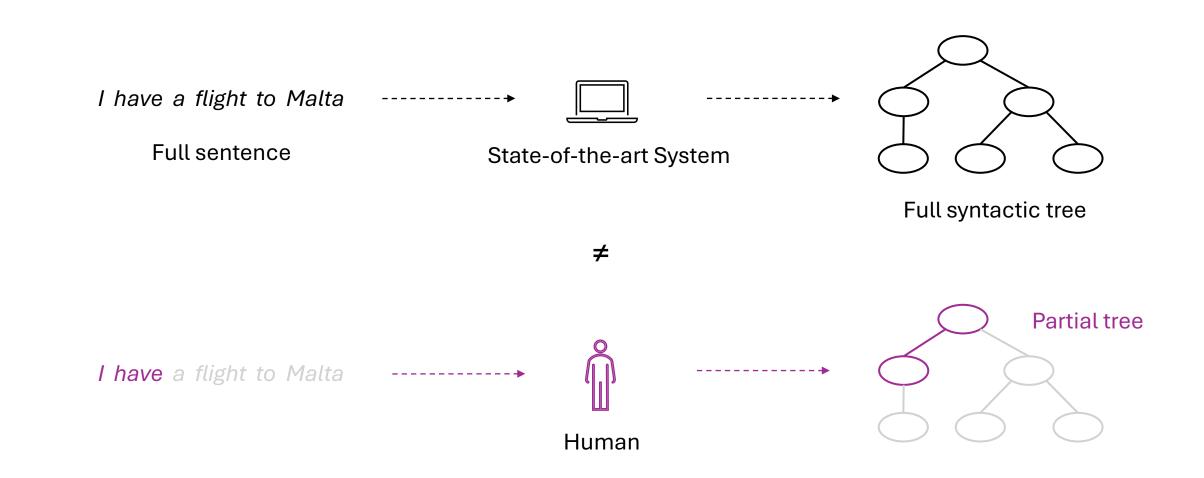


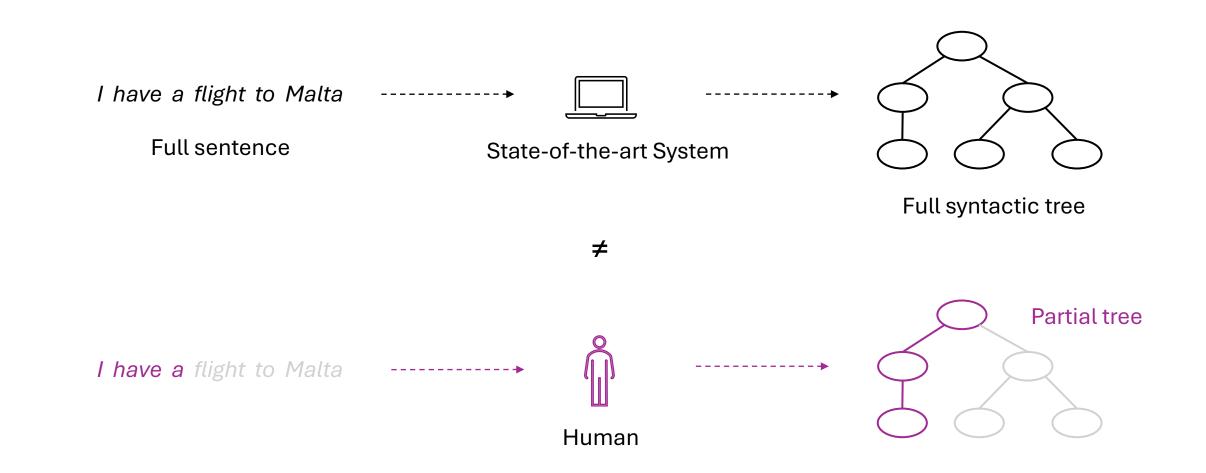


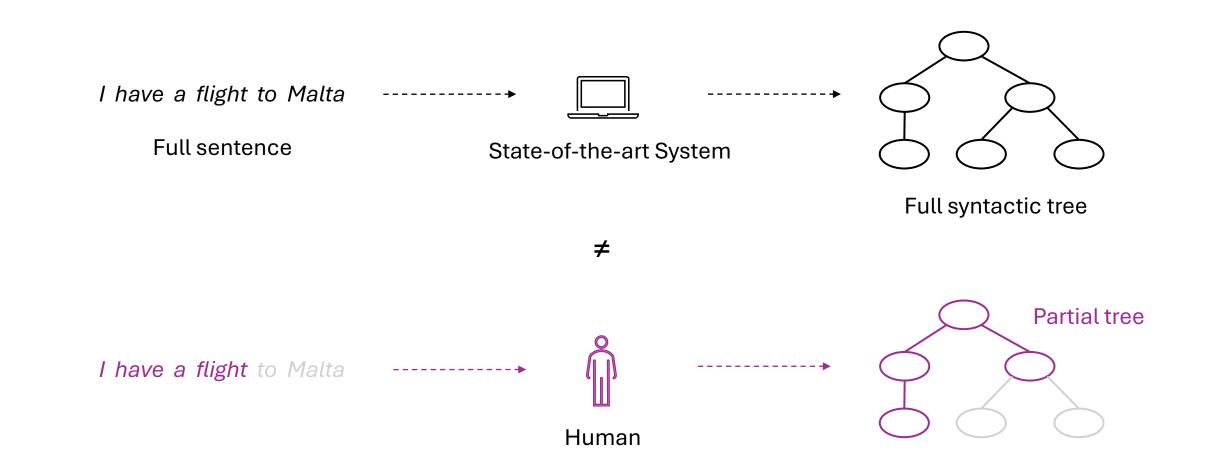
- Simultaneous access to all elements of the sentence.
- No reference of the information provided by each word to build the tree.

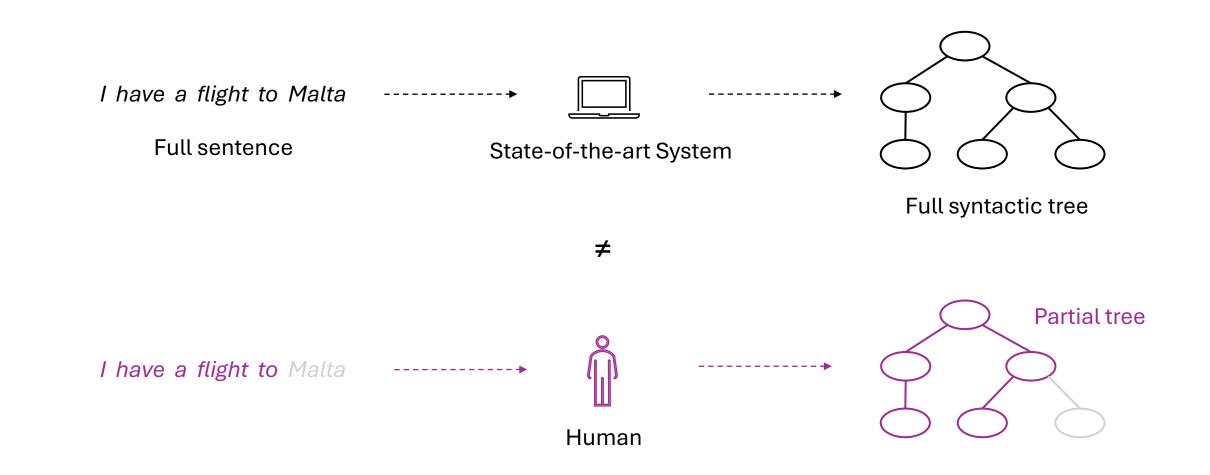


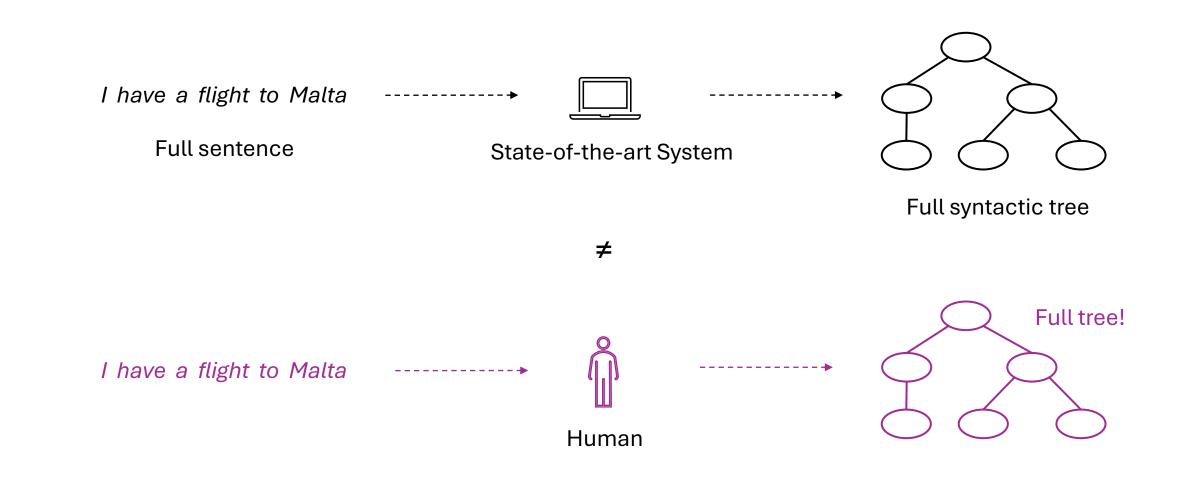


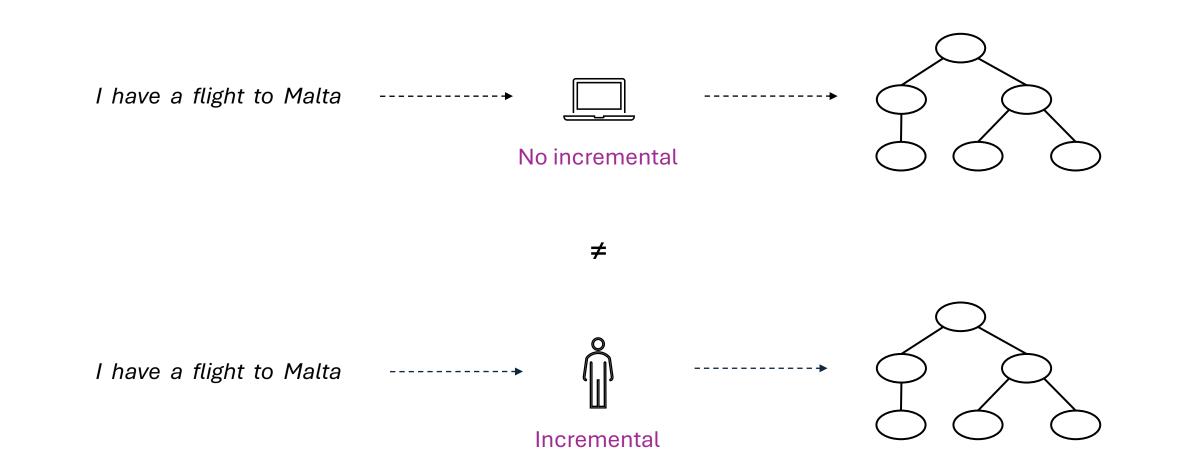








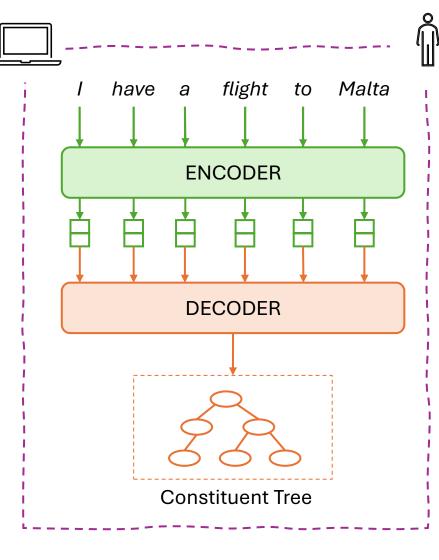




### Incremental Constituent Parsing

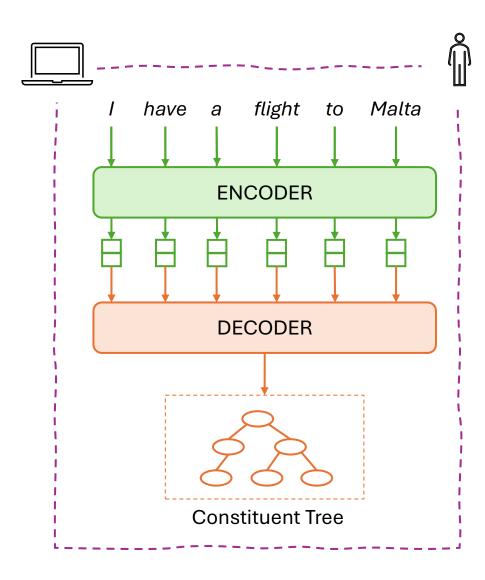
#### State-of-the-Art System

- Bidirectional encoder:
  - BERT & ELMo.
- Non-incremental decoder.
  - Kitaev & Klein (2018).



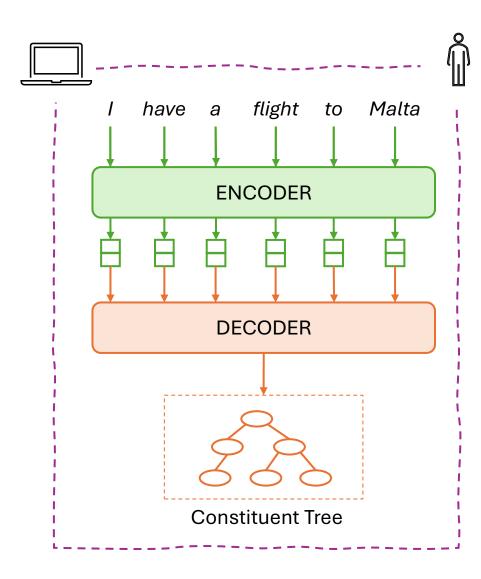
#### **Incremental System**

- Unidirectional encoder:
  - GPT & LSTMs.
- Incremental decoder.
  - Transition-based.
  - Sequence labeling.



Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

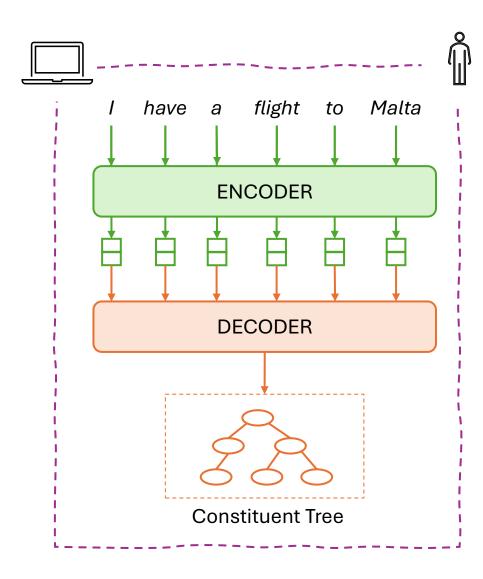


Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

How incremental?

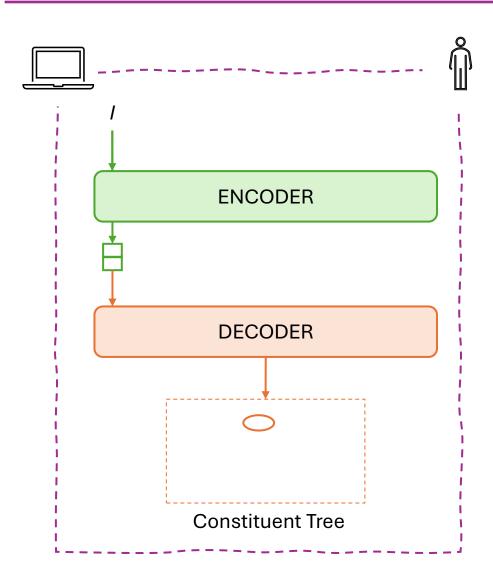




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

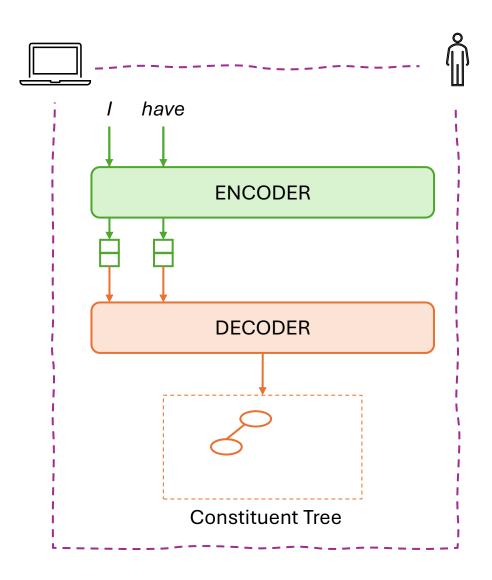




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

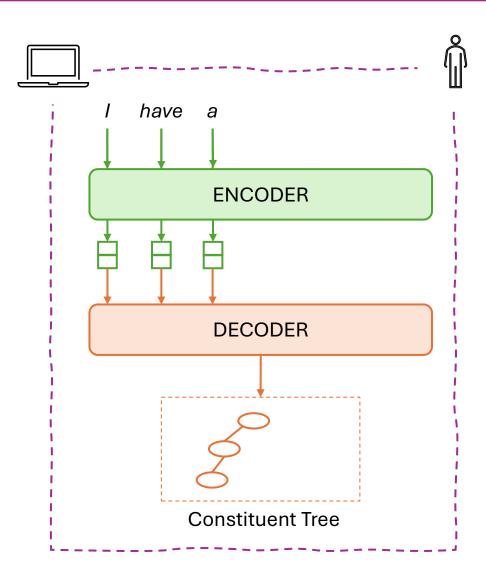




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

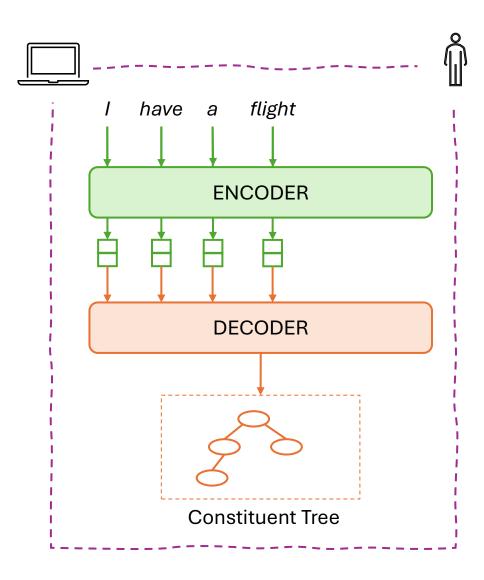




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

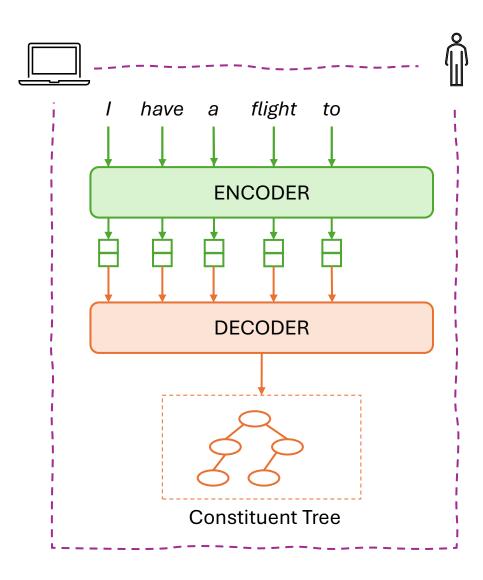




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

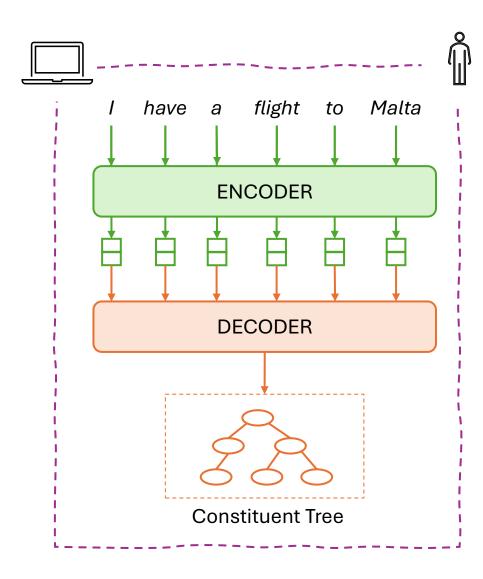




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

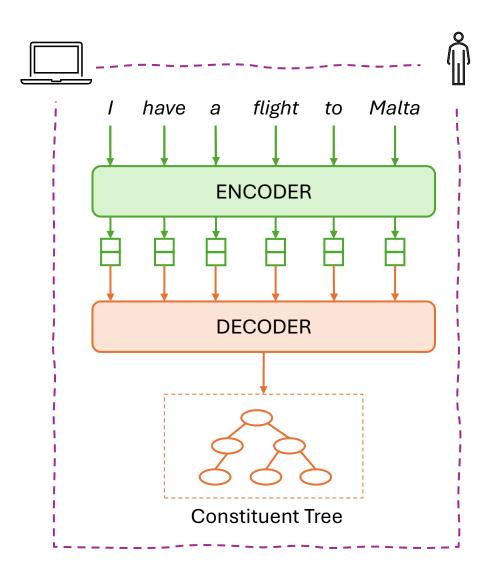




Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).





Incremental decoder

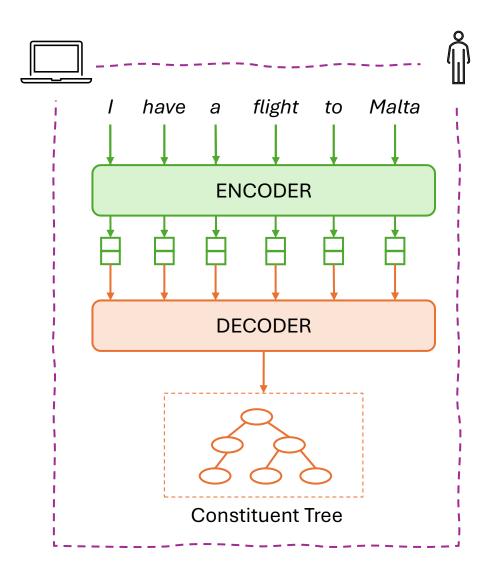
- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k (by default, k = 0).
- Incrementally encode each word  $w_i$  with  $w_1, \dots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.

What happens if delay k > 0?





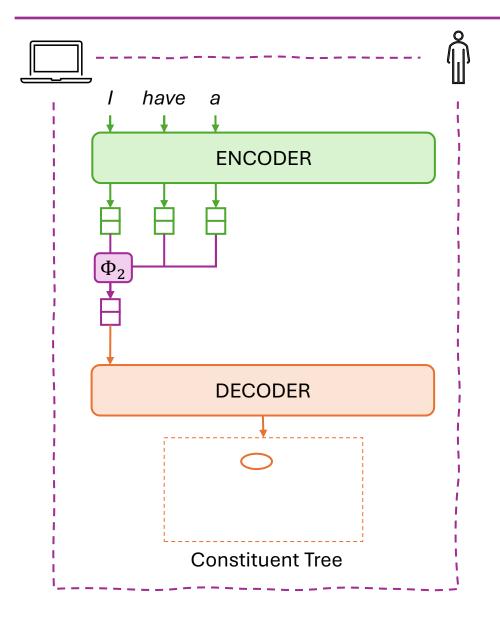
Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k (by default, k = 0).
- Incrementally encode each word  $w_i$  with  $w_1, \ldots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.





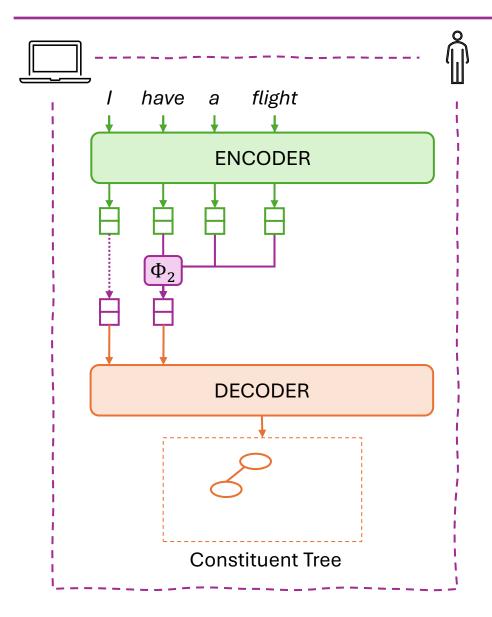
#### Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k (by default, k = 0).
- Incrementally encode each word  $w_i$  with  $w_1, \dots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.





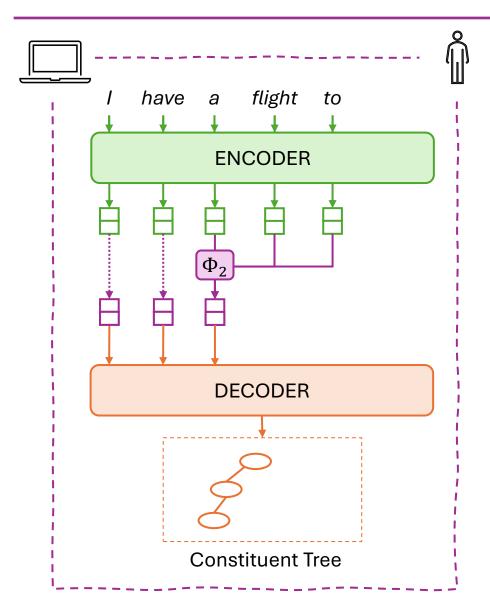
#### Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k.
- Incrementally encode each word  $w_i$  with  $w_1, \ldots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.





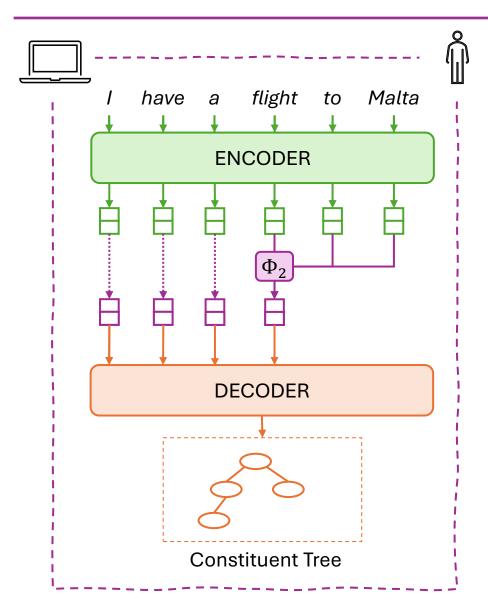
Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k.
- Incrementally encode each word  $w_i$  with  $w_1, \dots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.





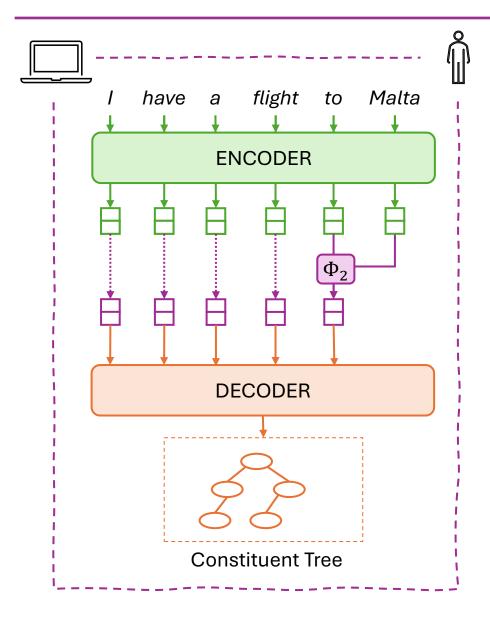
**Incremental decoder** 

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k.
- Incrementally encode each word  $w_i$  with  $w_1, \dots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.





#### Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

- Parameter k.
- Incrementally encode each word  $w_i$  with  $w_1, \dots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.



							ŝ
	, _↓	have ↓	a ↓	flight ↓	to ↓	Malta ↓	ш ¦
			EN	CODER	ł		
						Φ <sub>2</sub>	
			DE				
			Cons	tituent	С Гree		

#### Incremental decoder

- Attach-Juxtapose from Yang & Deng (2020).
- Sequence Labeling from Gómez-Rodríguez & Vilares (2018).

#### **Delayed incremental processing**

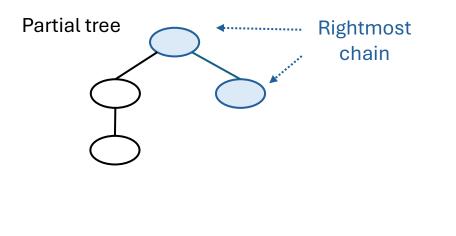
- Parameter *k*.
- Incrementally encode each word  $w_i$  with  $w_1, \dots, w_{i+k}$ .
- In practice:  $\Phi_k(\mathbf{h}_i \cdots \mathbf{h}_{i+k})$  where  $\Phi$  is a feed-forward network.

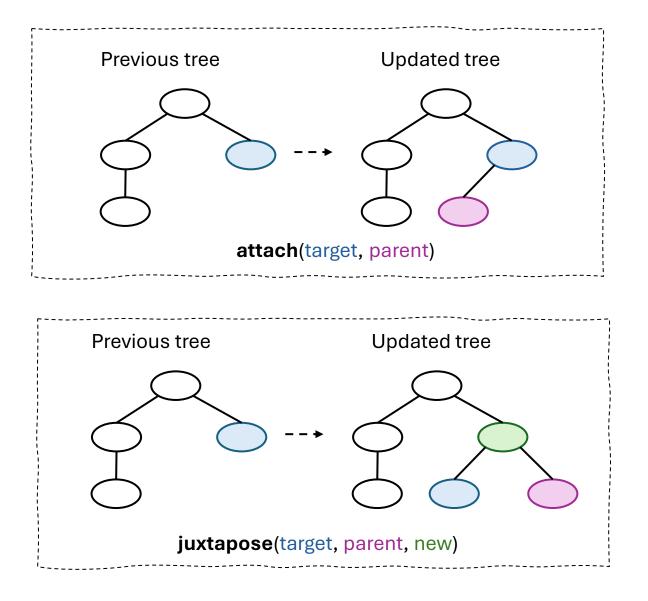


- Transition-based system.
- Two actions: **attach** & **juxtapose**.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.

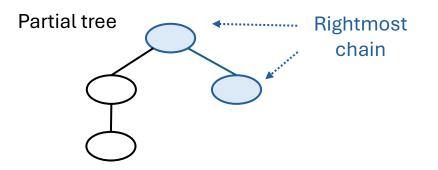


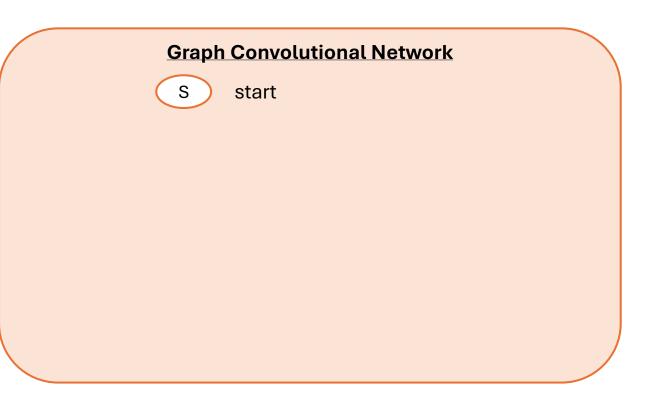


- Transition-based system.
- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.





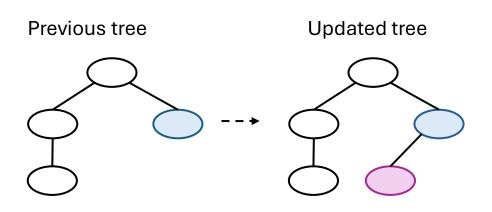
Transition-based system. • Two actions: attach & juxtapose. ٠ Sentence of *n* words to *n* transitions. ٠ **Graph Convolutional Network**  $W_1, \ldots, W_n \rightarrow t_1, \ldots, t_n$ attach(S, NP) S Graph Convolutional Network (GCN). ٠ NP Append subtrees to the rightmost chain. ٠ Updated tree Previous tree

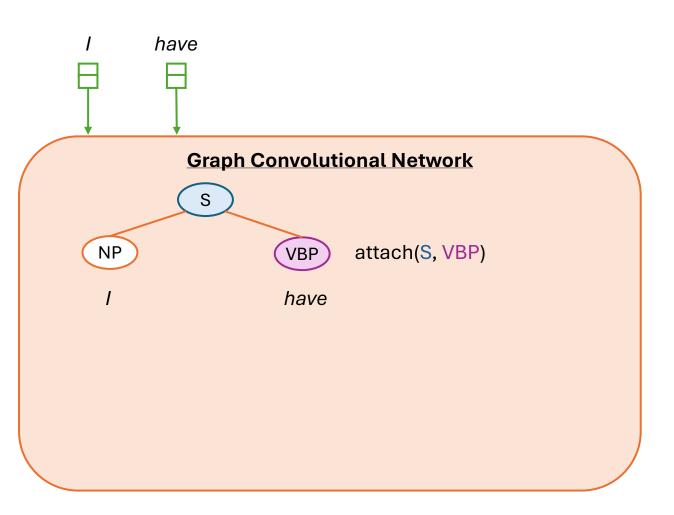
attach(target, parent)

- Transition-based system.
- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.





attach(target, parent)

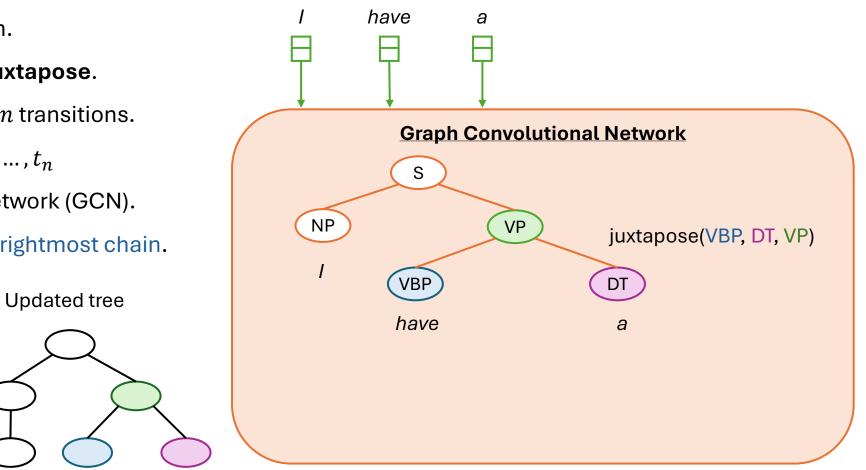
• Transition-based system.

Previous tree

- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.

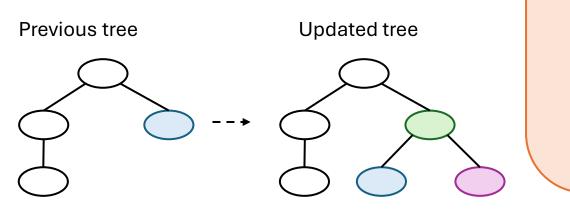


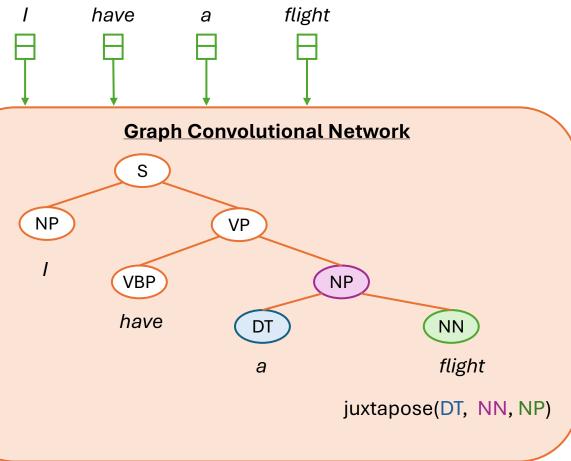
juxtapose(target, parent, new)

- Transition-based system.
- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.



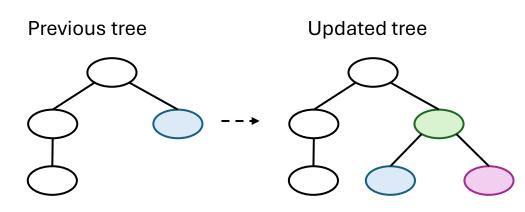


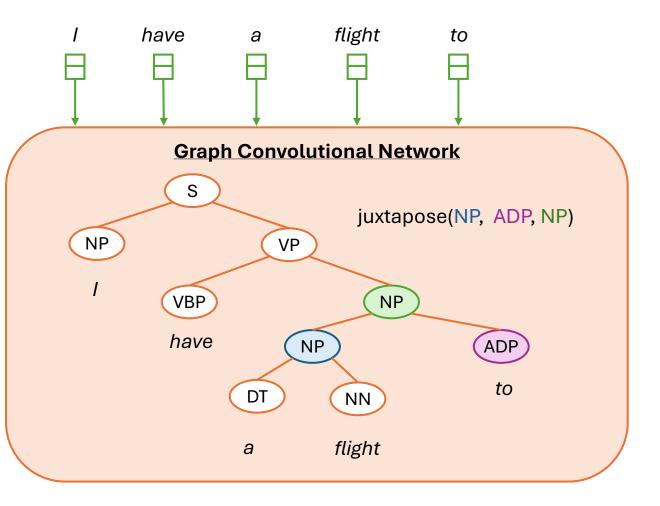
juxtapose(target, parent, new)

- Transition-based system.
- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.



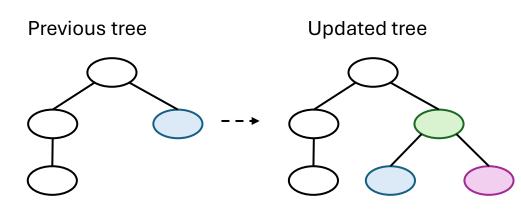


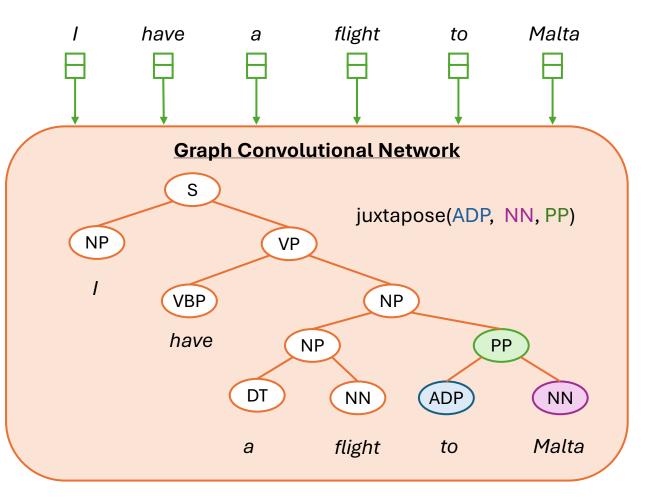
juxtapose(target, parent, new)

- Transition-based system.
- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.



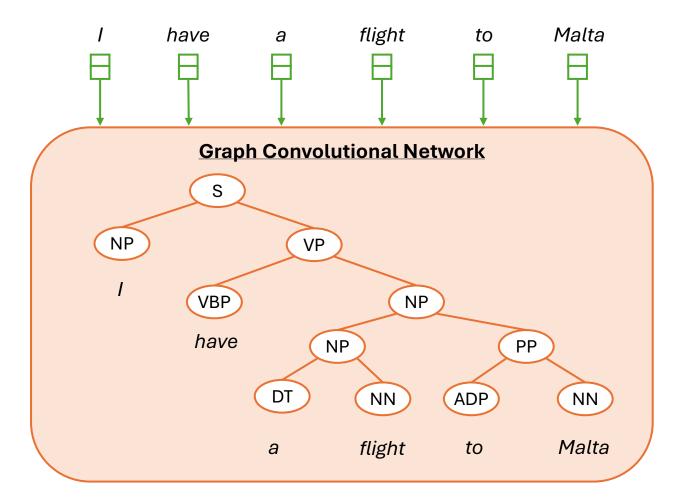


juxtapose(target, parent, new)

- Transition-based system.
- Two actions: attach & juxtapose.
- Sentence of *n* words to *n* transitions.

 $w_1, \ldots, w_n \to t_1, \ldots, t_n$ 

- Graph Convolutional Network (GCN).
- Append subtrees to the rightmost chain.

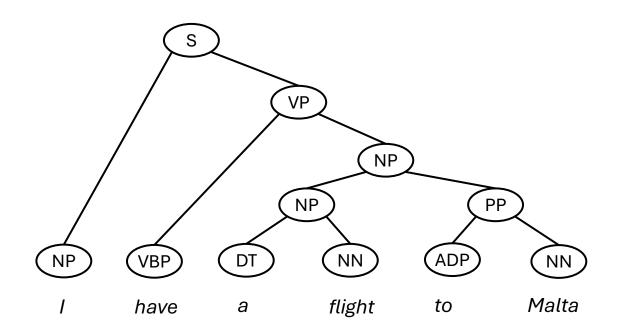


Full constituent tree!

- Sequence labeling method.
- Sentence of *n* words to *n* labels.

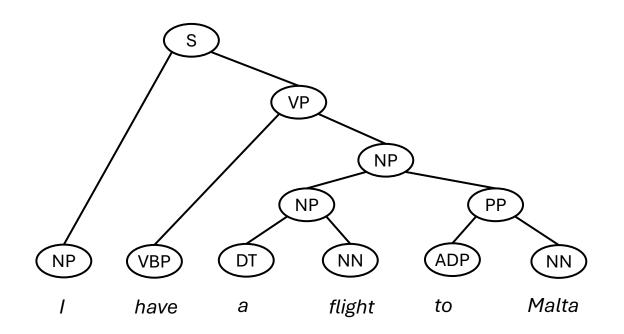
 $w_1, \dots, w_n \to \ell_1, \dots, \ell_n$ 

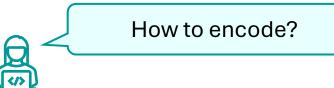
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .



- Sequence labeling method.
- Sentence of *n* words to *n* labels.

- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .



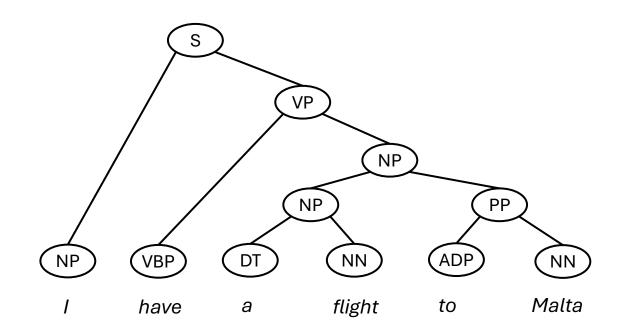


- Sequence labeling method.
- Sentence of *n* words to *n* labels.

 $w_1, \dots, w_n \to \ell_1, \dots, \ell_n$ 

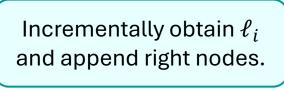
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

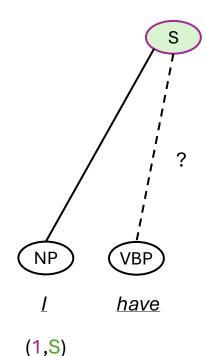
Incrementally obtain  $\ell_i$  and append right nodes.



- Sequence labeling method.
- Sentence of *n* words to *n* labels.

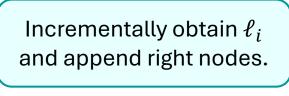
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

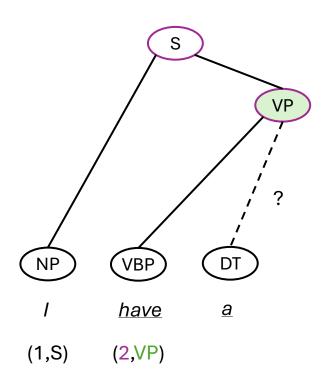




- Sequence labeling method.
- Sentence of *n* words to *n* labels.

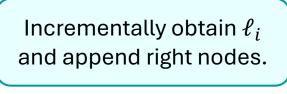
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

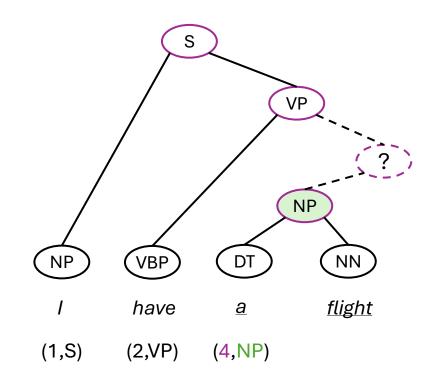




- Sequence labeling method.
- Sentence of *n* words to *n* labels.

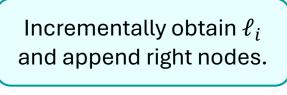
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

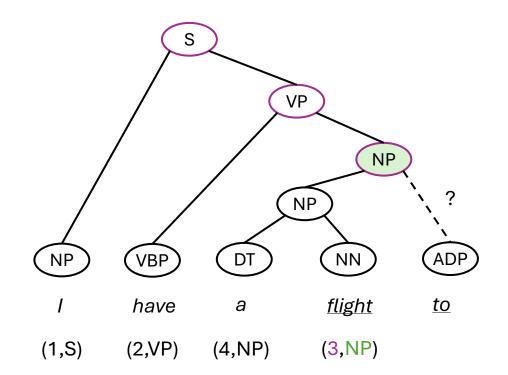




- Sequence labeling method.
- Sentence of *n* words to *n* labels.

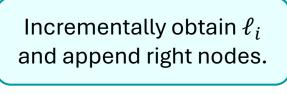
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

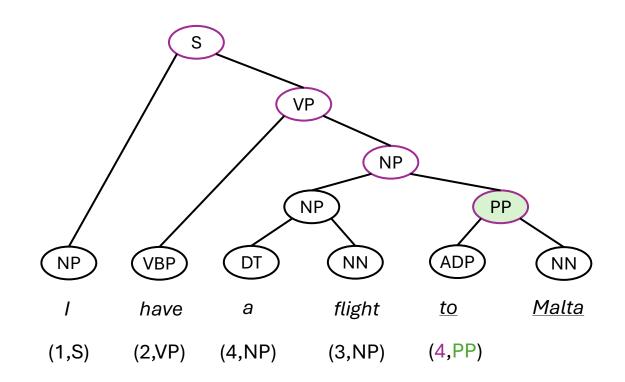




- Sequence labeling method.
- Sentence of *n* words to *n* labels.

- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .



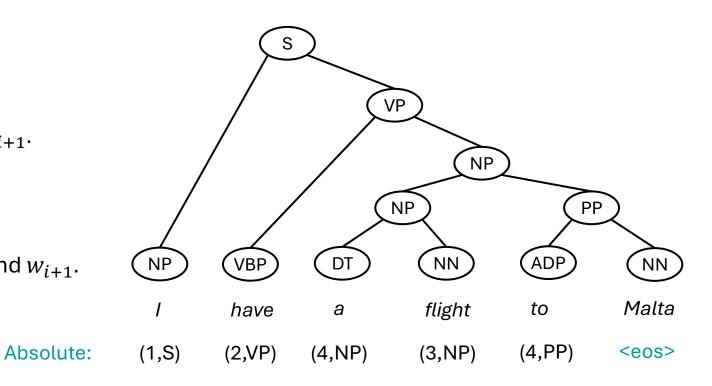


- Sequence labeling method.
- Sentence of *n* words to *n* labels.

 $w_1,\ldots,w_n \to \ell_1,\ldots,\ell_n$ 

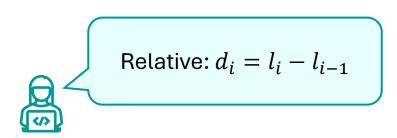
- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

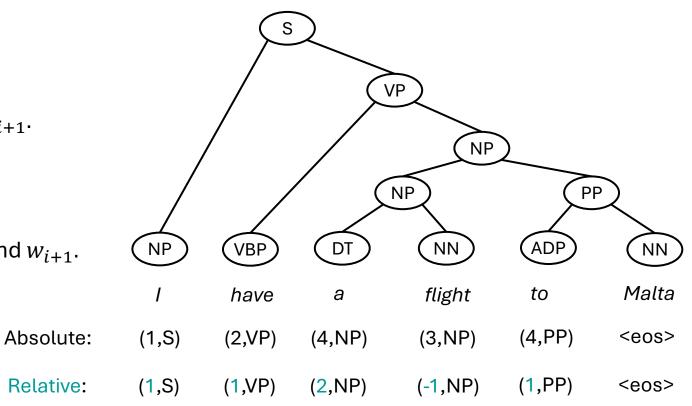
Absolute encoding!



- Sequence labeling method.
- Sentence of *n* words to *n* labels.

- Each label has 2 components:  $\ell_i = (d_i, c_i)$ .
  - $l_i$ : # common constituents of  $w_i$  and  $w_{i+1}$ .
  - Absolute:  $d_i = l_i$ .
  - Relative:  $d_i = l_i l_{i-1}$ .
  - $c_i$ : lowest common constituent of  $w_i$  and  $w_{i+1}$ .
- Two feed forward networks  $(d_i, c_i)$ .

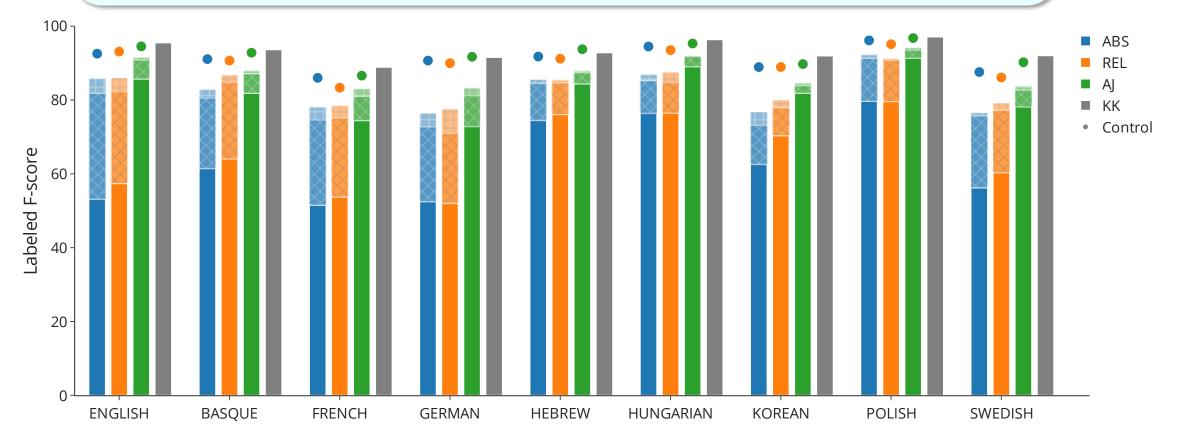




## **Experiments**

- **Multilingual benchmark:** PTB + SPMRL (wo. Arabic).
- Baseline: Kitaev & Klein (2018).
  - Bidirectional encoder: XLM-RoBERTa.
  - Non-incremental decoder: span-based.
- Encoders:
  - Bidirectional: 4-BiLSTM, XLM-RoBERTa.
  - Unidirectional: 4-LSTM, BLOOM-560M, mGPT.
- **Delay experiments:** k = 0, 1, 2.

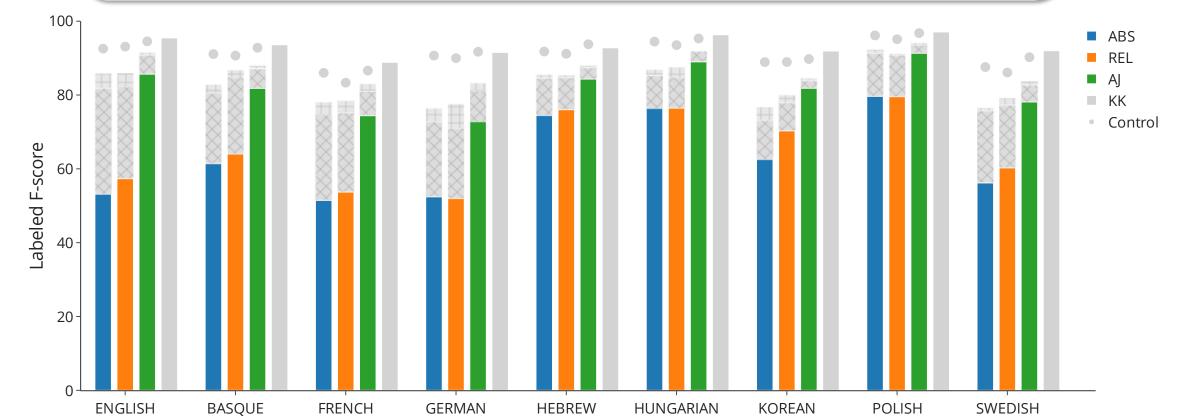
- Incremental (mGPT): absolute (■), relative (■) and attach-juxtapose (■).
- Control (XLM): absolute (
  ), relative (
  ) and attach-juxtapose (
  ).
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- **Delay 1** ( $\boxtimes$ ,  $\boxtimes$ ,  $\boxtimes$ ) and **Delay 2** ( $\boxplus$ ,  $\boxplus$ ,  $\boxplus$ ).





• Incremental (mGPT): absolute (■), relative (■) and attach-juxtapose (■).

- **Control** (XLM): absolute (●), relative (●) and attach-juxtapose (●).
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- **Delay 1**  $(\boxtimes, \boxtimes, \boxtimes)$  and **Delay 2**  $(\boxplus, \boxplus, \boxplus)$ .



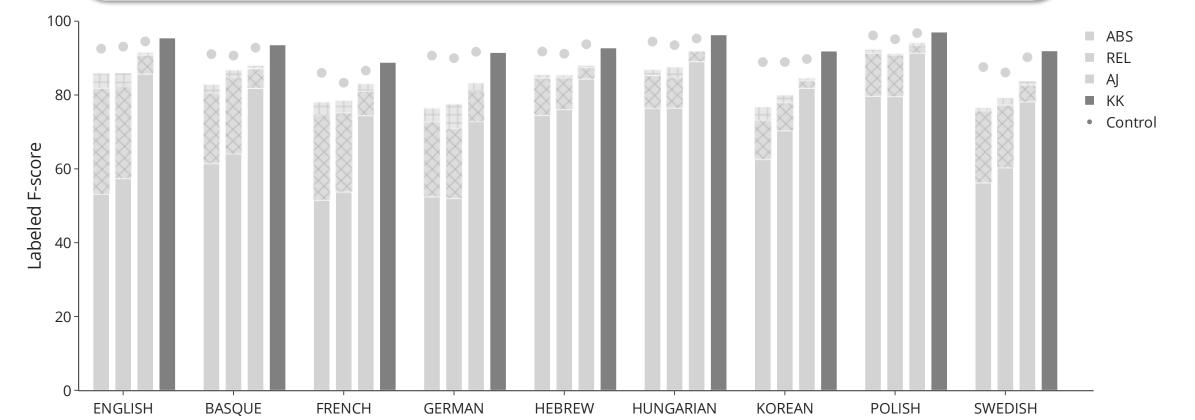


- Incremental (mGPT): absolute (■), relative (■) and attach-juxtapose (■).
- **Control** (XLM): absolute (**•**), relative (**•**) and attach-juxtapose (**•**).
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- Delay 1 ( $\boxtimes$ ,  $\boxtimes$ ,  $\boxtimes$ ) and Delay 2 ( $\boxplus$ ,  $\boxplus$ ,  $\boxplus$ ).

0 -	ENGLISH	BASQUE	FRENCH	GERMAN	HEBREW	HUNGARIAN	KOREAN	POLISH	SWEDISH	
20-										
Labeled F-score										
- 100 80 بو										<ul> <li>ABS</li> <li>REL</li> <li>AJ</li> <li>KK</li> <li>Control</li> </ul>

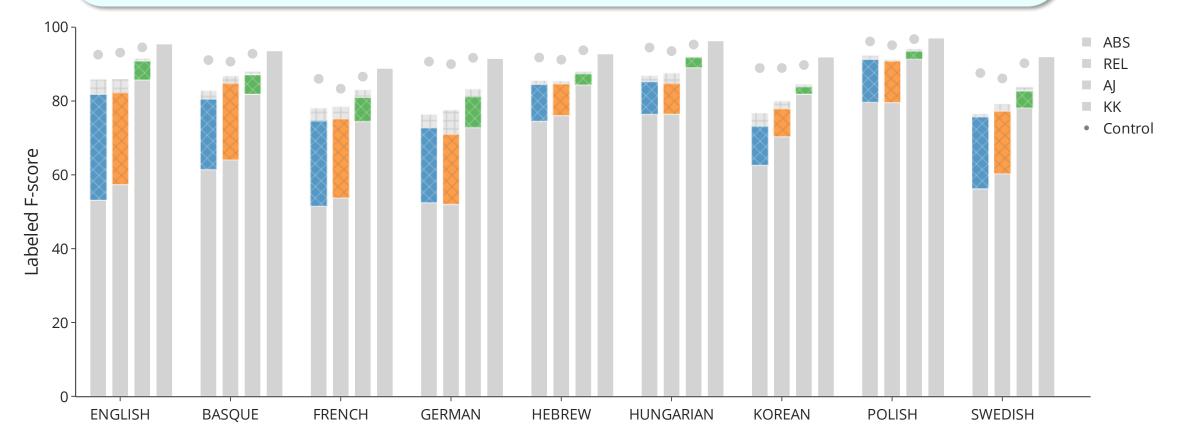


- Incremental (mGPT): absolute (■), relative (■) and attach-juxtapose (■).
- **Control** (XLM): absolute  $(\bullet)$ , relative  $(\bullet)$  and attach-juxtapose  $(\bullet)$ .
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- **Delay 1**  $(\boxtimes, \boxtimes, \boxtimes)$  and **Delay 2**  $(\boxplus, \boxplus, \boxplus)$ .





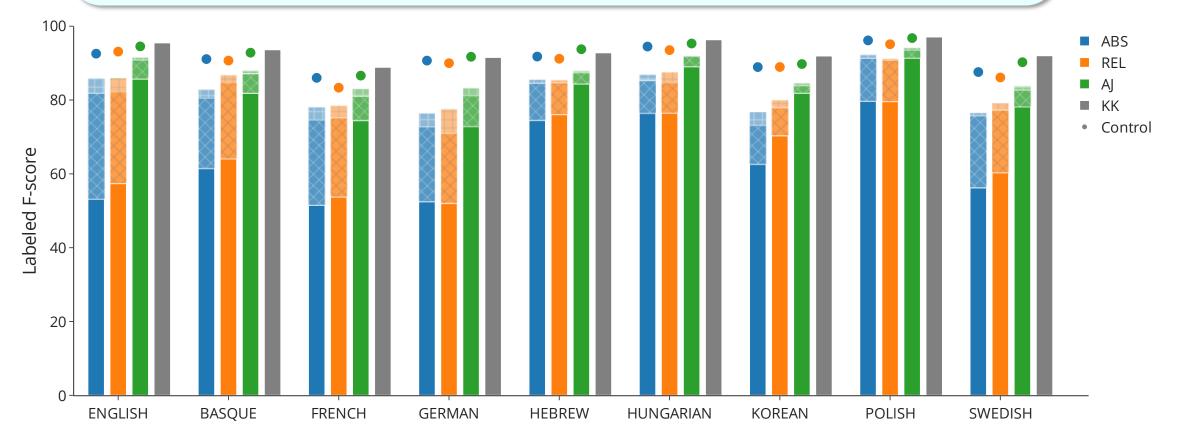
- Incremental (mGPT): absolute  $(\blacksquare)$ , relative  $(\blacksquare)$  and attach-juxtapose  $(\blacksquare)$ .
- **Control** (XLM): absolute (●), relative (●) and attach-juxtapose (●).
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- Delay 1 ( $\boxtimes$ ,  $\boxtimes$ ,  $\boxtimes$ ) and Delay 2 ( $\boxplus$ ,  $\boxplus$ ,  $\boxplus$ ).





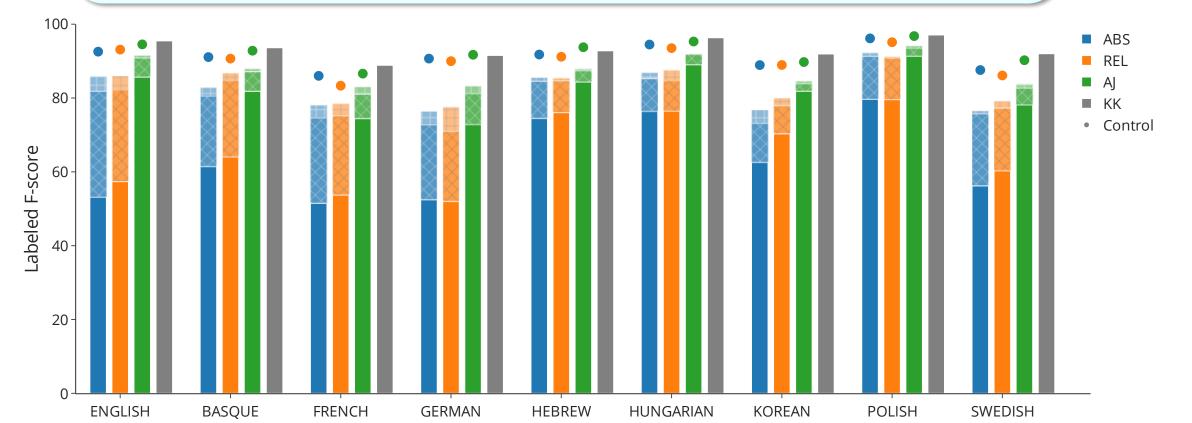
- Incremental (mGPT): absolute (■), relative (■) and attach-juxtapose (■).
- **Control** (XLM): absolute (●), relative (●) and attach-juxtapose (●).
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- Delay 1 ( $\boxtimes$ ,  $\boxtimes$ ,  $\boxtimes$ ) and Delay 2 ( $\boxplus$ ,  $\boxplus$ ,  $\boxplus$ ).

- Incremental (mGPT): absolute (■), relative (■) and attach-juxtapose (■).
- Control (XLM): absolute (
  ), relative (
  ) and attach-juxtapose (
  ).
- Non-incremental (XLM): Kitaev & Klein, 2018 (■).
- **Delay 1** ( $\boxtimes$ ,  $\boxtimes$ ,  $\boxtimes$ ) and **Delay 2** ( $\boxplus$ ,  $\boxplus$ ,  $\boxplus$ ).



### Conclusions

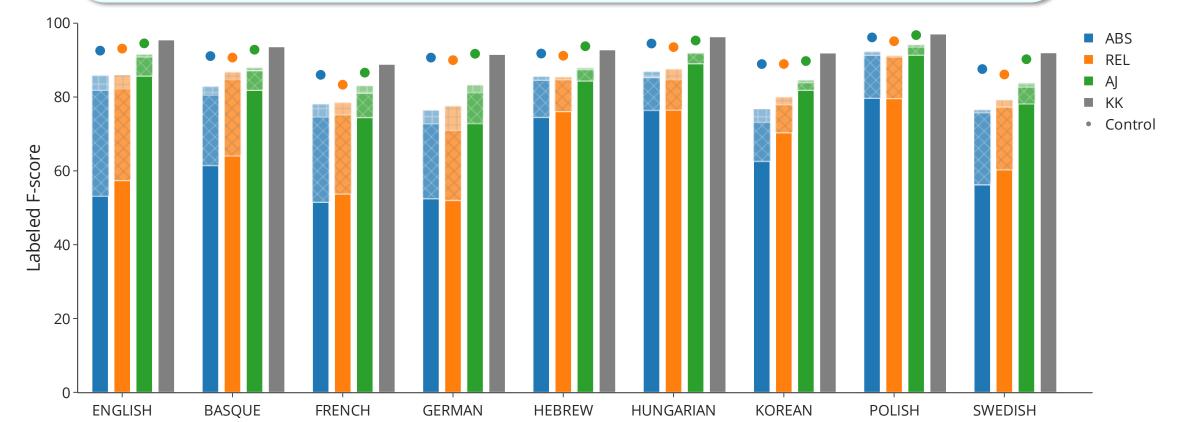
- 1. Control parsers ( $\bigcirc \bigcirc \bigcirc$ )  $\approx$  Kitaev & Klein, 2018 ( $\blacksquare$ ).
  - Meaning? State-of-the-art relies on a bidirectional encoder.
- 2. Incremental parsers (= = =) considerably worse than Control (< = =) and KK (=).
  - But introducing delay  $(\boxtimes, \boxtimes, \boxtimes)$  significantly improves the performance.



### Conclusions

</>

- 3. Sequence Labeling performance (■ ■) lags behind Attach-Juxtapose (■).
  - Attach-Juxtapose (■) relies on a powerful neural decoder (GCN).
  - Considering a larger decoder will improve the incremental results?
  - ABS (■) and REL (■) are more benefited of delayed processing than AJ (■).



## **Thanks for listening!**

#### Ana Ezquerro



#### Carlos Gómez-Rodríguez David Vilares





#### Acknowledgements:

This work has received funding by the European Research Council (ERC), under the Horizon Europe research and innovation programme (SALSA, grant agreement No 101100615), ERDF/MICINN-AEI (SCANNER-UDC, PID2020-113230RB-C21), Xunta de Galicia (ED431C 2020/11), Grant GAP (PID2022-139308OA-I00) funded by MCIN/AEI/10.13039/501100011033/ and by ERDF 'A way of making Europe', and Centro de Investigación de Galicia 'CITIC', funded by the Xunta de Galicia through the collaboration agreement between the Consellería de Cultura, Educación, Formación Profesional e Universidades and the Galician universities for the reinforcement of the research centers of the Galician University System (CIGUS).