Token and Type Constraints for Cross-Lingual Part-of-Speech Tagging

Oscar Täckström
Dipanjan Das, Slav Petrov, Ryan McDonald & Joakim Nivre

UPPSALA UNIVERSITET

SWEDISH ICT SICS

.	ADJ	ADP
ADV	CONJ	DET
NOUN	NUM	PRON
PRT	VERB	X

.	ADJ	ADP
ADV	CONJ	DET
NOUN	NUM	PRON
PRT	VERB	X

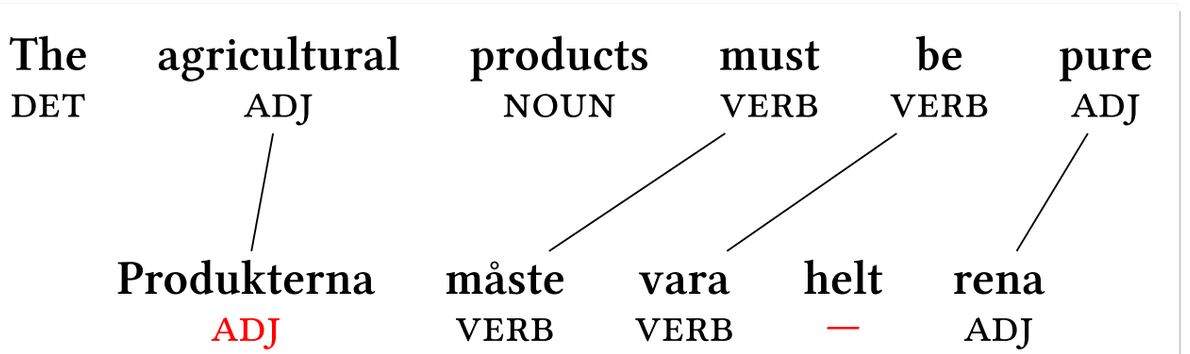
Wiktionary

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN, VERB
varade	VERB
…	…

.	ADJ	ADP
ADV	CONJ	DET
NOUN	NUM	PRON
PRT	VERB	X

Wiktionary

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN, VERB
varade	VERB
…	…

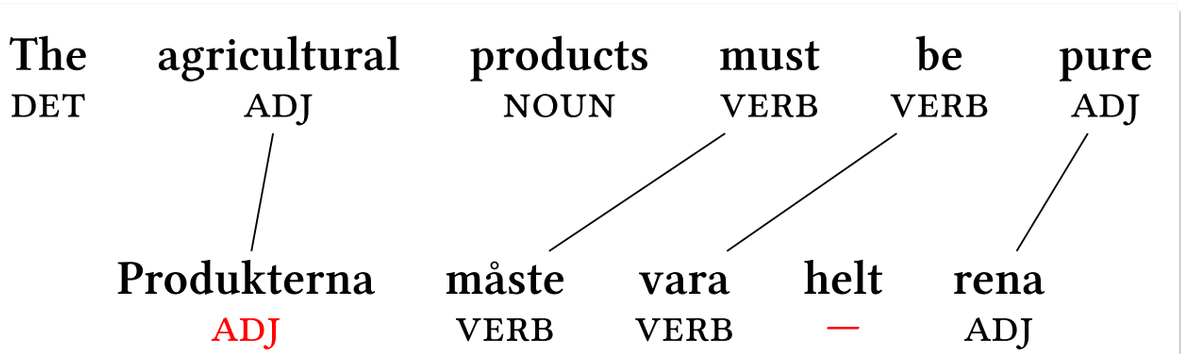


.	ADJ	ADP
ADV	CONJ	DET
NOUN	NUM	PRON
PRT	VERB	X

Wiktionary

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN, VERB
varade	VERB
…	…

+

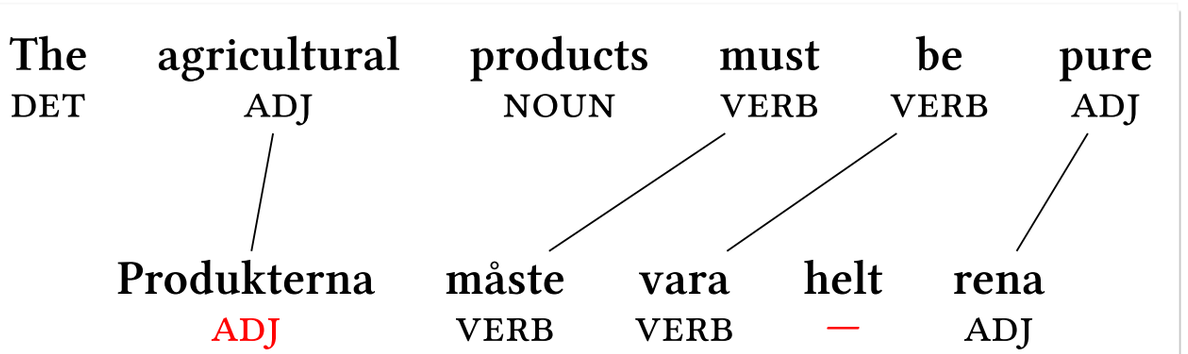


.	ADJ	ADP
ADV	CONJ	DET
NOUN	NUM	PRON
PRT	VERB	X

Wiktionary

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN, VERB
varade	VERB
…	…

+



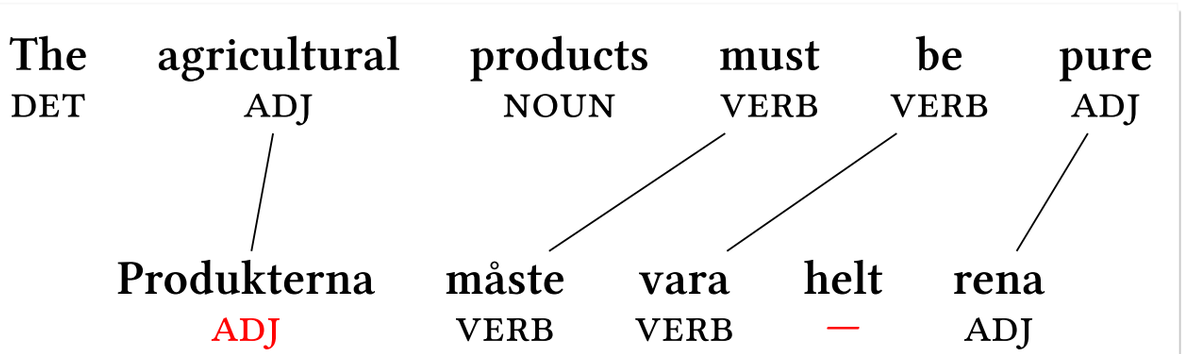
Latent variable CRF model

.	ADJ	ADP
ADV	CONJ	DET
NOUN	NUM	PRON
PRT	VERB	X

Wiktionary

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN, VERB
varade	VERB
…	…

+



Latent variable CRF model

>25% error reduction over state of the art

Part-of-Speech Tagging

Is	this	the	future	of	chamber	music	?

Part-of-Speech Tagging

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?

Part-of-Speech Tagging

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?

Måste	en	familj	ha	just	den	uppbyggnaden	?
MV	EN	NN	HV	AB	PO	NN	I?

Всеки	има	право	на	живот	.
PC	VPI	NC	R	NC	PUNCT

...

Part-of-Speech Tagging

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?

Måste	en	familj	ha	just	den	uppbyggnaden	?
MV	EN	NN	HV	AB	PO	NN	I?

Всеки	има	право	на	живот	.
PC	VPI	NC	R	NC	PUNCT

...

Supervised accuracies:

~95% in-domain (Petrov et al., 2012)

<90% out of domain (Blitzer et al., 2006)

Coarse "Universal" Tags

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?

Måste	en	familj	ha	just	den	uppbyggnaden	?
MV	EN	NN	HV	AB	PO	NN	I?

Всеки	има	право	на	живот	.
PC	VPI	NC	R	NC	PUNCT

Naseem et al., 2009; **Petrov et al., 2012**

Coarse "Universal" Tags

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?
VERB	DET	DET	NOUN	ADP	NOUN	NOUN	.

Måste	en	familj	ha	just	den	uppbyggnaden	?
MV	EN	NN	HV	AB	PO	NN	I?
VERB	NUM	NOUN	VERB	ADV	PRON	NOUN	.

Всеки	има	право	на	живот	.
PC	VPI	NC	R	NC	PUNCT
PRON	VERB	NOUN	ADP	NOUN	.

Naseem et al., 2009; **Petrov et al., 2012**

Coarse "Universal" Tags

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?
VERB	DET	DET	NOUN	ADP	NOUN	NOUN	.

Måste	en	familj	ha	just	den	uppbyggnaden	?
MV	EN	NN	HV	AB	PO	NN	I?
VERB	NUM	NOUN	VERB	ADV	PRON	NOUN	.

Всеки	има	право	на	живот	.
PC	VPI	NC	R	NC	PUNCT
PRON	VERB	NOUN	ADP	NOUN	.

Naseem et al., 2009; **Petrov et al., 2012**

Coarse "Universal" Tags

Is	this	the	future	of	chamber	music	?
VBZ	DT	DT	NN	IN	NN	NN	?
VERB	DET	DET	NOUN	ADP	NOUN	NOUN	.

Måste	en	familj	ha	just	den	uppbyggnaden	?
MV	EN	NN	HV	AB	PO	NN	I?
VERB	NUM	NOUN	VERB	ADV	PRON	NOUN	.

Всеки	има	право	на	живот	.
PC	VPI	NC	R	NC	PUNCT
PRON	VERB	NOUN	ADP	NOUN	.

Naseem et al., 2009; **Petrov et al., 2012**

Coarse "Universal" Tags

Is	this	the	future	of	chamber	music	?
VERB	DET	DET	NOUN	ADP	NOUN	NOUN	.

Måste	en	familj	ha	just	den	uppbyggnaden	?
VERB	NUM	NOUN	VERB	ADV	PRON	NOUN	.

Всеки	има	право	на	живот	.
PRON	VERB	NOUN	ADP	NOUN	.

Cross-lingual tagging (Das & Petrov, 2011)

Cross-lingual parsing (Zeman & Resnick, 2008; ...)

Constraints

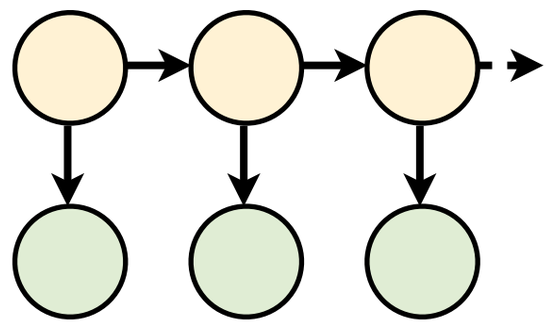
1. Token constraints
2. Type constraints
3. Coupled token and type constraints

Models

Models

1. Log-linear ("feature-based") HMM

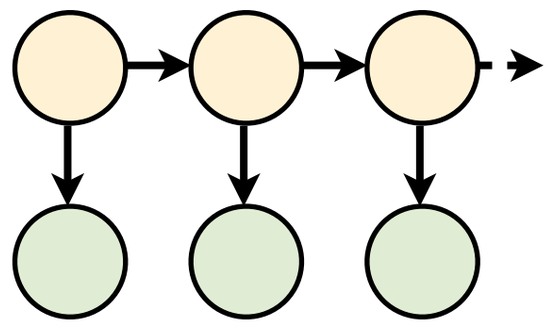
(Berg-Kirkpatrick et al., 2010; ...)



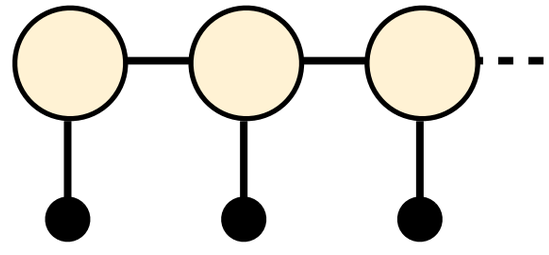
Models

1. Log-linear ("feature-based") HMM

(Berg-Kirkpatrick et al., 2010; ...)



2. Latent variable CRF



Complete Token Supervision

Produkterna	**måste**	**vara**	**helt**	**rena**
(The products)	(must)	(be)	(completely)	(pure)
NOUN	VERB	VERB	ADV	ADJ

En	**dyrbar**	**vara**
(An)	(expensive)	(commodity)
DET	ADJ	NOUN

...

Full disambiguation *in context*

Complete Token Supervision

Produkterna (The products) NOUN	**måste** (must) VERB	**vara** (be) VERB	**helt** (completely) ADV	**rena** (pure) ADJ

En (An) DET	**dyrbar** (expensive) ADJ	**vara** (commodity) NOUN

...

Full disambiguation *in context*

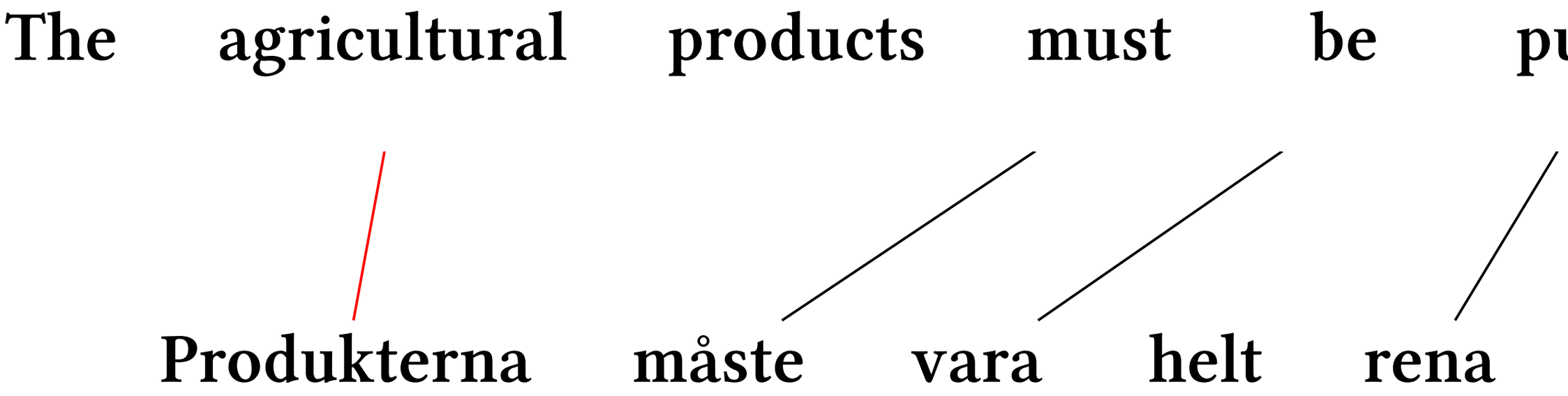
Projected Token Constraints

The agricultural products must be pure

Produkterna måste vara helt rena

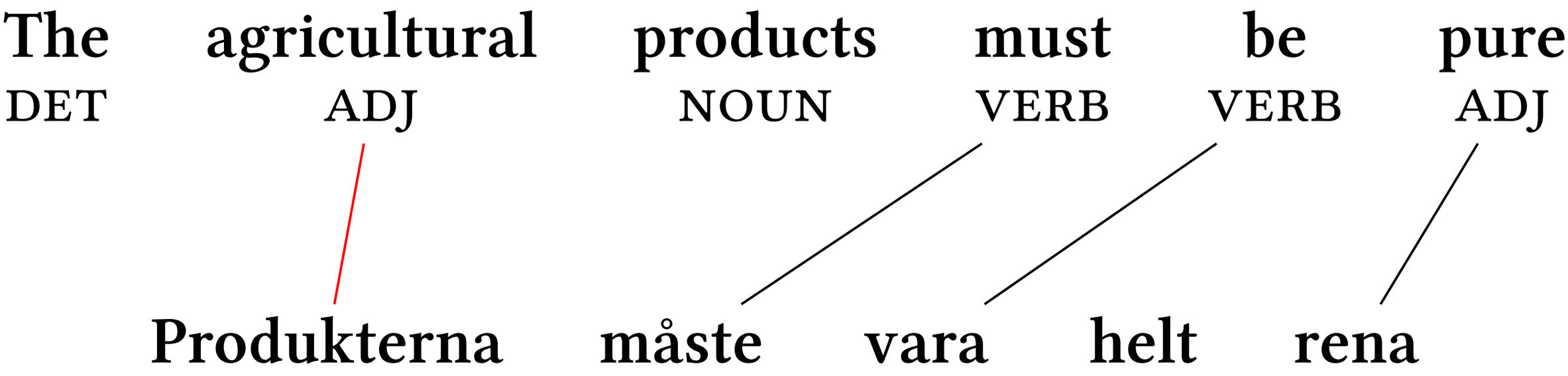
Yarowsky & Ngai, 2001; ...

Projected Token Constraints



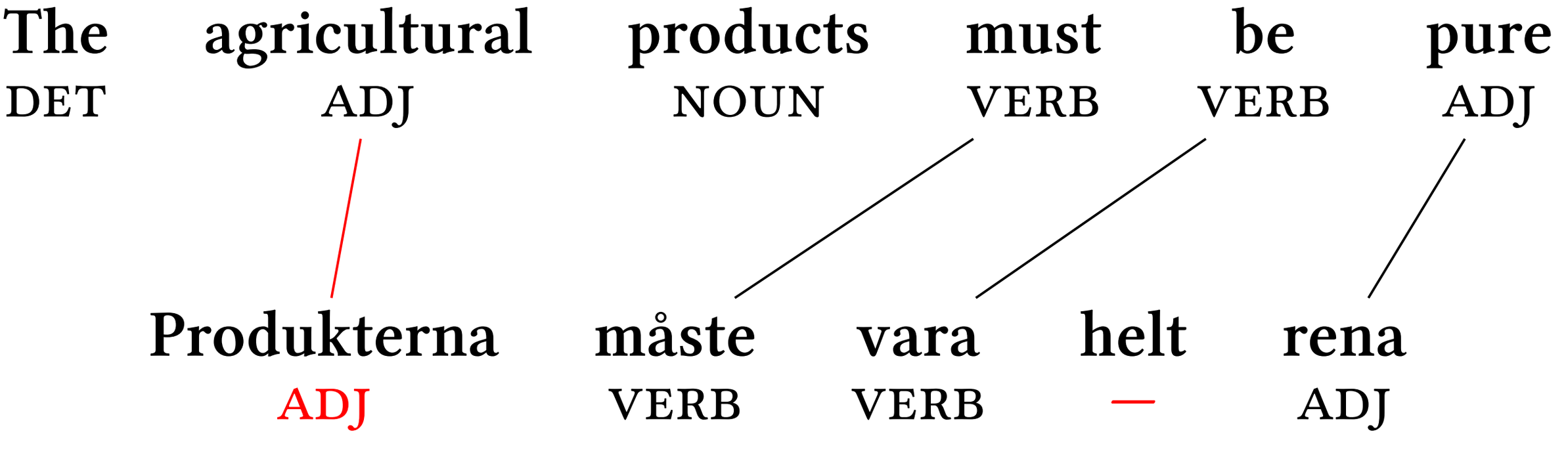
Yarowsky & Ngai, 2001; ...

Projected Token Constraints



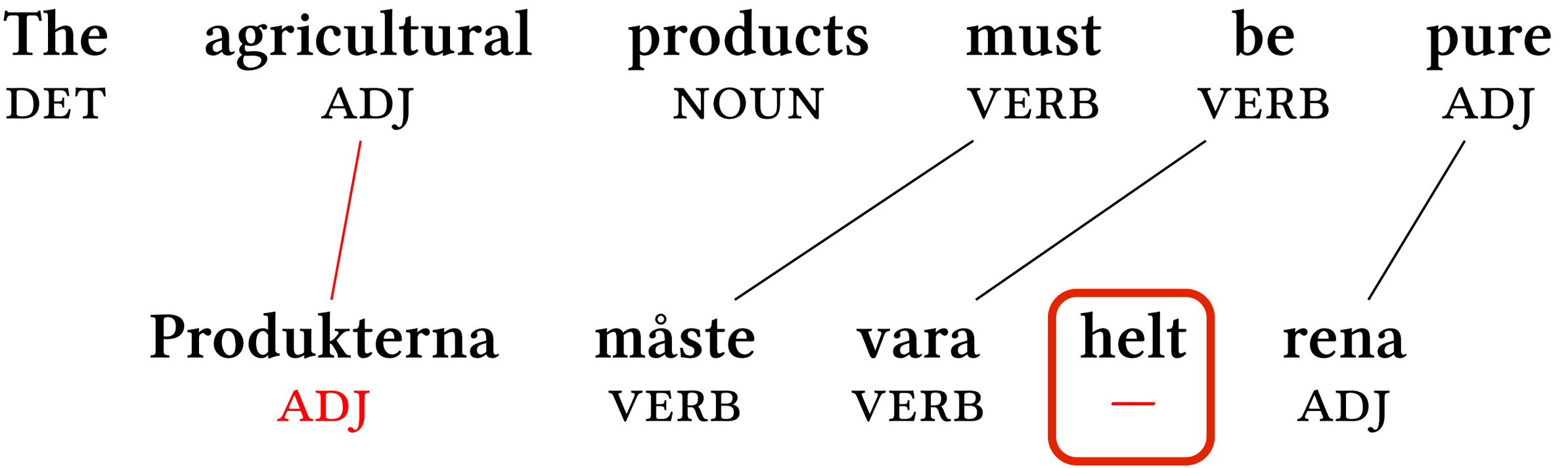
Yarowsky & Ngai, 2001; ...

Projected Token Constraints



Yarowsky & Ngai, 2001; ...

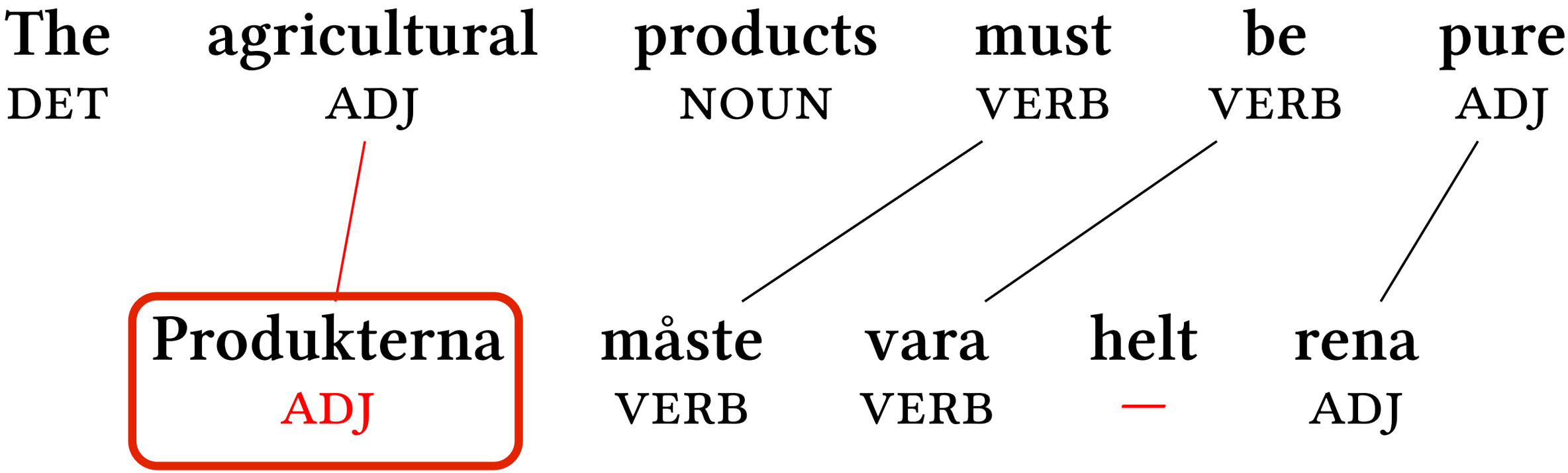
Projected Token Constraints



Incomplete

Yarowsky & Ngai, 2001; ...

Projected Token Constraints

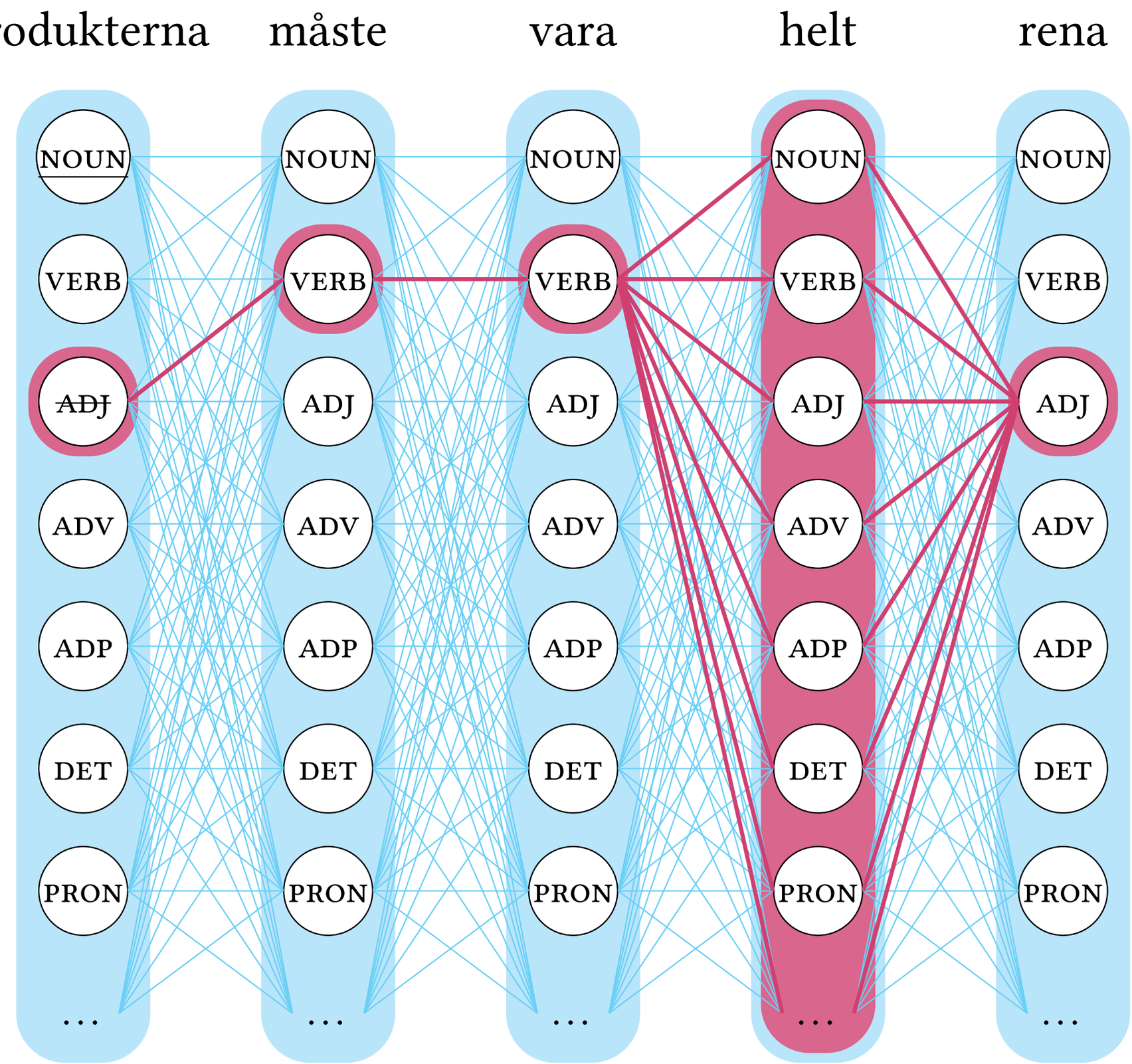


Incomplete

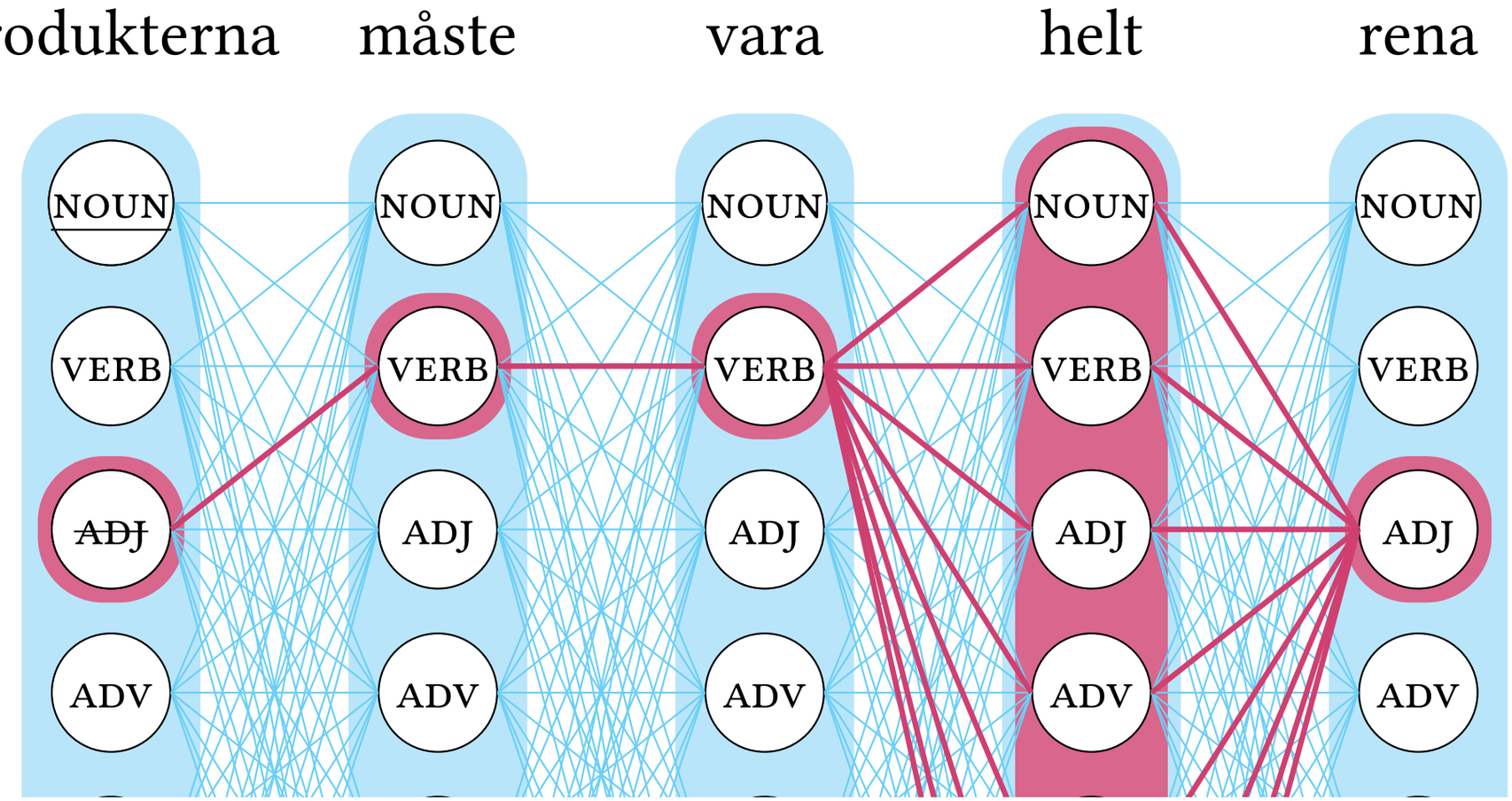
Noisy (systematic)

Yarowsky & Ngai, 2001; ...

Projected Token Constraints



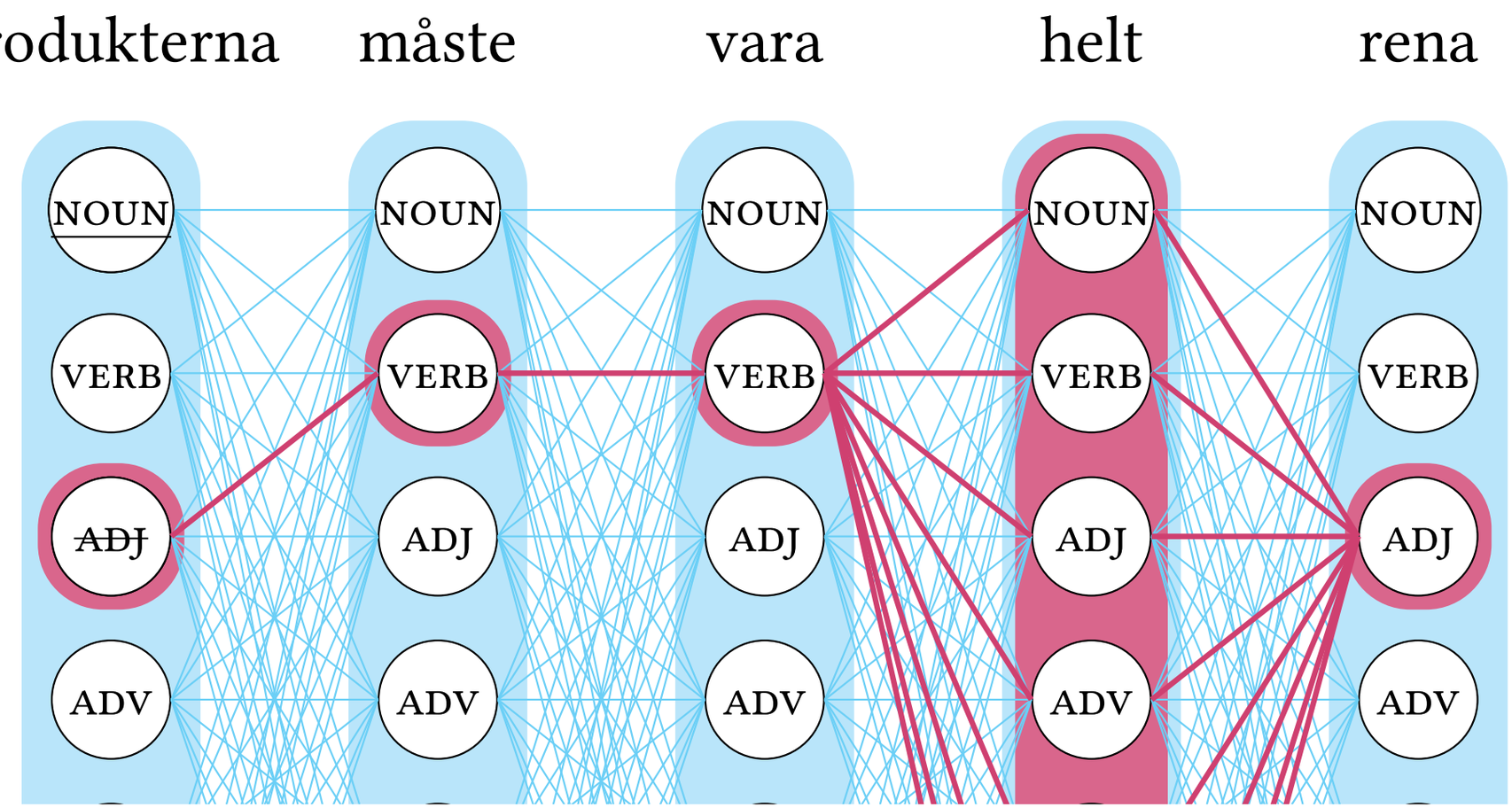
Projected Token Constraints



Marginal conditional likelihood

$$p_\theta(\tilde{\mathbf{y}} \mid x) = \frac{\sum_{y \in \tilde{\mathbf{y}}} \exp\left\{\theta^\top \Phi(x, y)\right\}}{\sum_{y' \in \mathcal{Y}(x)} \exp\left\{\theta^\top \Phi(x, y')\right\}}$$

Projected Token Constraints



Marginal conditional likelihood

$$p_\theta(\tilde{\mathbf{y}} \mid x) = \frac{\sum_{y \in \tilde{\mathbf{y}}} \exp\left\{\theta^\top \Phi(x, y)\right\}}{\sum_{y' \in \mathcal{Y}(x)} \exp\left\{\theta^\top \Phi(x, y')\right\}}$$

Type Constraints

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN (commodity), VERB (to be)
varade	VERB
…	…

(~1.4 tags / token)

Type Constraints

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN (commodity), VERB (to be)
varade	VERB
…	…

(~1.4 tags / token)

Cross-lingual projection (Das & Petrov, 2011)

Type Constraints

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN (commodity), VERB (to be)
varade	VERB
…	…

(~1.4 tags / token)

Cross-lingual projection (Das & Petrov, 2011)

Wiktionary (Li et al., 2012)

Type Constraints

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN (commodity), VERB (to be)
varade	VERB
…	…

(~1.4 tags / token)

Cross-lingual projection (Das & Petrov, 2011)

Wiktionary (Li et al., 2012)

“2 hours of annotation” (Garrette et al., 2013)

Type Constraints

…	…
var	ADV, CONJ, DET, NOUN, VERB
vara	NOUN (commodity), VERB (to be)
varade	VERB
…	…

(~1.4 tags / token)

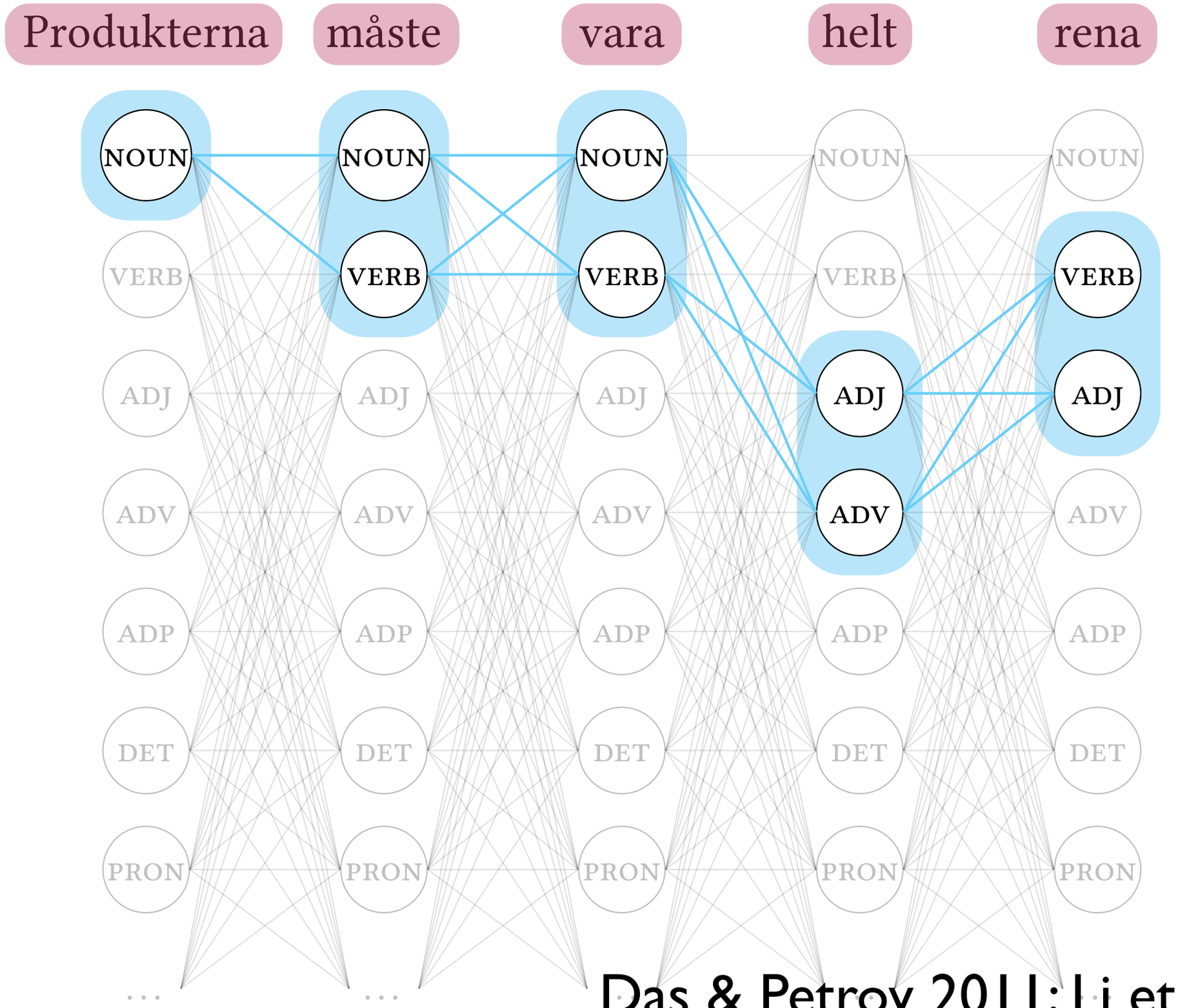
Cross-lingual projection (Das & Petrov, 2011)

Wiktionary (Li et al., 2012)

“2 hours of annotation” (Garrette et al., 2013)

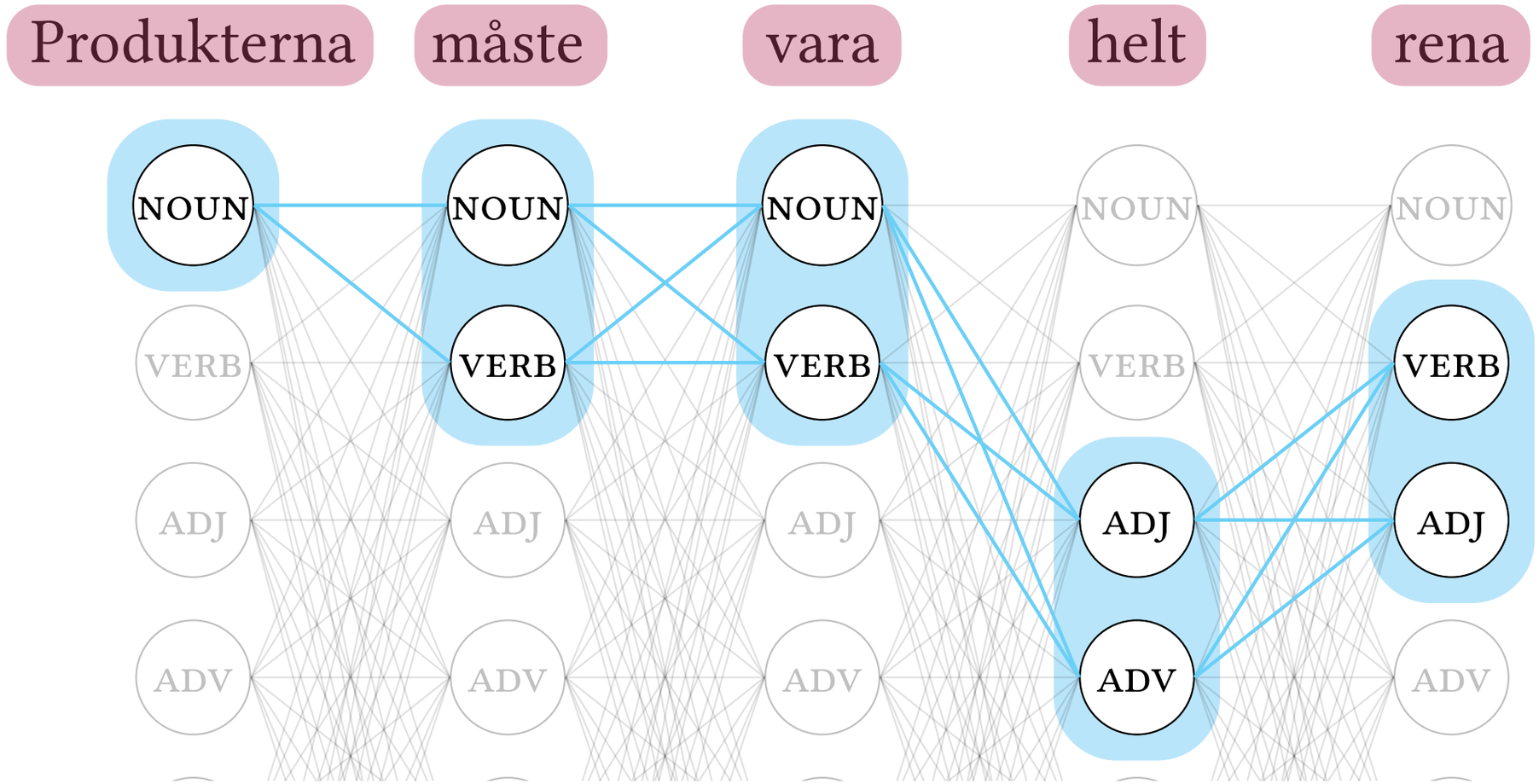
Only partial disambiguation

Type Constraints (HMM)



Das & Petrov 2011; Li et al., 2012

Type Constraints (HMM)



Marginal joint likelihood

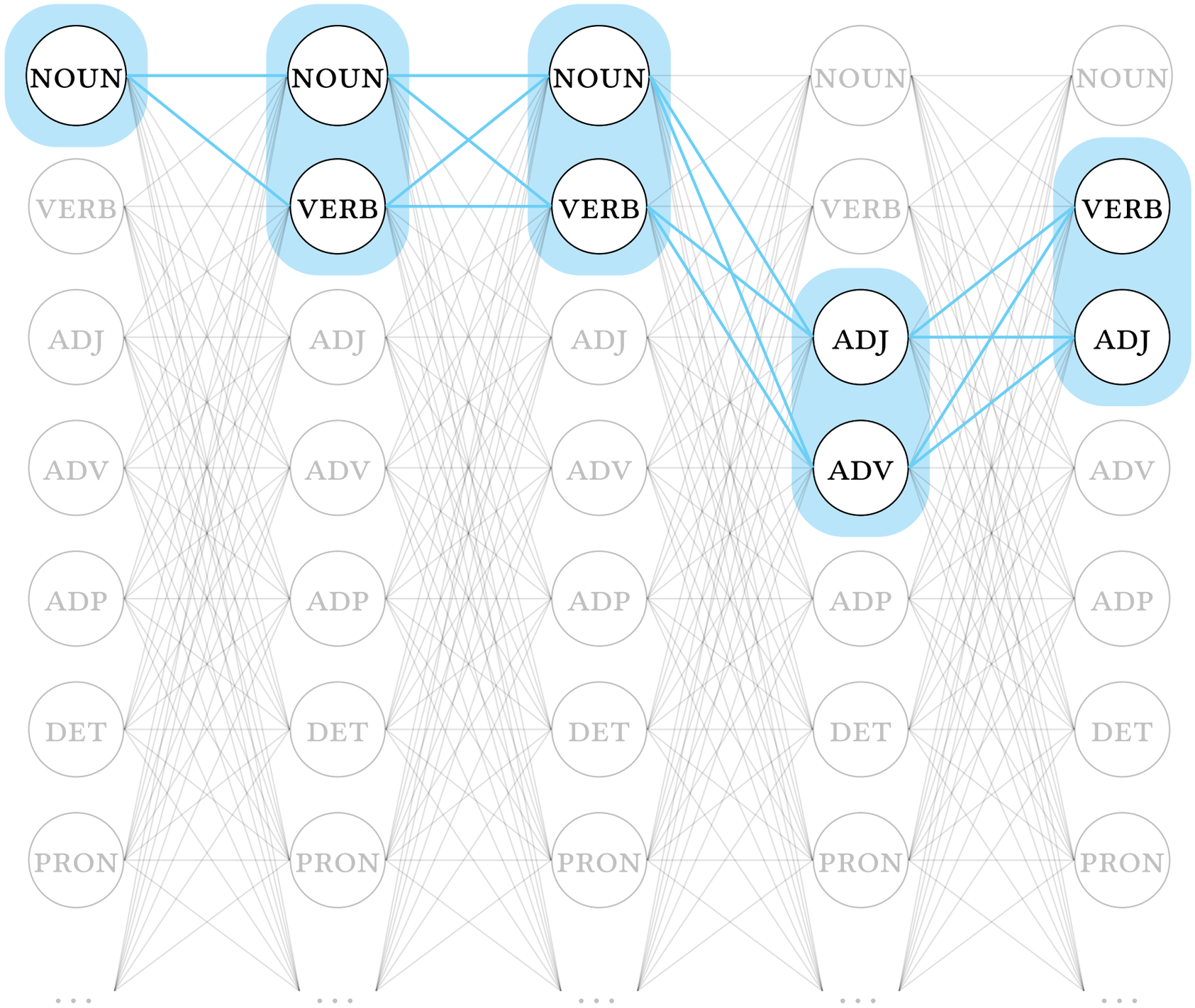
$$p_\beta(x) = \sum_{y \in \mathcal{Y}(x)} \prod_{i=1}^{|x|} \underbrace{p_\beta(x_i \mid y_i)}_{\text{emission}} \underbrace{p_\beta(y_i \mid y_{i-1})}_{\text{transition}}$$

Das & Petrov 2011, Li et al., 2012

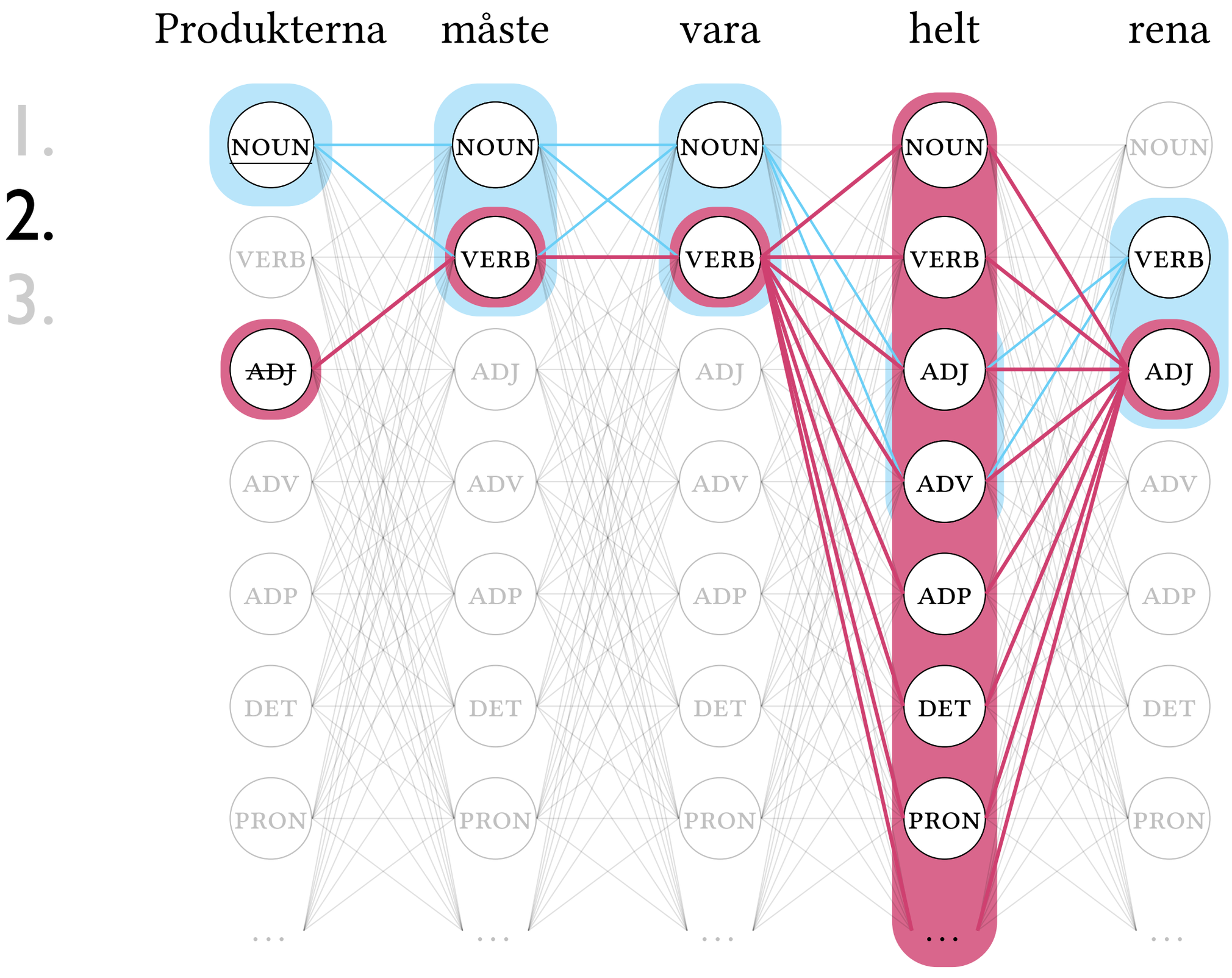
Token and Type Constraints

Produkterna måste vara helt rena

1.
2.
3.



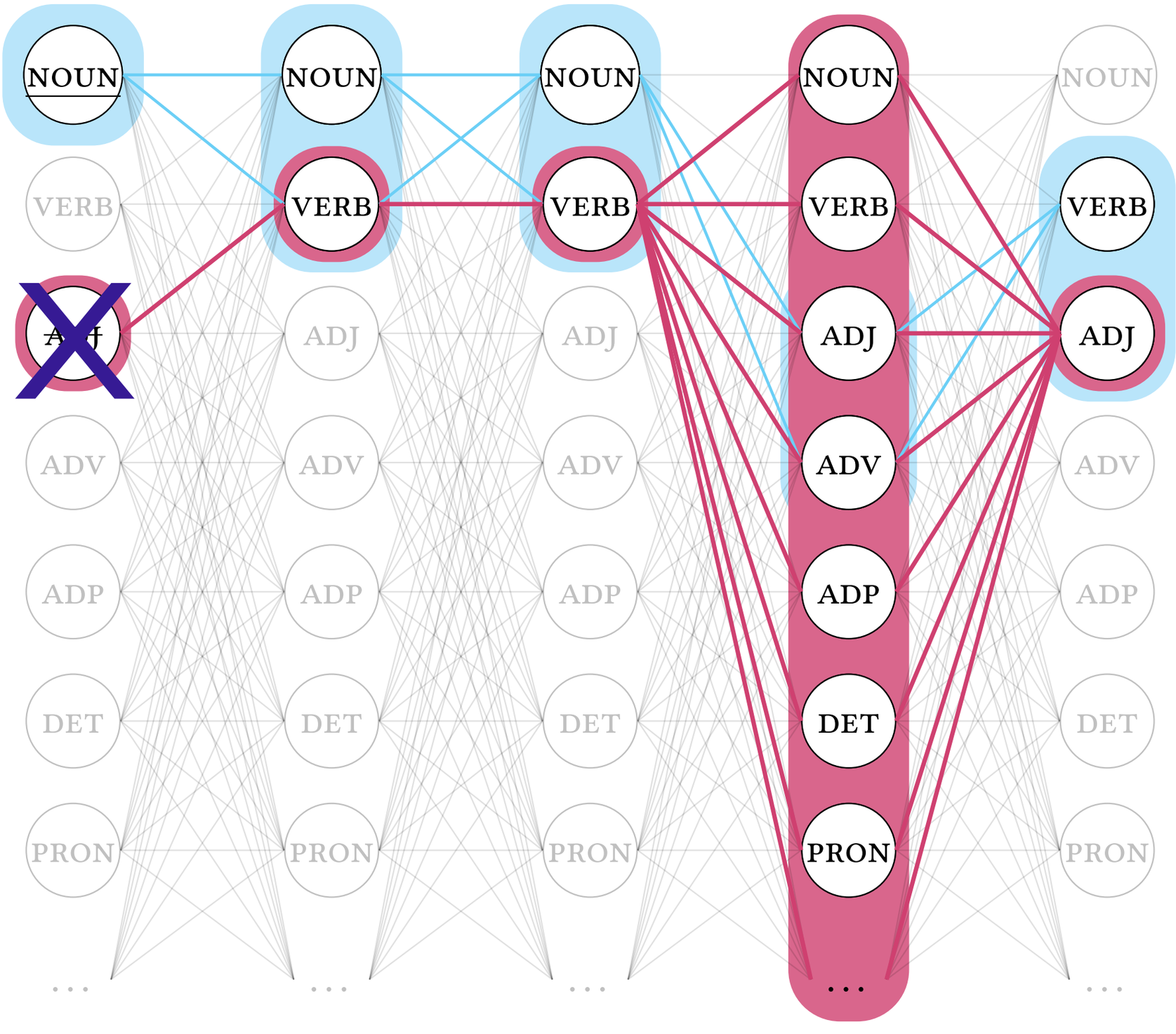
Token and Type Constraints



Token and Type Constraints

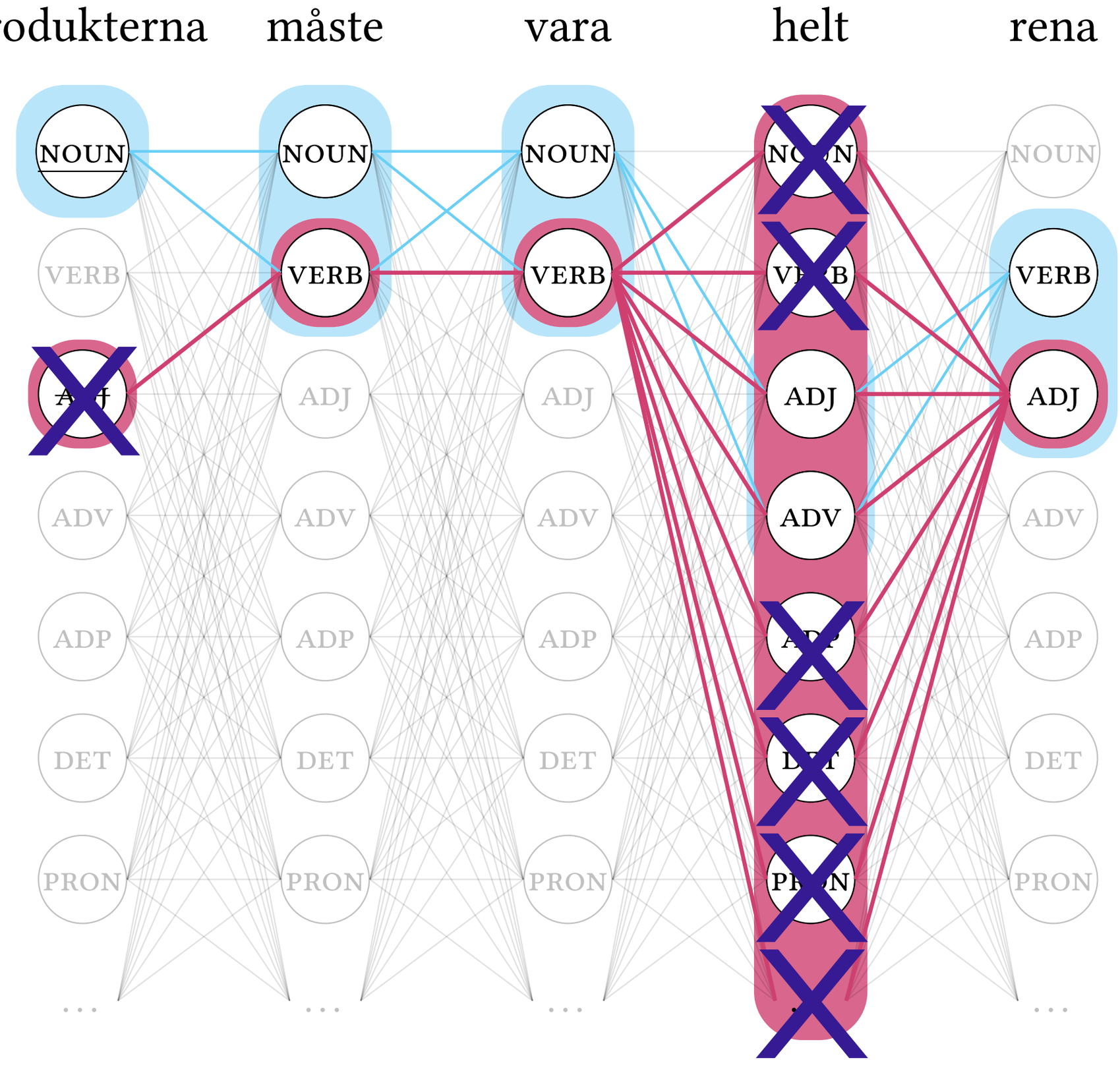
Produkterna måste vara helt rena

1.
2.
3.



Token and Type Constraints

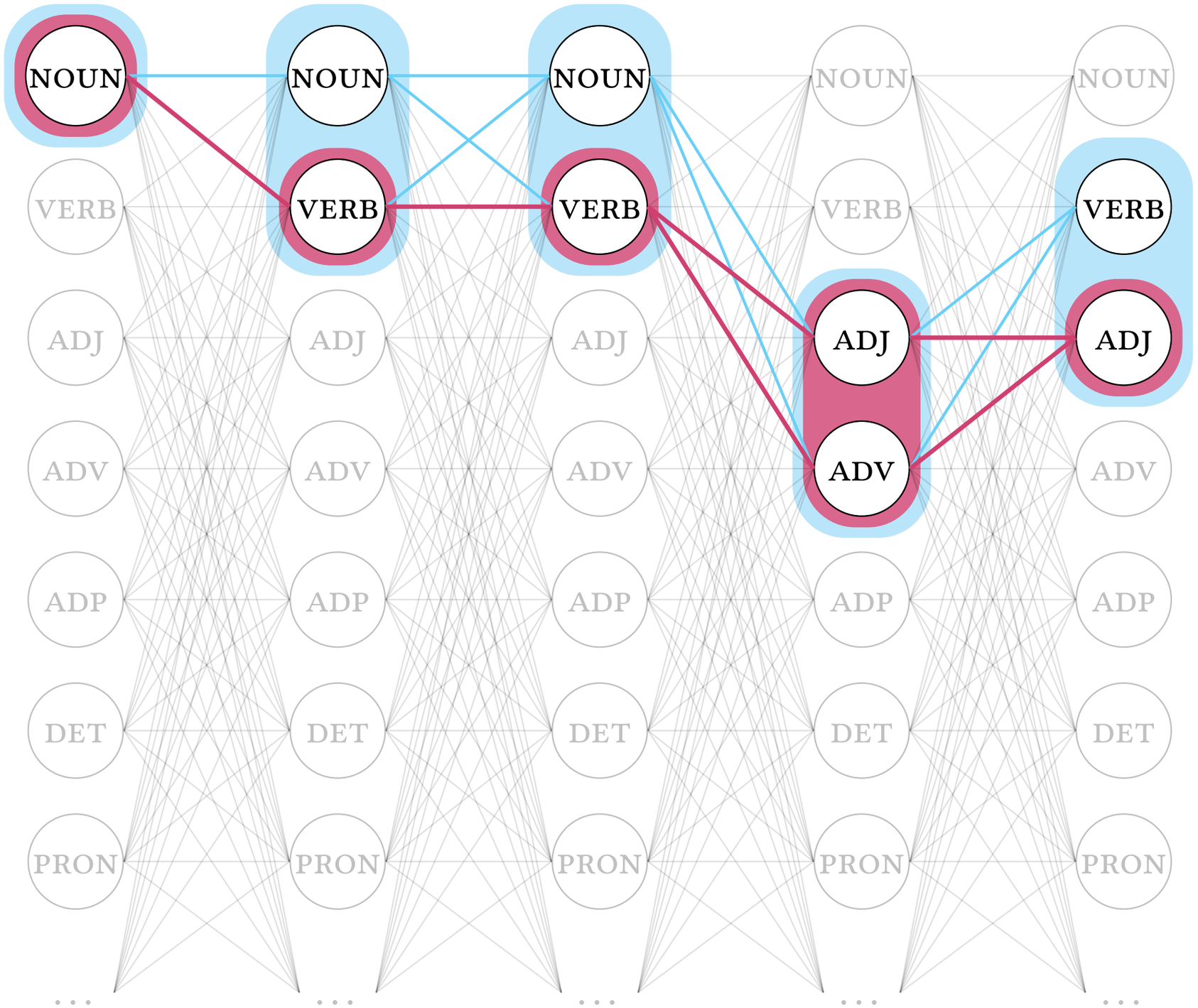
1.
2.
3.



Token and Type Constraints

Produkterna måste vara helt rena

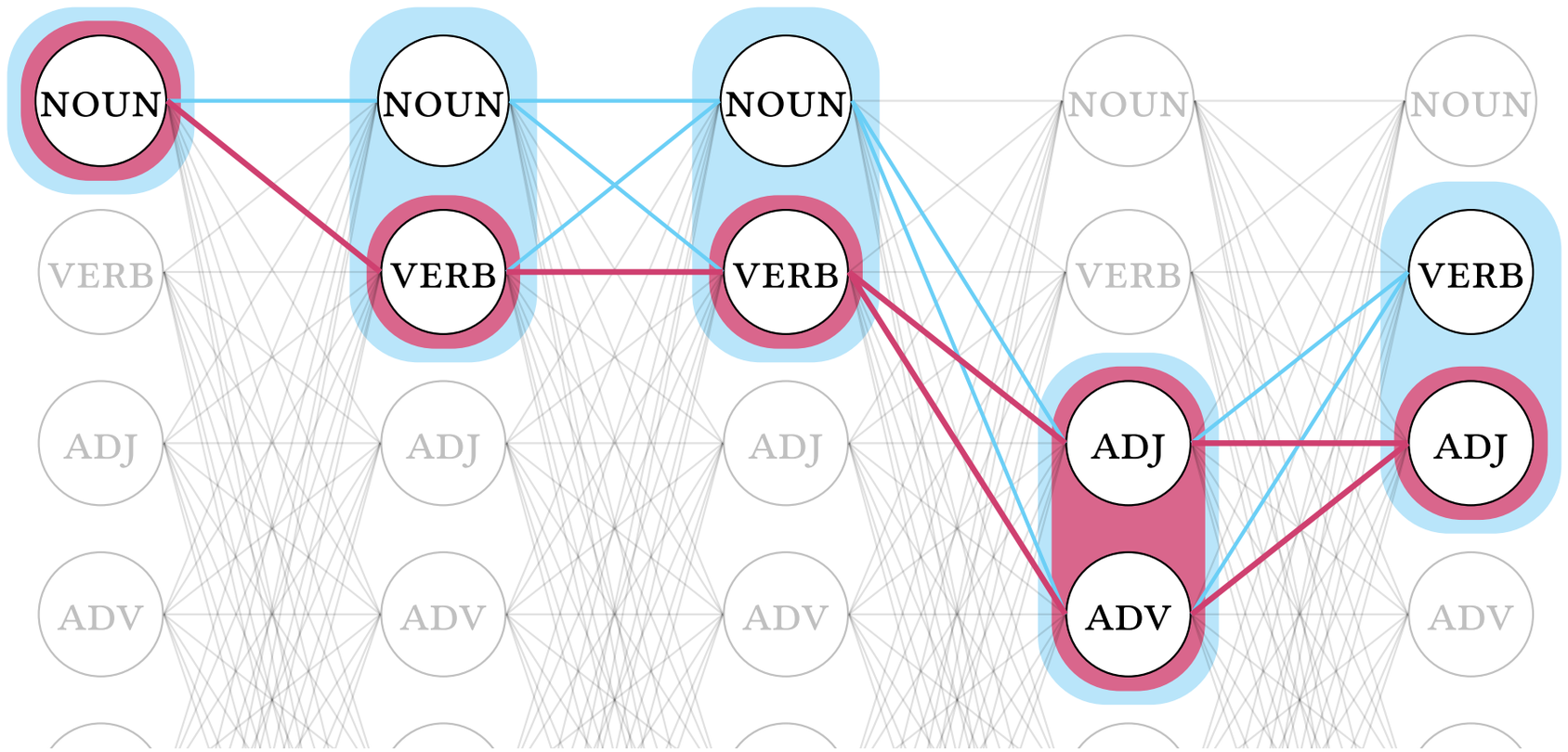
1.
2.
3.



Token and Type Constraints

Produkterna måste vara helt rena

1.
2.
3.



Marginal conditional likelihood

$$p_\theta(\widehat{\mathcal{Y}}(x, \tilde{\mathbf{y}}) \mid x) = \frac{\sum_{y \in \widehat{\mathcal{Y}}(x,\tilde{\mathbf{y}})} \exp\left\{\theta^\top \Phi(x, y)\right\}}{\sum_{y' \in \mathcal{Y}(x)} \exp\left\{\theta^\top \Phi(x, y')\right\}}$$

Token and Type Results

Type
(Wik)

Type
(Proj)

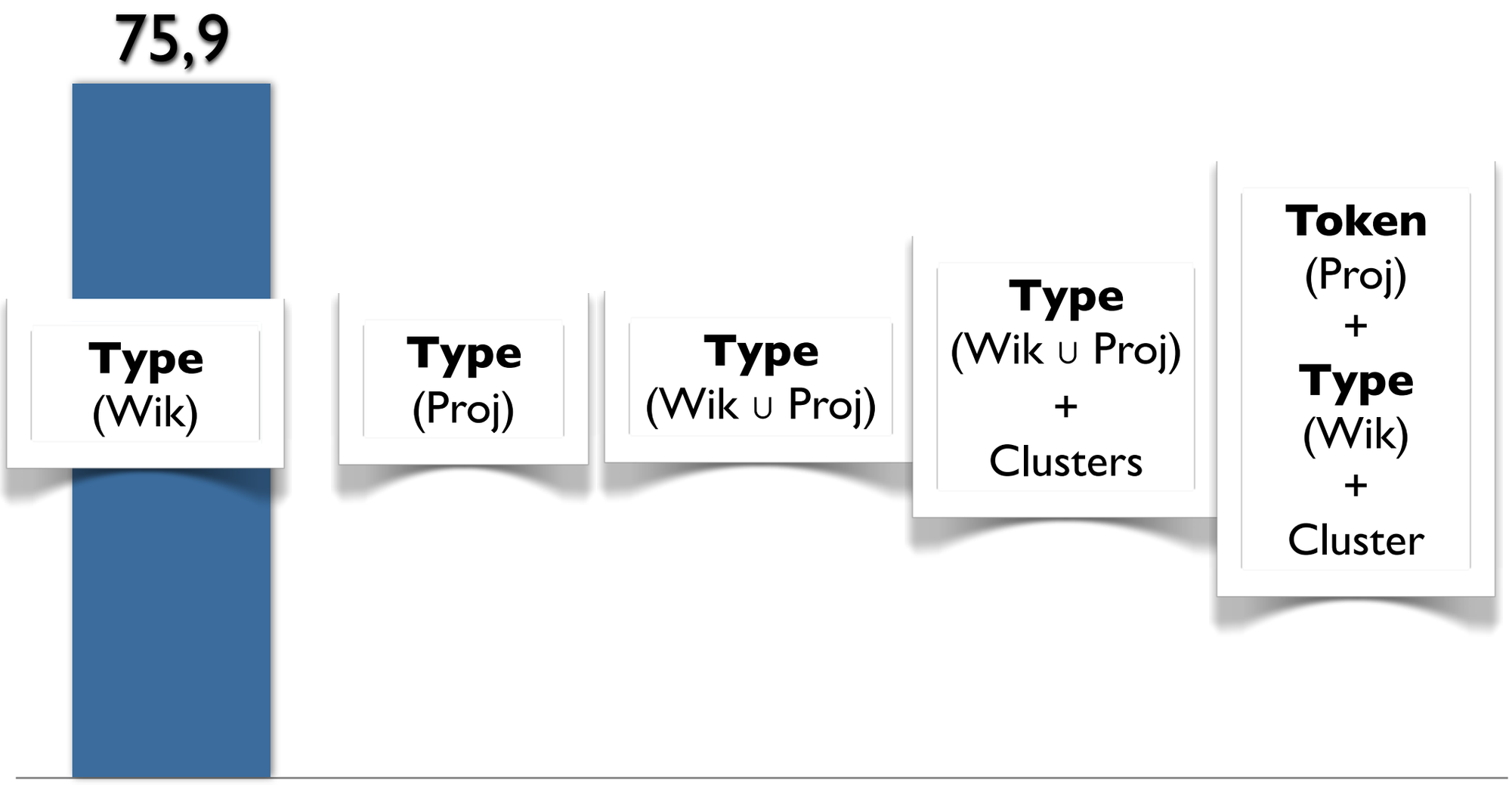
Type
(Wik ∪ Proj)

Type
(Wik ∪ Proj)
+
Clusters

Token
(Proj)
+
Type
(Wik)
+
Cluster

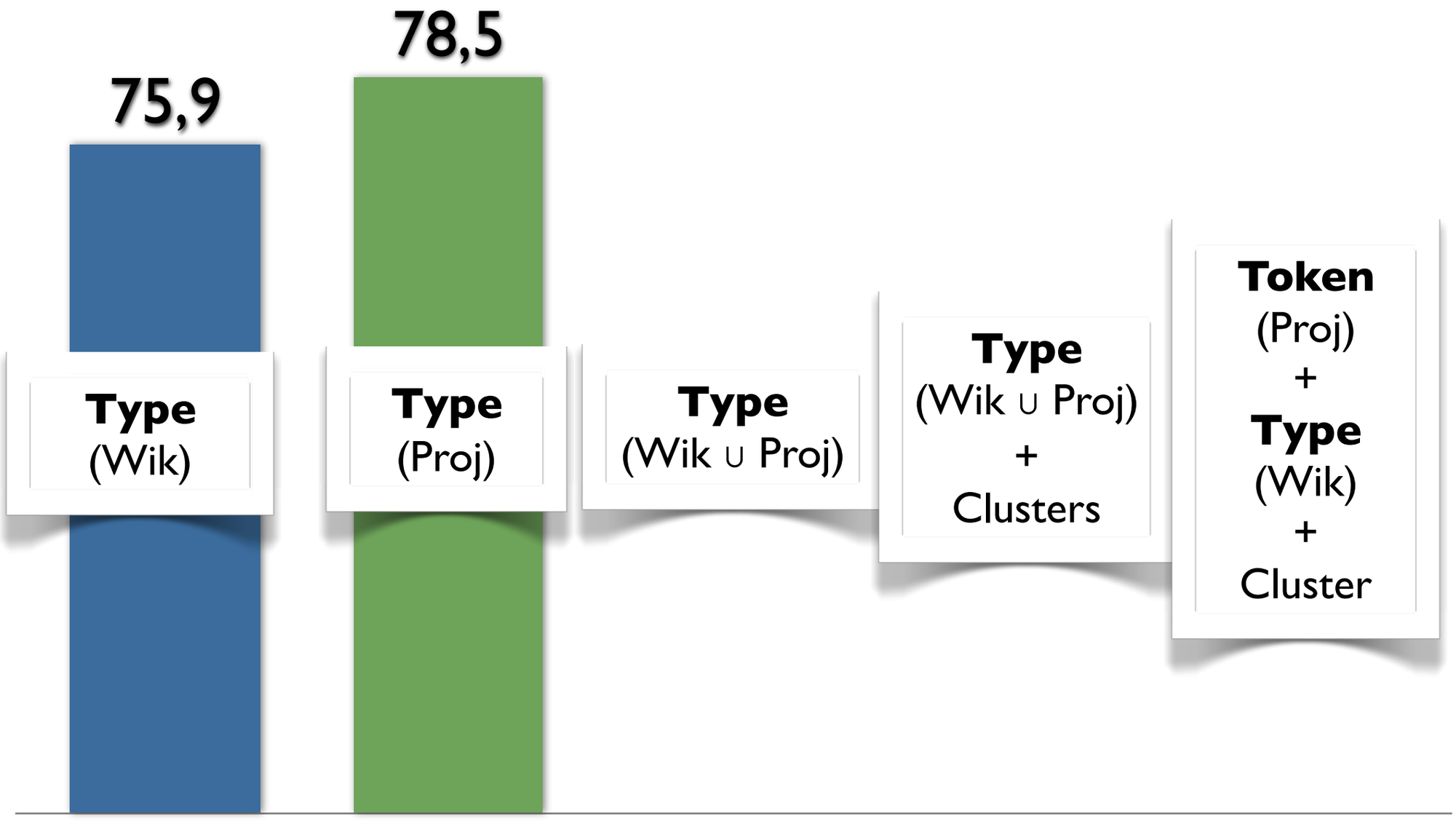
Average across 15 languages

Token and Type Results



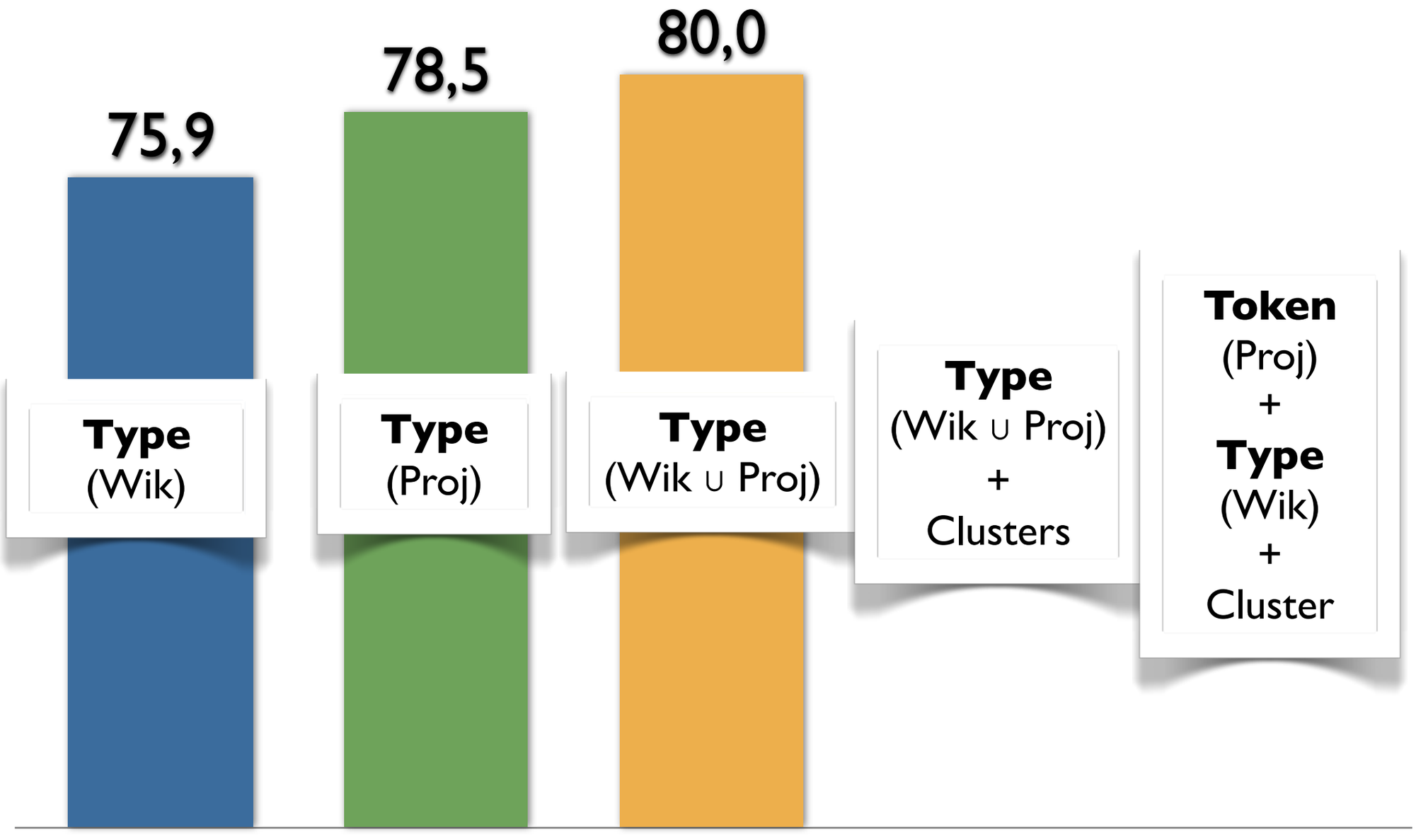
Average across 15 languages

Token and Type Results



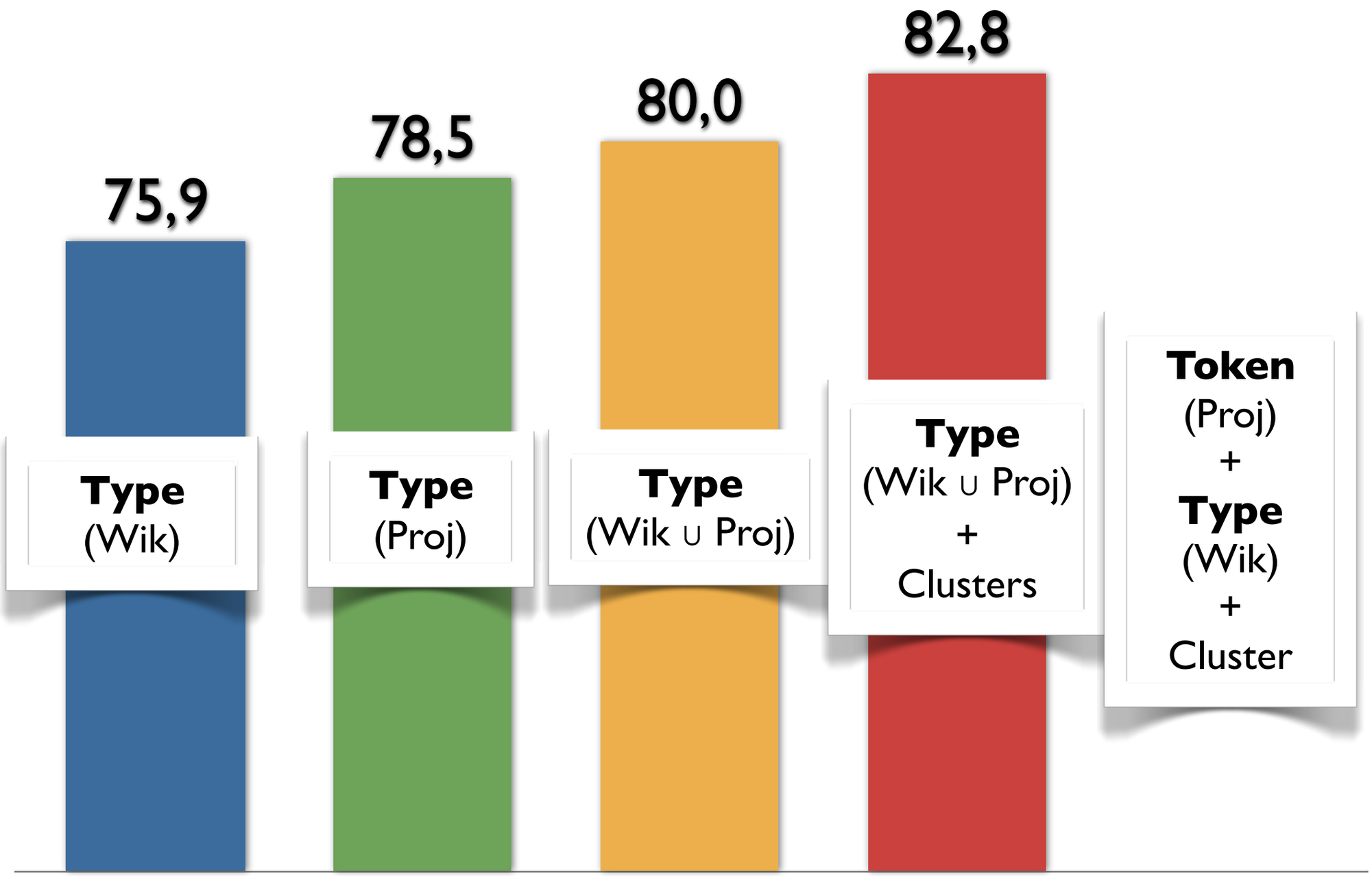
Average across 15 languages

Token and Type Results



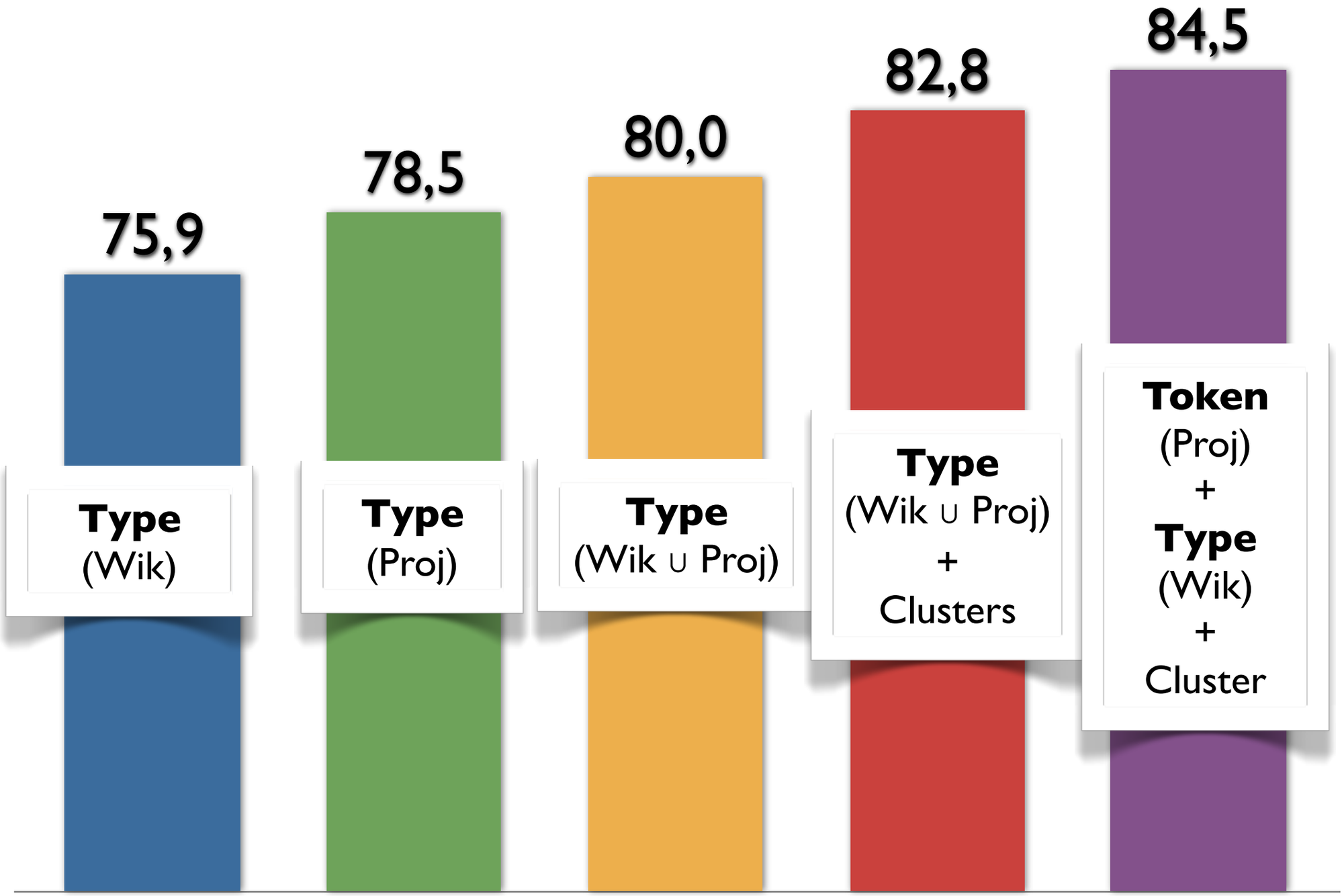
Average across 15 languages

Token and Type Results



Average across 15 languages

Token and Type Results



Average across 15 languages

Comparing with the State of the Art

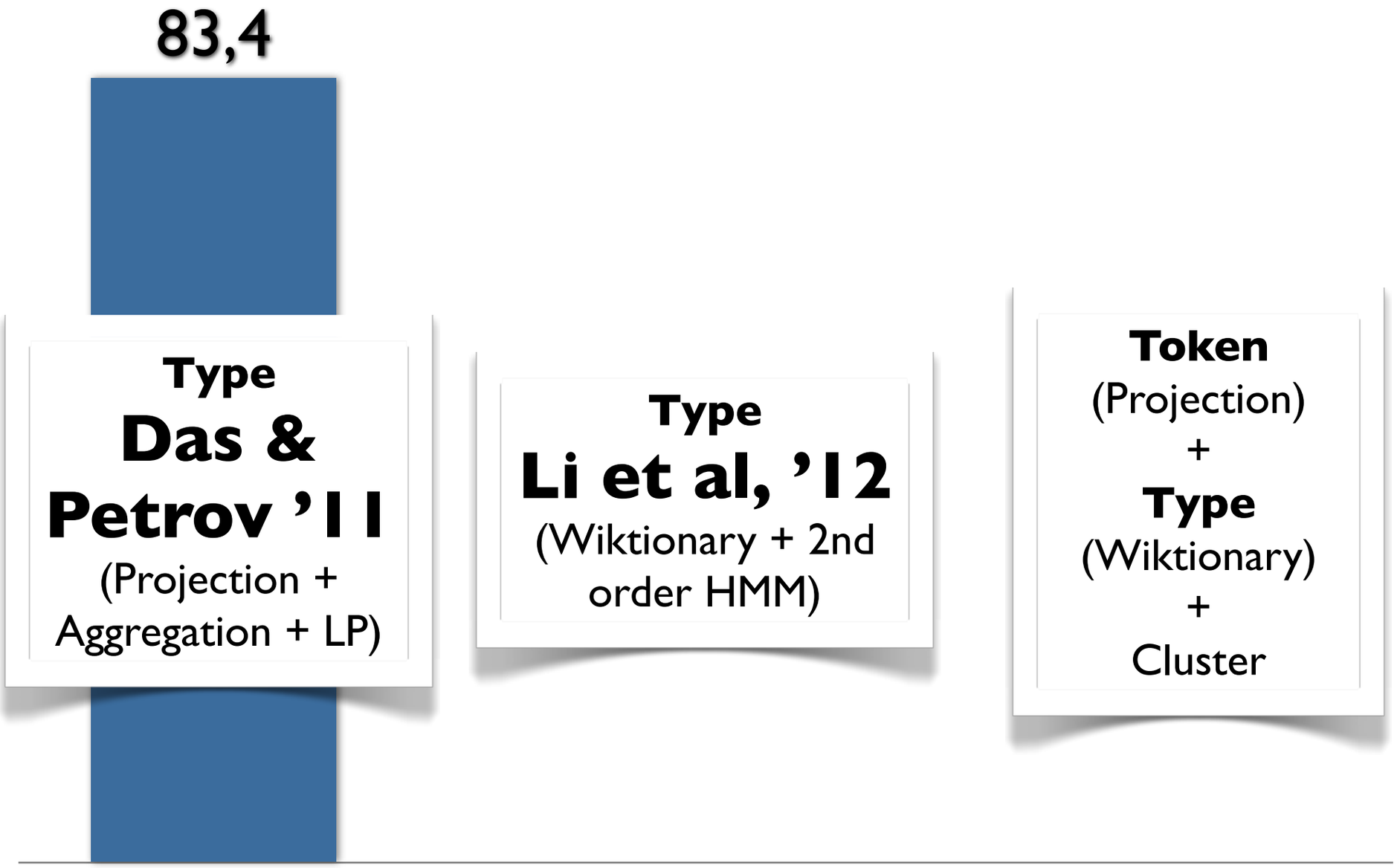
Type
Das & Petrov '11
(Projection + Aggregation + LP)

Type
Li et al, '12
(Wiktionary + 2nd order HMM)

Token
(Projection)
+
Type
(Wiktionary)
+
Cluster

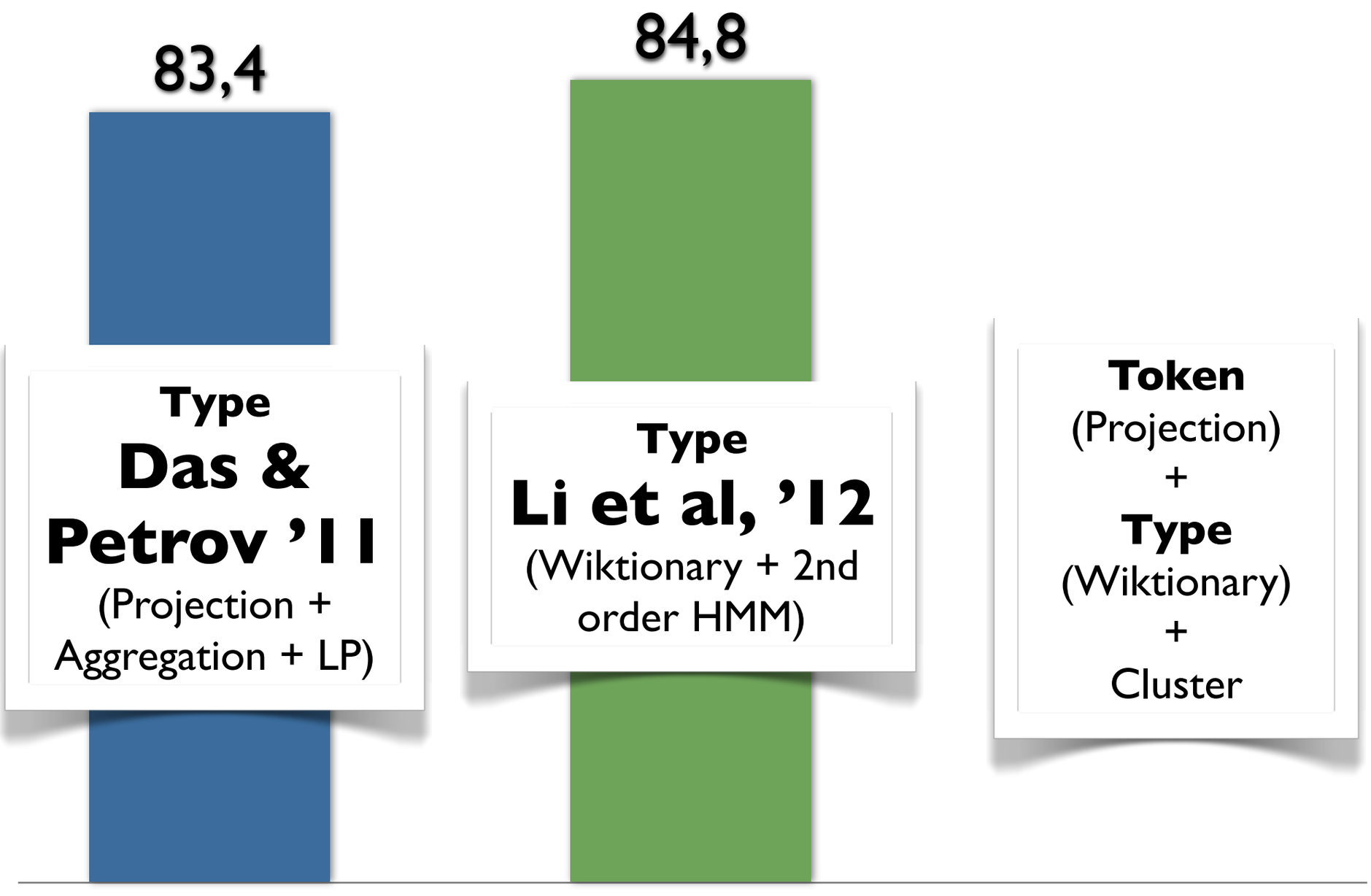
Average across 8 languages

Comparing with the State of the Art



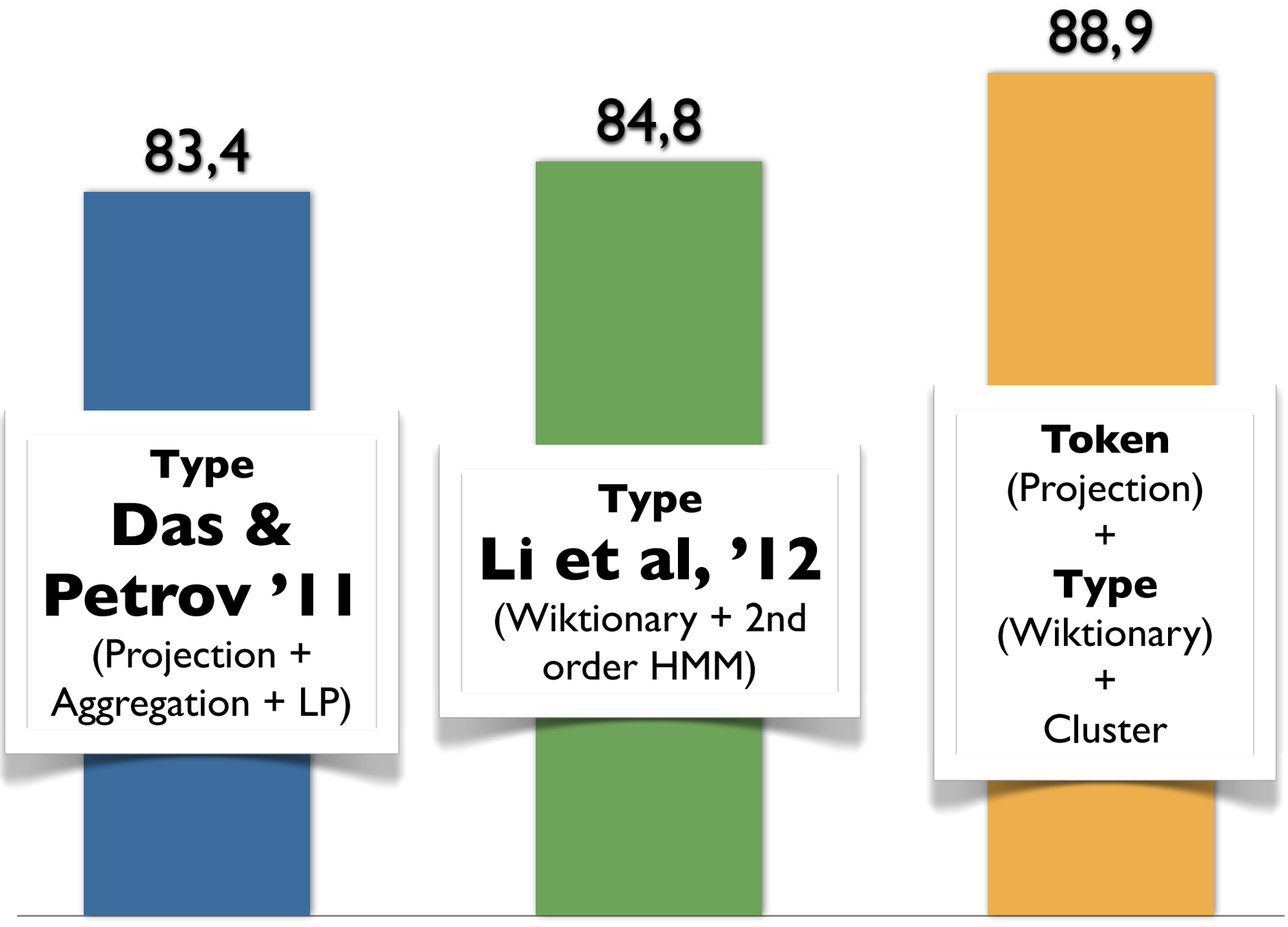
Average across 8 languages

Comparing with the State of the Art



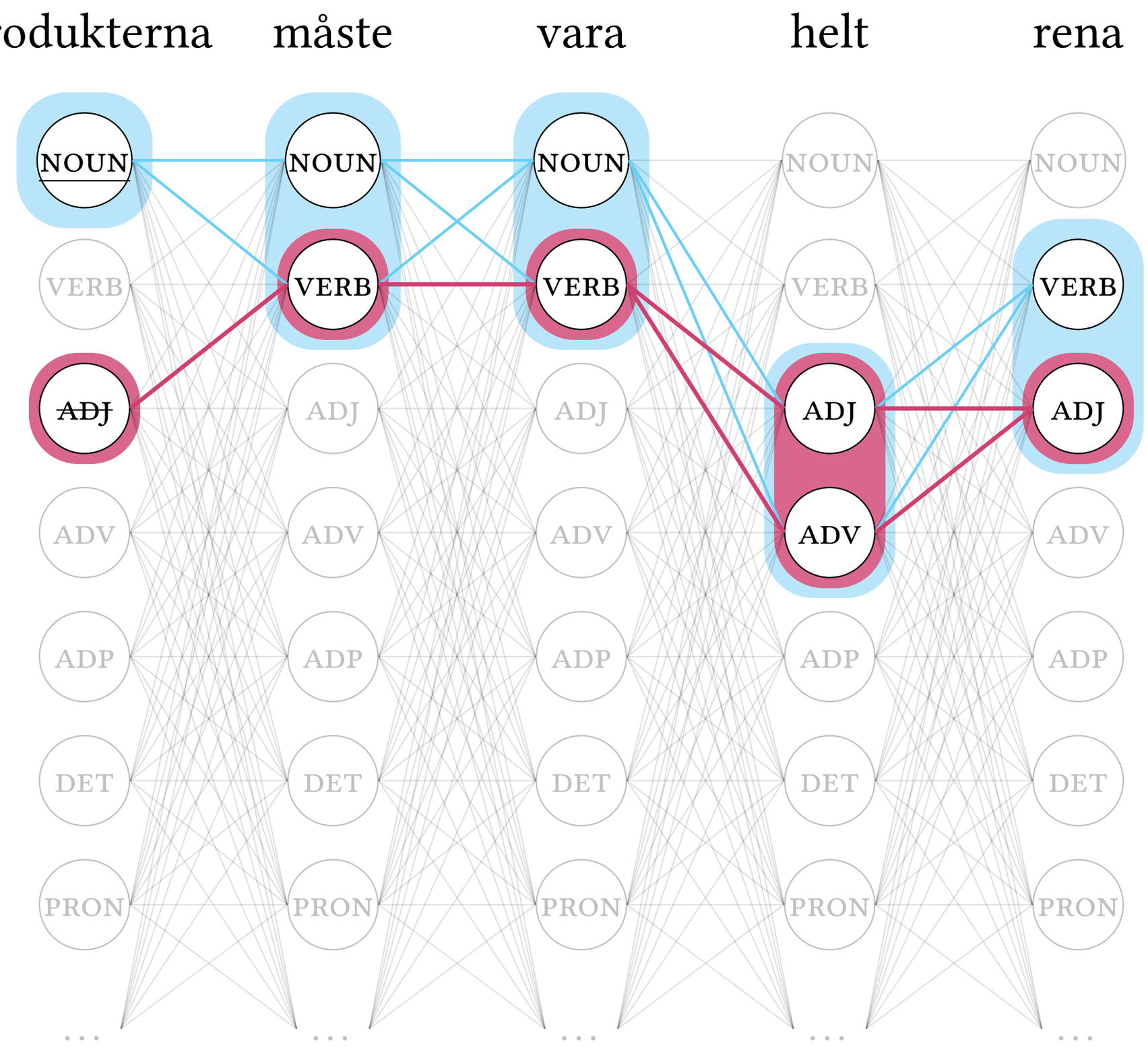
Average across 8 languages

Comparing with the State of the Art

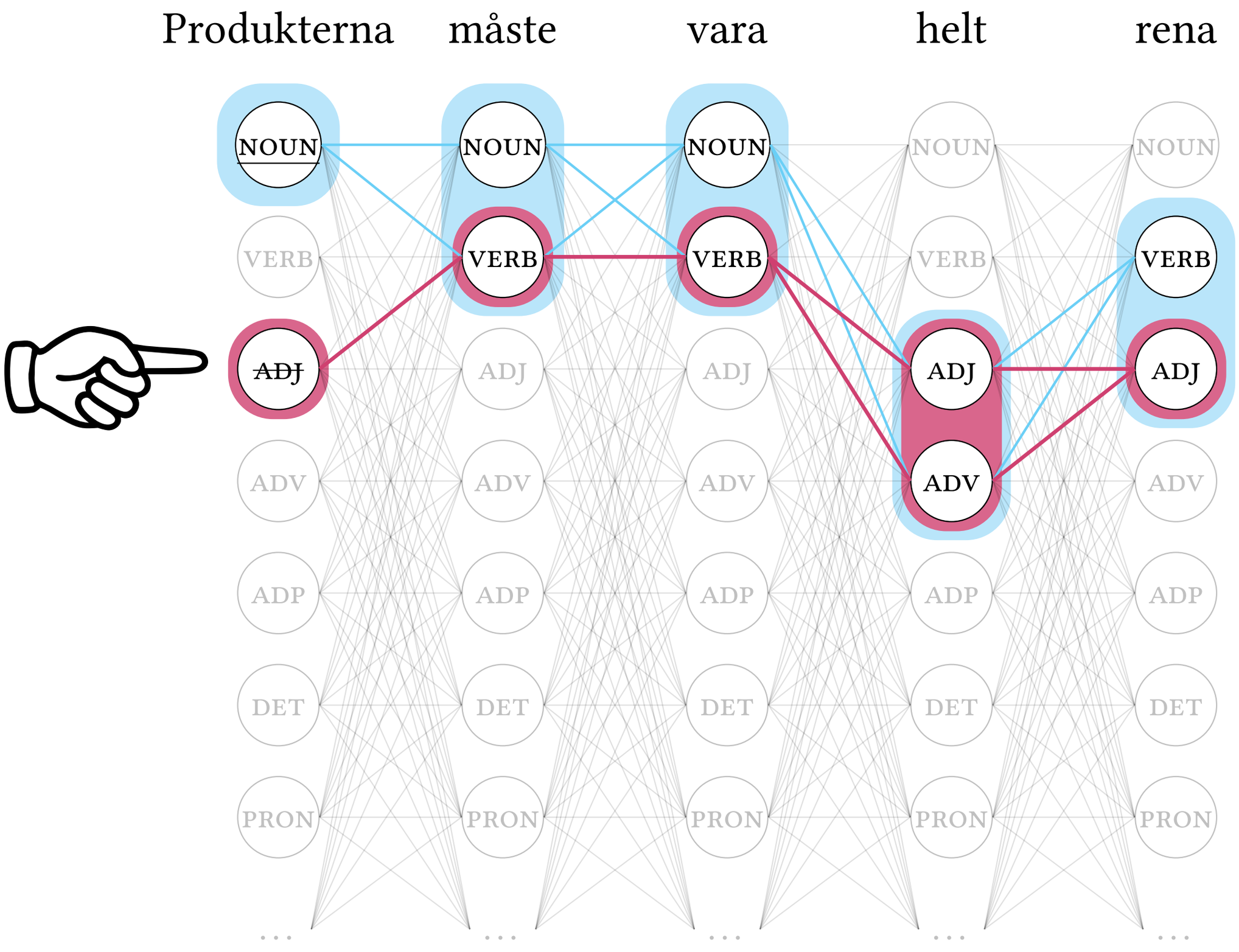


Average across 8 languages

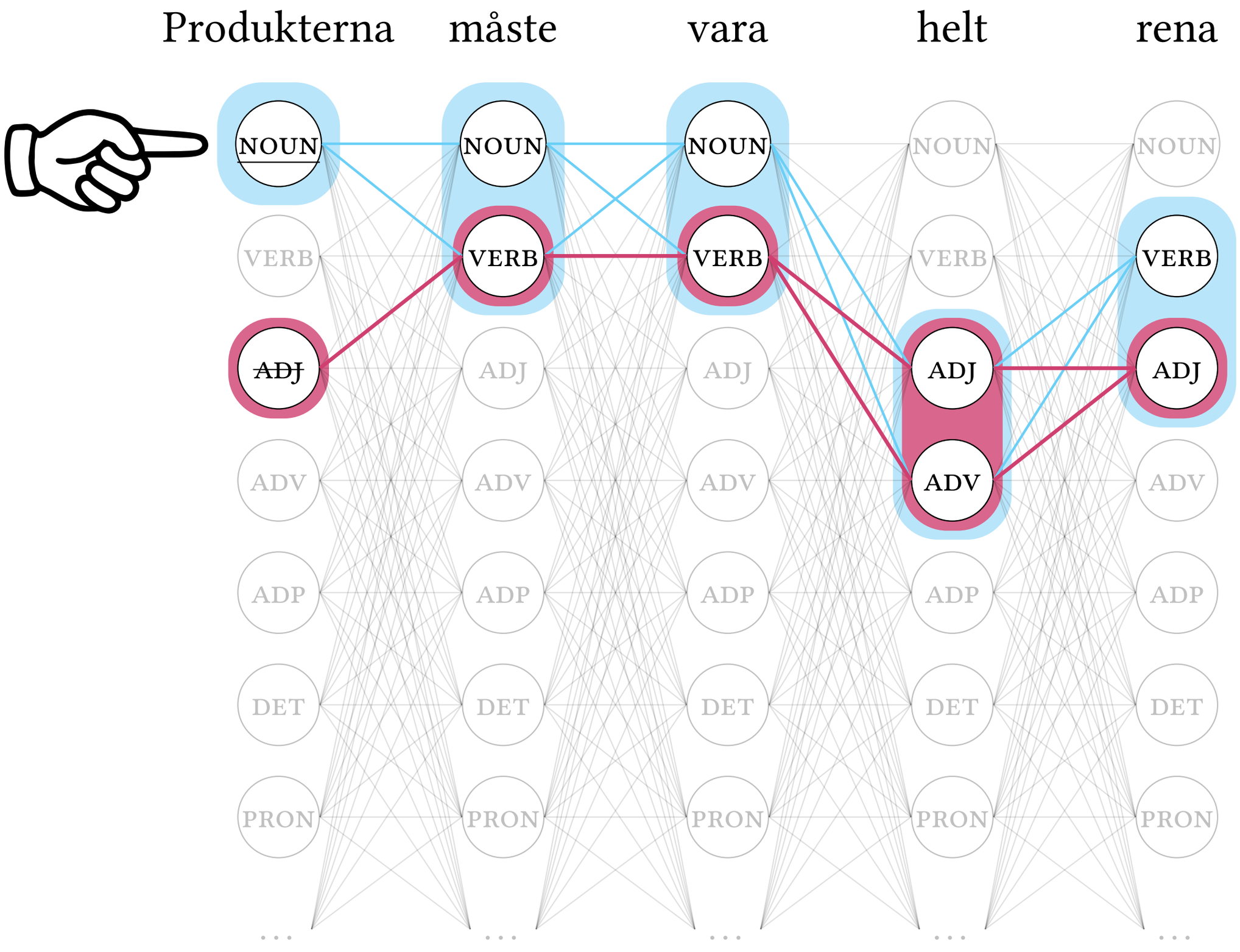
Projection Errors



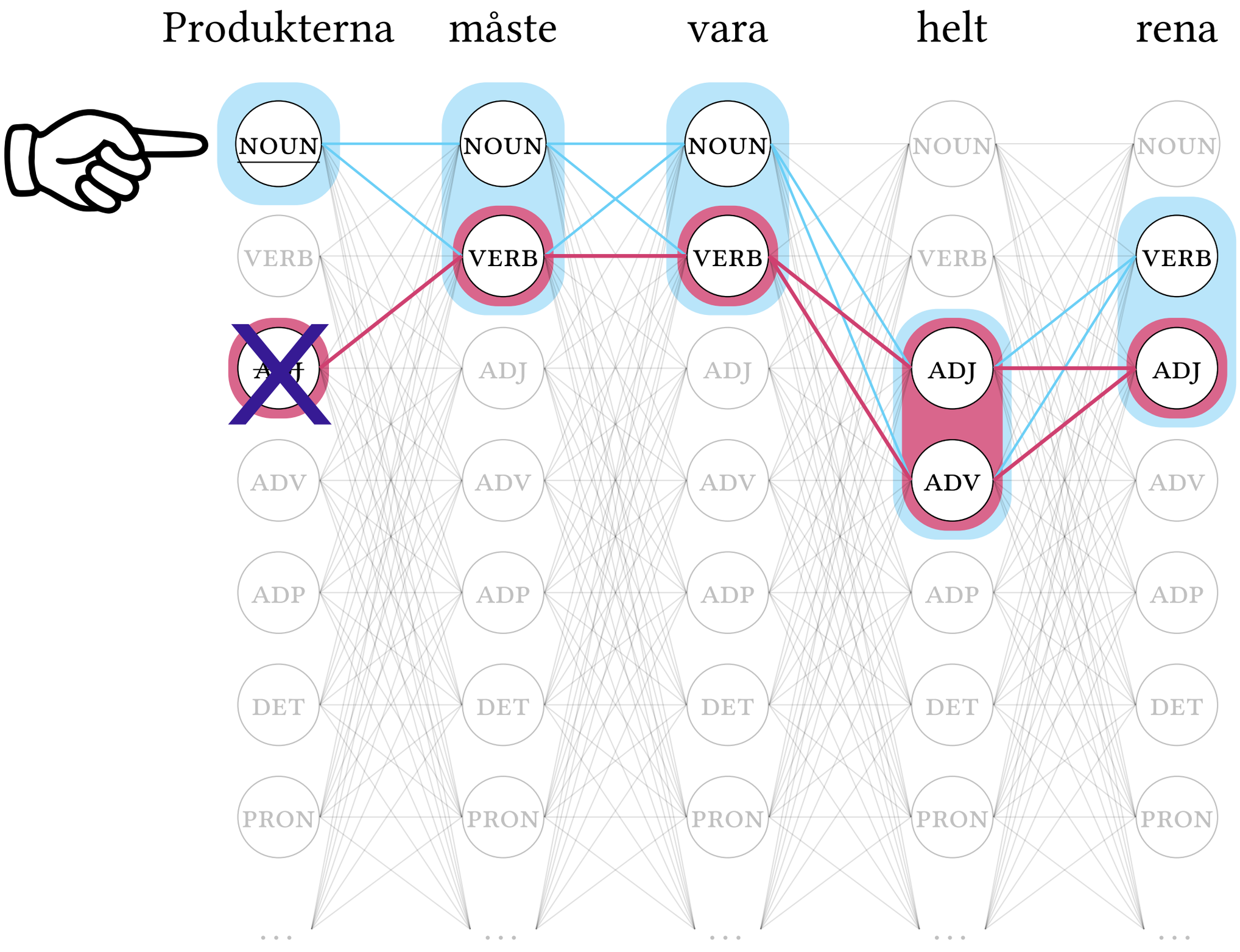
Projection Errors



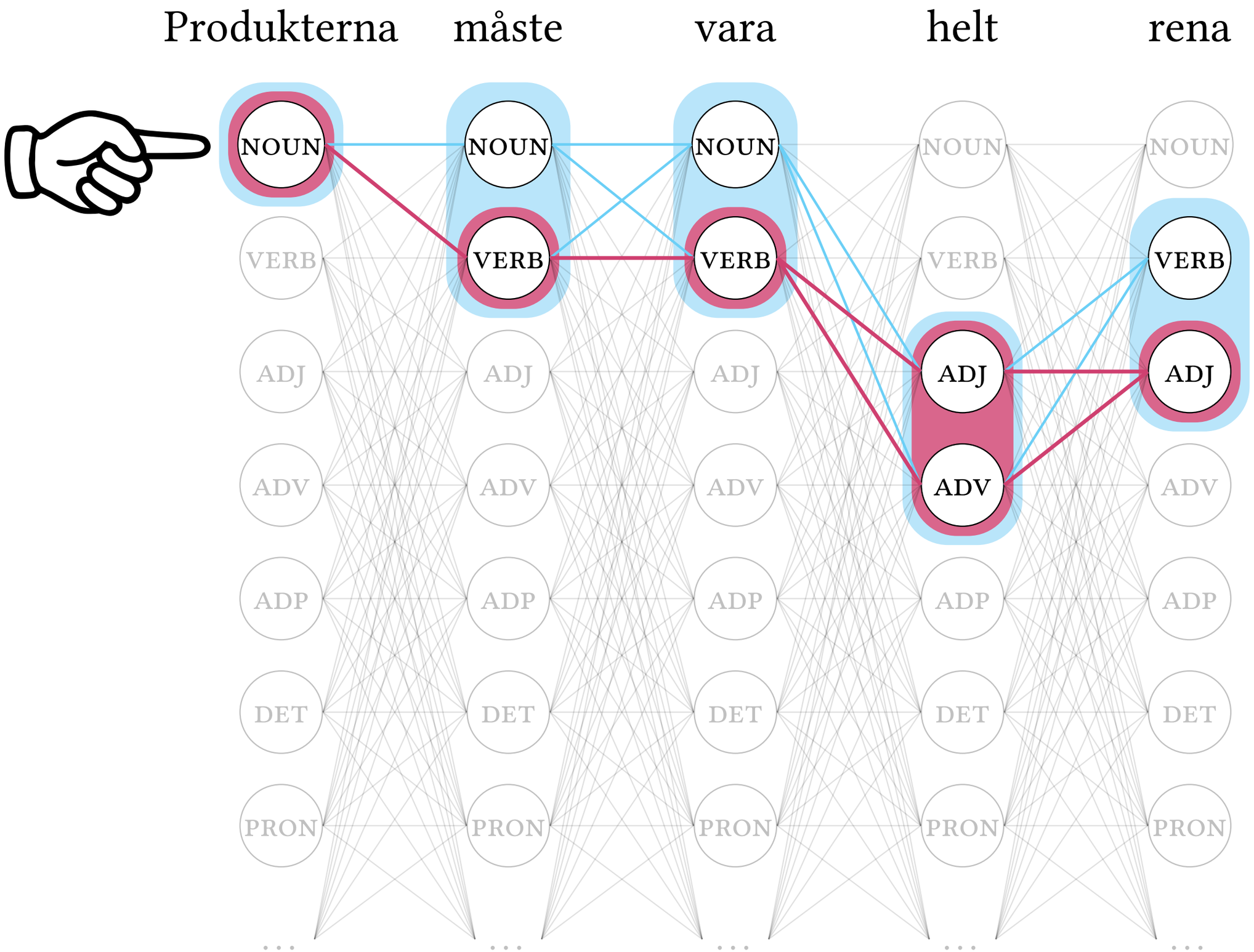
Projection Errors



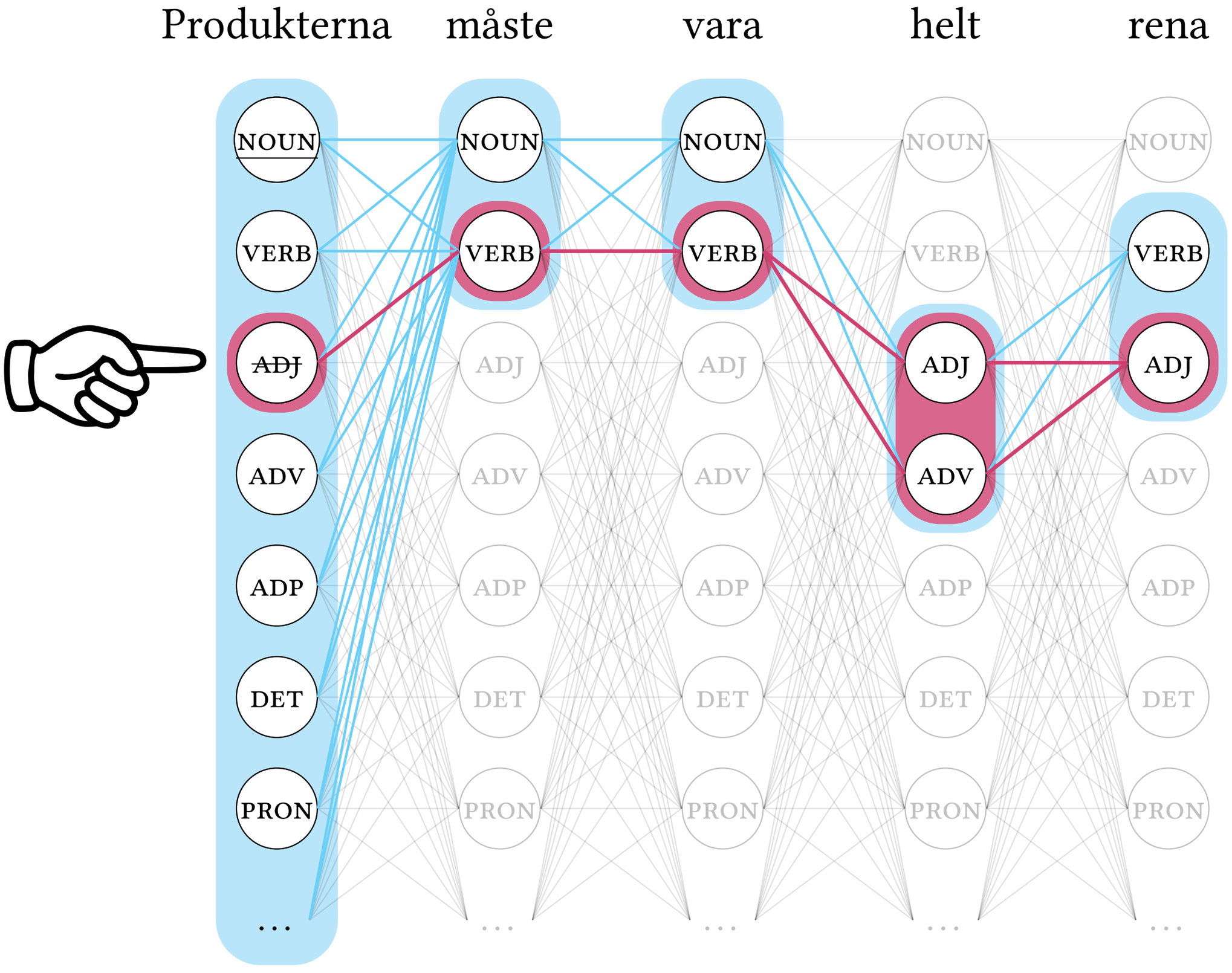
Projection Errors



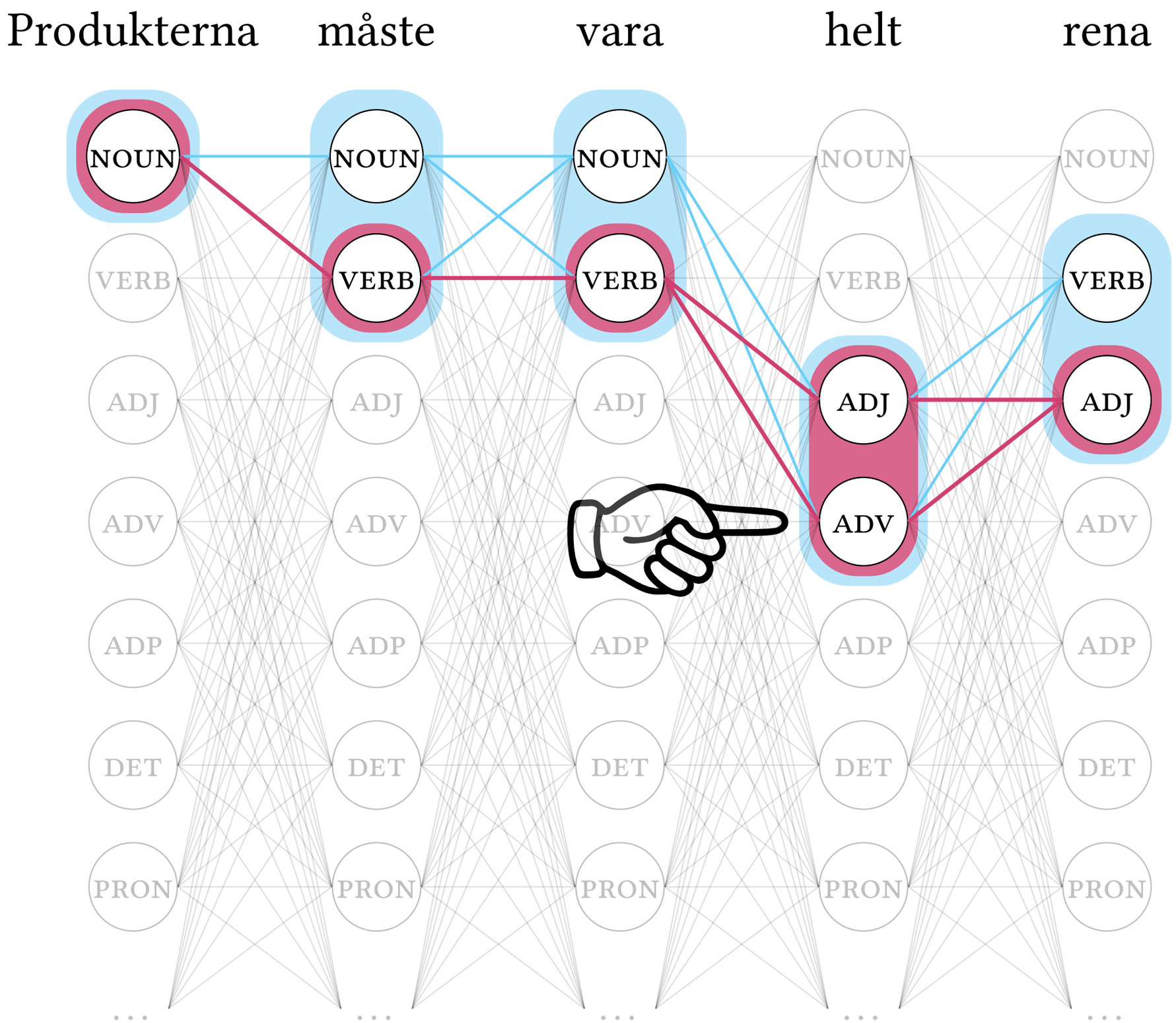
Projection Errors



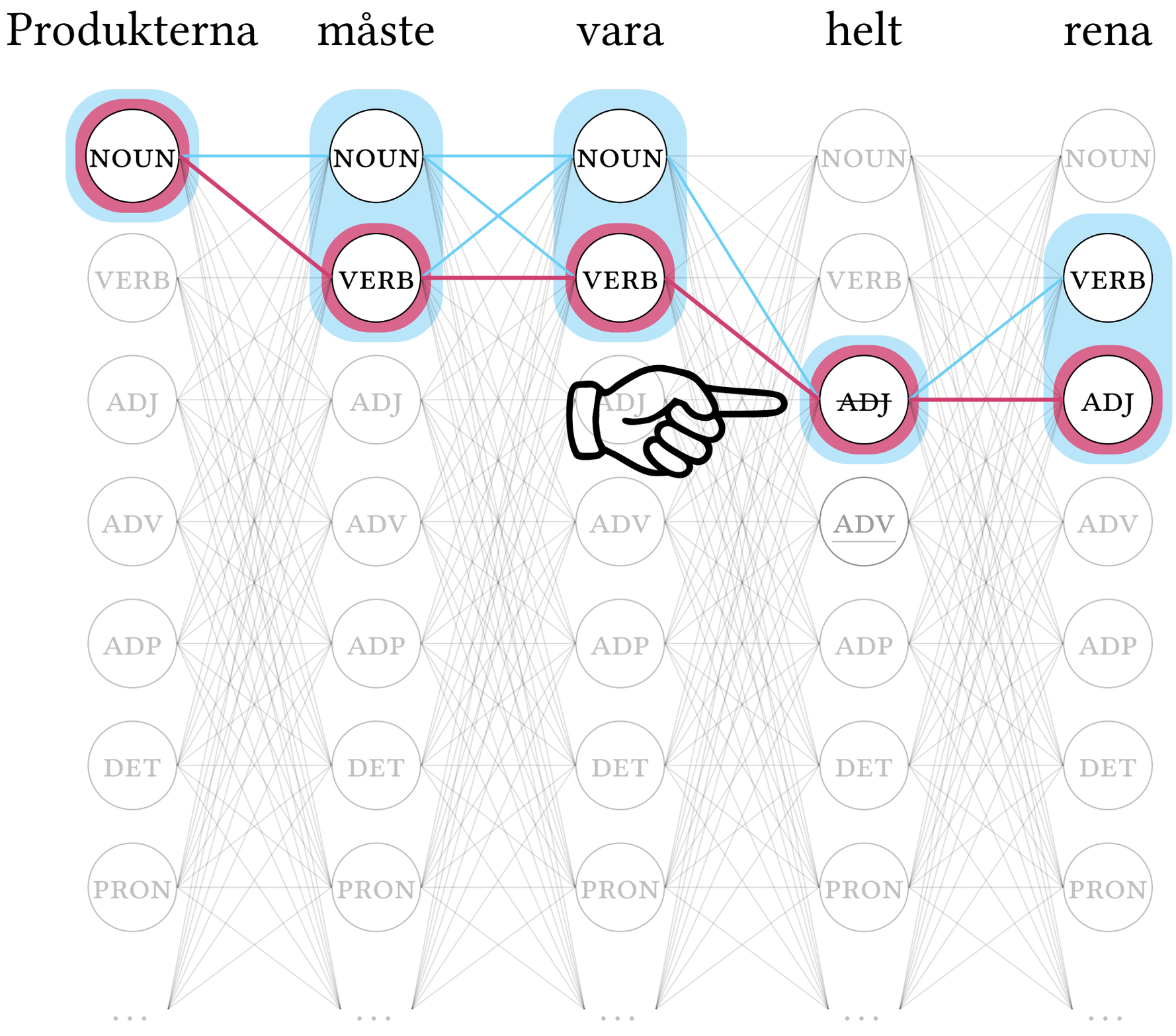
Projection Errors



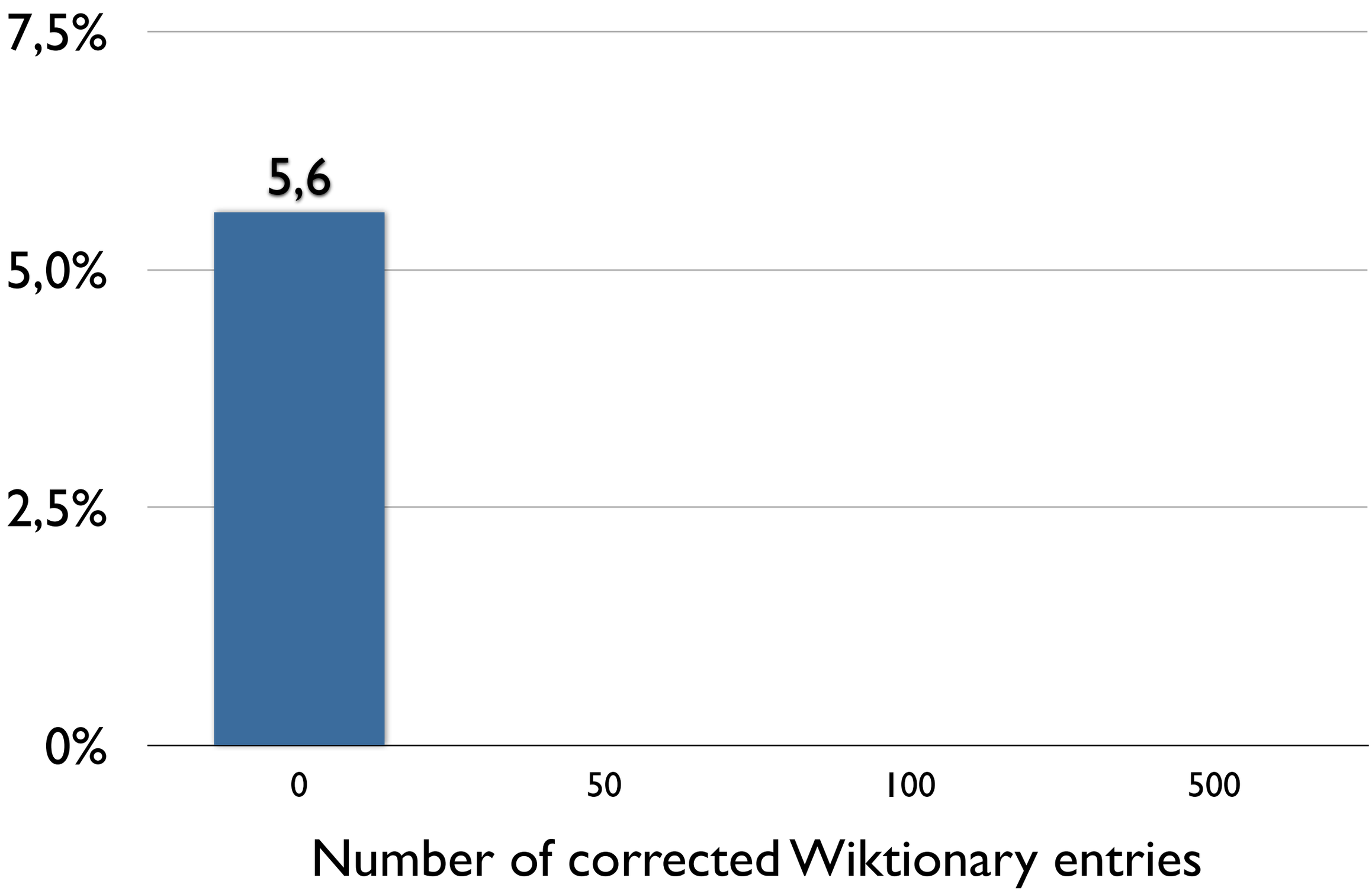
Pruning Errors



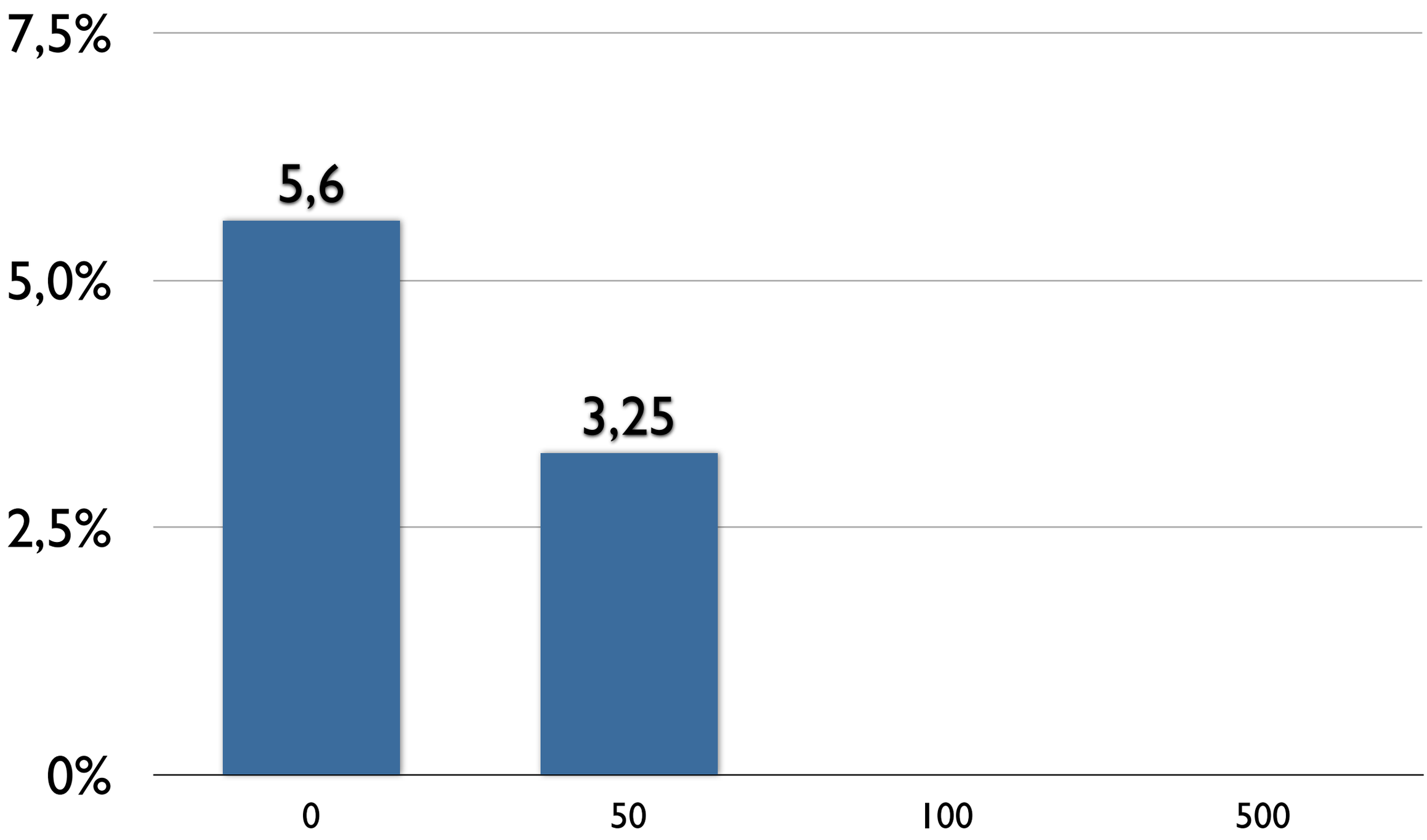
Pruning Errors



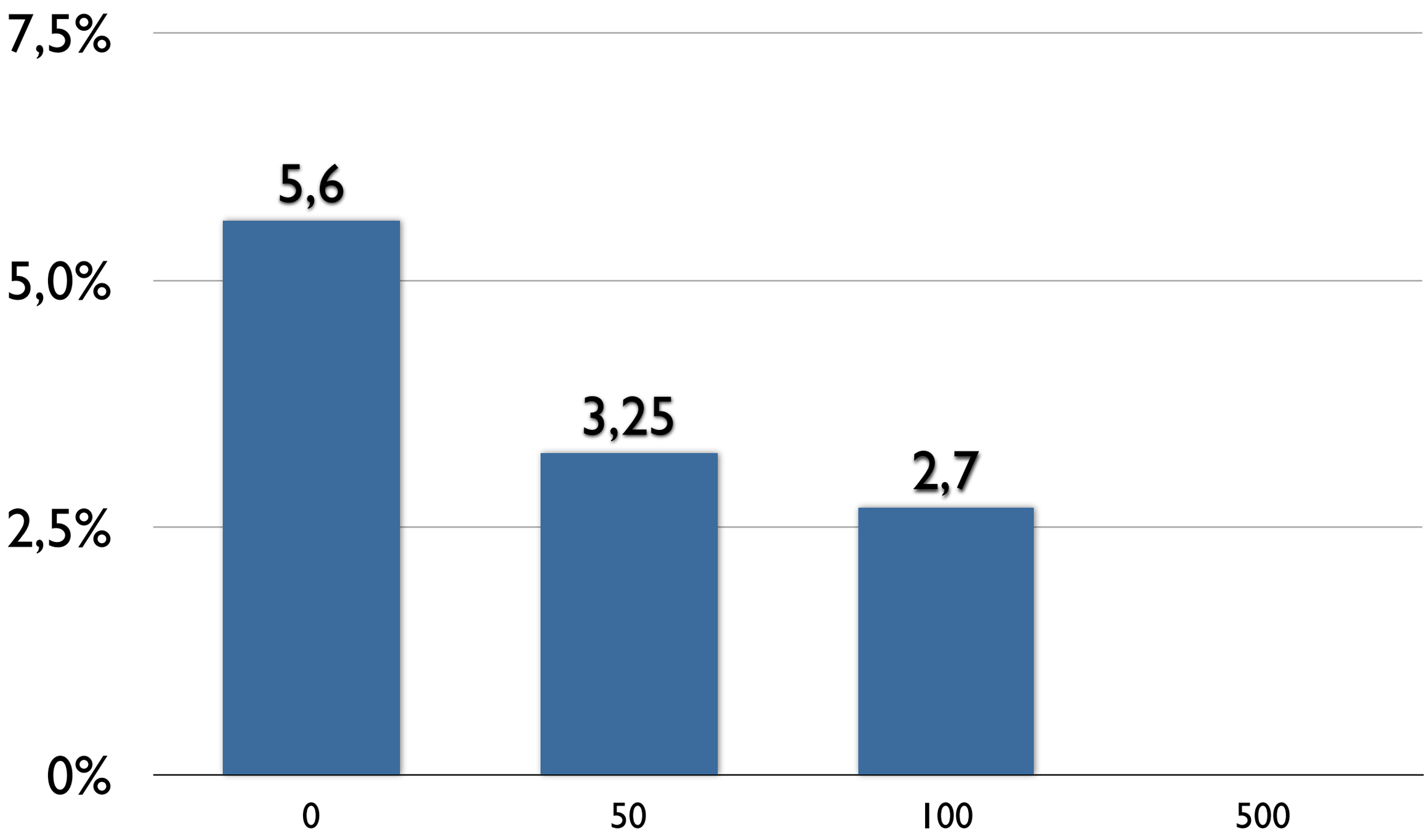
Pruning Errors



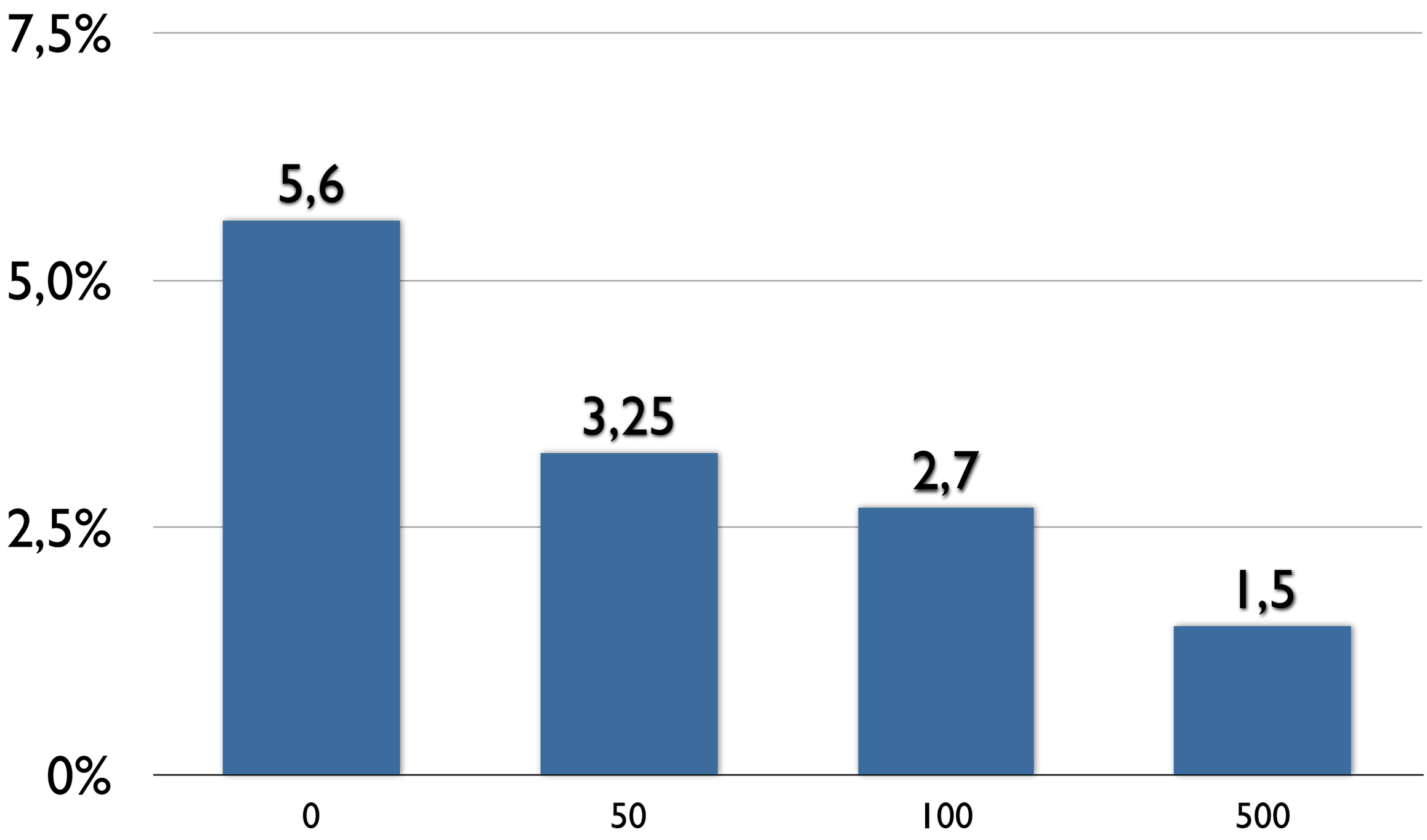
Pruning Errors



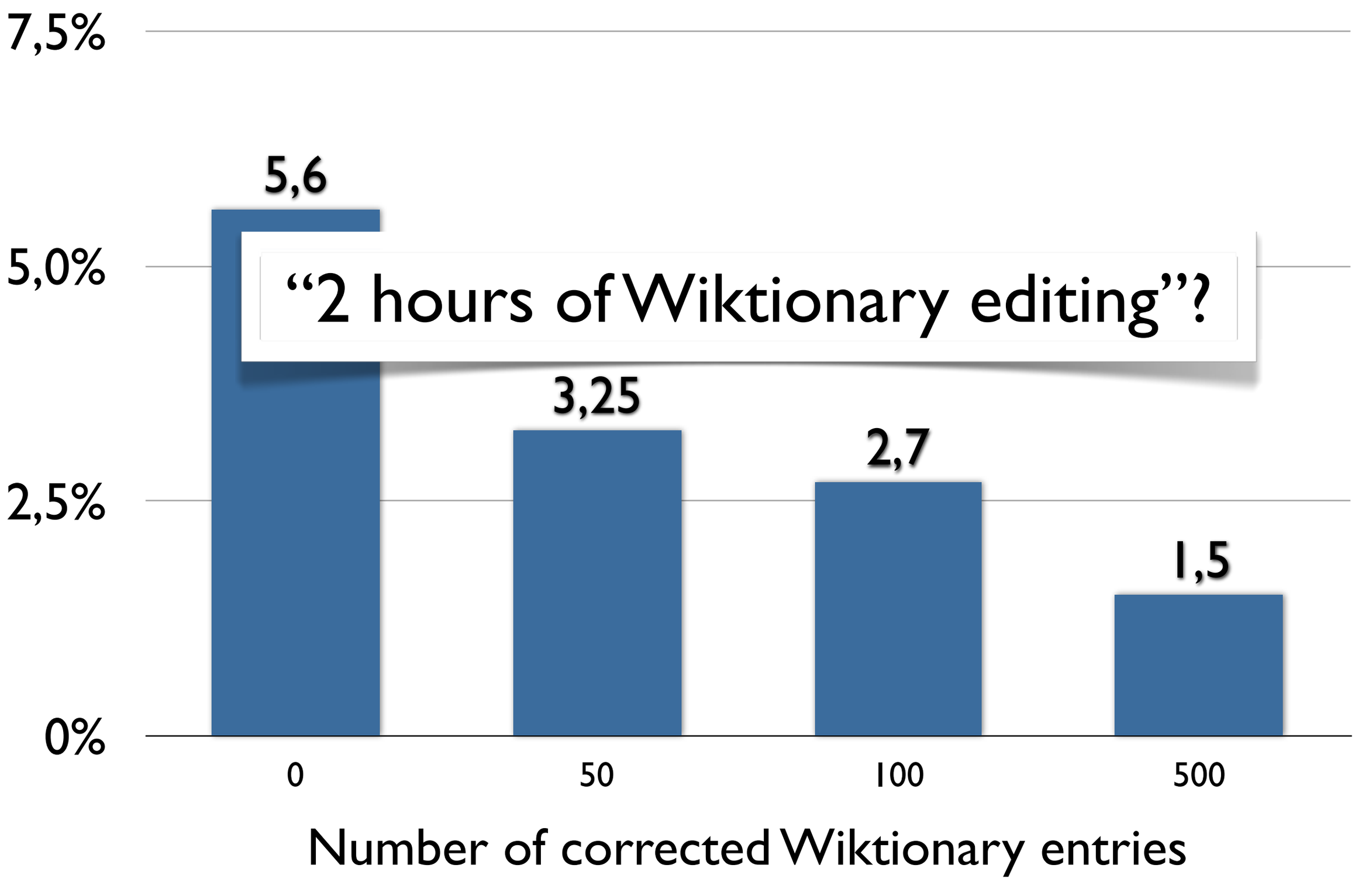
Pruning Errors



Pruning Errors



Pruning Errors



Conclusion

- Powerful discriminative latent variable models
- Coupled type and token constraints
- >25% error reduction over state of the art
- No label propagation, model minimization, etc.
- Potential extensions: NER, frame semantics, ...

Thanks!

Alexander Rush, Klaus Macherey
Yoav Goldberg, Keith Hall, Kuzman Ganchev and Hao Zhang