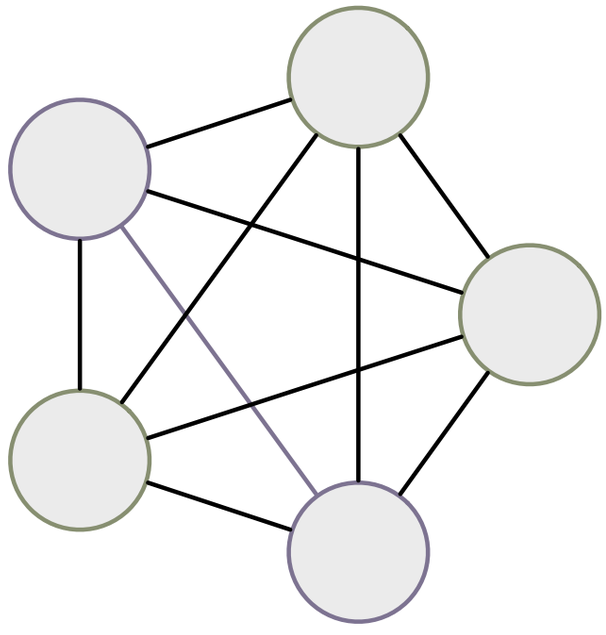
Joint part-of-speech and dependency projection from multiple sources

Anders Johannsen* Željko Agić* Anders Søgard

(formerly)* University of Copenhagen

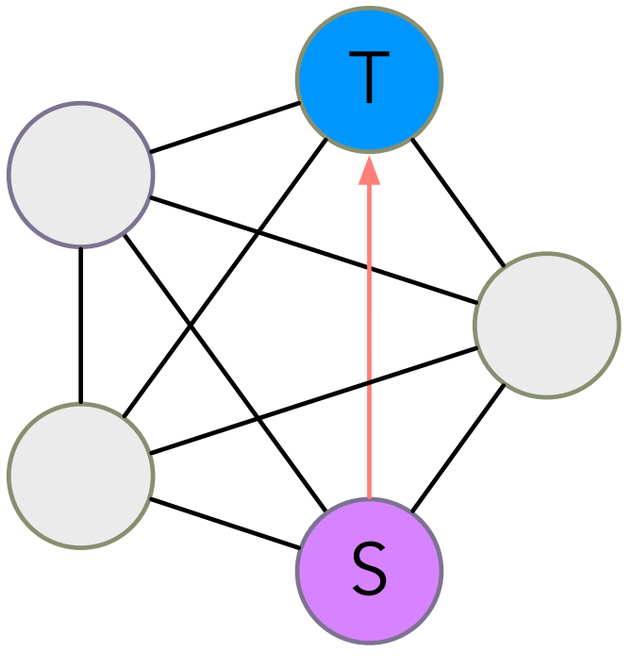
Annotation projection

Parallel corpora



Annotation projection

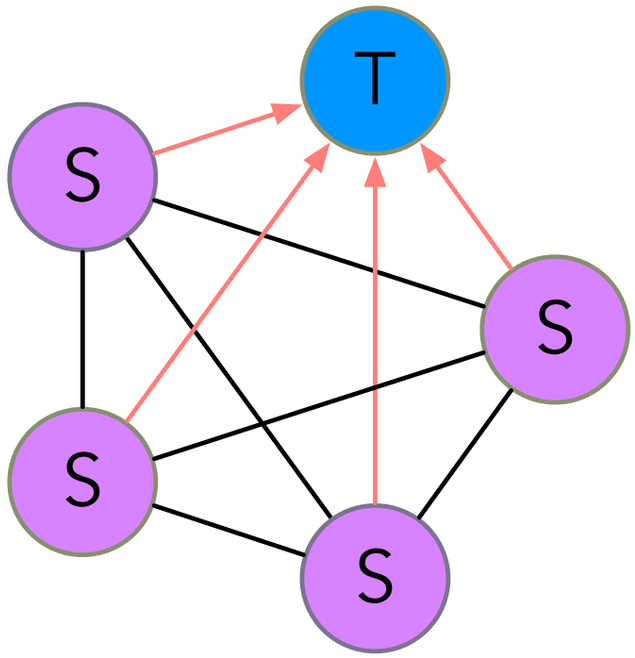
Parallel corpora



transfer annotation from source to target

Annotation projection

Parallel corpora

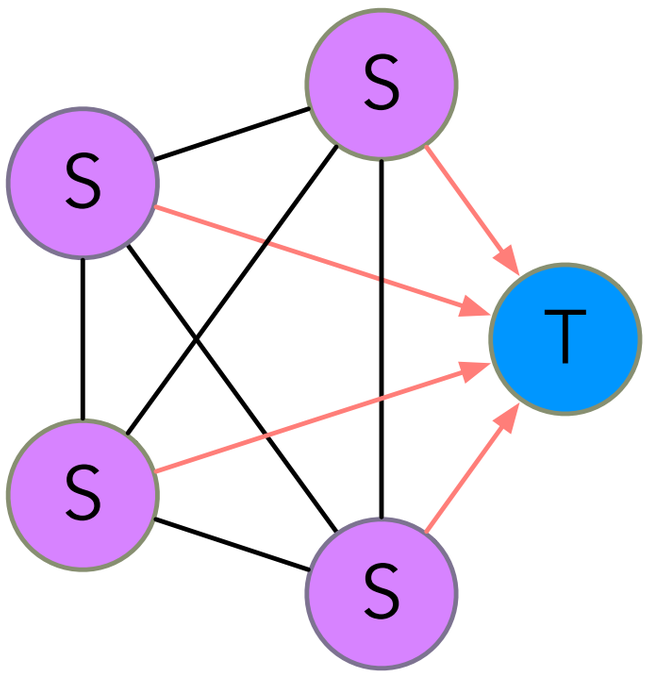


transfer annotation from source to target

may have multiple sources

Annotation projection

Parallel corpora



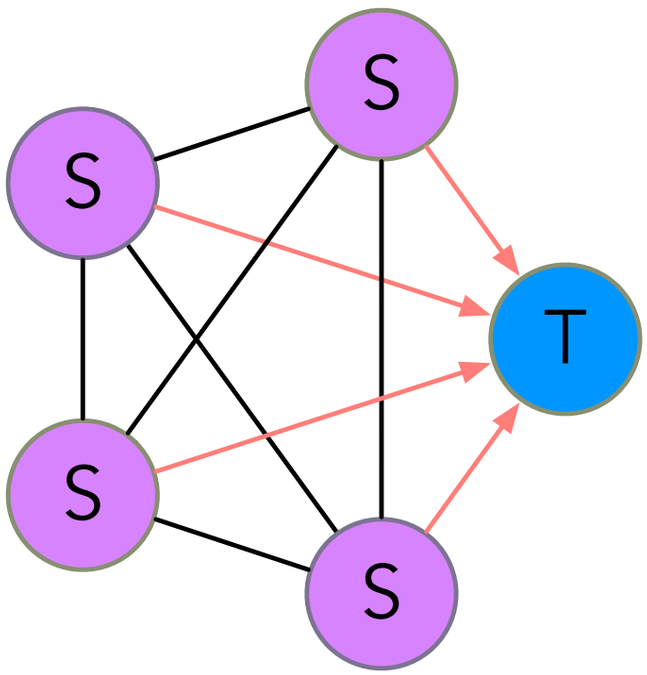
transfer annotation from source to target

may have multiple sources

evaluate by leave-one-out

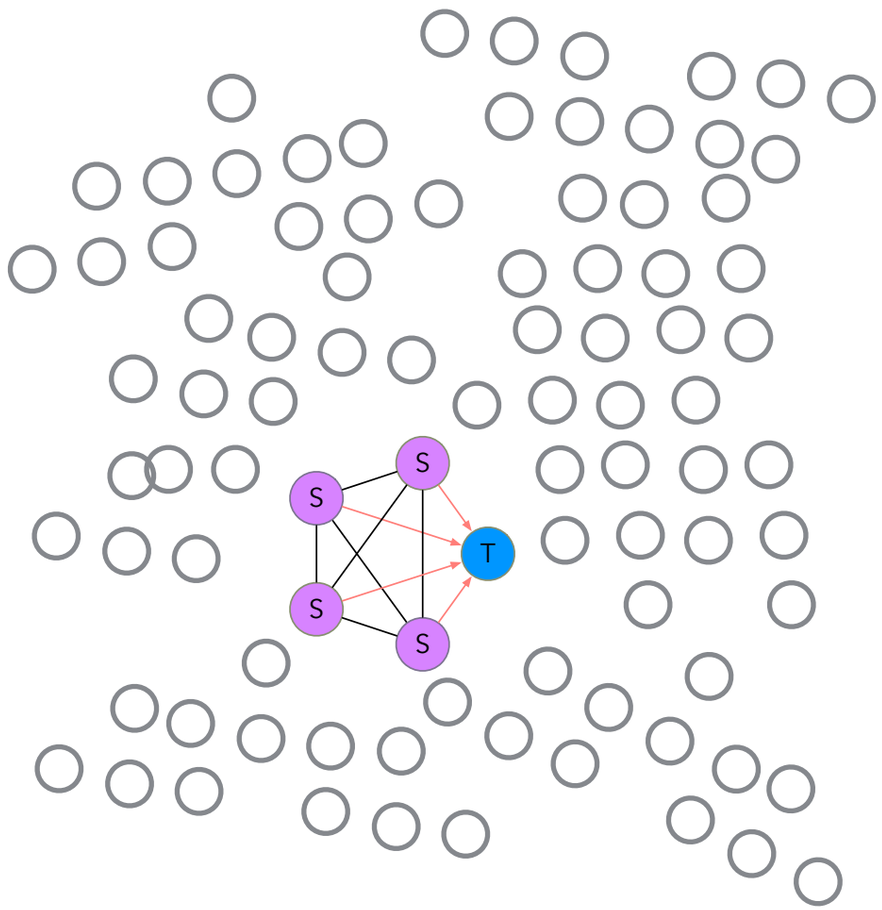
Annotation projection

Parallel corpora



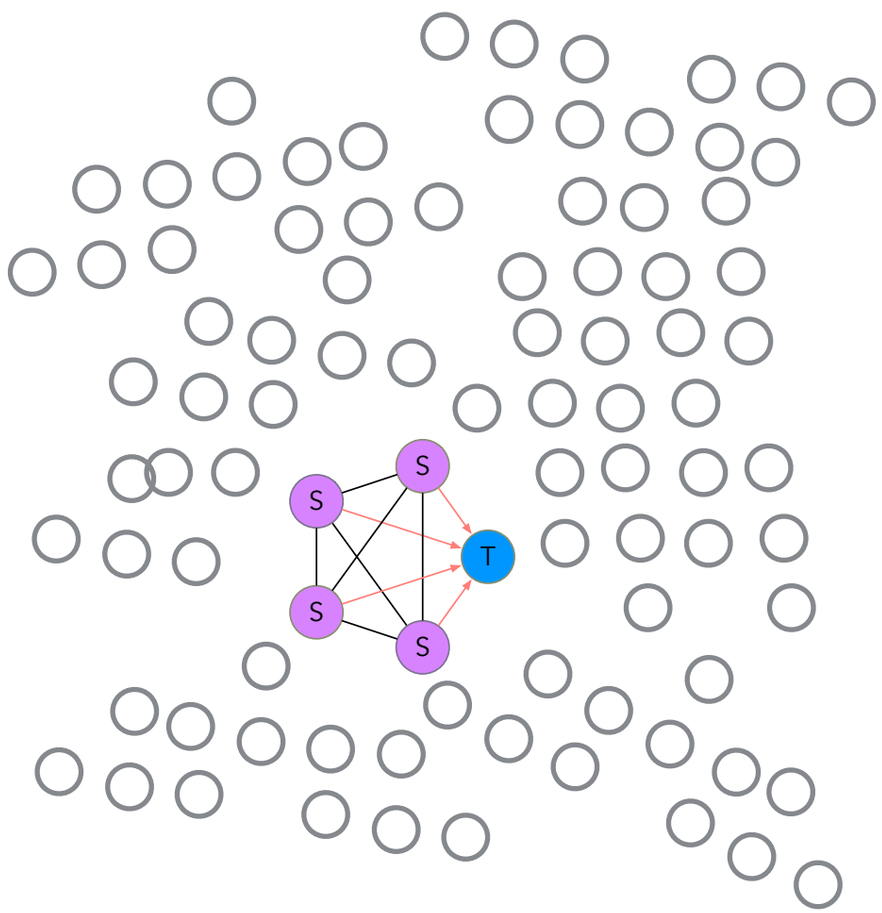
→ train parser → parse test set

The many languages of the world



cross-lingual parsing suffers a little from EUROPARLalism

The many languages of the world



cross-lingual parsing suffers a little from EUROPARLalism

This work extends Agić (2016):

train models for hundreds of languages

evaluate on 26 languages

Our corpora


BIBLIA
HEBRAICA
STUTTGARTENSIA
HOLY
BIBLE
Die Bibel
Die Heilige Schrift
SEPTUA
GINTA


THE WATCHTOWER
ANNOUNCING JEHOVAH'S KINGDOM
BORN AGAIN
SHAPES
VALUES?
GOOD NEWS
OF THE KINGDOM"
WHAT IS IT?
IS THE END N
GOD
WILL END ALL
SUFFERING!
WHEN? HOW?

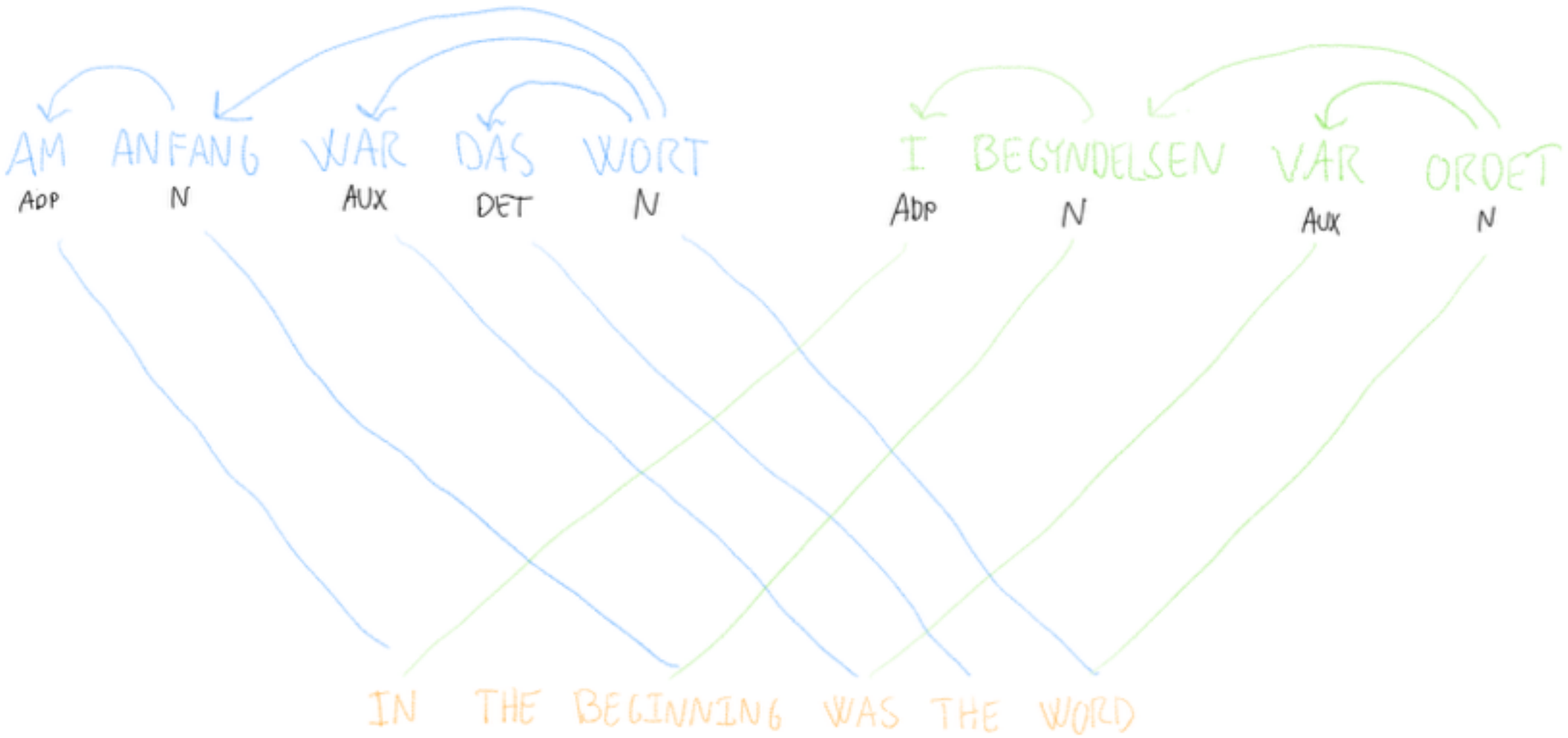
Our corpora


BIBLIA
HEBRAICA
STUTTGARTENSIA
HOLY
BIBLE
Die Bibel
Die Heilige Schrift
SEPTUA
GINTA


THE WATCHTOWER
ANNOUNCING JEHOVAH'S KINGDOM
BORN AGAIN
SHAPES
VALUES?
GOOD NEWS
OF THE KINGDOM"
WHAT IS IT?
IS THE END N
GOD
WILL END ALL
SUFFERING!
WHEN? HOW?


All we have to decide is what to do
with the time that is given to us.
Nosotros tenemos que decidir qué hacer
con el tiempo que se nos ha dado.

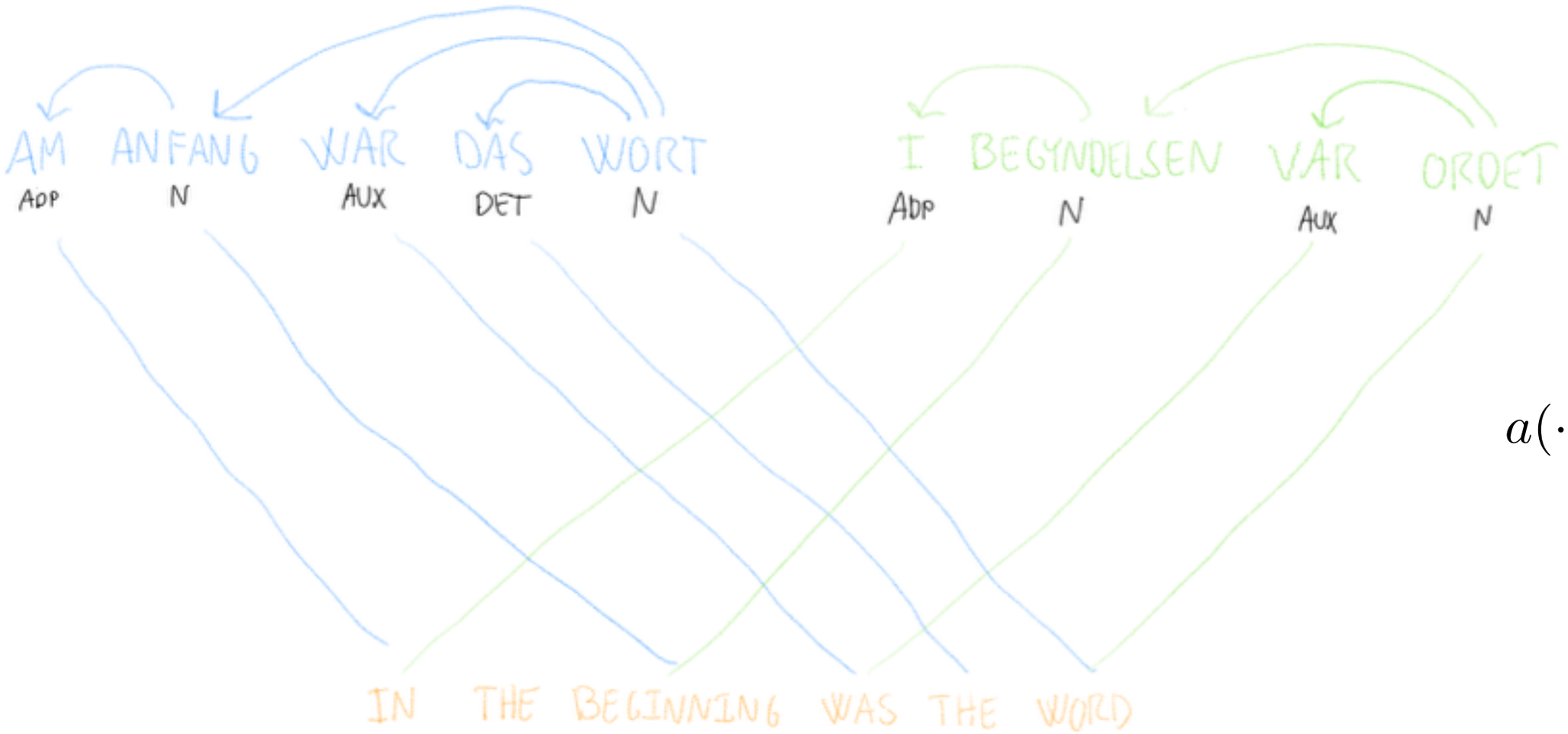
Example


AM ANFANG WAR DAS WORT
ADP N AUX DET N
I BEGYNDELSEN VAR ORDET
ADP N AUX N
IN THE BEGINNING WAS THE WORD

Example

score_{DE}

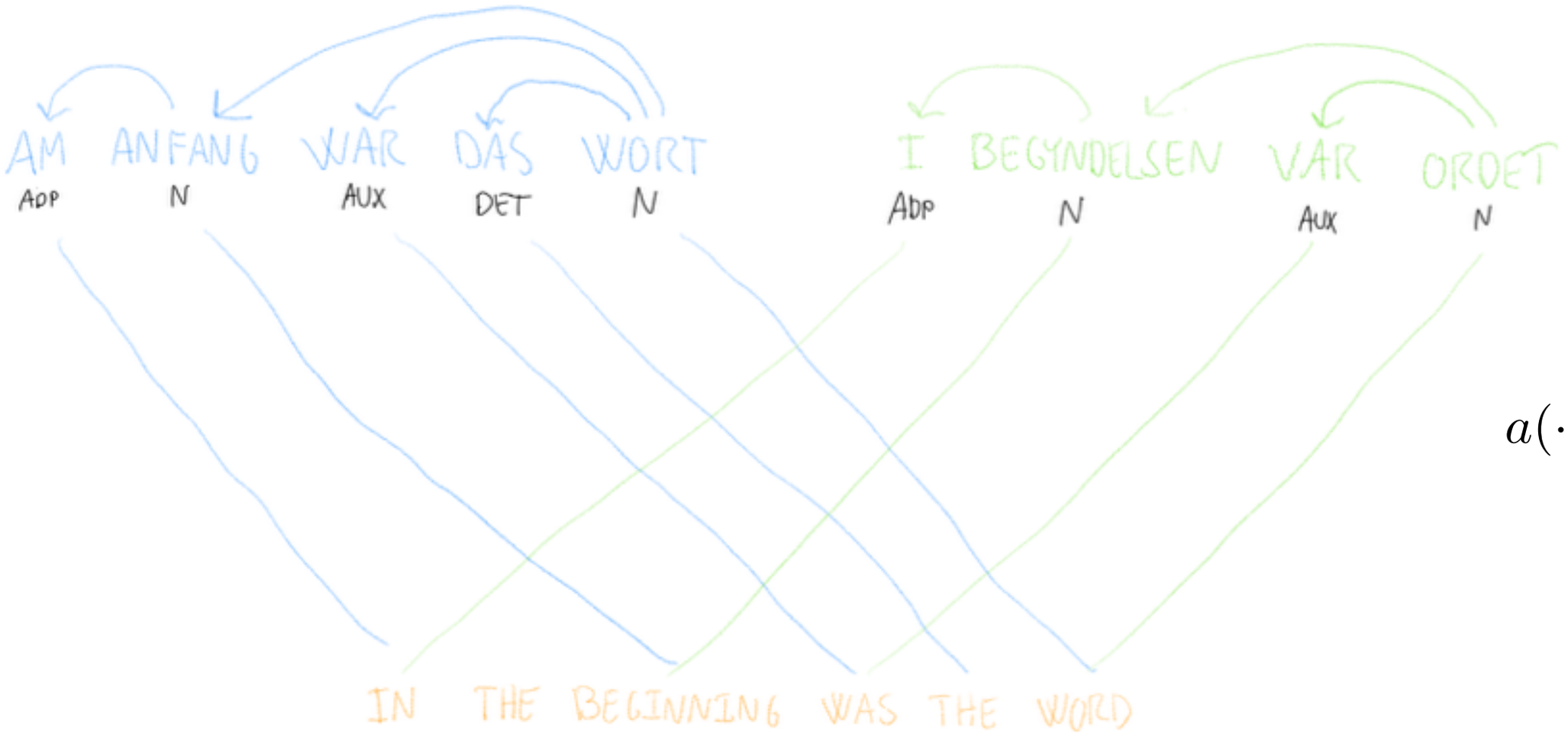
score_{DA}



$a(\cdot, \cdot)$

Example

score_{DE} score_{DA}

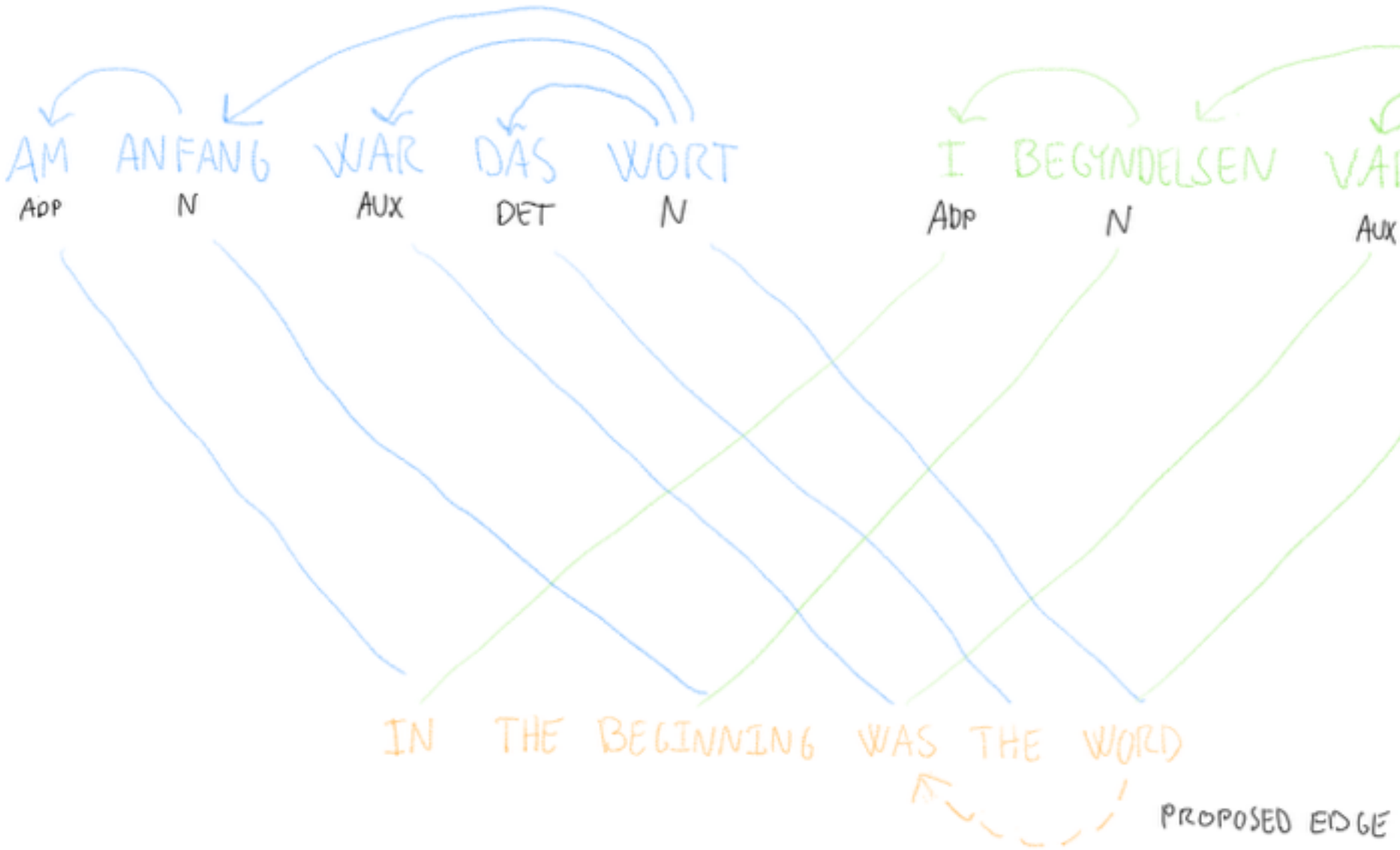


$a(\cdot, \cdot)$

$$\arg\max_{y} \sum_{(i,j) \in y} \text{score}_{\text{T}}(i, j) \quad \text{s.t. } y \text{ is a tree}$$

Example

score_{DE} score_{DA}

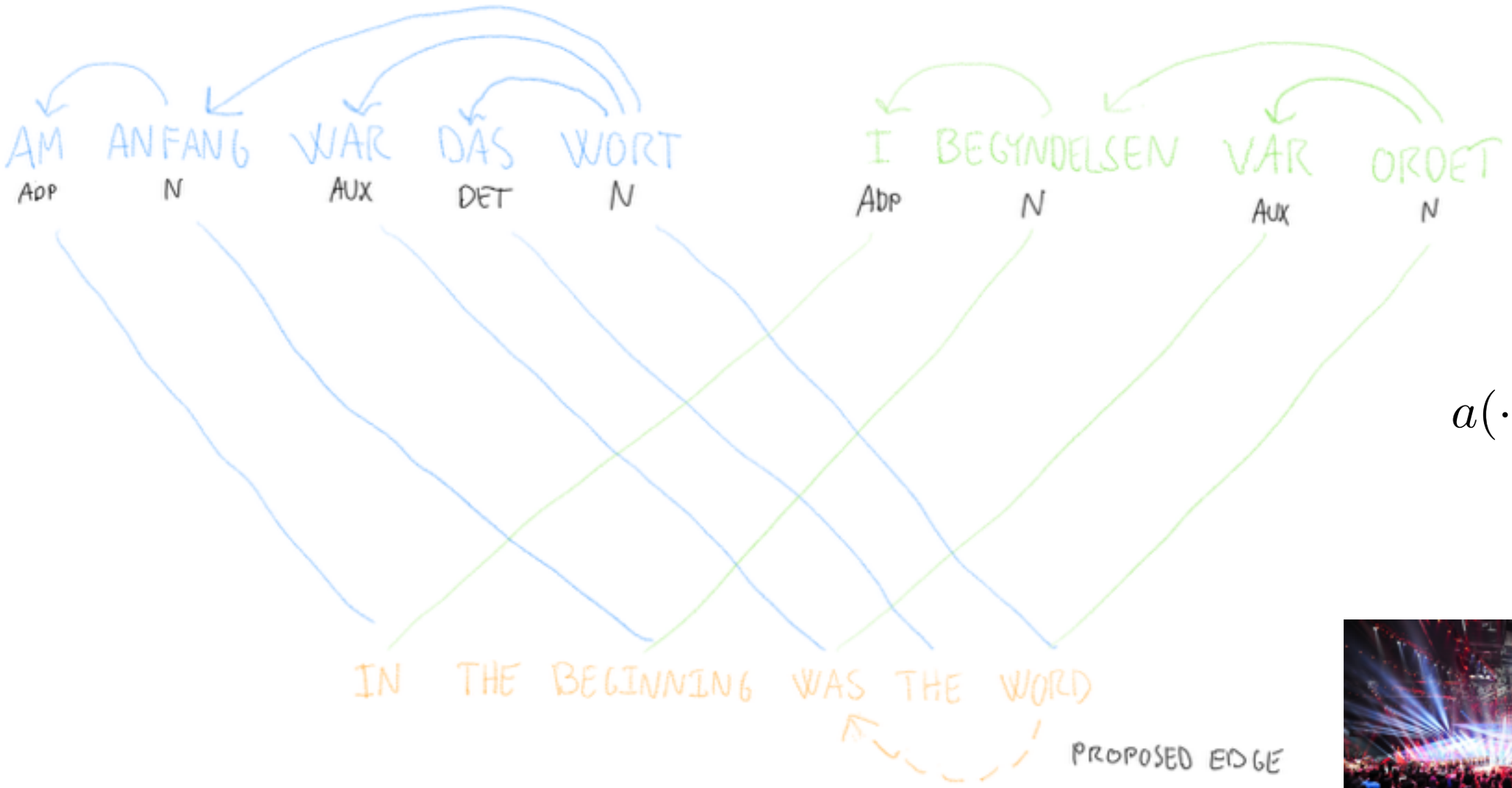


$a(\cdot, \cdot)$

$$\arg\max_{y} \sum_{(i,j) \in y} \text{score}_{\text{T}}(i, j) \quad \text{s.t. } y \text{ is a tree}$$

Example

score_{DE} score_{DA}

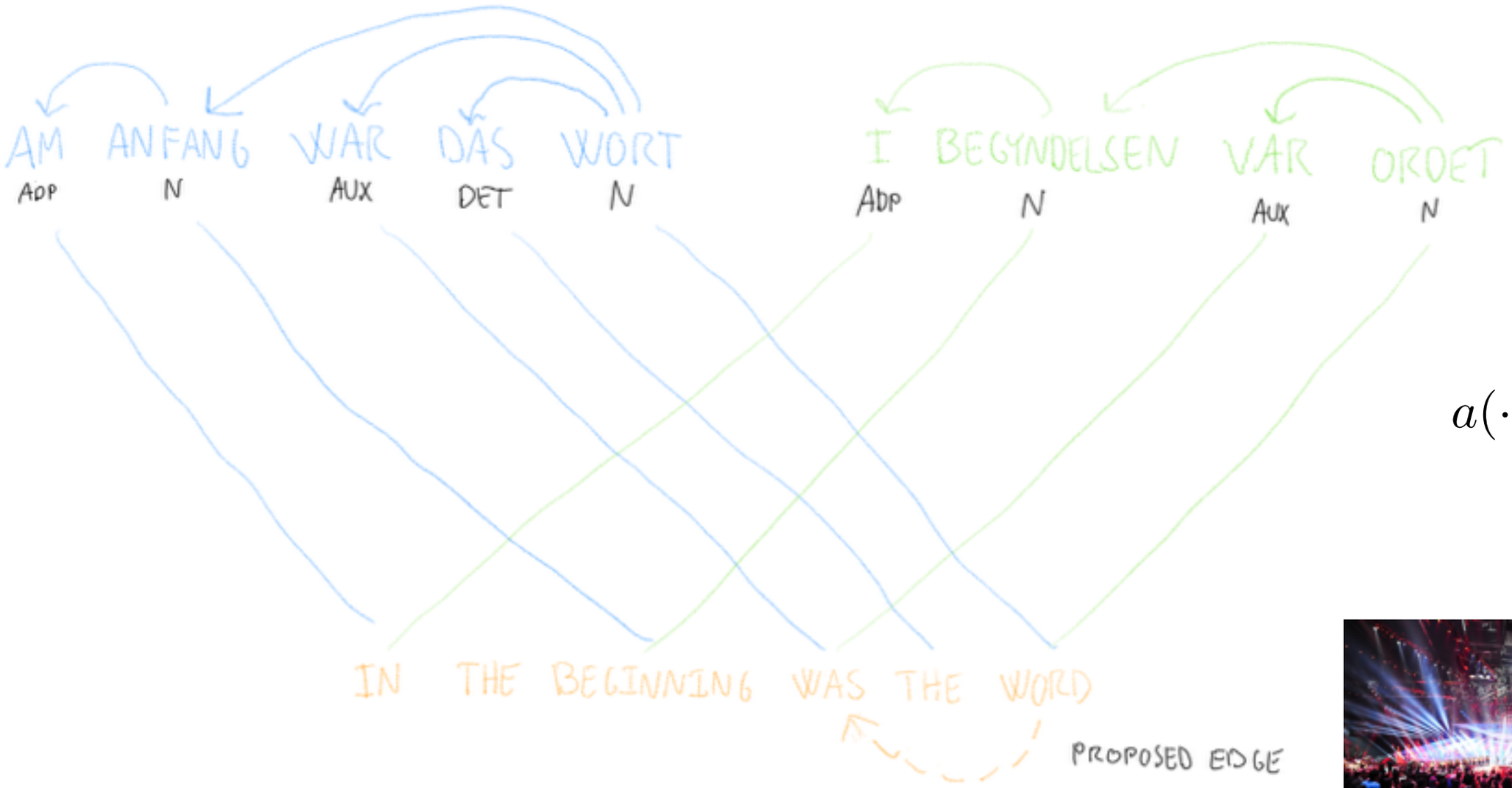


$a(\cdot, \cdot)$

$$\arg\max_{y} \sum_{(i,j)\in y} \text{score}_{\text{T}}(i, j) \quad \text{s.t. } y \text{ is a tree}$$

Example

score_{DE} score_{DA}



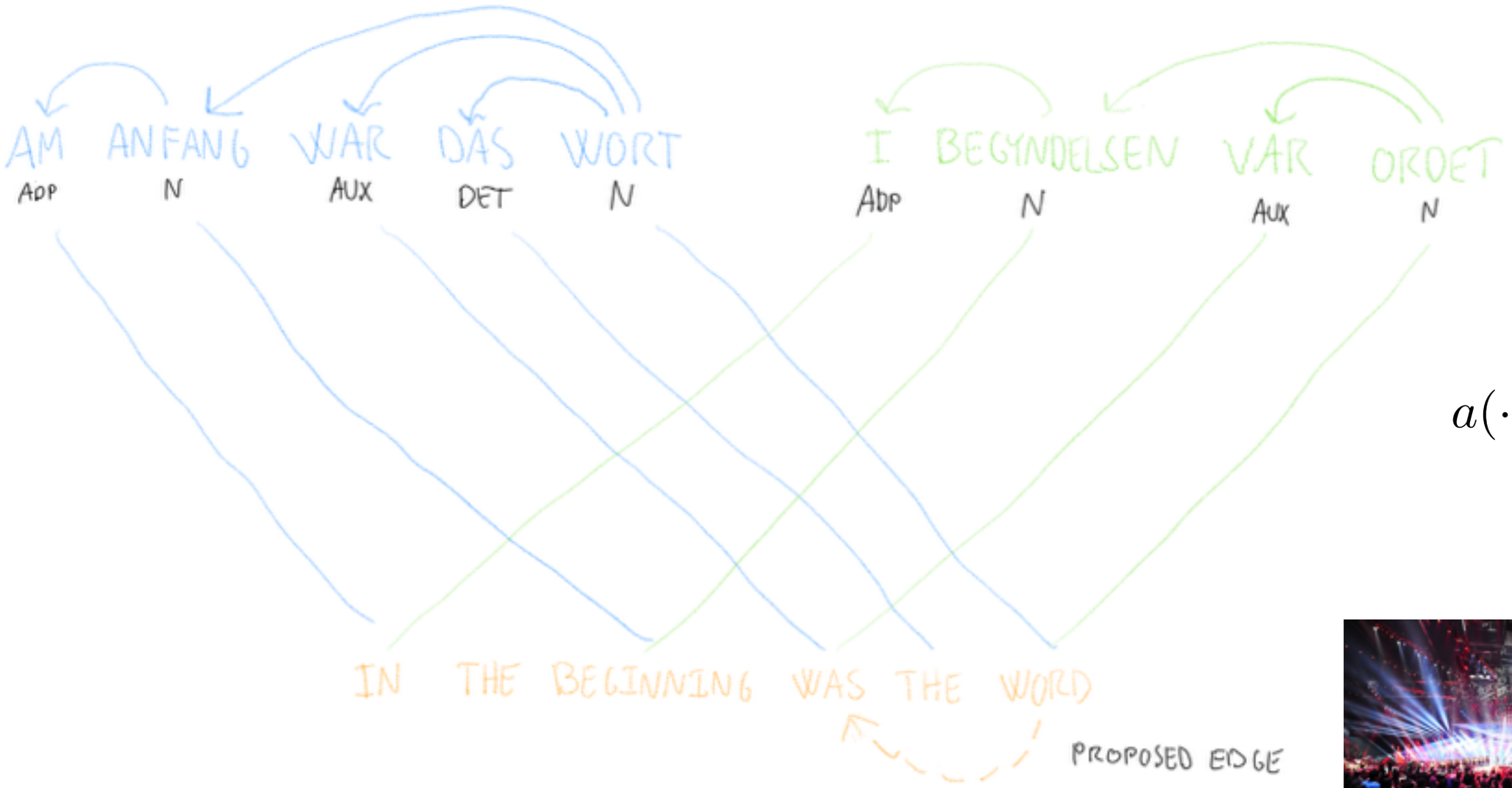
$a(\cdot, \cdot)$

$$\arg\max_{y} \sum_{(i,j)\in y} \text{score}_{\text{T}}(i, j) \quad \text{s.t. } y \text{ is a tree}$$

$$\text{score}_{\text{T}}(\text{word}, \text{was}) = \text{score}_{\text{DA}}(\text{ordet}, \text{var})$$

Example

score_{DE} score_{DA}



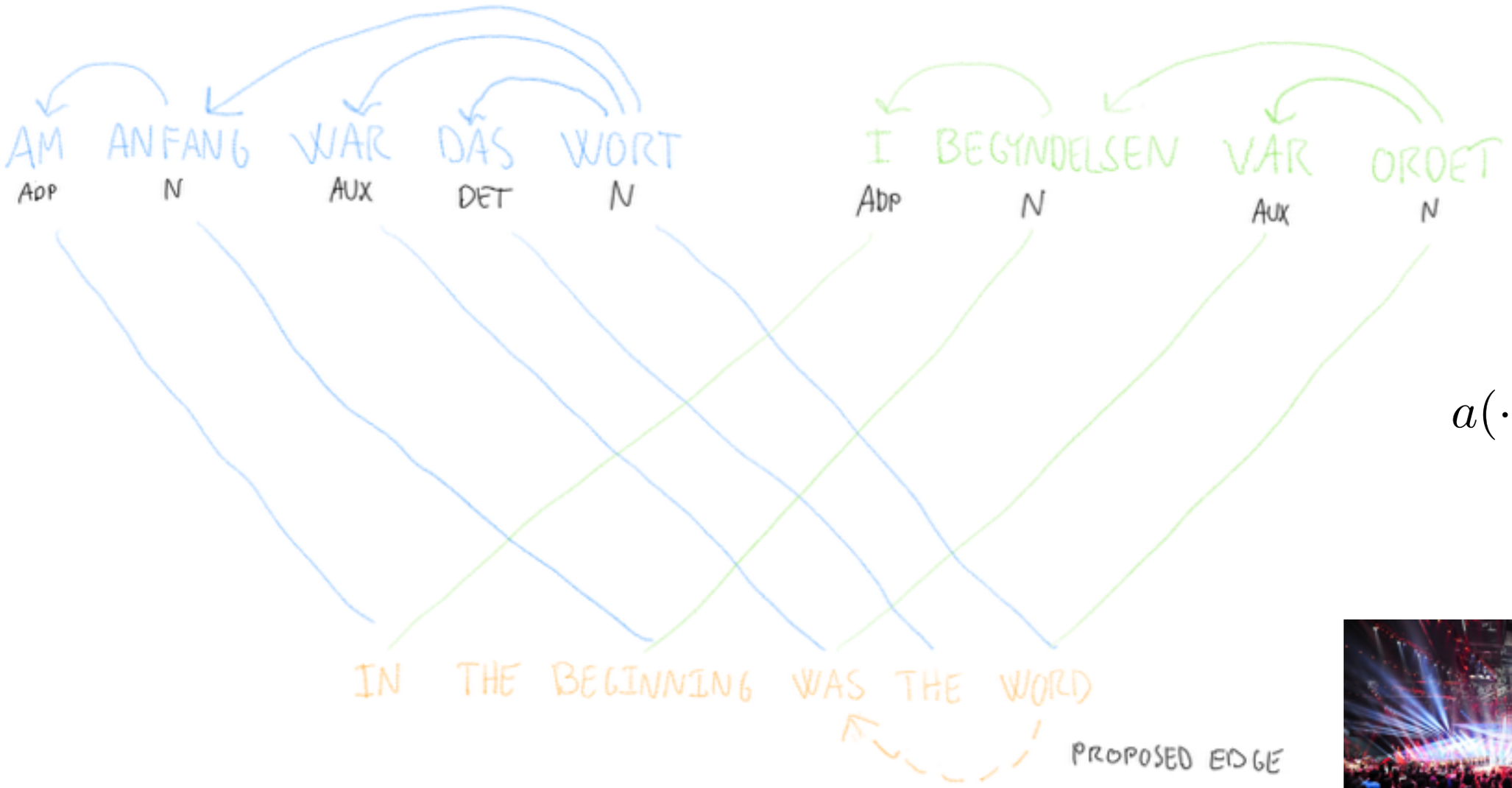
$a(\cdot, \cdot)$

$$\arg\max_{y} \sum_{(i,j)\in y} \text{score}_{\text{T}}(i, j) \quad \text{s.t. } y \text{ is a tree}$$

$$\text{score}_{\text{T}}(\text{word}, \text{was}) = \text{score}_{\text{DA}}(\text{ordet}, \text{var})\, a(\text{var}, \text{was})\, a(\text{ordet}, \text{word})$$

Example

score_{DE} score_{DA}

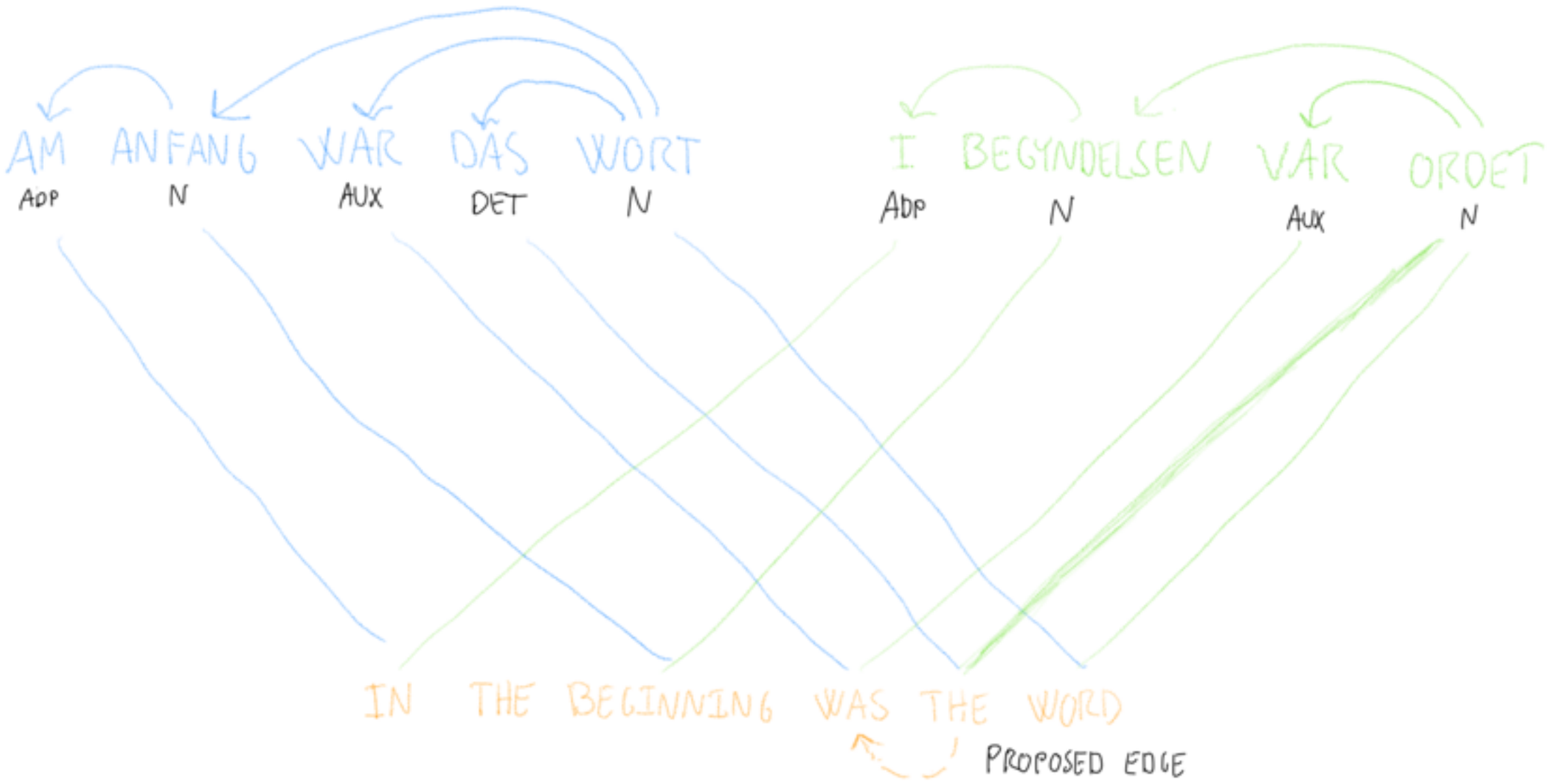


$a(\cdot, \cdot)$

$$\arg\max_{y} \sum_{(i,j)\in y} \text{score}_{\text{T}}(i,j) \quad \text{s.t. } y \text{ is a tree}$$

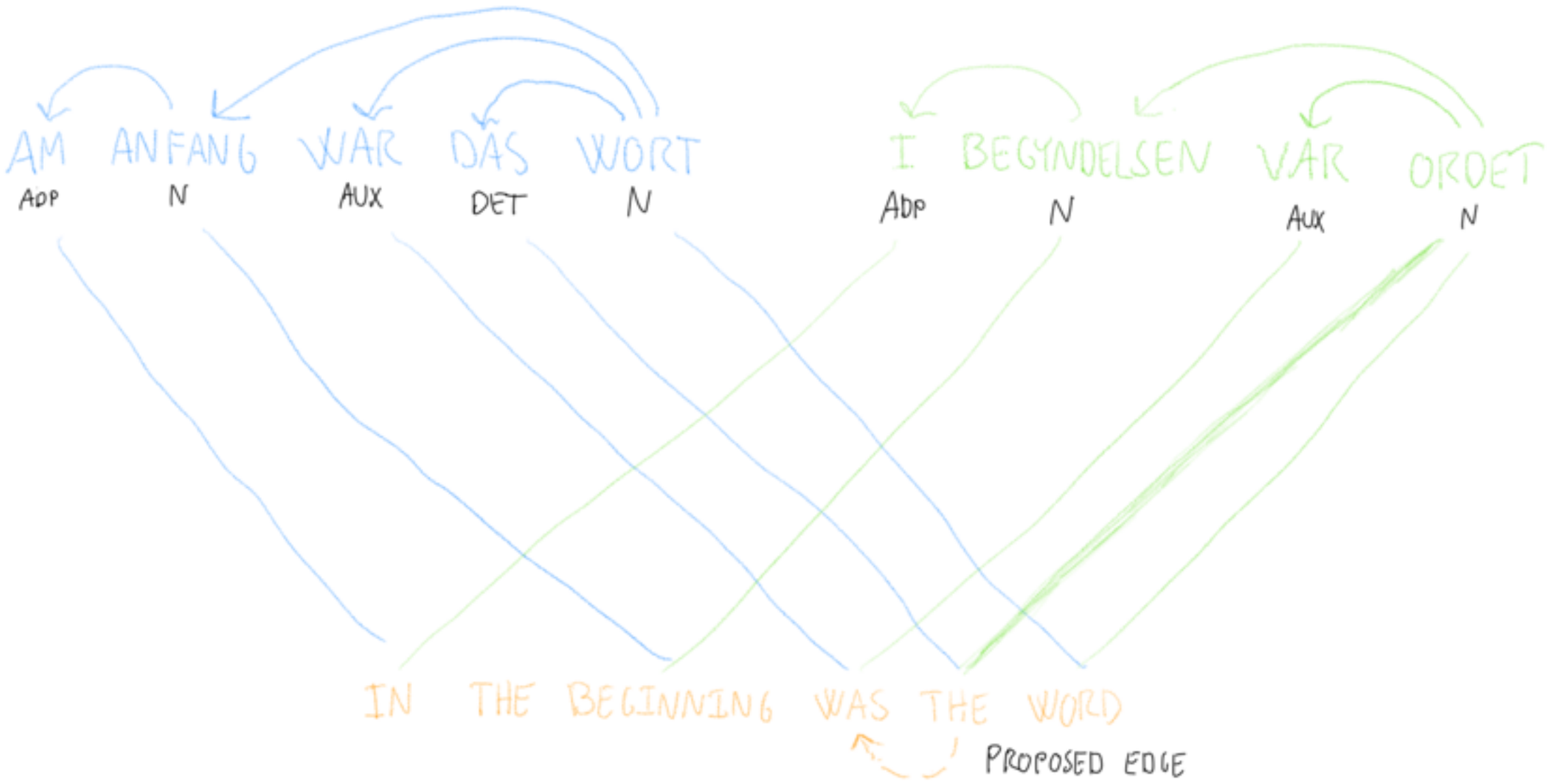
$$\begin{aligned}\text{score}_{\text{T}}(\text{word}, \text{was}) = &\ \text{score}_{\text{DA}}(\text{ordet}, \text{var})\, a(\text{var}, \text{was})\, a(\text{ordet}, \text{word}) \\ &+\text{score}_{\text{DE}}(\text{wort}, \text{war})\, a(\text{war}, \text{was})\, a(\text{wort}, \text{word})\end{aligned}$$

Example gone bad



$$\text{score}_{\text{T}}(\text{the}, \text{was}) = \text{score}_{\text{DA}}(\text{ordet}, \text{var})\, a(\text{var}, \text{was})\, a(\text{ordet}, \text{word})$$

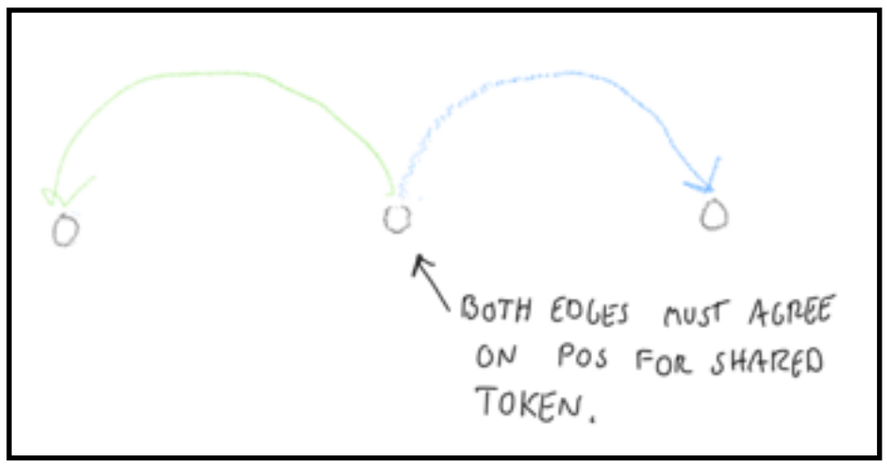
Example gone bad



$$\text{score}_{\text{T}}(\text{the}, \text{was}) = \cancel{\text{score}_{\text{BA}}(\text{ordet}, \text{var}) \, a(\text{var}, \text{was}) \, a(\text{ordet}, \text{word})}$$

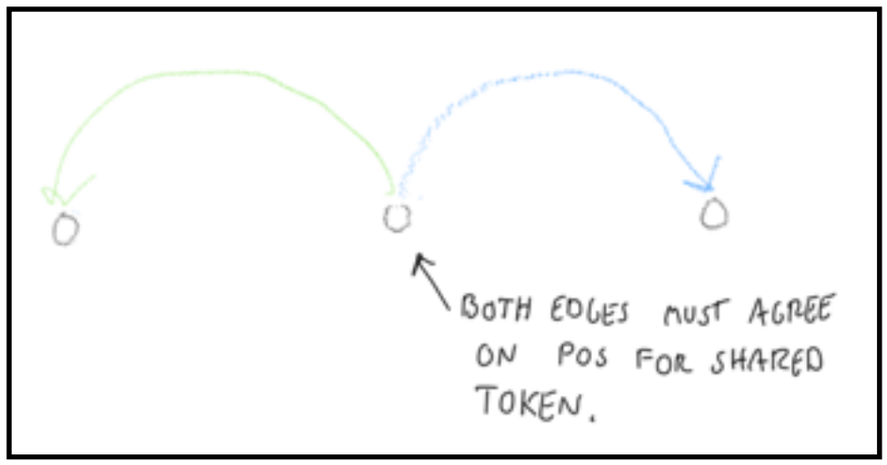
Yes, but only if “was” is AUX and “the” is N

Projecting layers of annotation



$$\arg\max_{y} \sum_{(i,k,j,l)\in y} \text{score}_{\text{T}}(i,k,j,l) \quad \text{s.t. } y \text{ is a tree}$$

Projecting layers of annotation

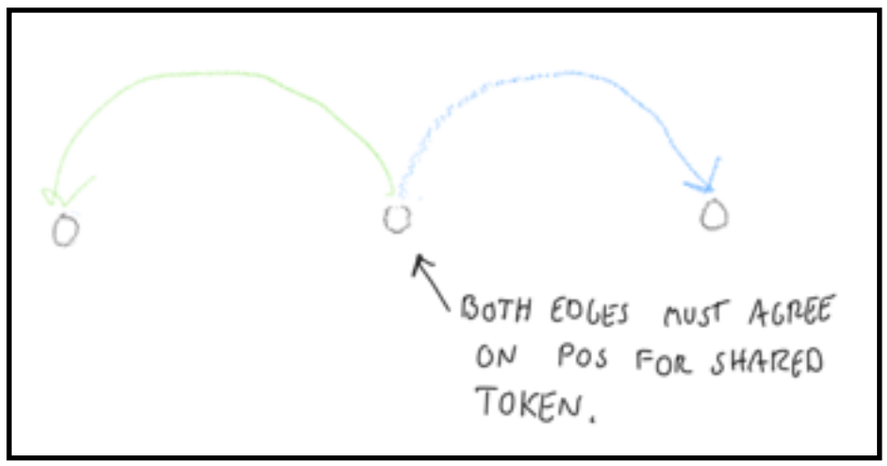


$$\arg\max_{y} \sum_{(i,k,j,l)\in y} \text{score}_{\text{T}}(i,k,j,l)$$

the edge (i, j)

s.t. *y* is a tree

Projecting layers of annotation



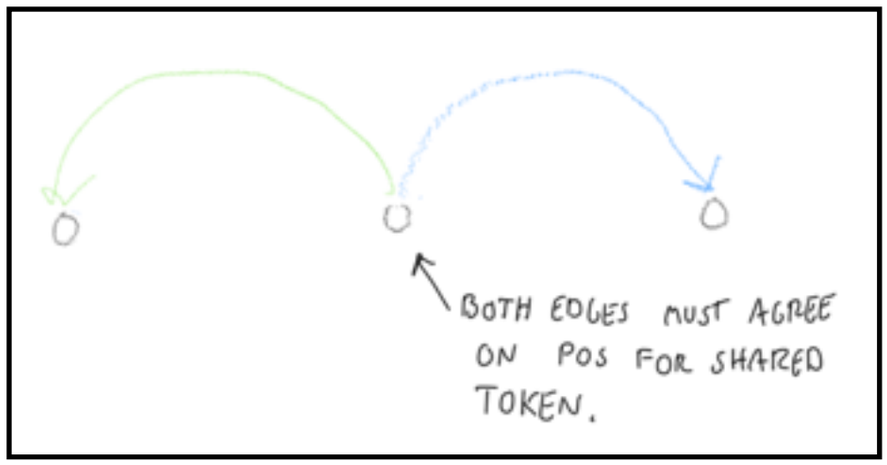
$$\arg\max_{y} \sum_{(i,k,j,l)\in y} \text{score}_{\text{T}}(i,k,j,l)$$

the edge (i, j)

s.t. *y* is a tree

tags for i and j

Projecting layers of annotation



$$\arg\max_{y} \sum_{(i,k,j,l)\in y} \text{score}_{\text{T}}(i,k,j,l)$$

the edge (i, j)

s.t. *y* is a tree

tags for i and j

more difficult

ILP model

Edges	$e_{i,k,j,l}$	$\in \{0,1\}$
Vertices	$v_{i,k}$	$\in \{0,1\}$
Flow	$\phi_{i,k,j,l}$	$\in \mathbb{R}^+$

Maximize $\sum_{i,k,j,l} e_{i,k,j,l}\, w_{i,k,j,l}$

One parent per token

$$\sum_{i,k,l} e_{i,k,j,l} = 1 \qquad \forall j \neq 0$$

The root token (index 0) sends n flow

$$\sum_{j,l} \phi_{0,0,j,l} = n$$

Each token consumes one unit of flow

$$\sum_{i,k,l} \phi_{i,k,x,l} - \sum_{k,j,l} \phi_{x,k,j,l} = 1 \qquad \forall x \neq 0$$

One POS per token

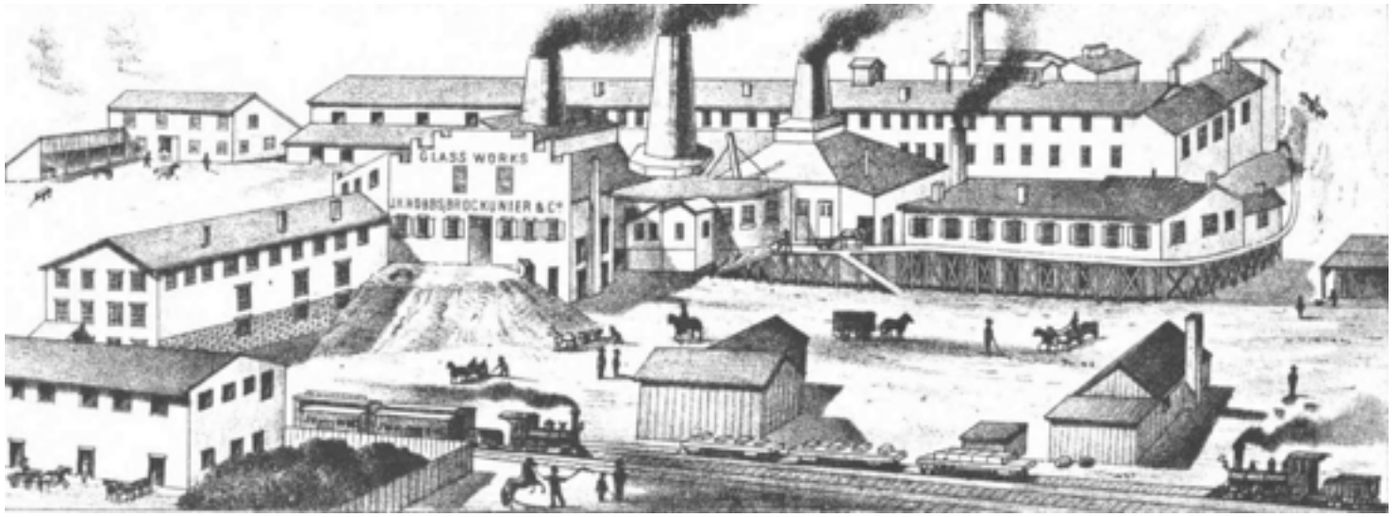
$$\sum_{k} v_{i,k} = 1 \qquad \forall i \neq 0$$

Active edges choose token POS

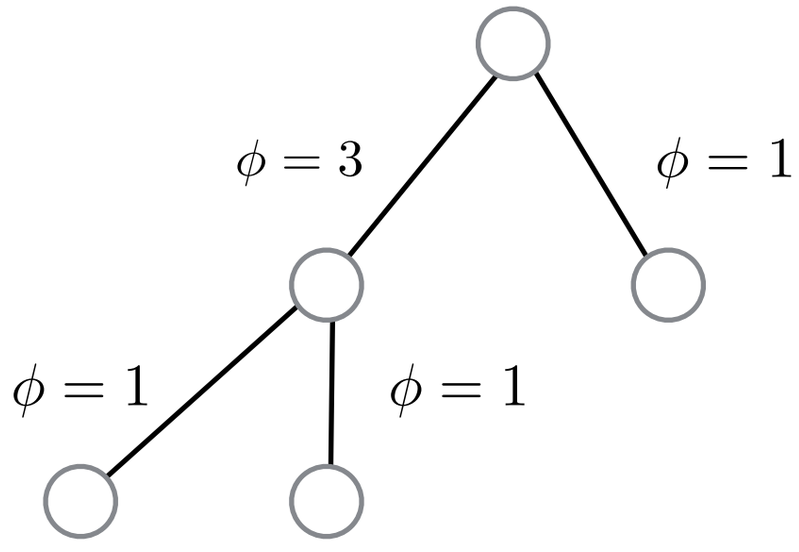
$$v_{i,k} \geq e_{i,k,j,l} \qquad \forall i \neq 0, j, k, l$$

$$v_{i,l} \geq e_{i,k,j,l} \qquad \forall i, j, k, l$$

Above, i, j, and x are token indices, while k and l refer to POS. Quantification over these symbols in the equations are always with respect to a given target graph.



Root produces *n* flow



Each node consumes one flow

(Martins, 2012)

Results

	Approach		
Predicted POS	ILP	DCA	DELEX
EBC	**51.62** (18)	48.39 (8)	42.44 (1)
WTC	**53.58** (20)	48.40 (0)	47.35 (3)
Gold POS			
EBC	**65.43** (25)	59.94 (2)	64.13 (–)
WTC	**66.51** (23)	55.73 (0)	66.68 (–)

POS tagging	
EBC	WTC
69.40	73.05

Conclusion

We extended Agić (2016) to project multiple layers of annotation jointly.

Approach stays simple and heuristics-free.

The initial experiments show promising results.

Future work

Project higher/lower layers of annotation, or larger tree parts.

Penalise inconsistent structures instead of disallowing.

Questions?