GRAPH-BASED POSTERIOR REGULARIZATION FOR SEMI-SUPERVISED STRUCTURED PREDICTION

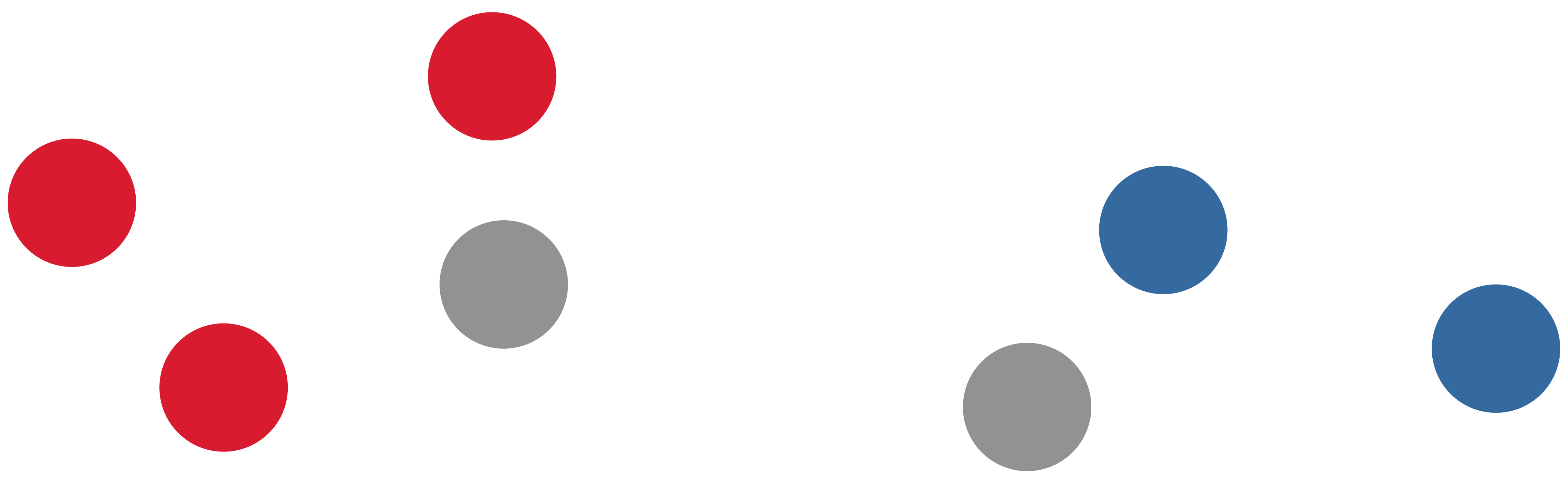
Luheng He Jennifer Gillenwater
University of Pennsylvania

Ben Taskar
University of Washington

GRAPH-BASED LEARNING

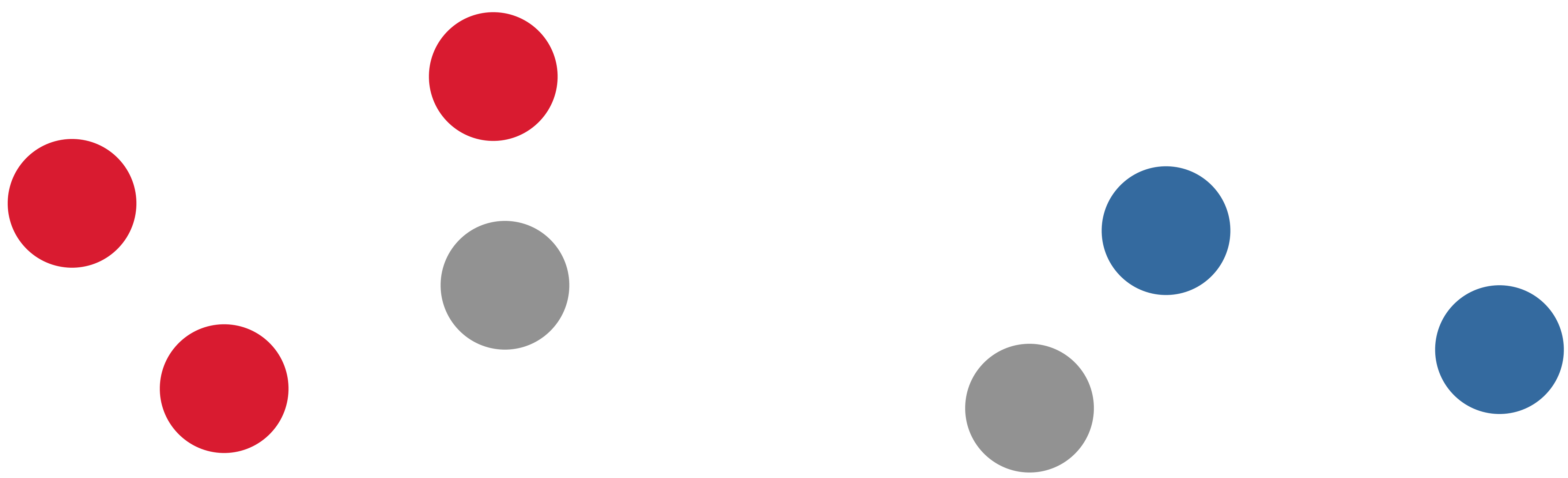
GRAPH-BASED LEARNING

GRAPH-BASED LEARNING



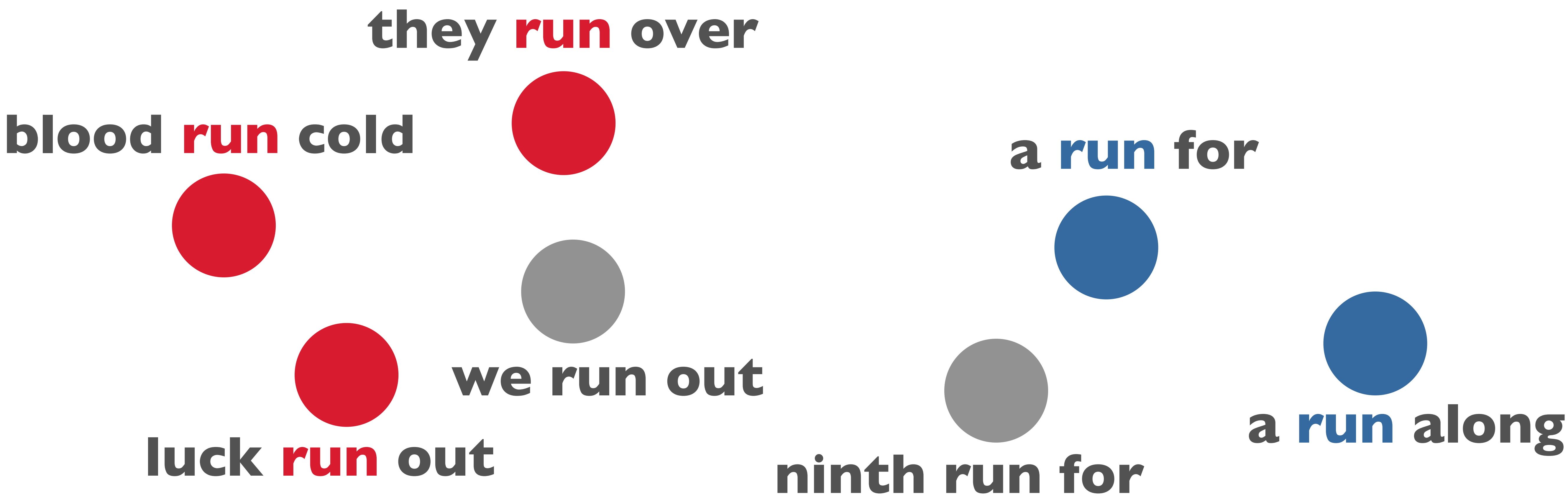
GRAPH-BASED LEARNING

Labels: verb (V), noun (N), etc.



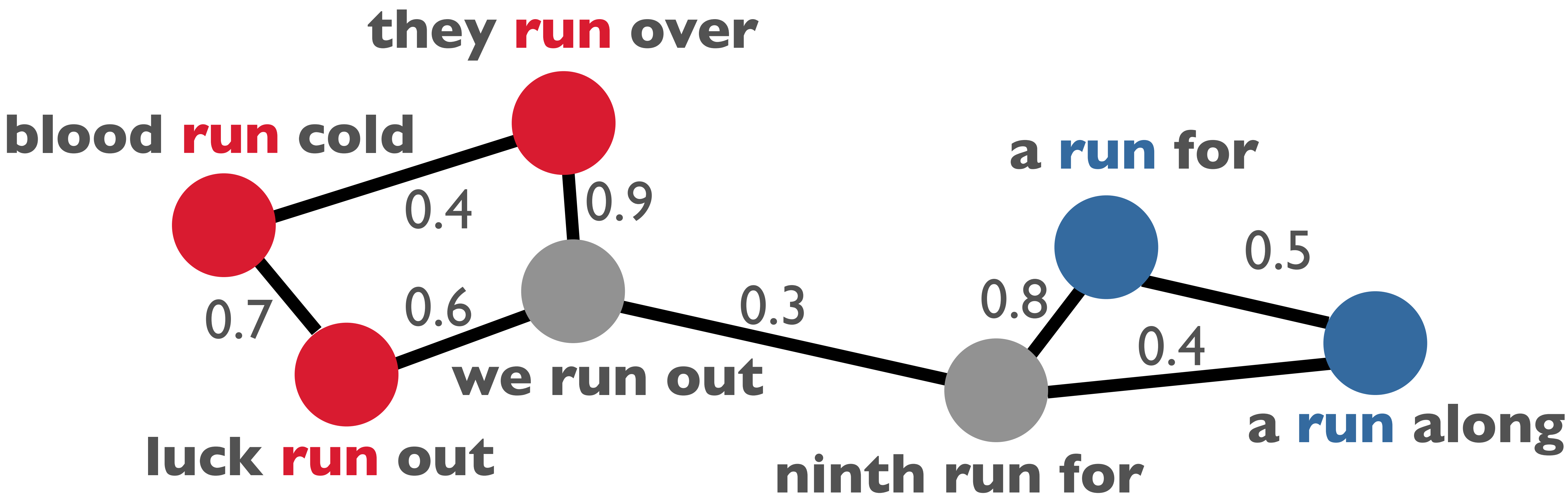
GRAPH-BASED LEARNING

Labels: verb (V), noun (N), etc.



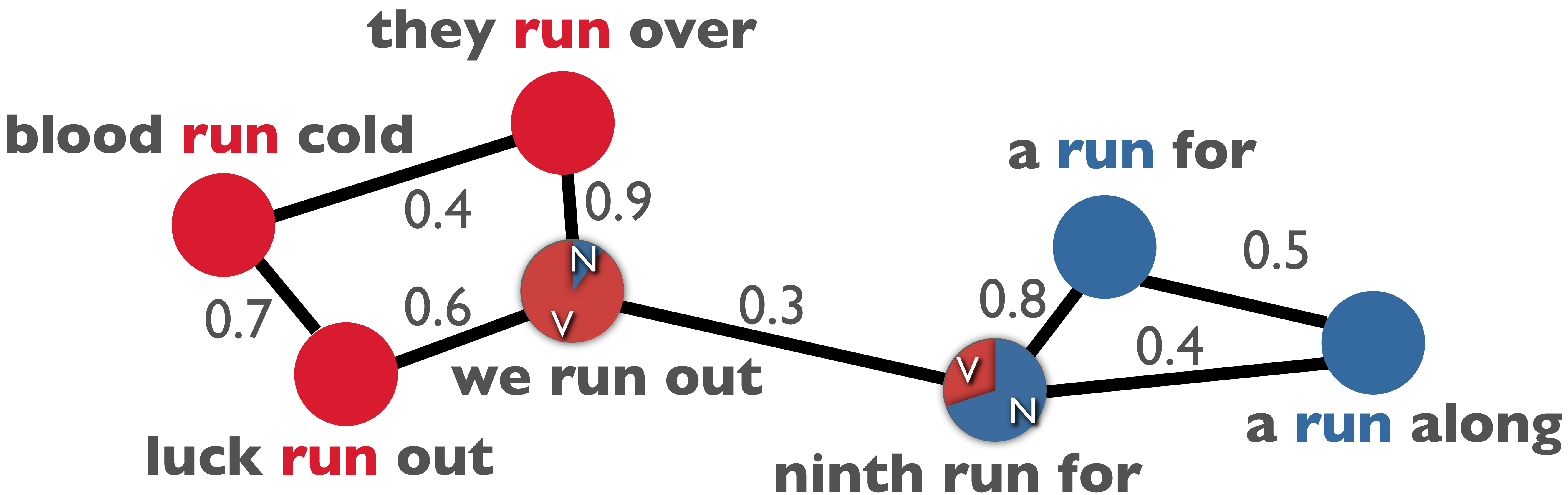
GRAPH-BASED LEARNING

Labels: verb (V), noun (N), etc.



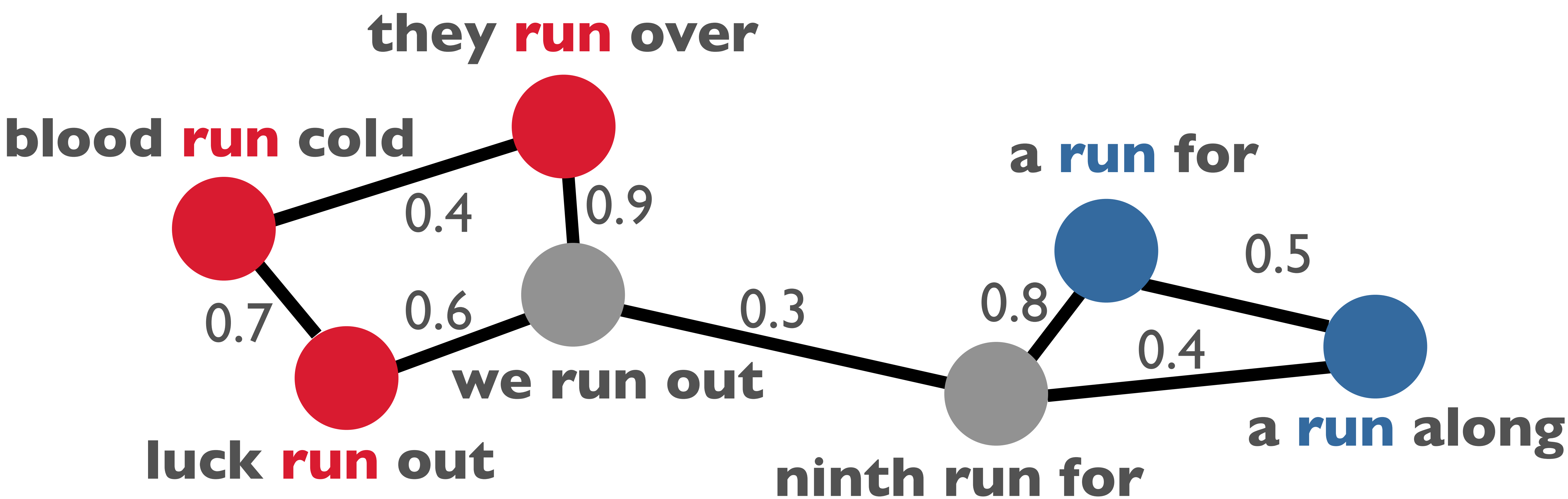
GRAPH-BASED LEARNING

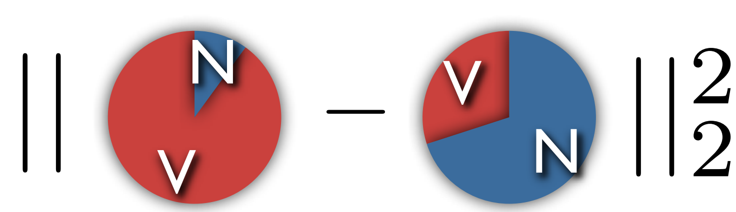
Labels: verb (V), noun (N), etc.



GRAPH-BASED LEARNING

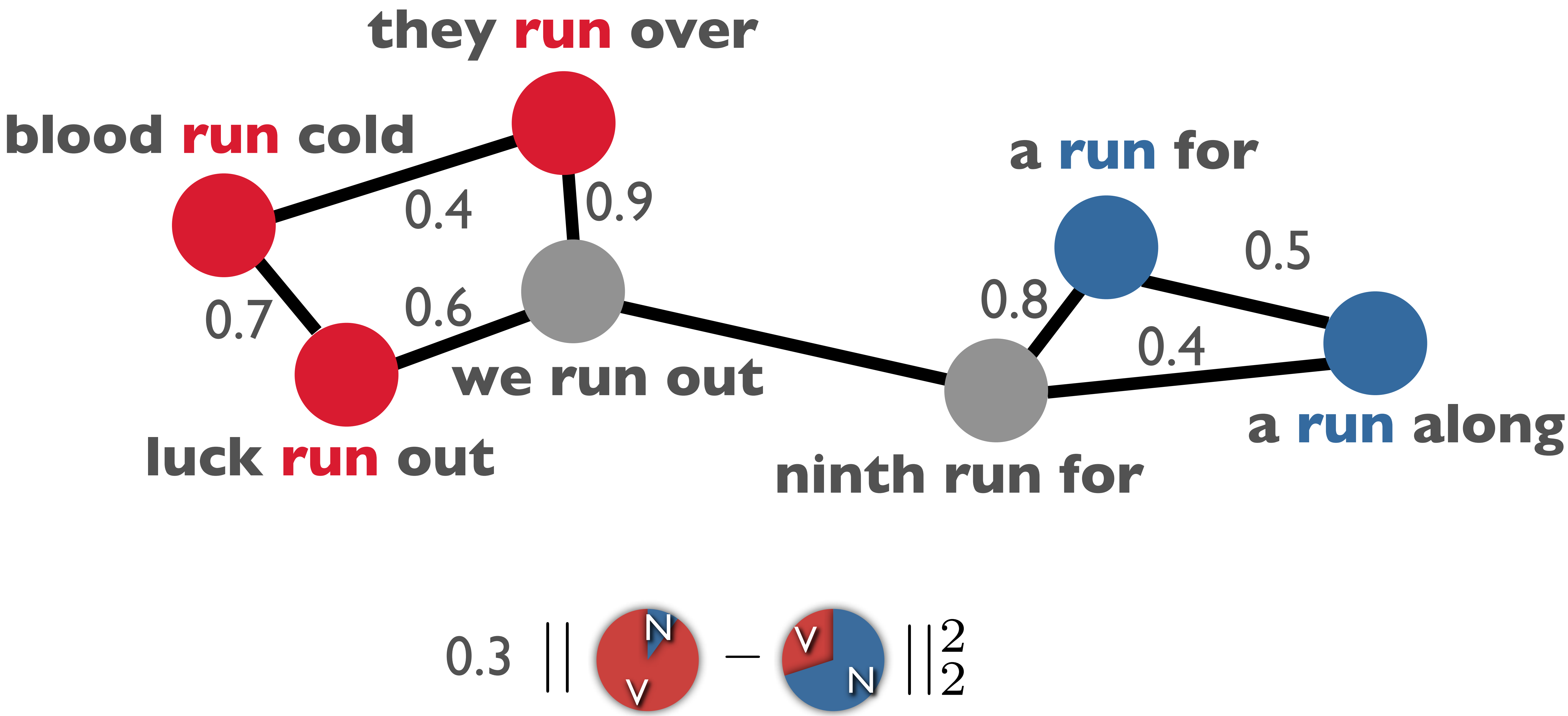
Labels: verb (V), noun (N), etc.





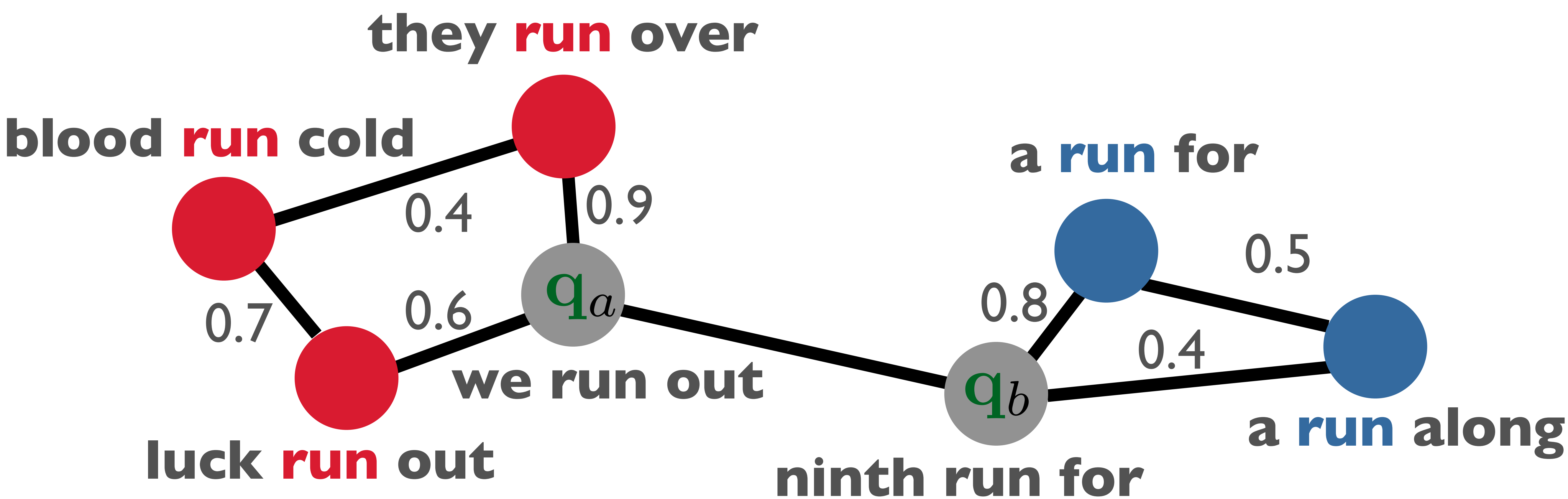
GRAPH-BASED LEARNING

Labels: verb (V), noun (N), etc.



GRAPH-BASED LEARNING

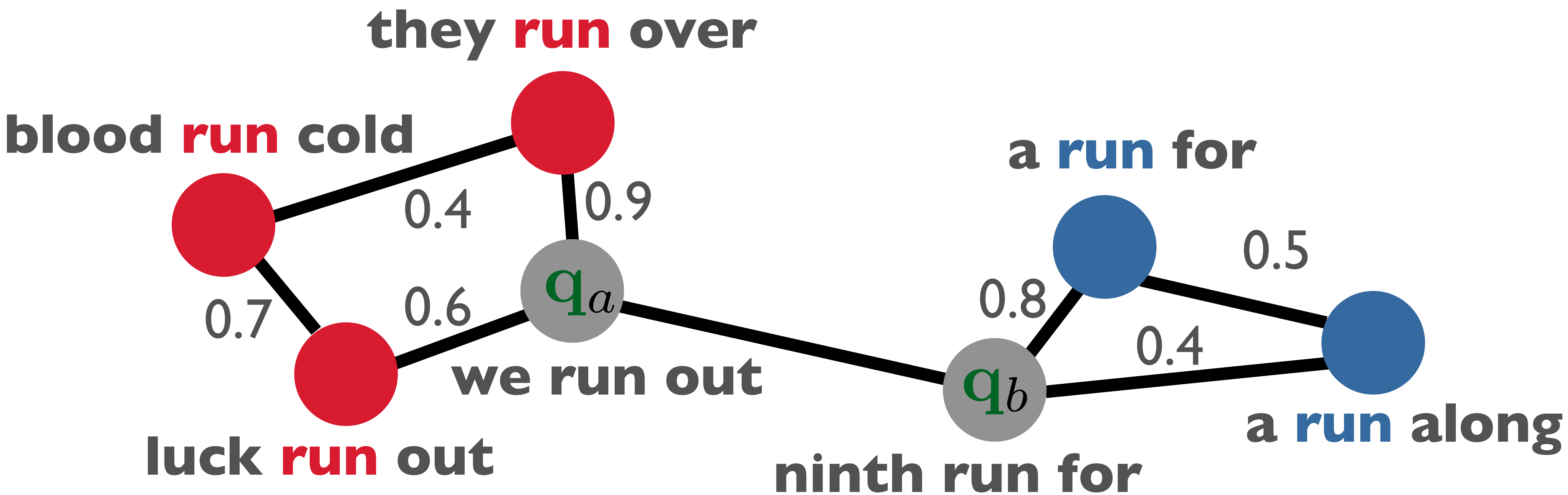
Labels: verb (V), noun (N), etc.



$$w_{ab}||\mathbf{q}_a - \mathbf{q}_b||_2^2$$

GRAPH-BASED LEARNING

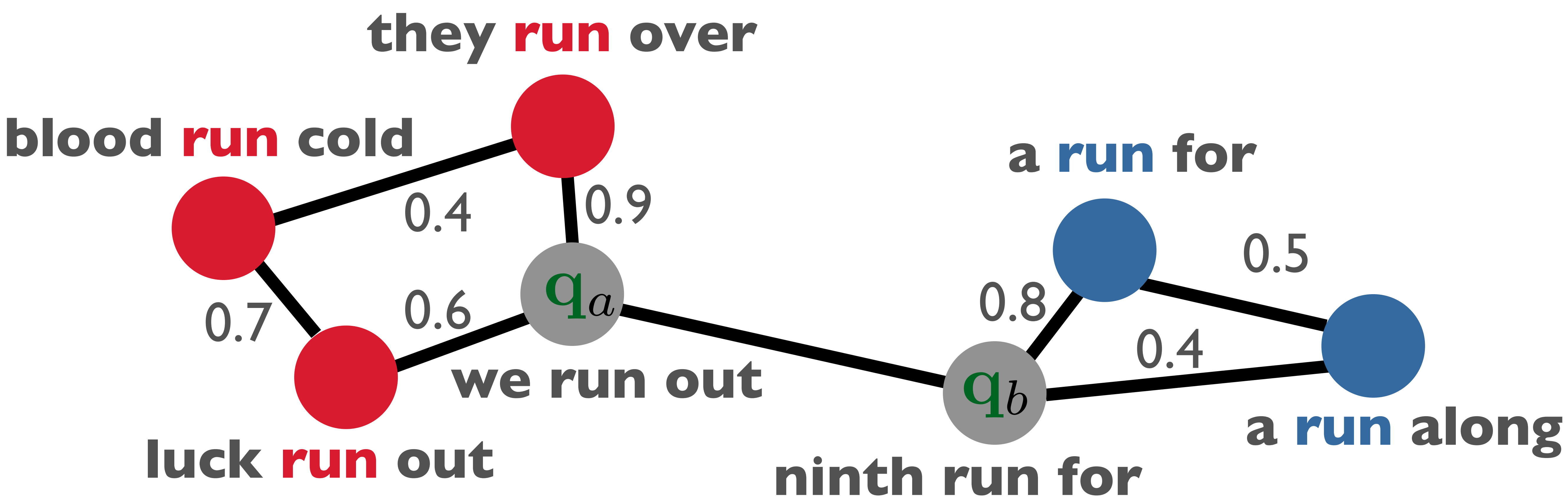
Labels: verb (V), noun (N), etc.



$$\mathrm{Lap}(q) = \quad w_{ab}||\mathbf{q}_a - \mathbf{q}_b||_2^2$$

GRAPH-BASED LEARNING

Labels: verb (V), noun (N), etc.



$$\mathrm{Lap}(q) = \sum_{a=1}^{N} \sum_{b=L+1}^{N} w_{ab} ||\mathbf{q}_a - \mathbf{q}_b||_2^2$$

STRUCTURED PREDICTION

ninth run for

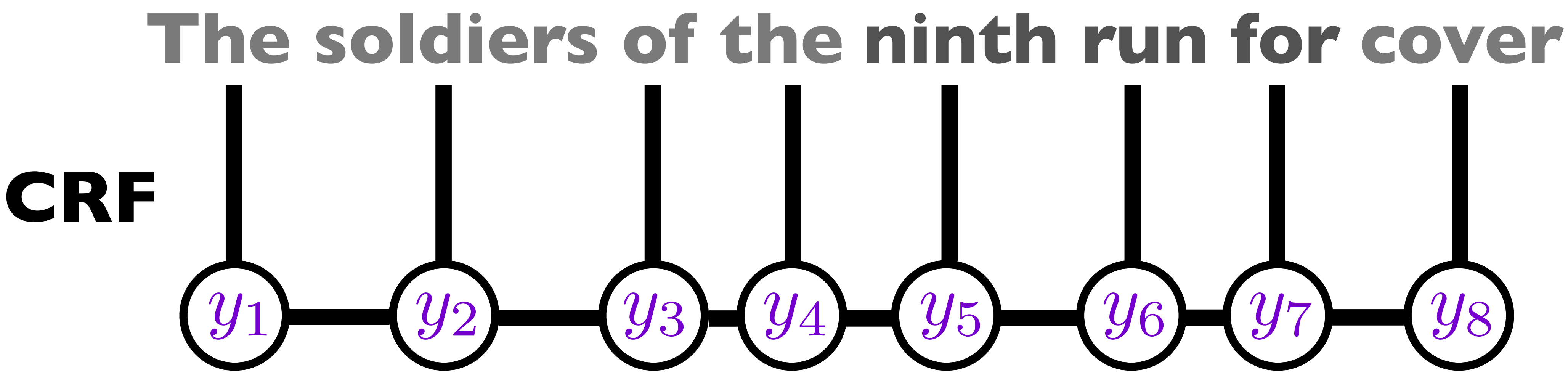
STRUCTURED PREDICTION

ninth run for

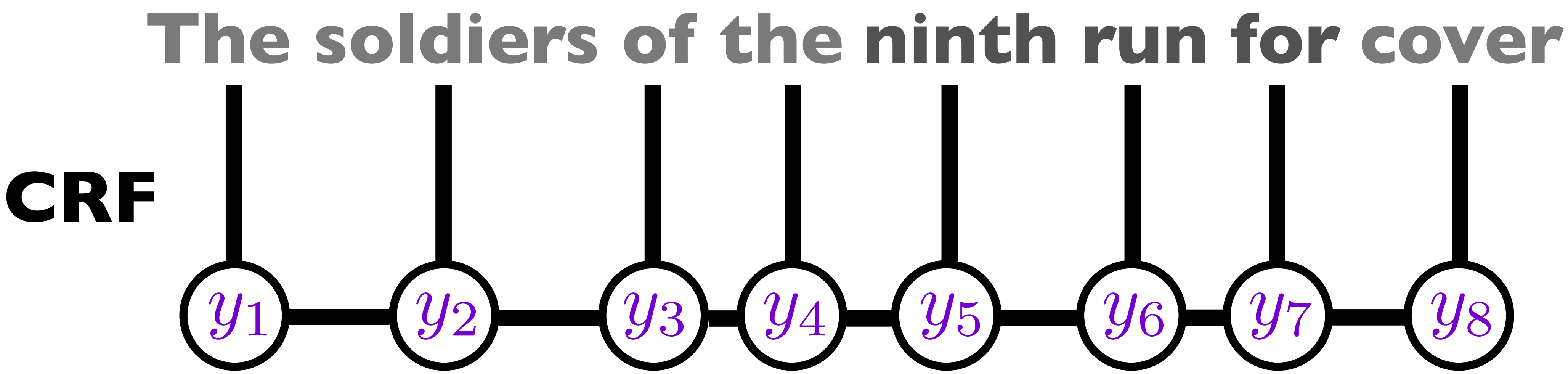
STRUCTURED PREDICTION

The soldiers of the ninth run for cover

STRUCTURED PREDICTION

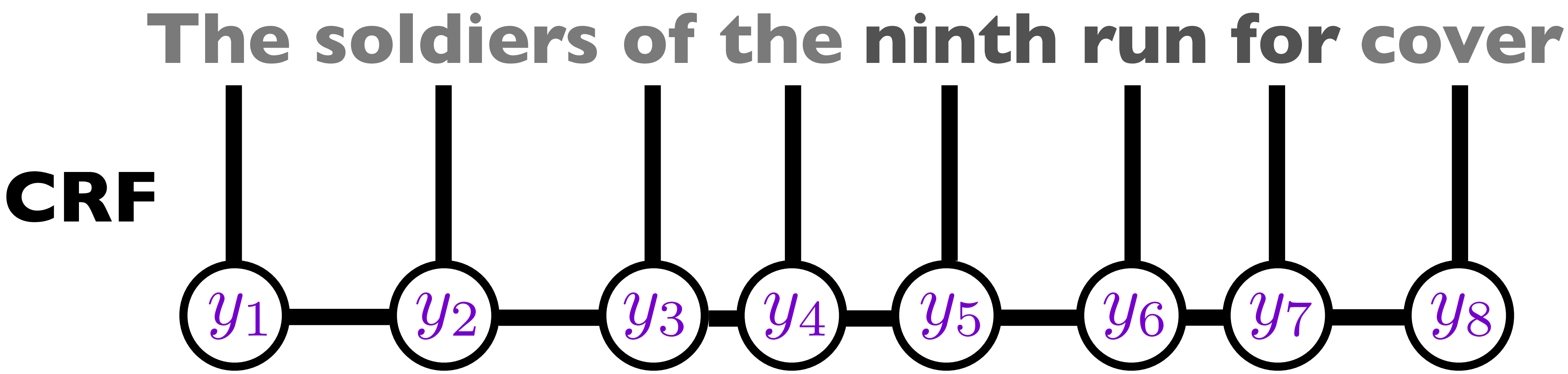


STRUCTURED PREDICTION



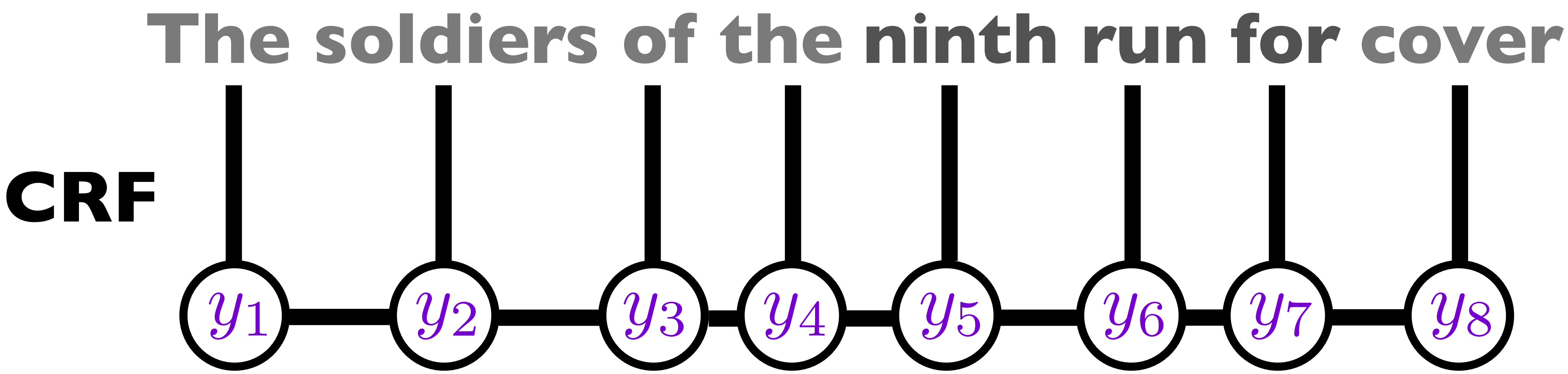
$$\mathbf{f}\left(\qquad\qquad\qquad \right)$$

STRUCTURED PREDICTION



$$\mathbf{f}(\ y_t \qquad\qquad)$$

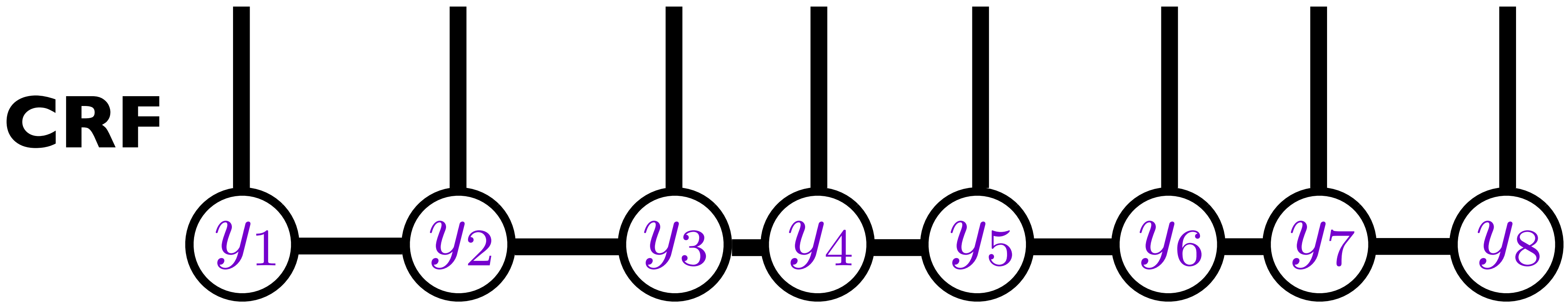
STRUCTURED PREDICTION



$$\mathbf{f}(\,y_t\,, y_{t-1}\qquad)$$

STRUCTURED PREDICTION

$\mathbf{x} =$ **The soldiers of the ninth run for cover**

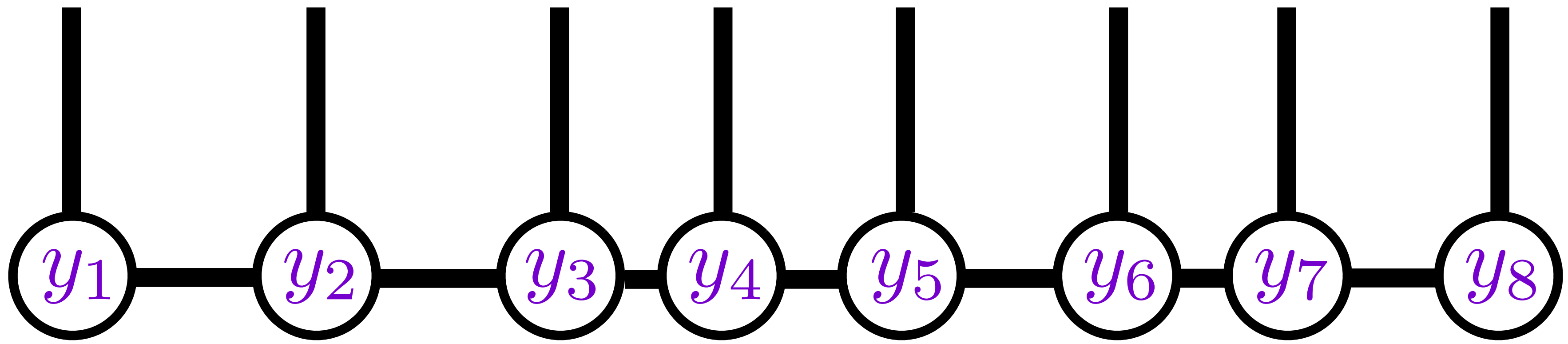


$$\mathbf{f}(\, y_t \,, y_{t-1}, \mathbf{x} \,)$$

STRUCTURED PREDICTION

$\mathbf{x} =$ **The soldiers of the ninth run for cover**

CRF

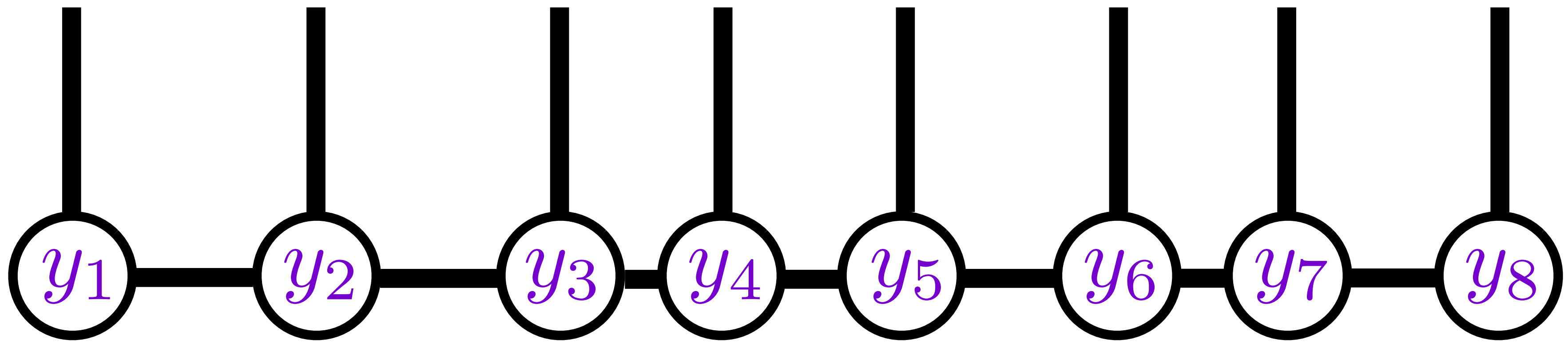


$$\mathbf{f}(\underbrace{y_t, y_{t-1}}_{p\text{-factor}}, \mathbf{x})$$

STRUCTURED PREDICTION

$\mathbf{x} =$ **The soldiers of the ninth run for cover**

CRF

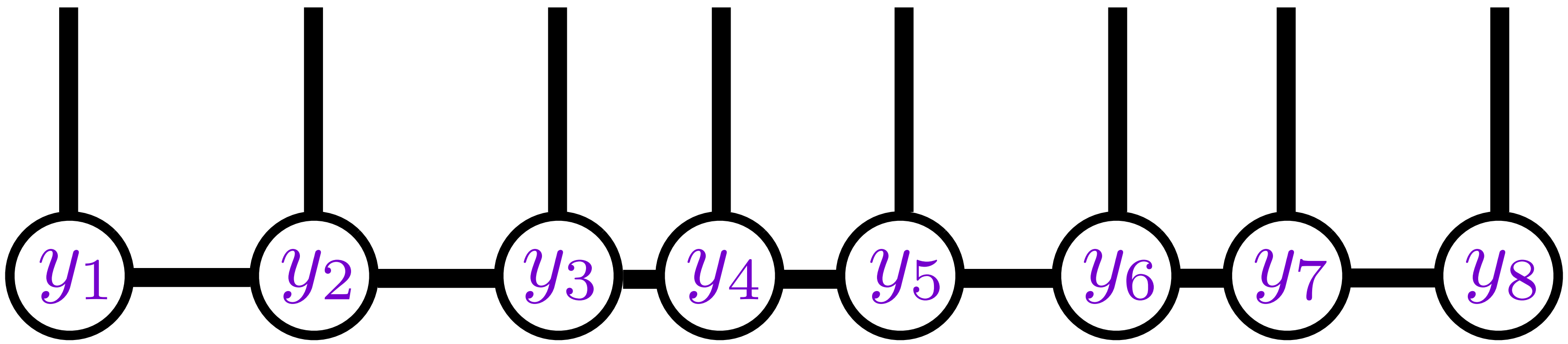


$$p_\theta(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z_\theta(\mathbf{x})} \exp\left[\sum_{t=1}^{T} \theta^\top \mathbf{f}(\underbrace{y_t, y_{t-1}, \mathbf{x}}_{p\text{-factor}})\right]$$

STRUCTURED PREDICTION

$\mathbf{x} =$ **The soldiers of the ninth run for cover**

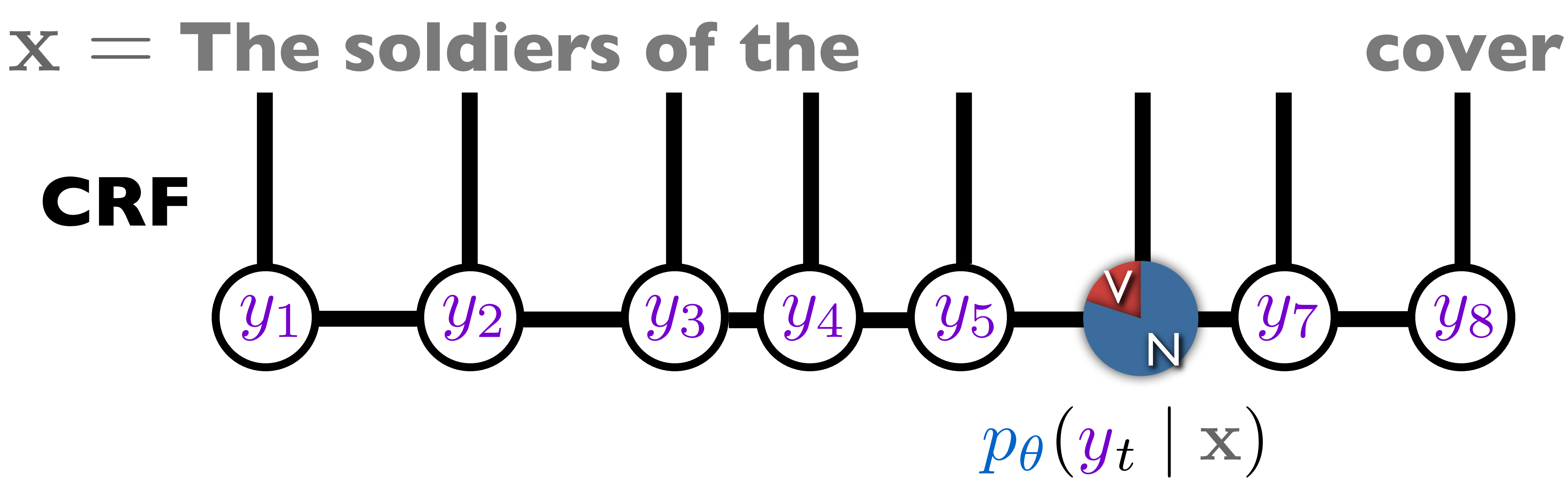
CRF



$$p_\theta(\mathbf{y} \mid \mathbf{x}) = \tfrac{1}{Z_\theta(\mathbf{x})} \exp\left[\sum_{t=1}^{T} \theta^\top \mathbf{f}(\underbrace{y_t, y_{t-1}, \mathbf{x}}_{p\text{-factor}})\right]$$

$$\mathrm{NLik}(p_\theta) = -\sum_{i=1}^{\ell} \log p_\theta(\mathbf{y}^i \mid \mathbf{x}^i)$$

STRUCTURED PREDICTION

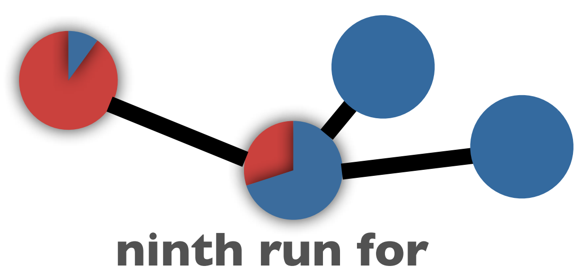
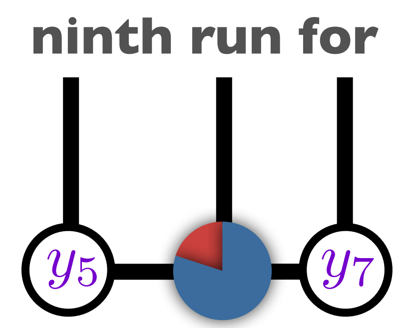


WHY COMBINE?

Each type of learning incorporates different information

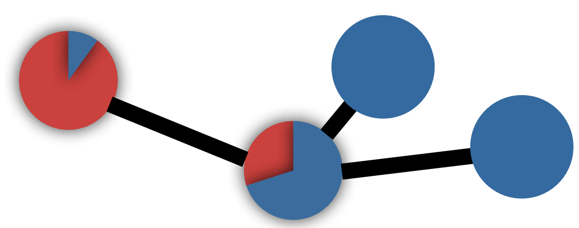
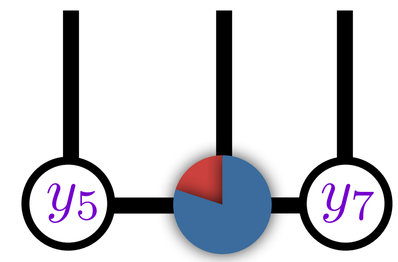
WHY COMBINE?

Each type of learning incorporates different information

	 graph-propagation	 CRF estimation

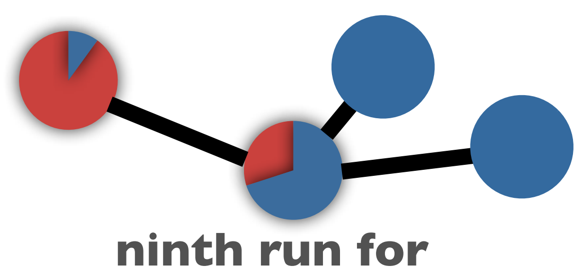
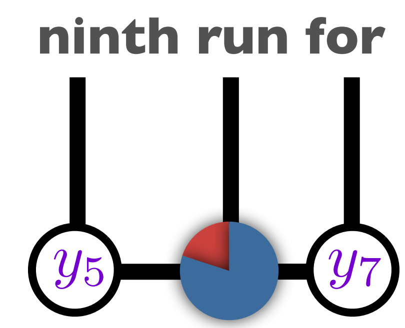
WHY COMBINE?

Each type of learning incorporates different information

	 graph-propagation	 CRF estimation
Data	**unlabeled**	**labeled**

WHY COMBINE?

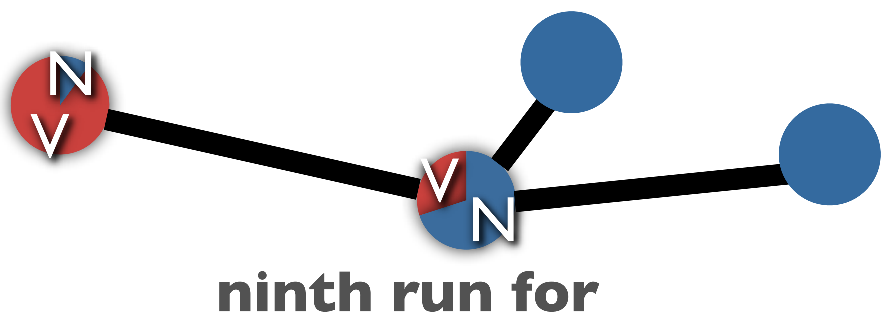
Each type of learning incorporates different information

	 graph-propagation	 CRF estimation
Data	**unlabeled**	**labeled**
Context	**trigram**	**sentence**

PRIOR WORK

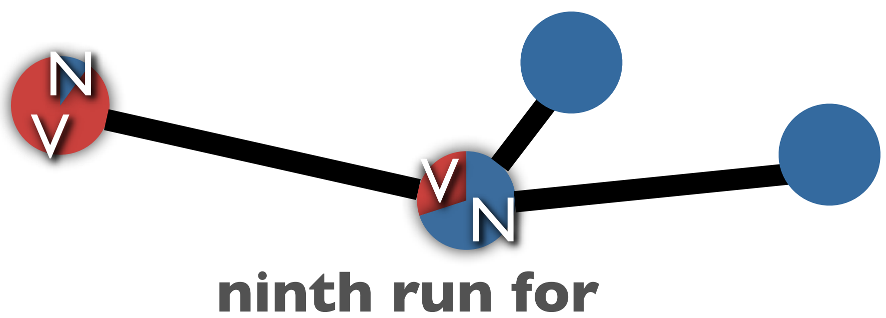
PRIOR WORK

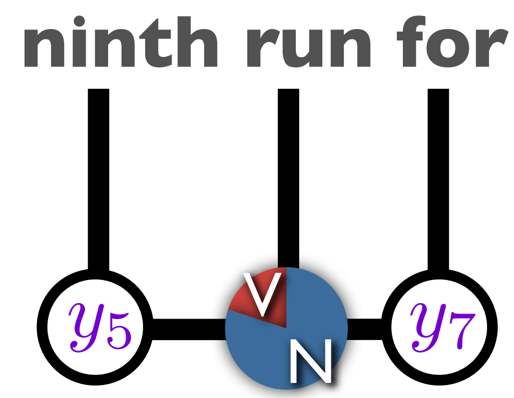
$\mathrm{Lap}(q)$ graph-propagation



PRIOR WORK

$\mathrm{Lap}(q)$ graph-propagation $+$ CRF estimation $\mathrm{NLik}(p_\theta)$

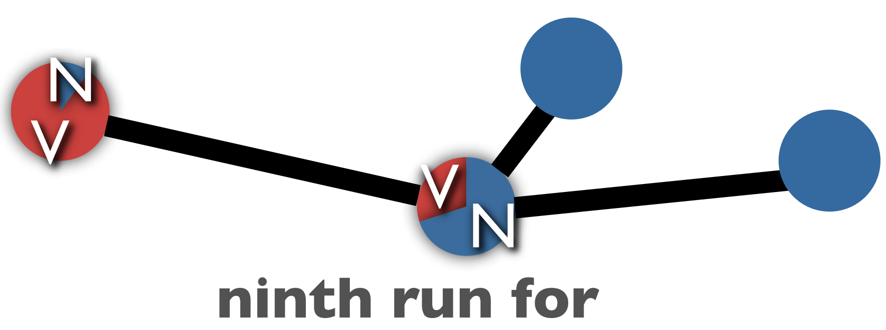


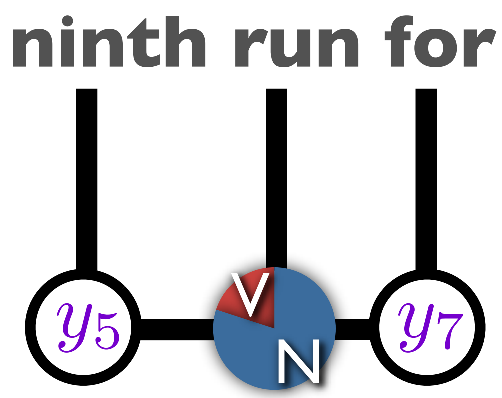


PRIOR WORK

Subramanya et al. (EMNLP 2010)

$\mathrm{Lap}(q)$ graph-propagation + CRF estimation $\mathrm{NLik}(p_\theta)$

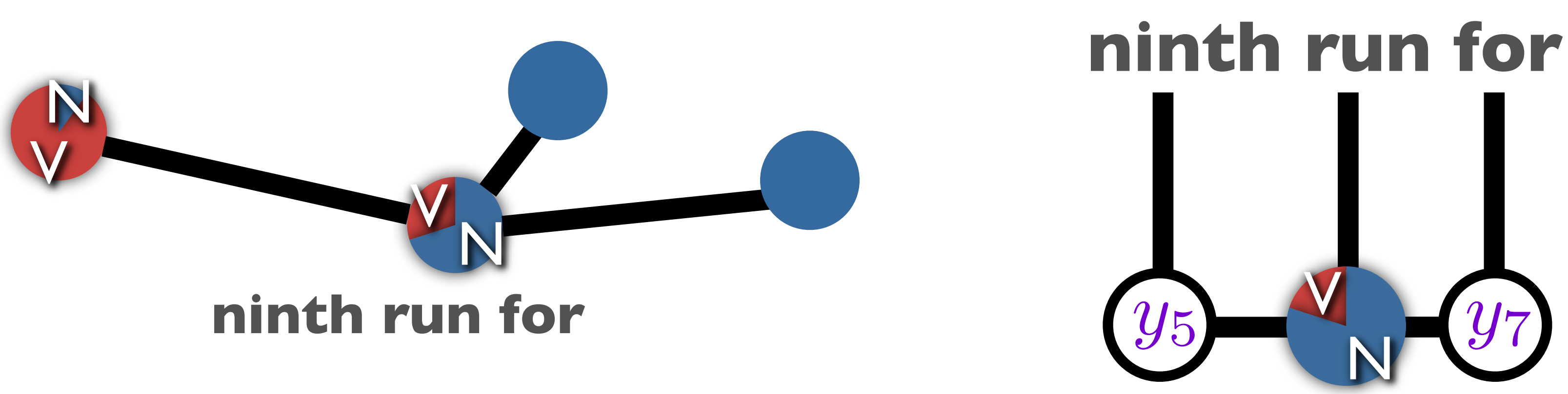




PRIOR WORK

Subramanya et al. (EMNLP 2010)

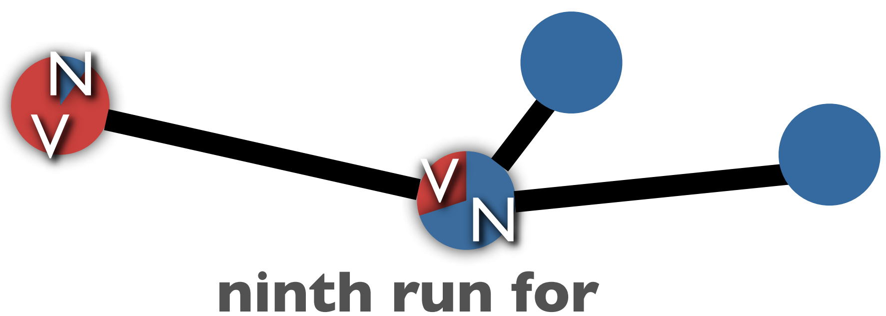
$\mathrm{Lap}(q)$ graph-propagation + CRF estimation $\mathrm{NLik}(p_\theta)$

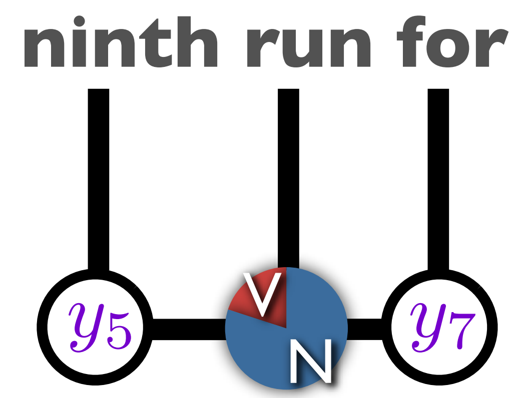


PRIOR WORK

Subramanya et al. (EMNLP 2010)

$\mathrm{Lap}(q)$ graph-propagation $+$ CRF estimation $\mathrm{NLik}(p_\theta)$

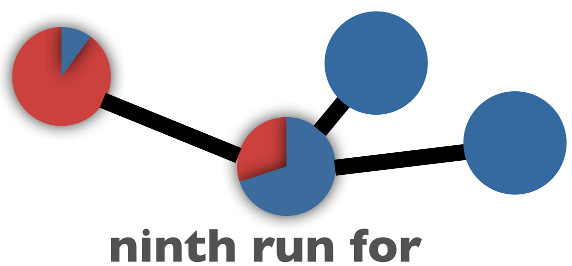




This work: retains <u>efficiency</u> while optimizing an <u>extendible, joint</u> objective.

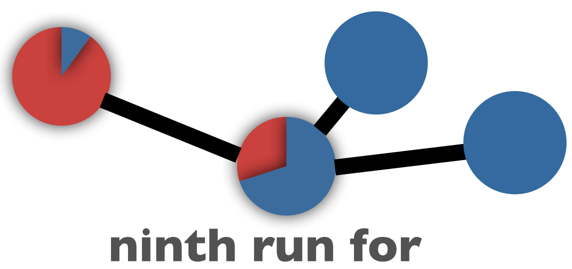
JOINT OBJECTIVE

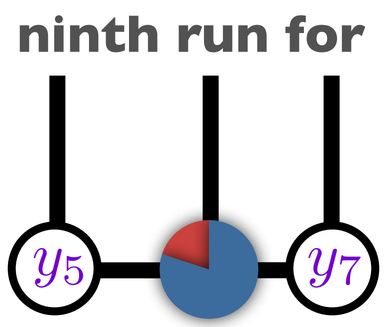
JOINT OBJECTIVE



$$\mathrm{Lap}(q)$$

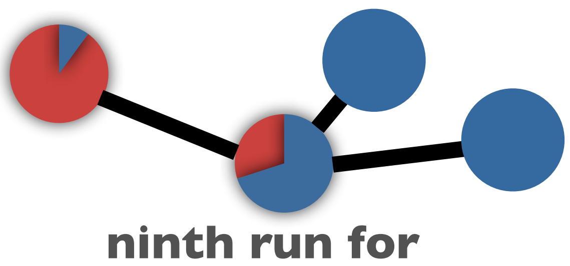
JOINT OBJECTIVE

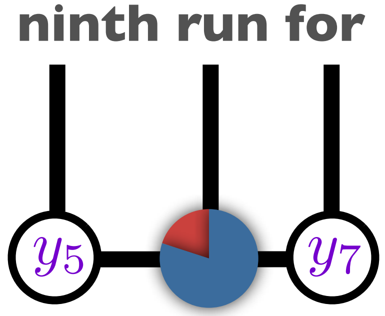




$$\mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta)$$

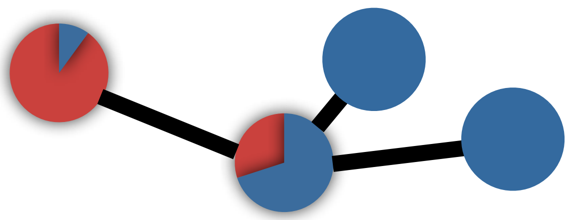
JOINT OBJECTIVE

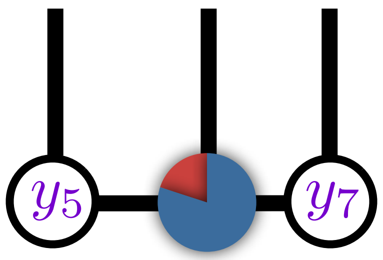




$$\mathcal{J}(q, p_\theta) = \mathrm{Lap}(q) + \mathrm{NLik}(p_\theta)$$

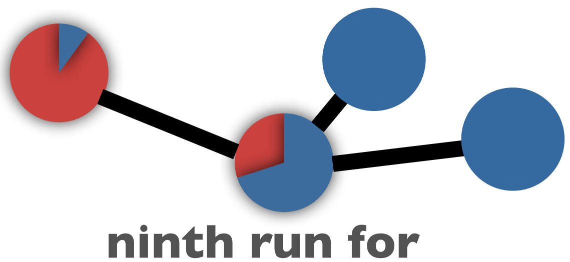
JOINT OBJECTIVE

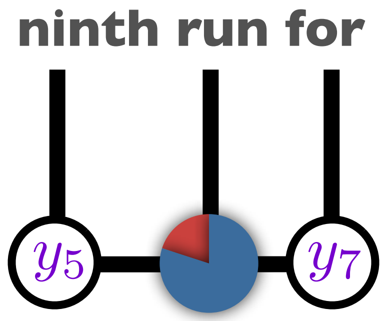




$$\mathcal{J}(q, p_\theta) \;=\; \mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta) \;+\; \mathrm{KL}(q \,\|\, p_\theta)$$

JOINT OBJECTIVE

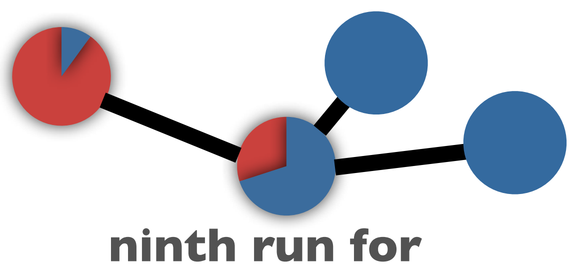


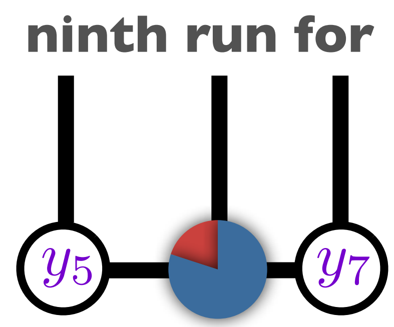


$$\mathcal{J}(q, p_\theta) \;=\; \mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta) \;+\; \mathrm{KL}(q \,\|\, p_\theta)$$

The soldiers of the ninth run for cover

JOINT OBJECTIVE

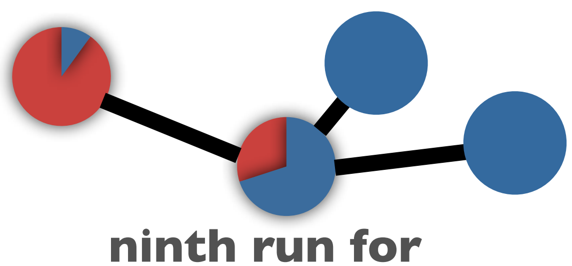


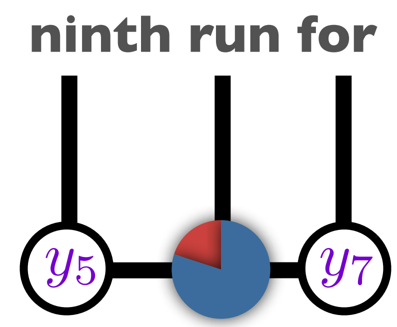


$$\mathcal{J}(q, p_\theta) \; = \; \mathrm{Lap}(q) \; + \; \mathrm{NLik}(p_\theta) \; + \; \mathrm{KL}(q \parallel p_\theta)$$

	The	soldiers	of	the	ninth	run	for	cover
$(\#\ \text{tags})^8$	N	N	N	N	N	N	N	N
	N	N	N	N	N	N	N	V
					…			

JOINT OBJECTIVE





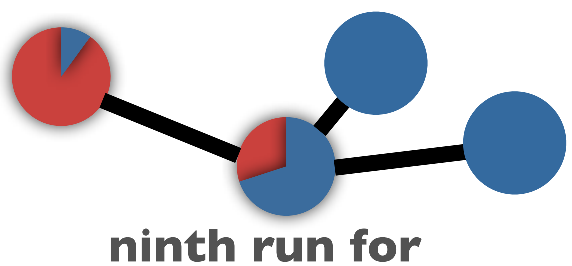
$$\mathcal{J}(q, p_\theta) \;=\; \mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta) \;+\; \mathrm{KL}(q \,\|\, p_\theta)$$

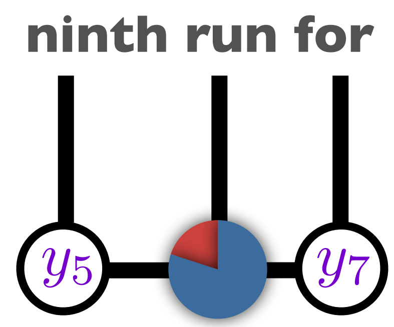
q | p_θ

The	soldiers	of	the	ninth	run	for	cover
N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	V

. . .

JOINT OBJECTIVE





$$\mathcal{J}(q, p_\theta) = \mathrm{Lap}(q) + \mathrm{NLik}(p_\theta) + \mathrm{KL}(q \,\|\, p_\theta)$$

q	p_θ	The	soldiers	of	the	ninth	run	for	cover
7e-5	2e-5	N	N	N	N	N	N	N	N
3e-6	8e-6	N	N	N	N	N	N	N	V
...	...				...				

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

unconstrained

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

Δ

unconstrained

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

Δ unconstrained

p update:

$$\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

Δ unconstrained

p update:

$$\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$$

Next 3 slides: Why several common techniques don't work for updating q

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

q update:

$$q_{\mathbf{y}}^{i}{}' = q_{\mathbf{y}}^{i} - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}$$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

q update:

$$q_{\mathbf{y}}^{i}{}' = \mathrm{proj}_\Delta\Big(q_{\mathbf{y}}^{i} - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}} \Big)$$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

q update:

$$q_{\mathbf{y}}^{i\,\prime} = \mathrm{proj}_\Delta\Big(q_{\mathbf{y}}^i - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^i} \Big)$$

$$q^i \in \Delta \text{ of dimension } (\#\text{ tags})^{(i\text{'s length})}$$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

q update:

$$q_{\mathbf{y}}^{i}{}' = \mathrm{proj}_\Delta\Big(q_{\mathbf{y}}^{i} - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}} \Big)$$

$$q^i \in \Delta \text{ of dimension } (\#\text{ tags})^{(i\text{'s length})}$$

-Problem 1: projection is hard $q^i \notin \Delta$

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

q update:

$${q_{\mathbf{y}}^{i}}' = \text{proj}_\Delta\left(q_{\mathbf{y}}^{i} - \eta \frac{\partial \mathcal{J}(q,p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right)$$

$$q^i \in \Delta \text{ of dimension } (\#\text{ tags})^{(i\text{'s length})}$$

-Problem 1: projection is hard $q^i \notin \Delta$

-Problem 2: no compact form $(\#\text{ tags})^{(i\text{'s length})}$ values

OPTIMIZATION

$$\min_{q,\theta} \mathcal{J}(q, p_\theta)$$

q update:

$$\cancel{{q^i_{\mathbf{y}}}' = \text{proj}_\Delta\left(q^i_{\mathbf{y}} - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^i_{\mathbf{y}}}\right)}$$

$$q^i \in \Delta \text{ of dimension } (\#\text{ tags})^{(i\text{'s length})}$$

-Problem 1: projection is hard $\quad q^i \notin \Delta$

-Problem 2: no compact form $\quad (\#\text{ tags})^{(i\text{'s length})}$ values

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta)$$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma\left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1\right)$$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathbf{Lap}(q)$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathrm{Lap}(q)$ $\longrightarrow$ Standard PR: $\mathrm{simpler}$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathrm{Lap}(q)$ $\longrightarrow$ Standard PR: $\mathbf{simpler}$

$\uparrow$

p-factors

$y_t, y_{t-1}, \mathbf{x}$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathrm{\mathbf{Lap}}(q)$ $\longrightarrow$ Standard PR: $\mathrm{\mathbf{Linear}}(m)$

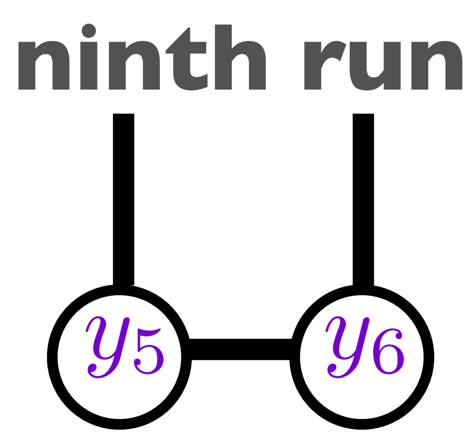
DUAL OPTIMIZATION

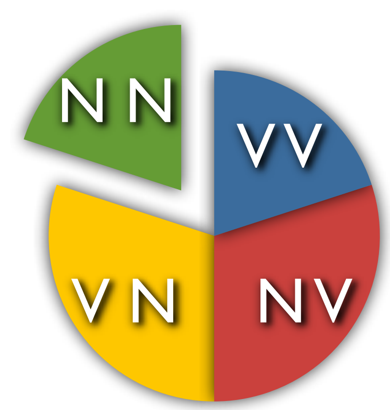
$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\textbf{Lap}(q)$ $\longrightarrow$ Standard PR: $\textbf{Linear}(m)$

m-entry:





DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathbf{Lap}(q)$ $\longrightarrow$ Standard PR: $\mathbf{Linear}(m)$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathrm{Lap}(q) \longrightarrow$ Standard PR: $\mathbf{Linear}(m)$

$\mathrm{Lap}(m)$

DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathbf{Lap}(q)$ $\longrightarrow$ Standard PR: $\mathbf{Linear}(m)$

$\mathbf{Lap}(m)$, a quadratic function

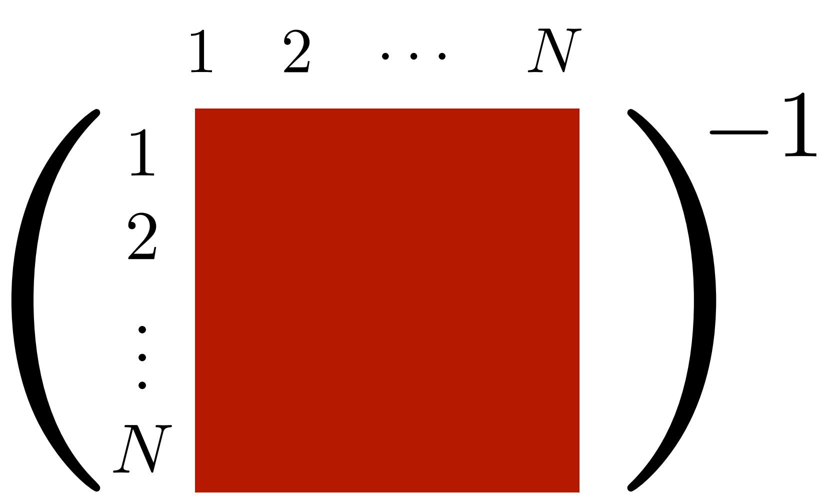
DUAL OPTIMIZATION

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathbf{Lap}(q)$ $\longrightarrow$ Standard PR: $\mathbf{Linear}(m)$

$\mathbf{Lap}(m)$, a quadratic function
Dual of quadratic requires:



DUAL ~~OPTIMIZATION~~

$$\mathcal{J}(q, p_\theta) + \gamma \left(\sum_{\mathbf{y}} q^i_{\mathbf{y}} - 1 \right)$$

Posterior Regularization (PR) uses dual
Ganchev et al. (JMLR 2010)

This work: $\mathbf{Lap}(q)$ $\longrightarrow$ Standard PR: $\mathbf{Linear}(m)$

$\mathbf{Lap}(m)$, a quadratic function
Dual of quadratic requires:

$$\begin{array}{cc} & \begin{array}{cccc} 1 & 2 & \cdots & N \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ N \end{array} & \left(\begin{array}{cccc} & & & \\ & & & \\ & & & \\ & & & \end{array} \right)^{-1} \end{array}$$

EXPONENTIATED GRADIENT

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

Collins et al. (JMLR 2008): Exponentiated gradient for CRFs

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

$$\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m_{\mathbf{y}}^{i})}{\partial m_{t,y_t,y_{t-1}}^{i}}\right]$$

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

$$\underbrace{\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m_{\mathbf{y}}^{i})}{\partial m_{t,y_t,y_{t-1}}^{i}}\right]}_{\text{product of p-factors}}$$

EXPONENTIATED GRADIENT

$$q^{i}_{\mathbf{y}}{}' \propto q^{i}_{\mathbf{y}} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^{i}_{\mathbf{y}}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^{i}_{\mathbf{y}}}\right] =$$

$$\underbrace{\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m^{i}_{\mathbf{y}})}{\partial m^{i}_{t,y_t,y_{t-1}}}\right]}_{\text{product of p-factors}} p_\theta(\mathbf{y} \mid \mathbf{x}^{i})^{\eta} (q^{i}_{\mathbf{y}})^{-\eta} e$$

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

$$\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m_{\mathbf{y}}^{i})}{\partial m_{t, y_t, y_{t-1}}^{i}}\right] p_\theta(\mathbf{y} \mid \mathbf{x}^{i})^{\eta} (q_{\mathbf{y}}^{i})^{-\eta} e$$

product of p-factors

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i\,\prime} \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q,p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

$$\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m_{\mathbf{y}}^{i})}{\partial m^{i}_{t,y_t,y_{t-1}}}\right] p_\theta(\mathbf{y} \mid \mathbf{x}^i)^\eta (q_{\mathbf{y}}^{i})^{-\eta} e$$

product of p-factors

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i}{}' \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

$$\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m_{\mathbf{y}}^{i})}{\partial m_{t,y_t,y_{t-1}}^{i}}\right] p_\theta(\mathbf{y} \mid \mathbf{x}^i)^\eta (q_{\mathbf{y}}^{i})^{-\eta} e$$

product of p-factors

$$\mathrm{proj}_\Delta$$

EXPONENTIATED GRADIENT

$$q^{i}_{\mathbf{y}}{}' \propto q^{i}_{\mathbf{y}} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^{i}_{\mathbf{y}}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^{i}_{\mathbf{y}}}\right] =$$

$$\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m^{i}_{\mathbf{y}})}{\partial m^{i}_{t,y_t,y_{t-1}}}\right] p_\theta(\mathbf{y} \mid \mathbf{x}^i)^{\eta} (q^{i}_{\mathbf{y}})^{-\eta} e$$

product of p-factors

$$\mathrm{proj}_\Delta \longrightarrow Z_q(\mathbf{x}^i)$$

EXPONENTIATED GRADIENT

$$q_{\mathbf{y}}^{i\,\prime} \propto q_{\mathbf{y}}^{i} \exp\left[-\eta \frac{\partial \mathcal{J}(q,p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right]$$

$$\exp\left[-\eta \frac{\partial \mathcal{J}(q,p_\theta)}{\partial q_{\mathbf{y}}^{i}}\right] =$$

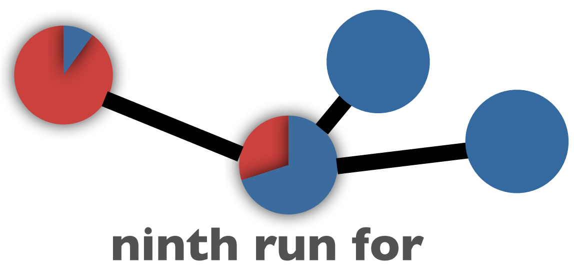
$$\exp\left[-\eta \sum_{t=1}^{T} \frac{\partial \mathrm{Lap}(m_{\mathbf{y}}^{i})}{\partial m^{i}_{t,y_t,y_{t-1}}}\right] p_\theta(\mathbf{y} \mid \mathbf{x}^{i})^{\eta} (q_{\mathbf{y}}^{i})^{-\eta} e$$

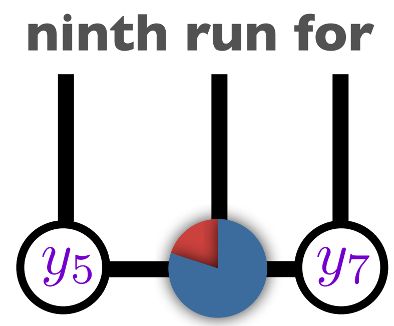
product of p-factors

$\mathrm{proj}_\Delta \longrightarrow Z_q(\mathbf{x}^i)$, computable via forward-backward

SUMMARY

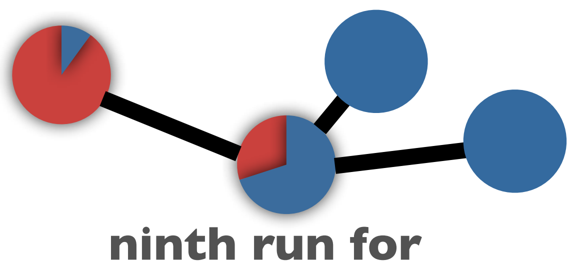
SUMMARY

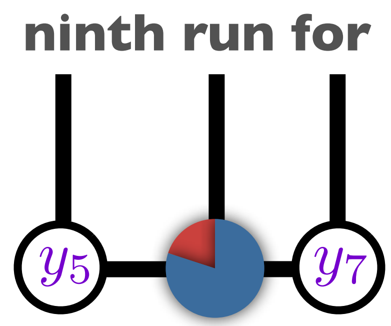




$$\mathcal{J}(q, p_\theta) \;=\; \mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta) \;+\; \mathrm{KL}(q \,\|\, p_\theta)$$

SUMMARY

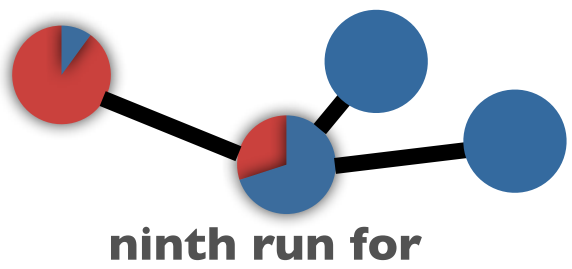


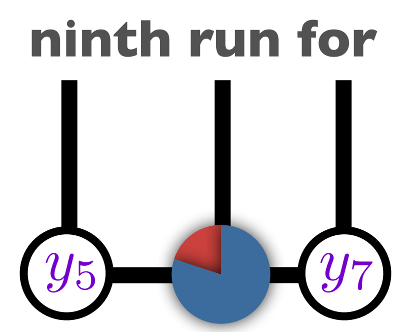


$$\mathcal{J}(q, p_\theta) = \mathrm{Lap}(q) + \mathrm{NLik}(p_\theta) + \mathrm{KL}(q \,\|\, p_\theta)$$

$$\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$$

SUMMARY



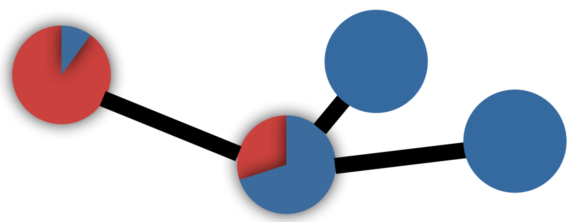


$$\mathcal{J}(q, p_\theta) = \mathrm{Lap}(q) + \mathrm{NLik}(p_\theta) + \mathrm{KL}(q \,\|\, p_\theta)$$

$$\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$$

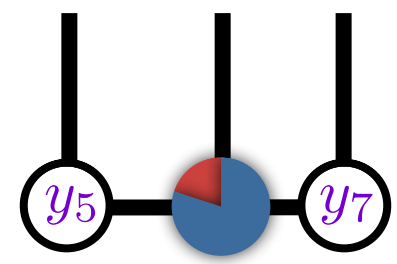
$$q^{i}_{\mathbf{y}}{}' = \frac{1}{Z_q(\mathbf{x}^i)} q^{i}_{\mathbf{y}} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^{i}_{\mathbf{y}}}\right]$$

SUMMARY



ninth run for

ninth run for

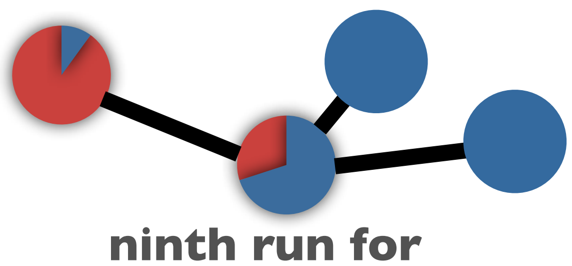


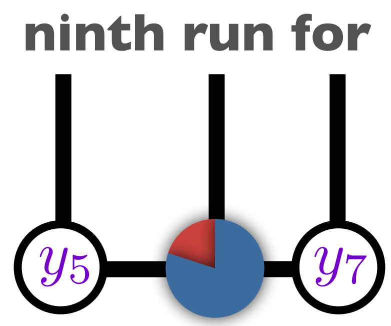
$$\mathcal{J}(q, p_\theta) \;=\; \mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta) \;+\; \mathrm{KL}(q \,\|\, p_\theta)$$

M-step: $\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$

E-step: $q_{\mathbf{y}}^{i\,\prime} = \frac{1}{Z_q(\mathbf{x}^i)} q_{\mathbf{y}}^i \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q_{\mathbf{y}}^i}\right]$

SUMMARY





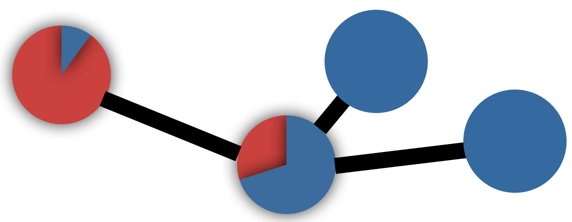
$$\mathcal{J}(q, p_\theta) = \mathrm{Lap}(q) + \mathrm{NLik}(p_\theta) + \mathrm{KL}(q \,\|\, p_\theta)$$

M-step: $\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$

E-step: ${q^i_{\mathbf{y}}}' = \frac{1}{Z_q(\mathbf{x}^i)} q^i_{\mathbf{y}} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^i_{\mathbf{y}}}\right]$

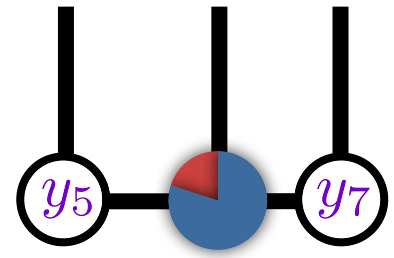
$\boxed{\begin{array}{c}\textbf{Theorem:} \\ \text{Converges to a local} \\ \text{optimum of} \\ \mathcal{J}(q, p_\theta)\end{array}}$

SUMMARY



ninth run for

ninth run for



$$\mathcal{J}(q, p_\theta) \;=\; \mathrm{Lap}(q) \;+\; \mathrm{NLik}(p_\theta) \;+\; \mathrm{KL}(q \,\|\, p_\theta)$$

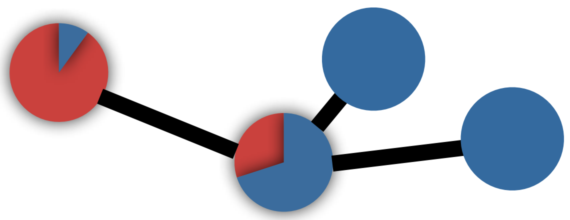
any convex, differentiable $g(m)$

M-step: $\theta' = \theta - \eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial \theta}$

E-step: $q^{i}_{\mathbf{y}}{}' = \frac{1}{Z_q(\mathbf{x}^i)} q^{i}_{\mathbf{y}} \exp\left[-\eta \frac{\partial \mathcal{J}(q, p_\theta)}{\partial q^{i}_{\mathbf{y}}}\right]$

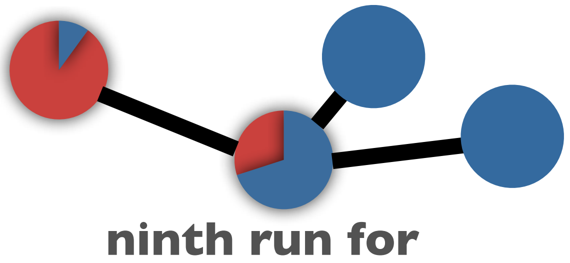
Theorem:
Converges to a local optimum of
$\mathcal{J}(q, p_\theta)$

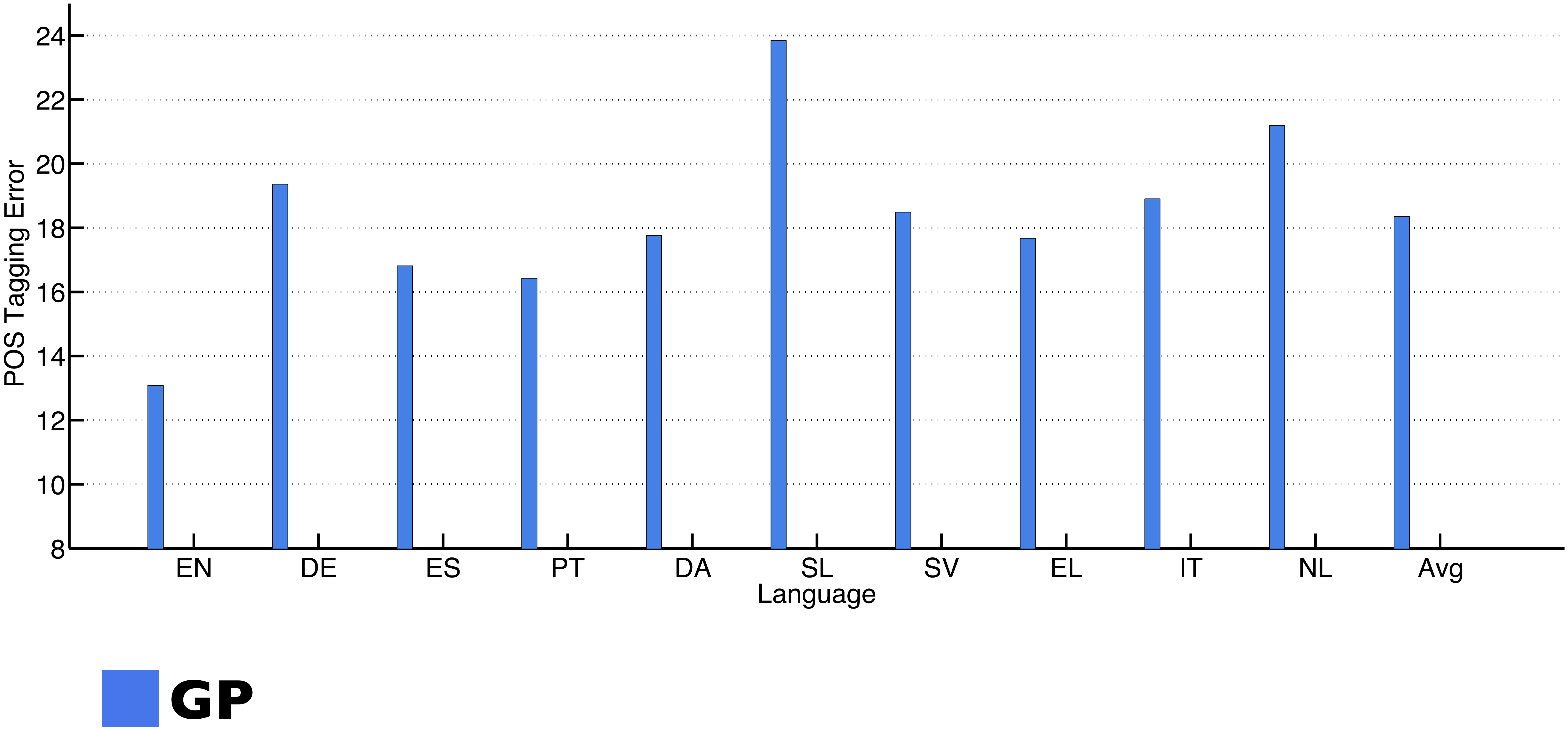
graph-propagation



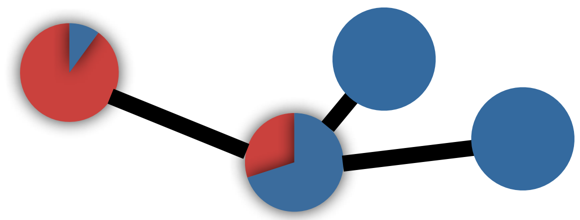
ninth run for

graph-propagation

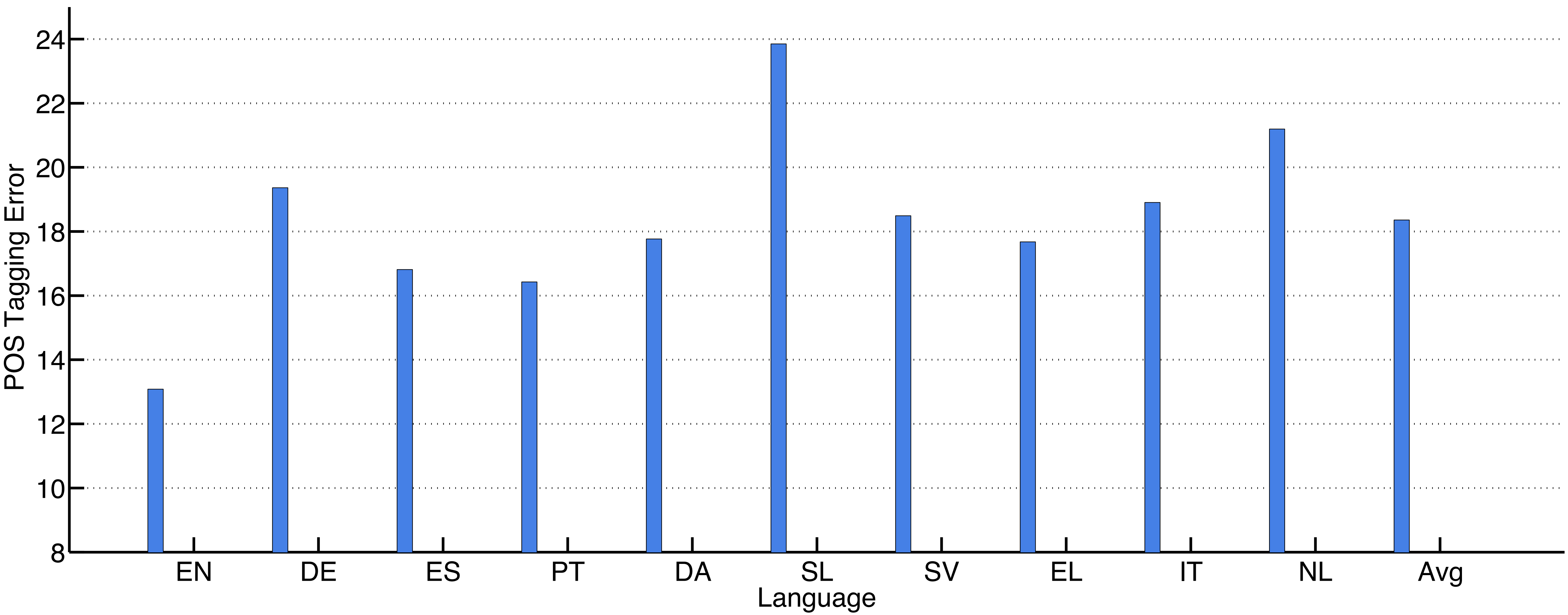




graph-propagation

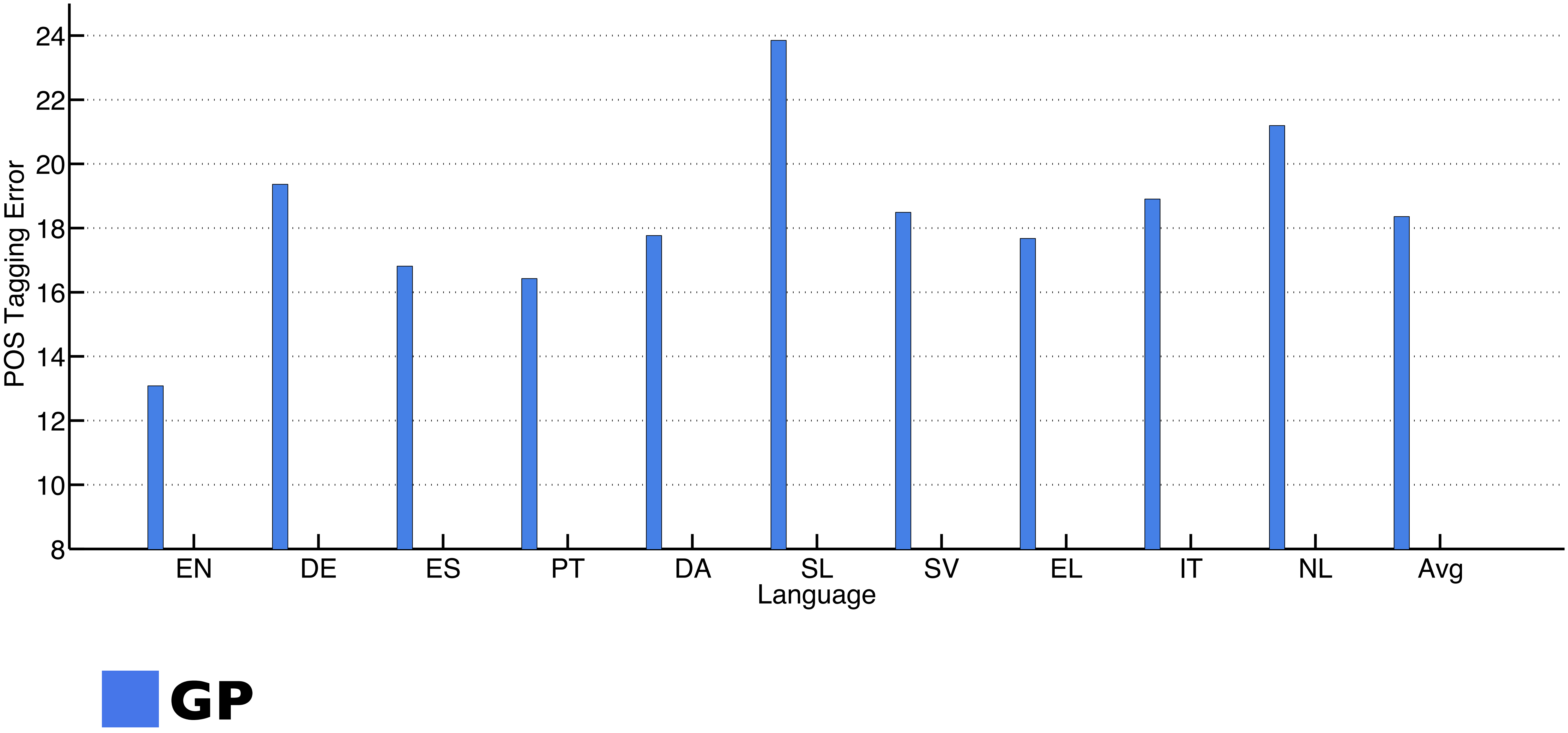


ninth run for

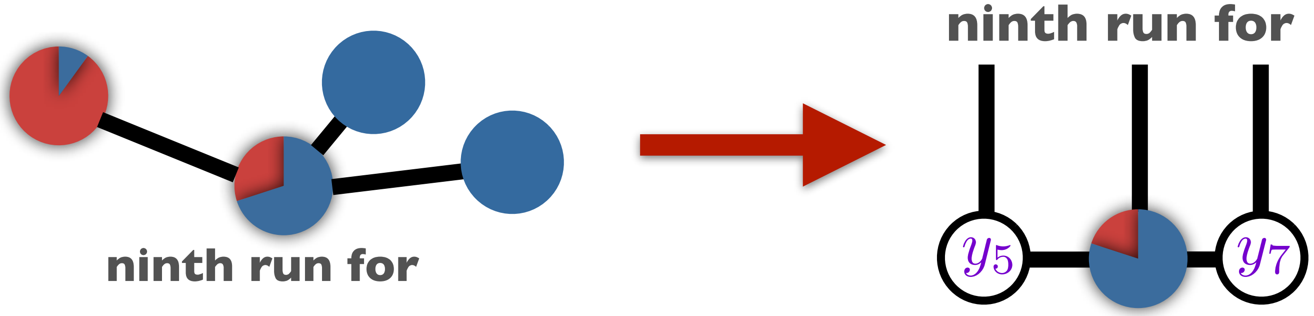


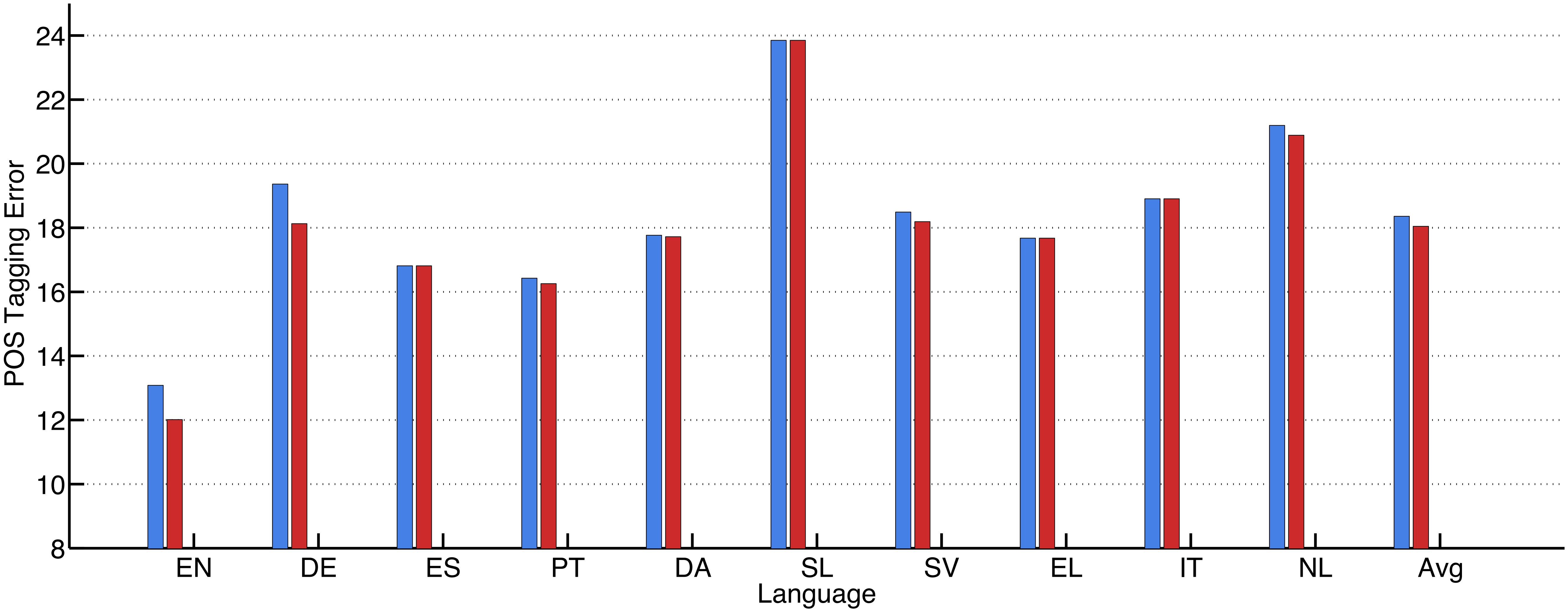
GP

100 labeled sentences


24
22
20
18
16
14
12
10
8
POS Tagging Error
EN
DE
ES
PT
DA
SL
SV
EL
IT
NL
Avg
Language
GP

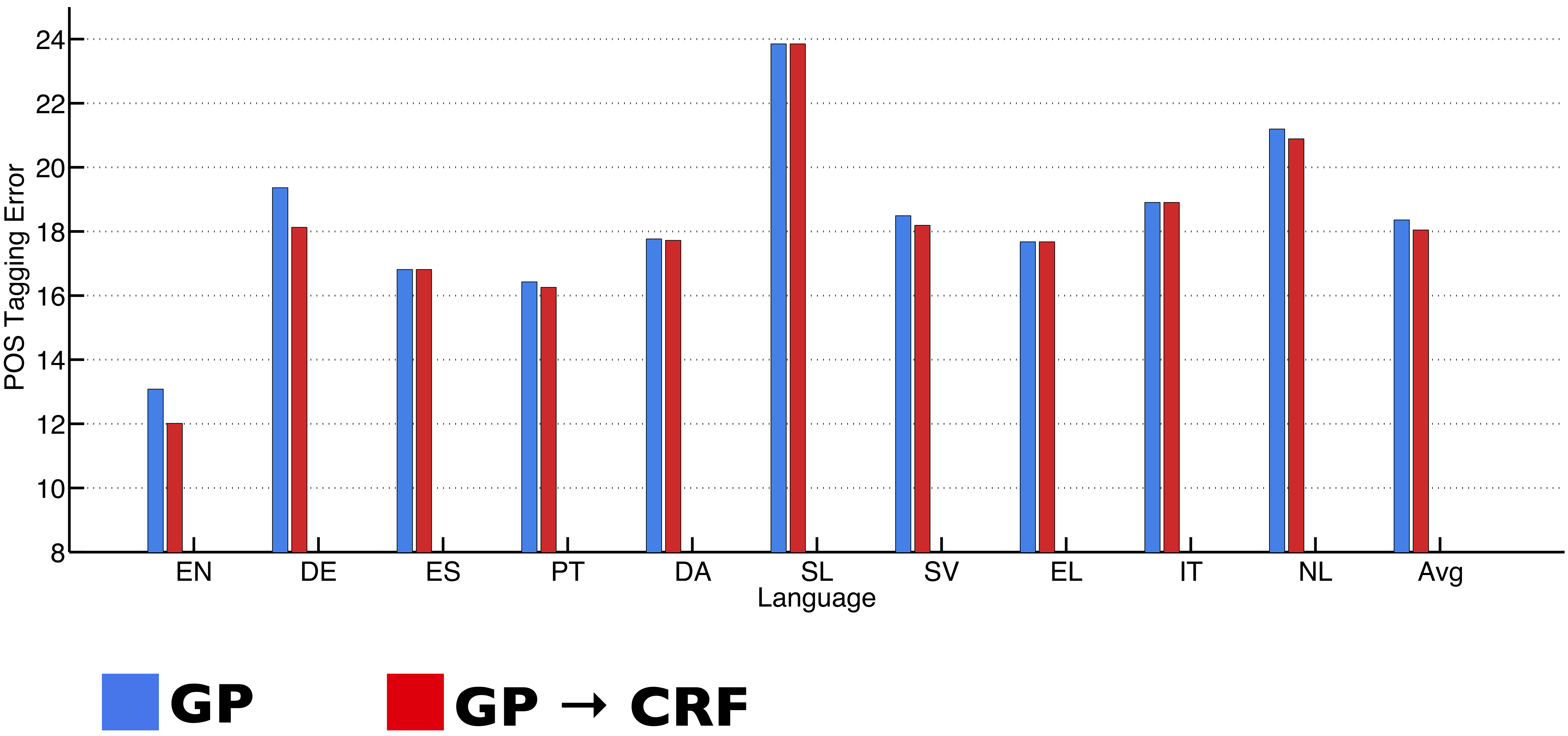
GP → CRF



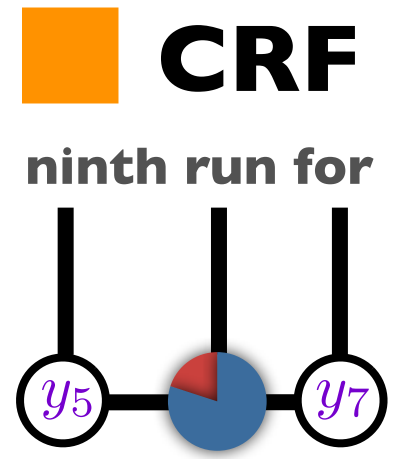


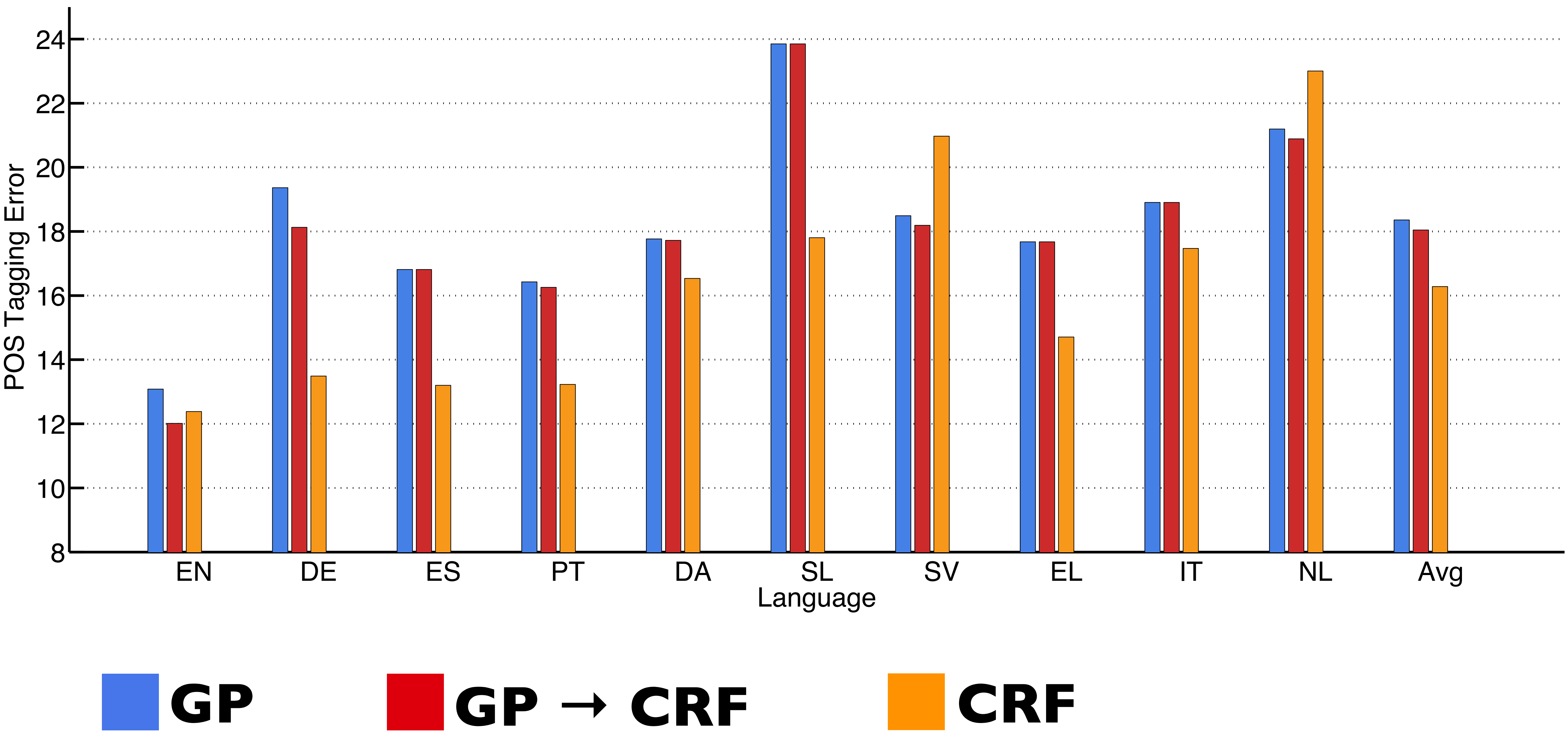
GP

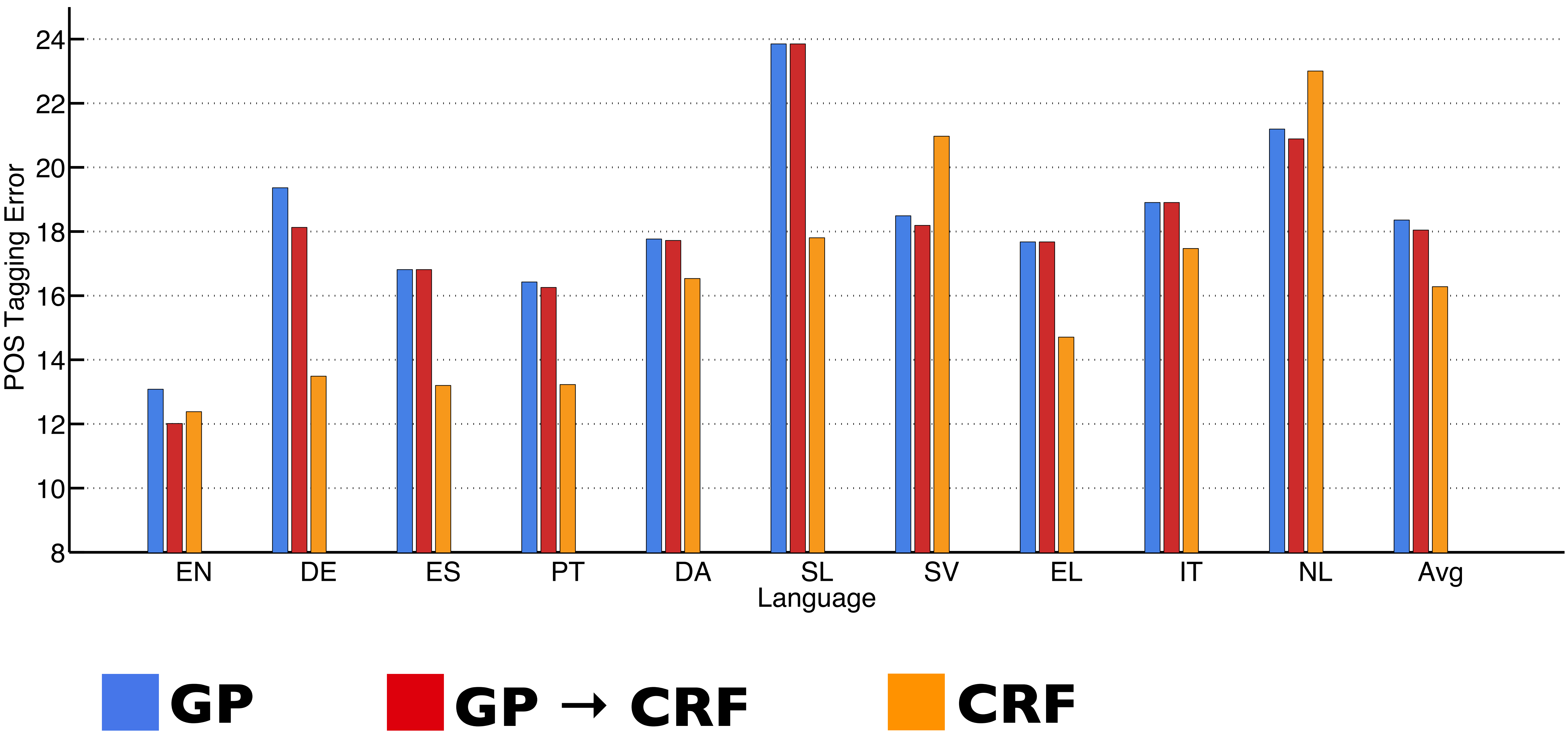
GP → CRF



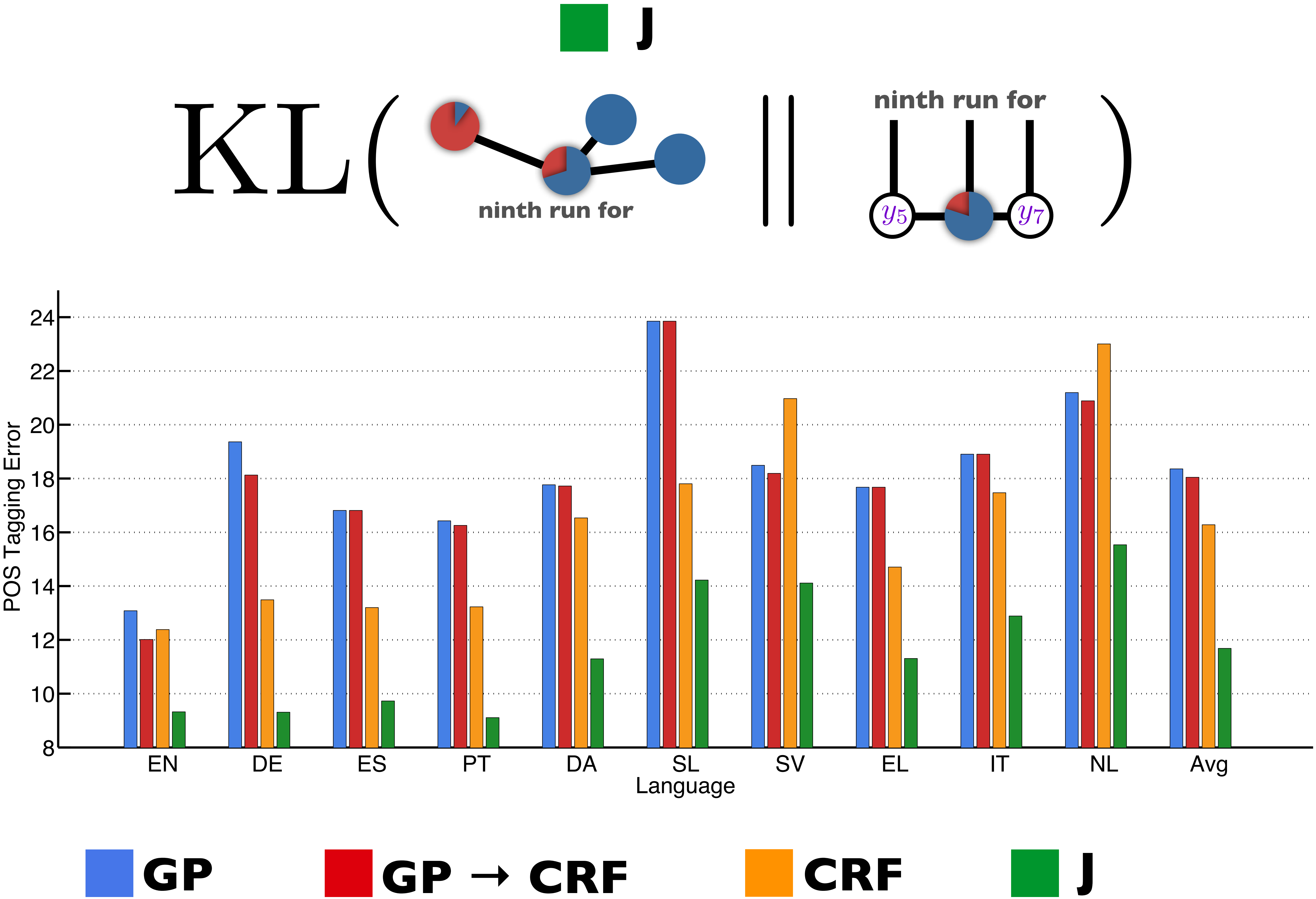
POS Tagging Error
24
22
20
18
16
14
12
10
8
EN
DE
ES
PT
DA
SL
SV
EL
IT
NL
Avg
Language
GP
GP → CRF


CRF
ninth run for
y_5
y_7

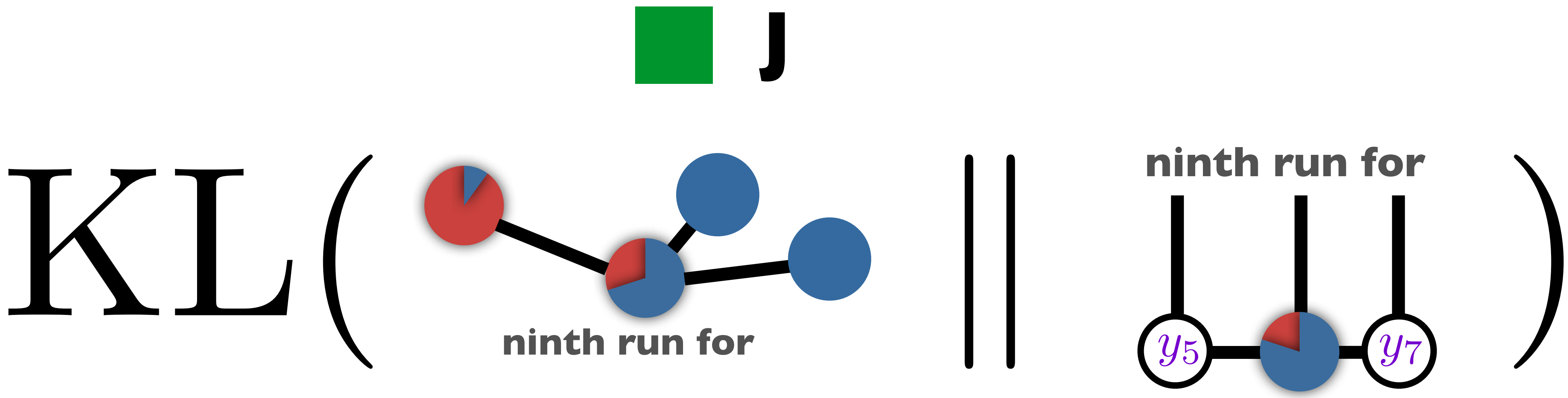

POS Tagging Error
24
22
20
18
16
14
12
10
8
EN
DE
ES
PT
DA
SL
SV
EL
IT
NL
Avg
Language
GP
GP → CRF
CRF

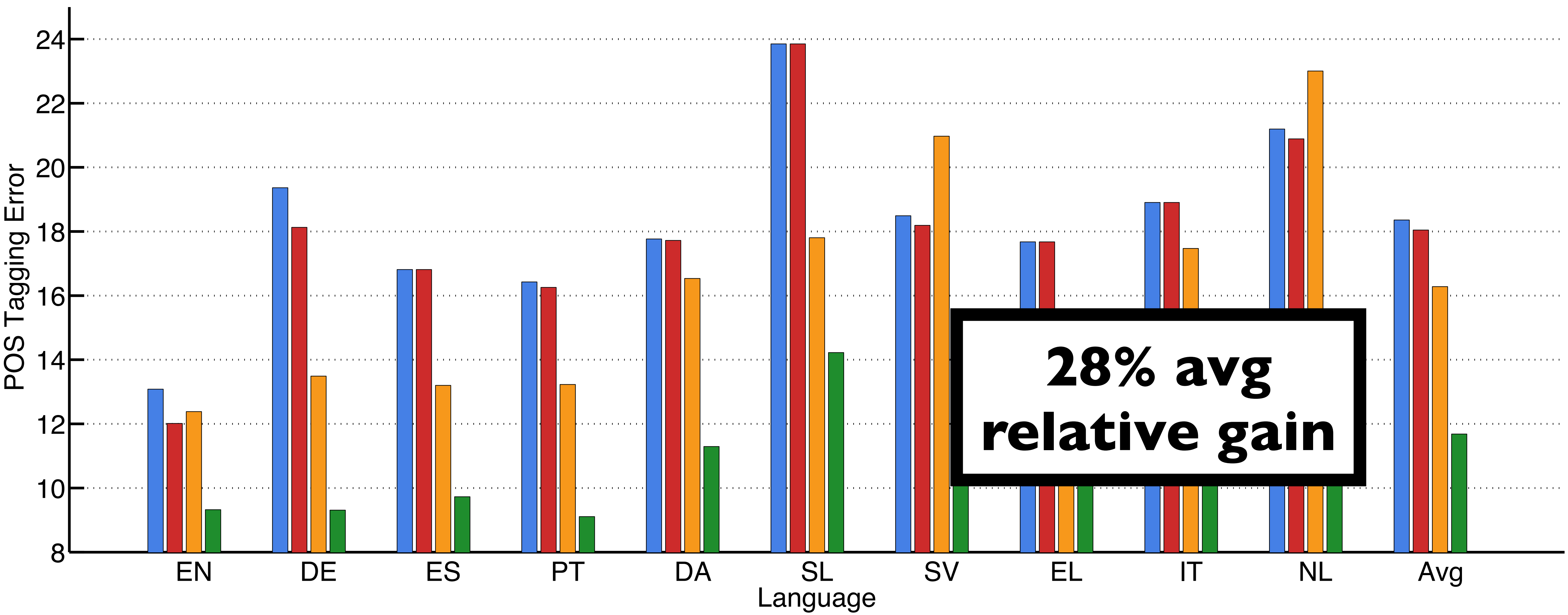


POS Tagging Error
24
22
20
18
16
14
12
10
8
EN
DE
ES
PT
DA
SL
SV
EL
IT
NL
Avg
Language
GP
GP → CRF
CRF



J
KL(
ninth run for
||
ninth run for
y5
y7
)
POS Tagging Error
24
22
20
18
16
14
12
10
8
EN
DE
ES
PT
DA
SL
SV
EL
IT
NL
Avg
Language
GP
GP → CRF
CRF
J





GP **GP → CRF** **CRF** **J**

QUESTIONS?

QUESTIONS?

Code: https://code.google.com/p/pr-graph/