Learning a Part-of-Speech Tagger from Minimal Annotation

Dan Garrette

University of Texas at Austin

Low-Resource Languages

Supervised training is not an option.

We do semi-supervised training.

→ Annotate some data by hand

... cheaply

Semi-Supervised Training

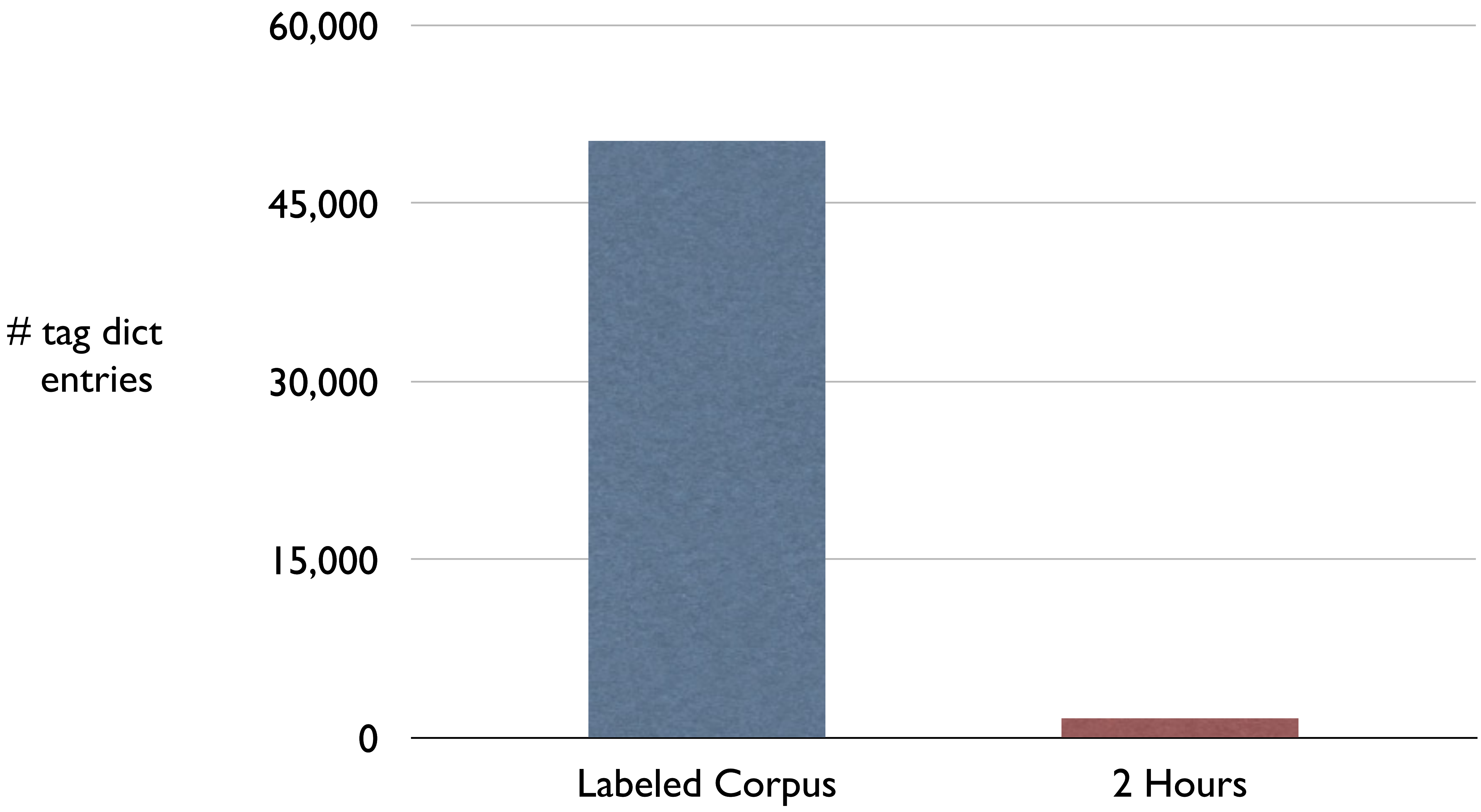
HMM with Expectation-Maximization (EM)

Need:

- Large **raw** corpus ← know how to get this
- Tag dictionary ← where is this from?

[Kupiec, 1992]
[Merialdo, 1994]

A **Real** Tag Dictionary

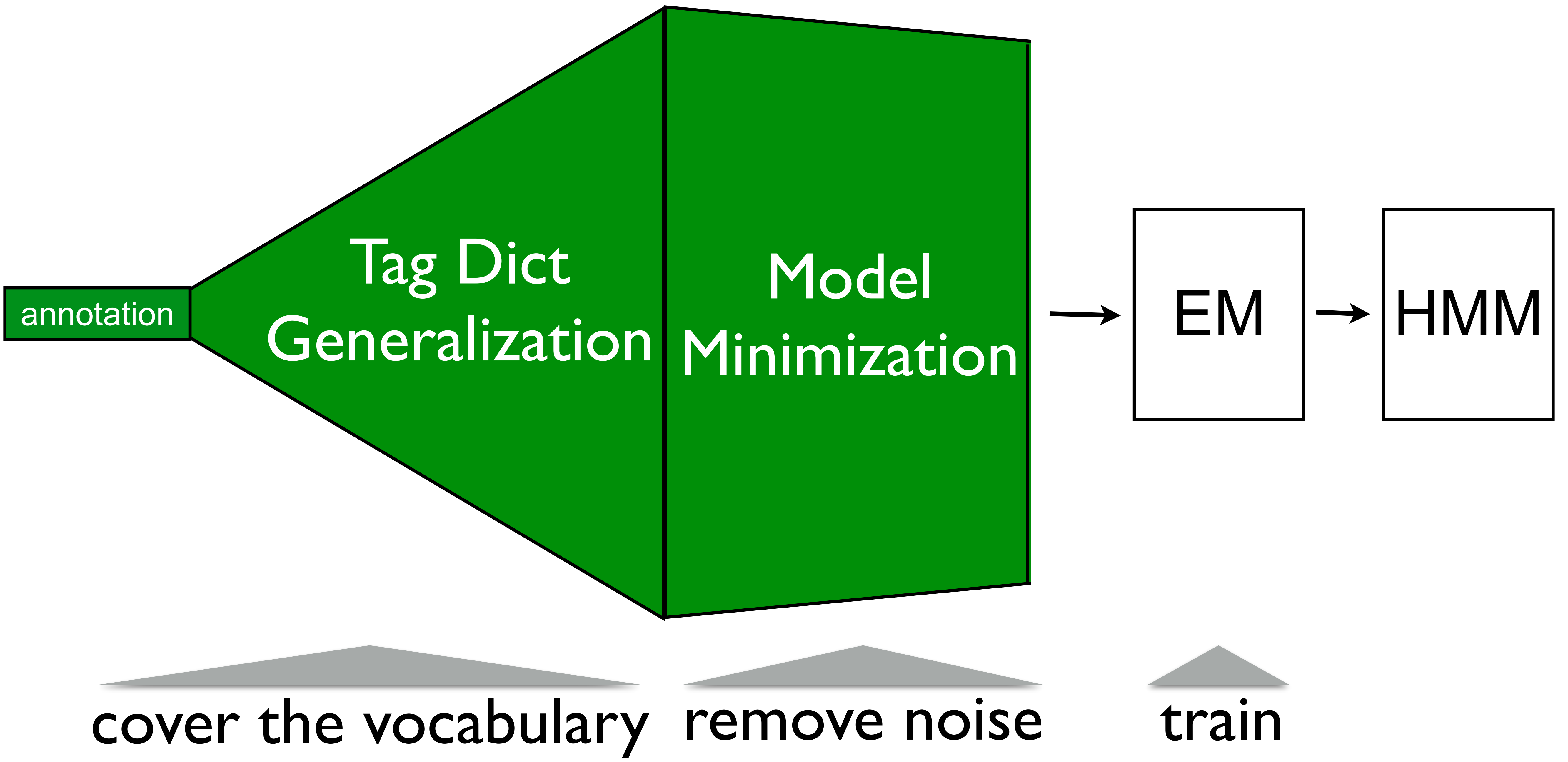


A **Real** Tag Dictionary

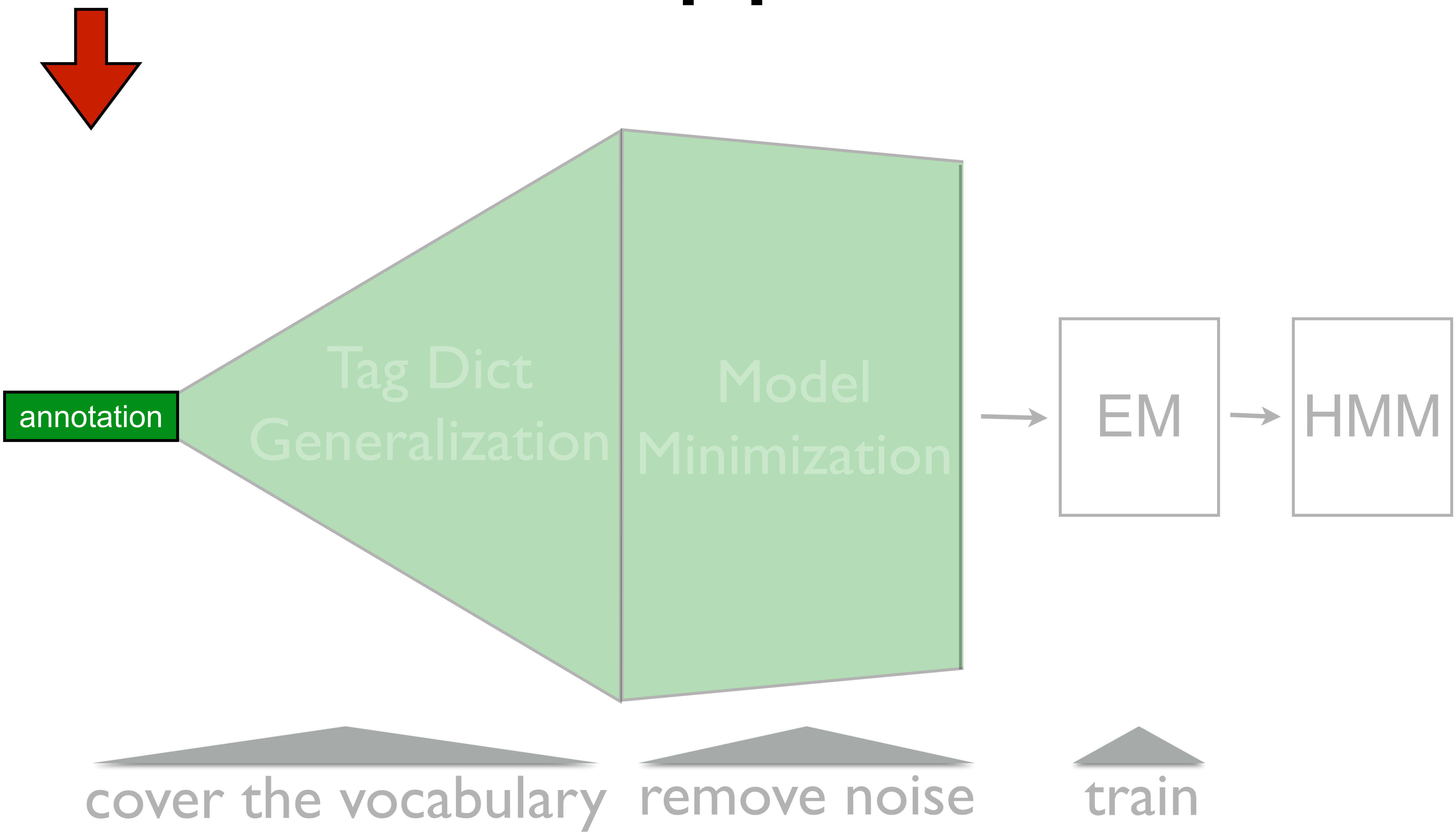
Extremely low coverage means **most** words are **unknown**

⇒ **Bad for learning** (poorly constrained)

Our Approach



Our Approach



Collecting Annotations

Task #1

Up to 4 hours to create a **tag dictionary**

ordered by frequency

,	,		
the	DT		
.	.		
of	IN	RP	
to	TO	RP	
a	DT		
and	CC		
⋮	⋮		
only	RB		
can	VB	VBP	MD
York	NNP		
into	IN	RP	
after	IN	RP	
president	NN		
⋮	⋮		

Collecting Annotations

Task #2

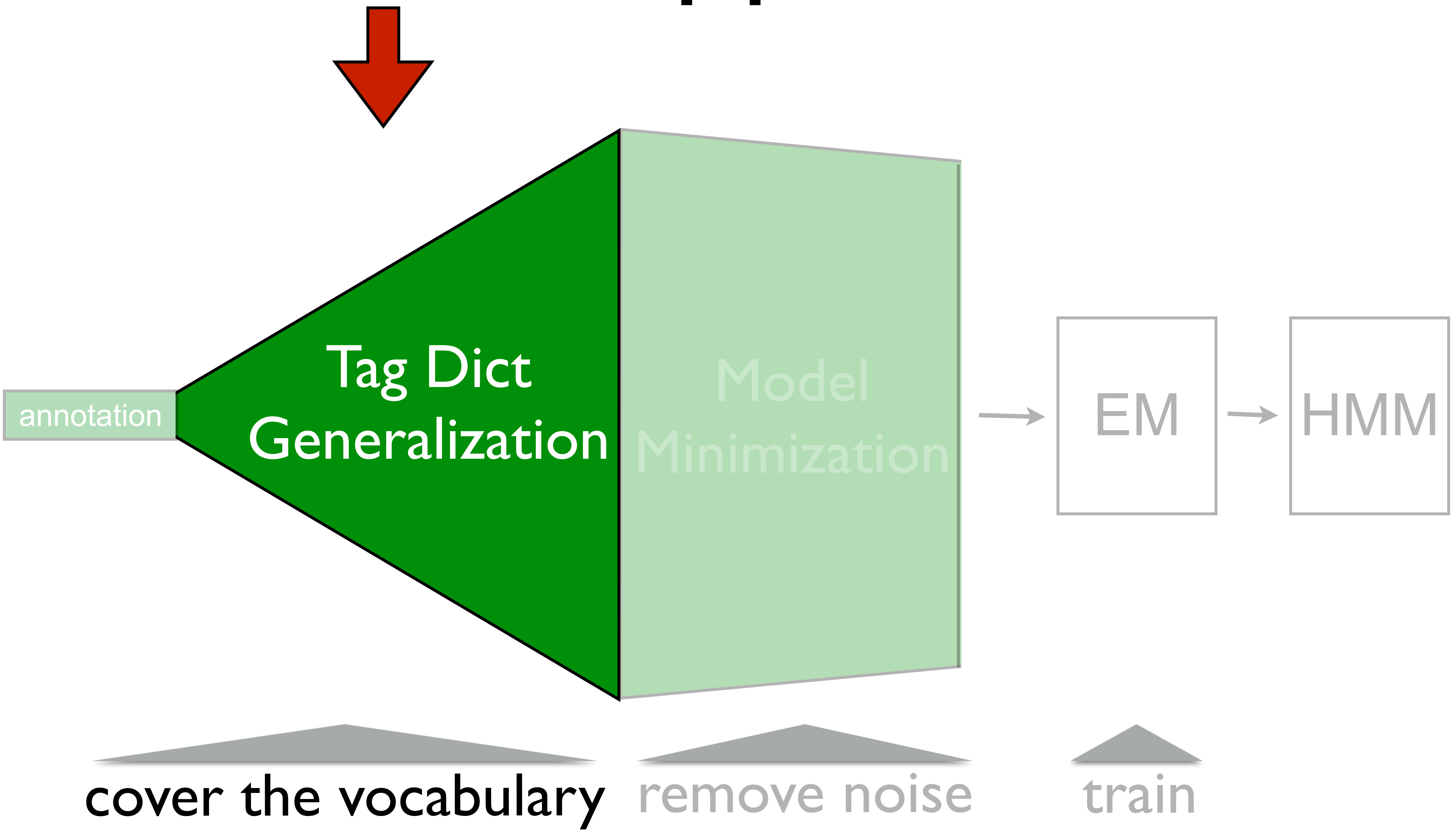
Up to 4 hours to annotate **full sentences**

Pierre	Vinken	,	61	years	old	,	will	join	the	board	as	a	nonexecutive	director	Nov.	29	.
NNP	NNP	,	CD	NNS	JJ	,	MD	VB	DT	NN	IN	DT	JJ	NN	NNP	CD	.

Mr.	Vinken	is	chairman	of	Elsevier	N.V.	,	the	Dutch	publishing	group	.
NNP	NNP	VB	NN	IN	NNP	NNP	,	DT	JJ	JJ	NN	.

⋮

Our Approach



Tag Dict Generalization

These annotations are too sparse!

→ Generalize to the entire vocabulary

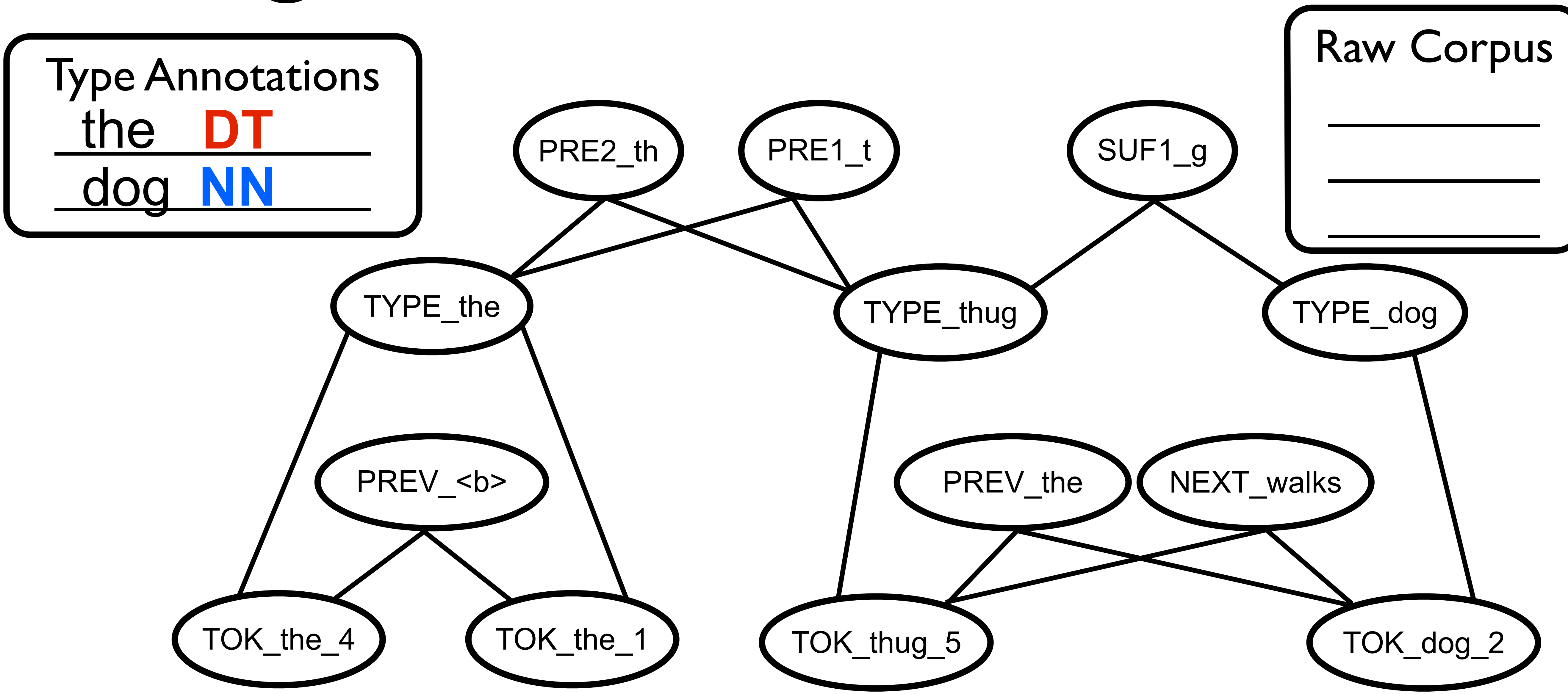
Tag Dict Generalization

Our strategy: Label Propagation

- **Connect** annotations to raw corpus tokens
- Push tag labels to **entire corpus**

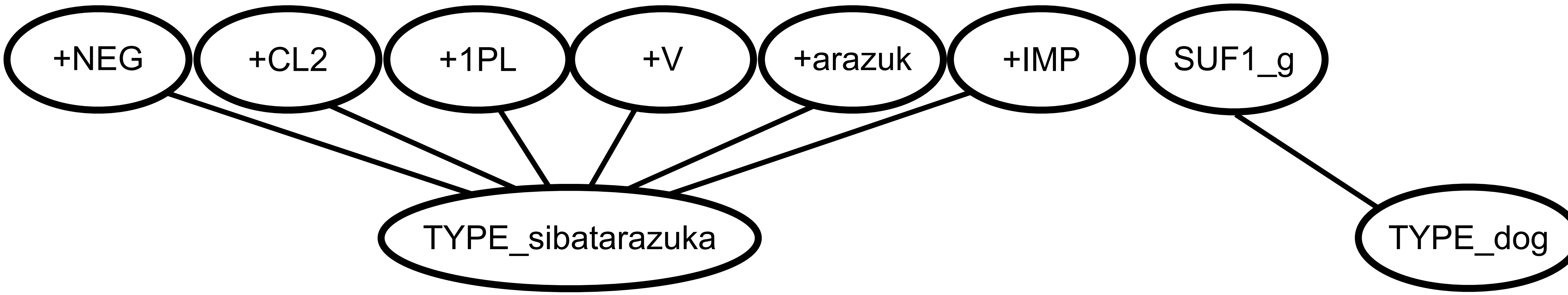
[Talukdar and Crammer. 2009]

Tag Dict Generalization



Any arbitrary features could be used

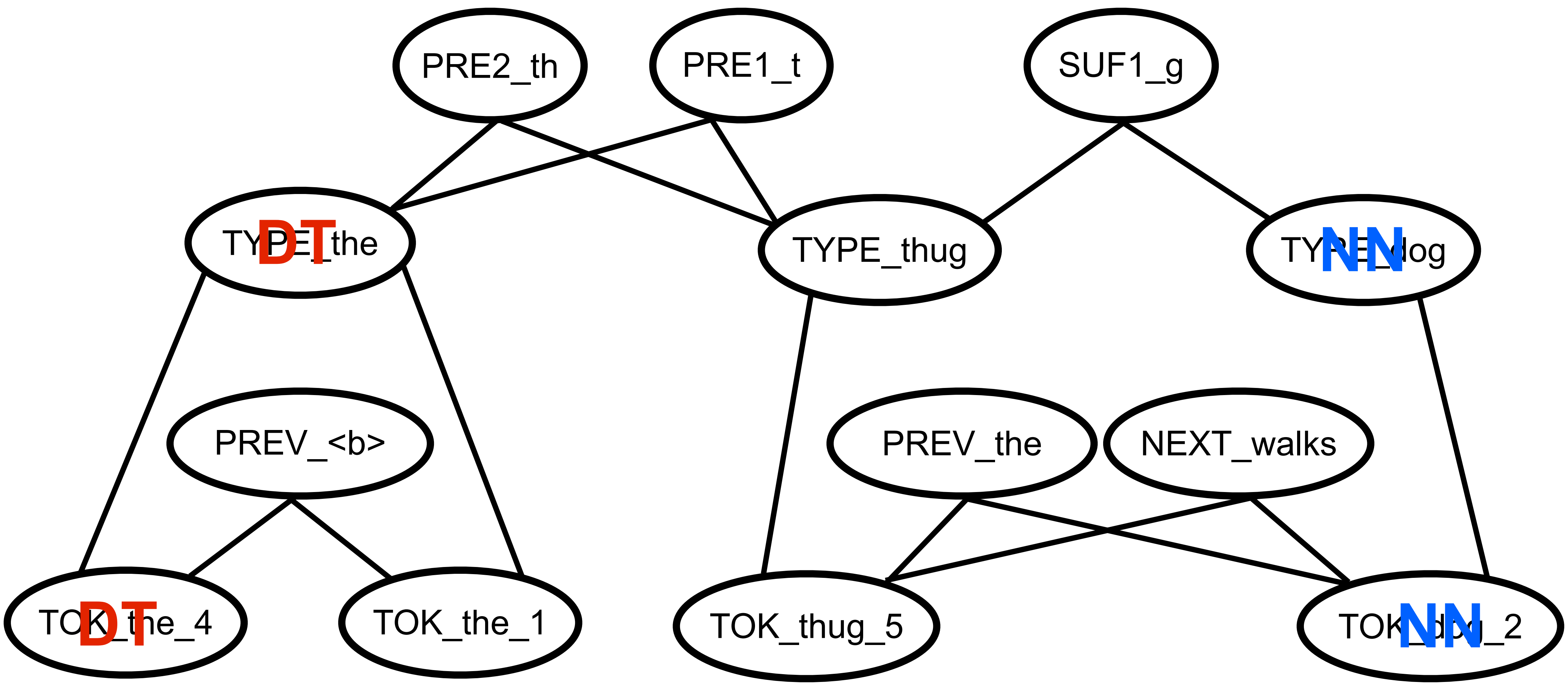
Tag Dict Generalization



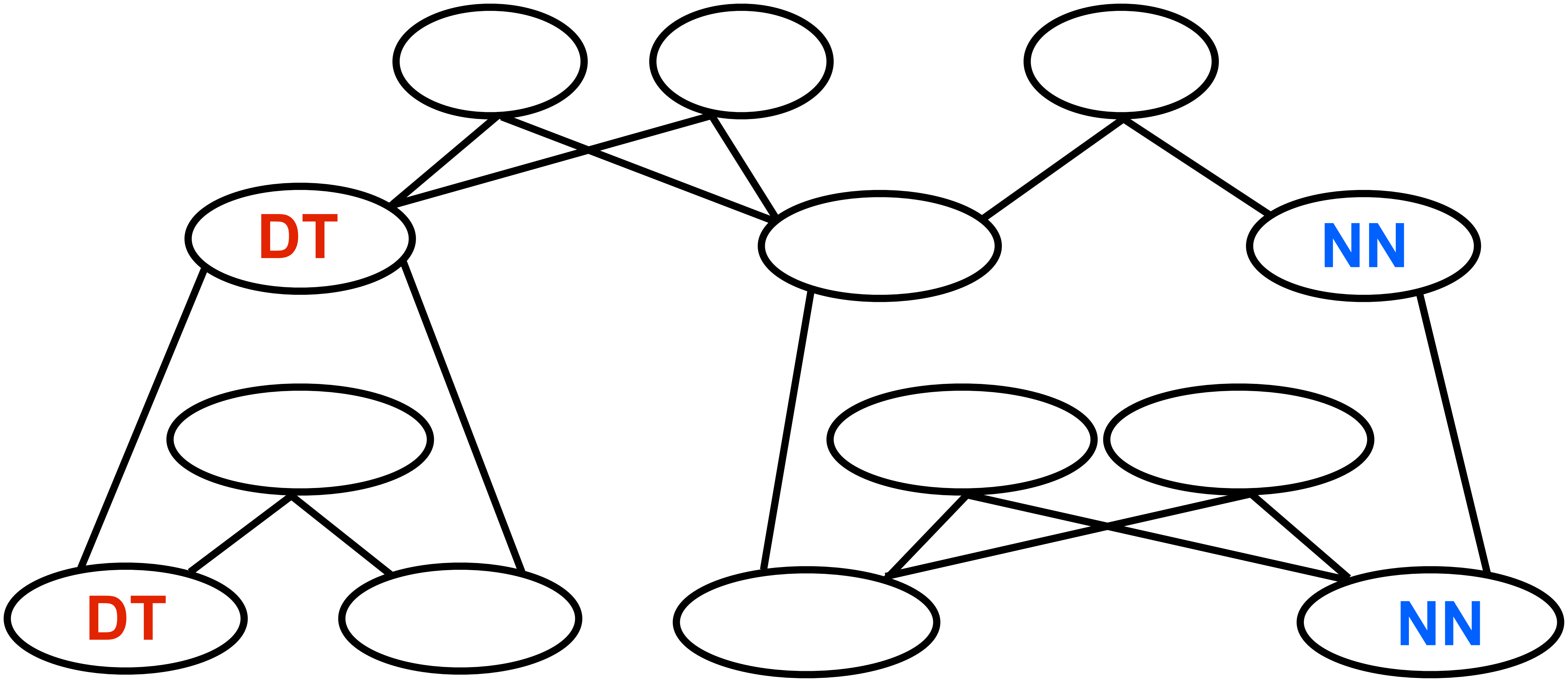
Finite-State Transducer (FST)

- Generates morphological analysis
- Hand-built by a linguist in 10 hours

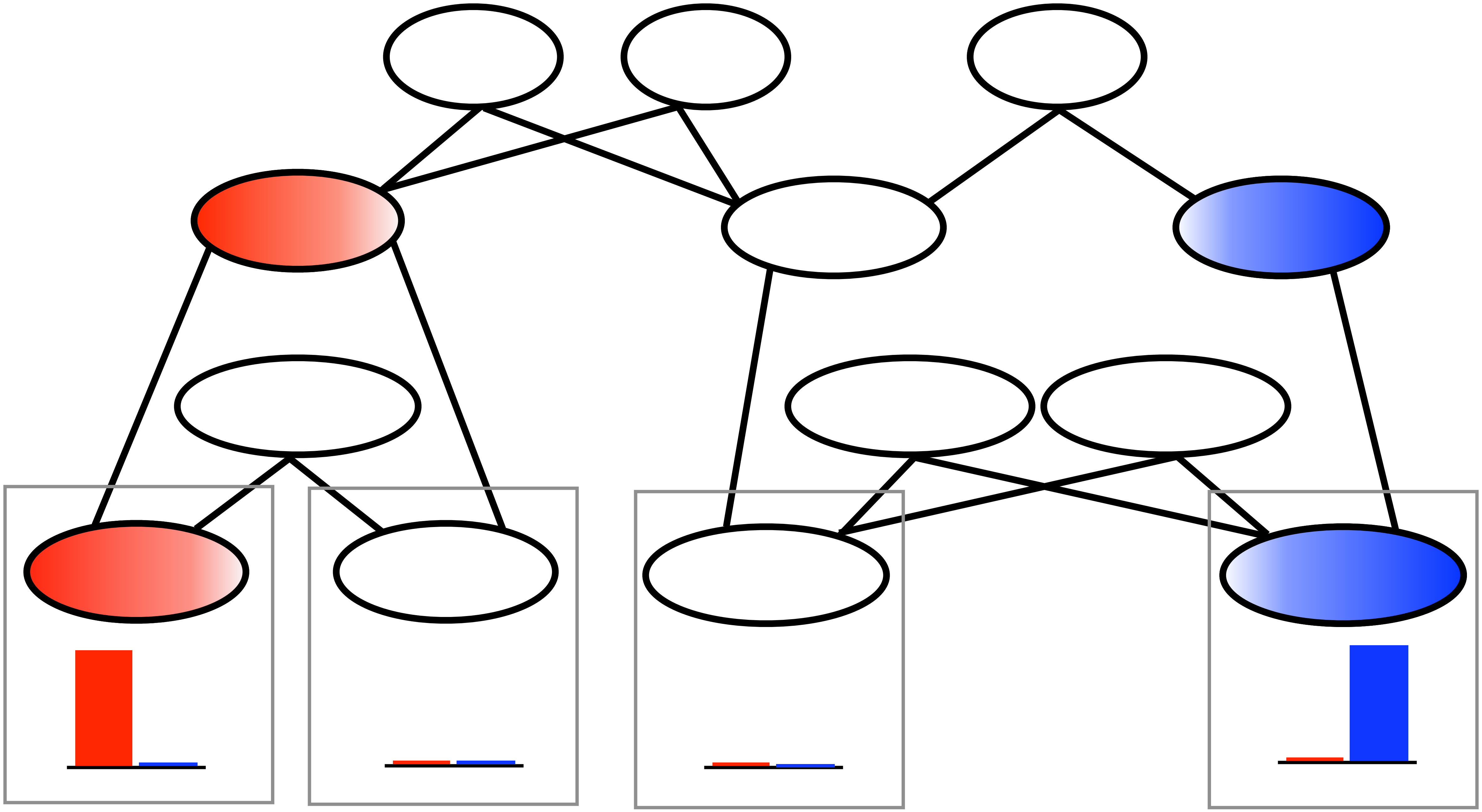
Tag Dict Generalization



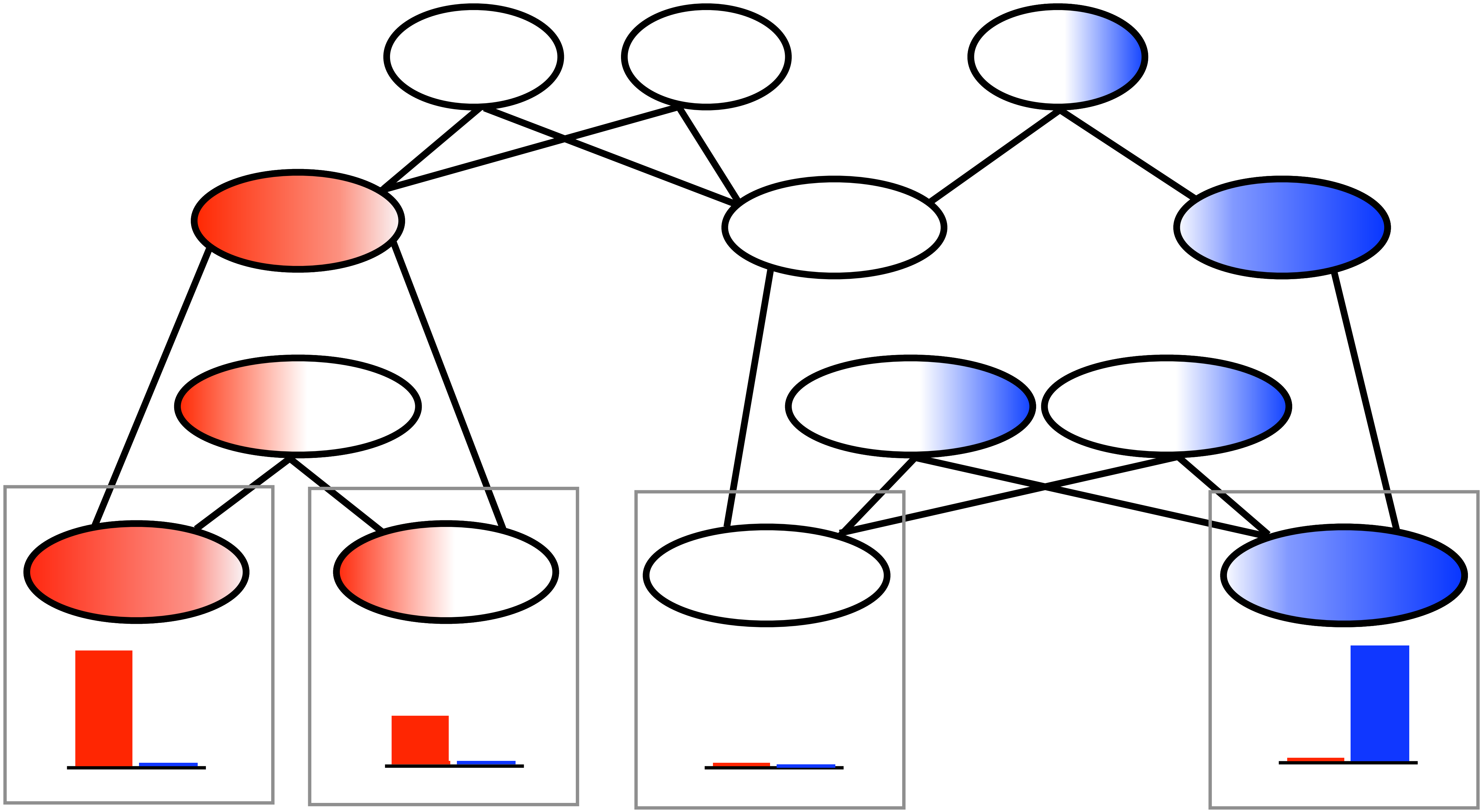
Tag Dict Generalization



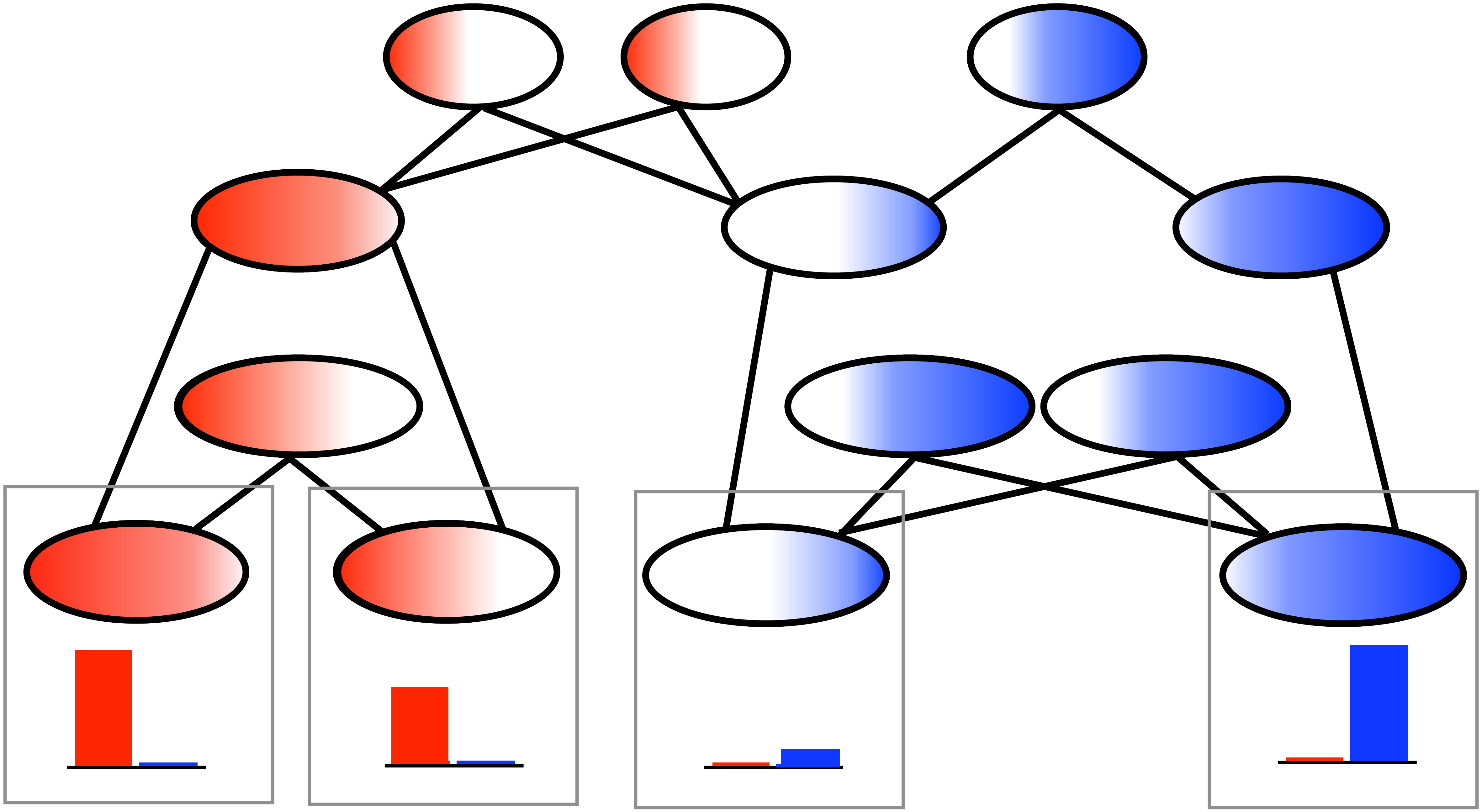
Tag Dict Generalization



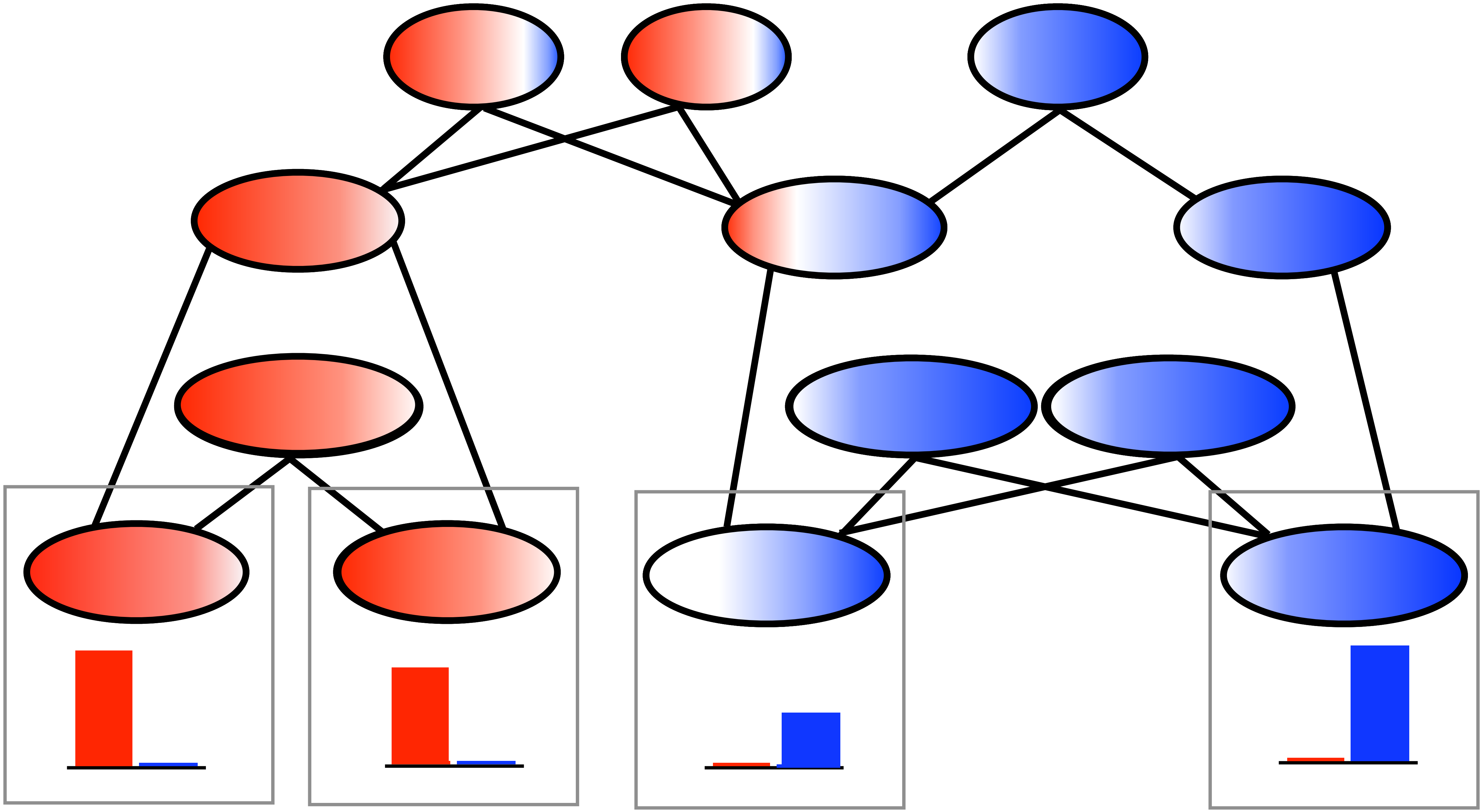
Tag Dict Generalization



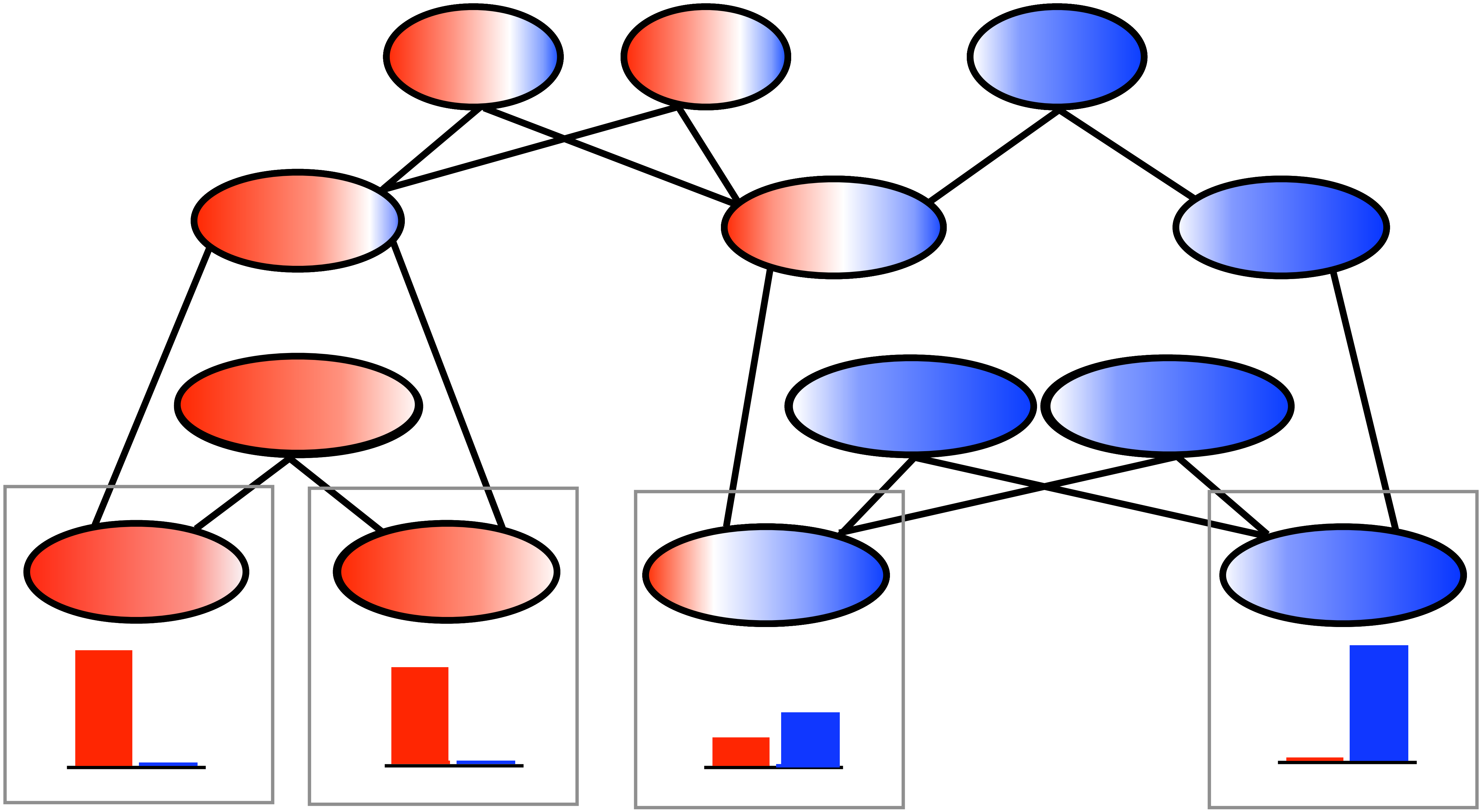
Tag Dict Generalization



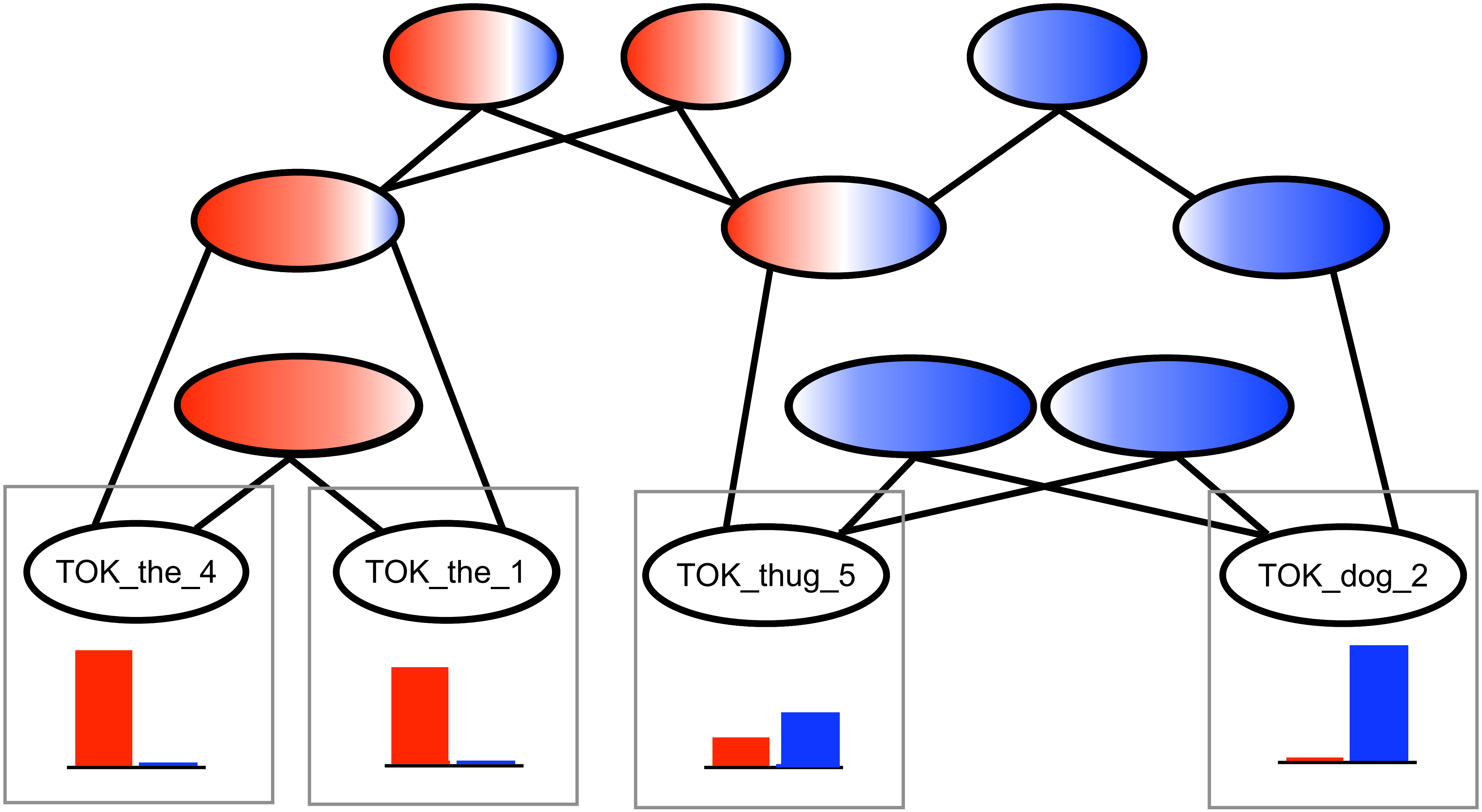
Tag Dict Generalization



Tag Dict Generalization



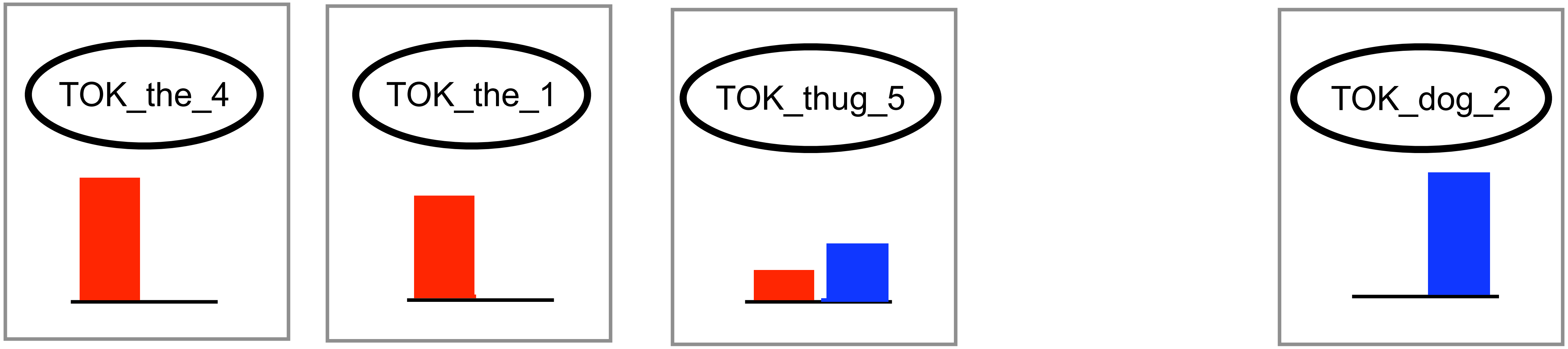
Tag Dict Generalization



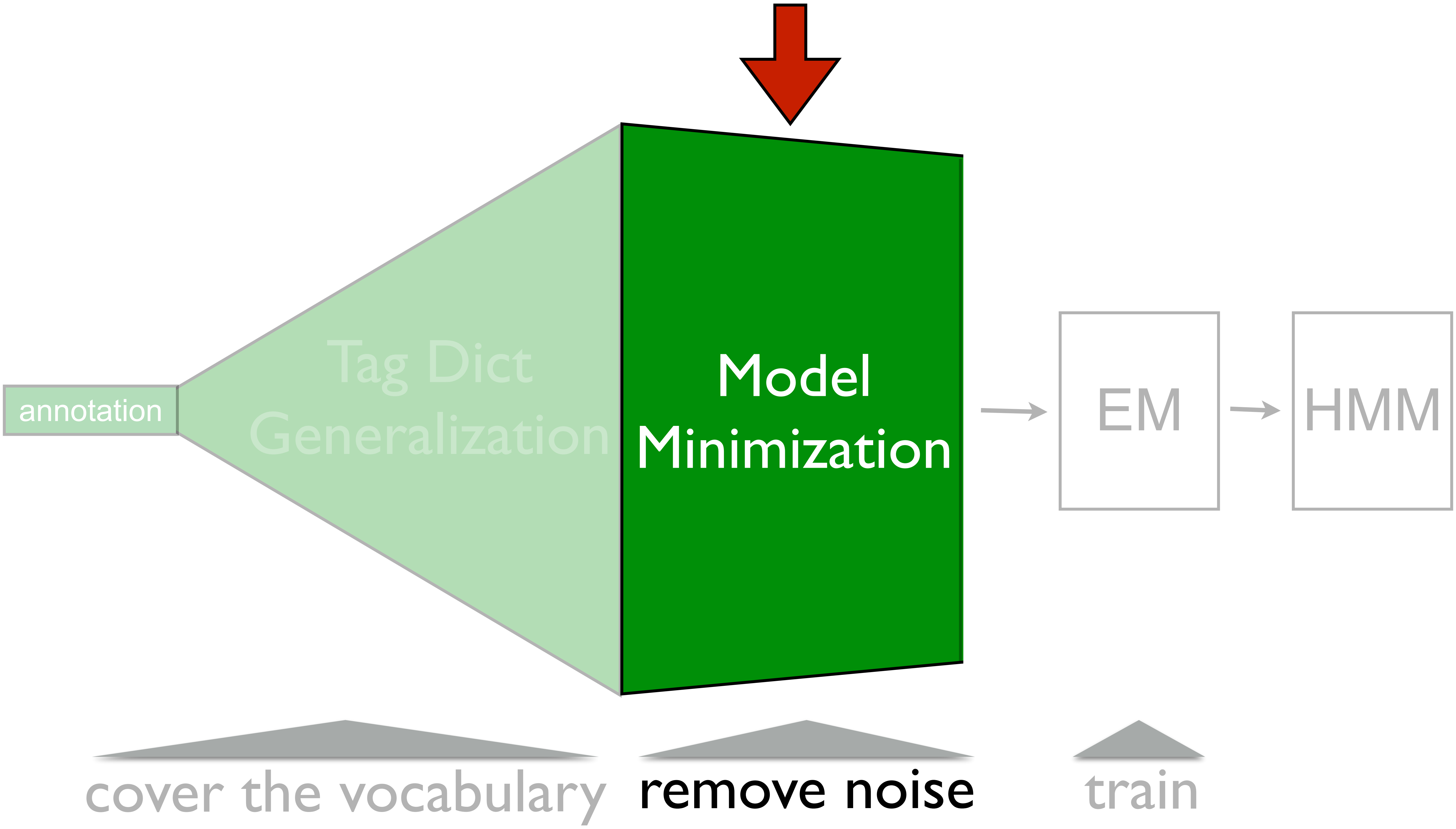
Tag Dict Generalization

Result:

- a tag distribution on every token (soft tagging)
- an expanded tag dictionary (non-zero tags)



Our Approach

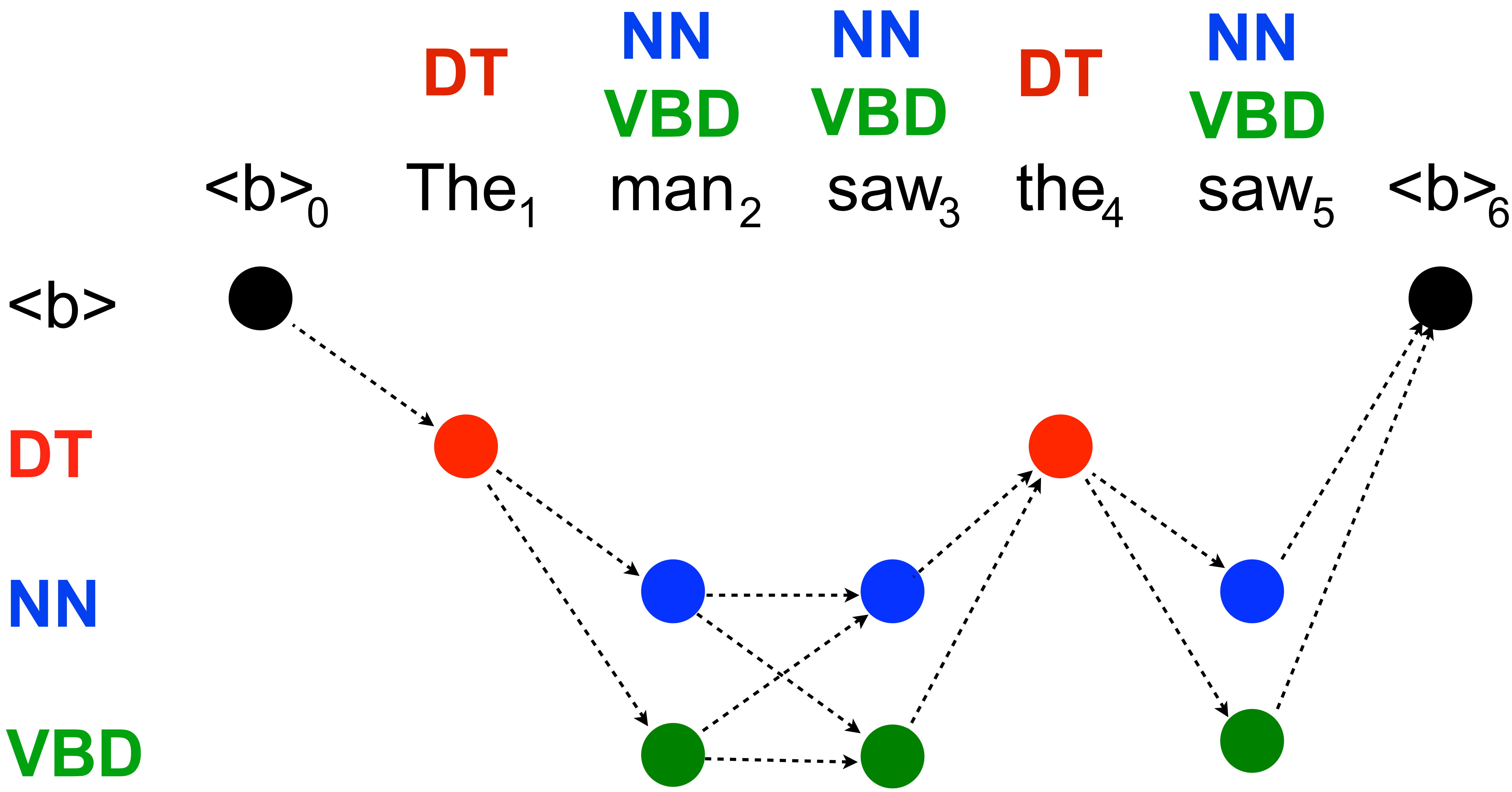


Model Minimization

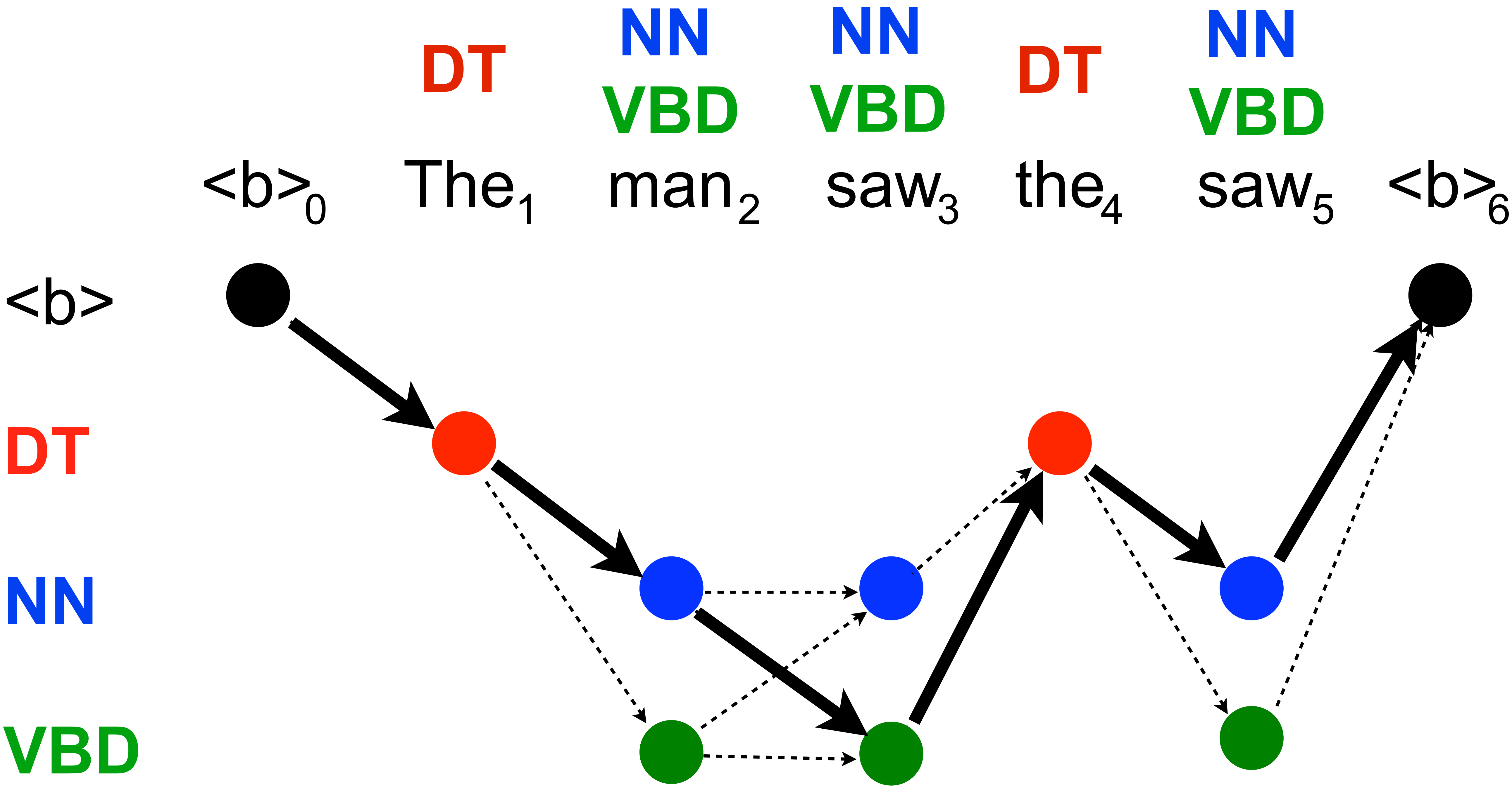
- Induce a cleaner hard tagging from a noisy soft tagging.
- Approach based on work by Sujith Ravi and Kevin Knight (ISI)

[Ravi et al., 2010; Garrette and Baldridge, 2012]

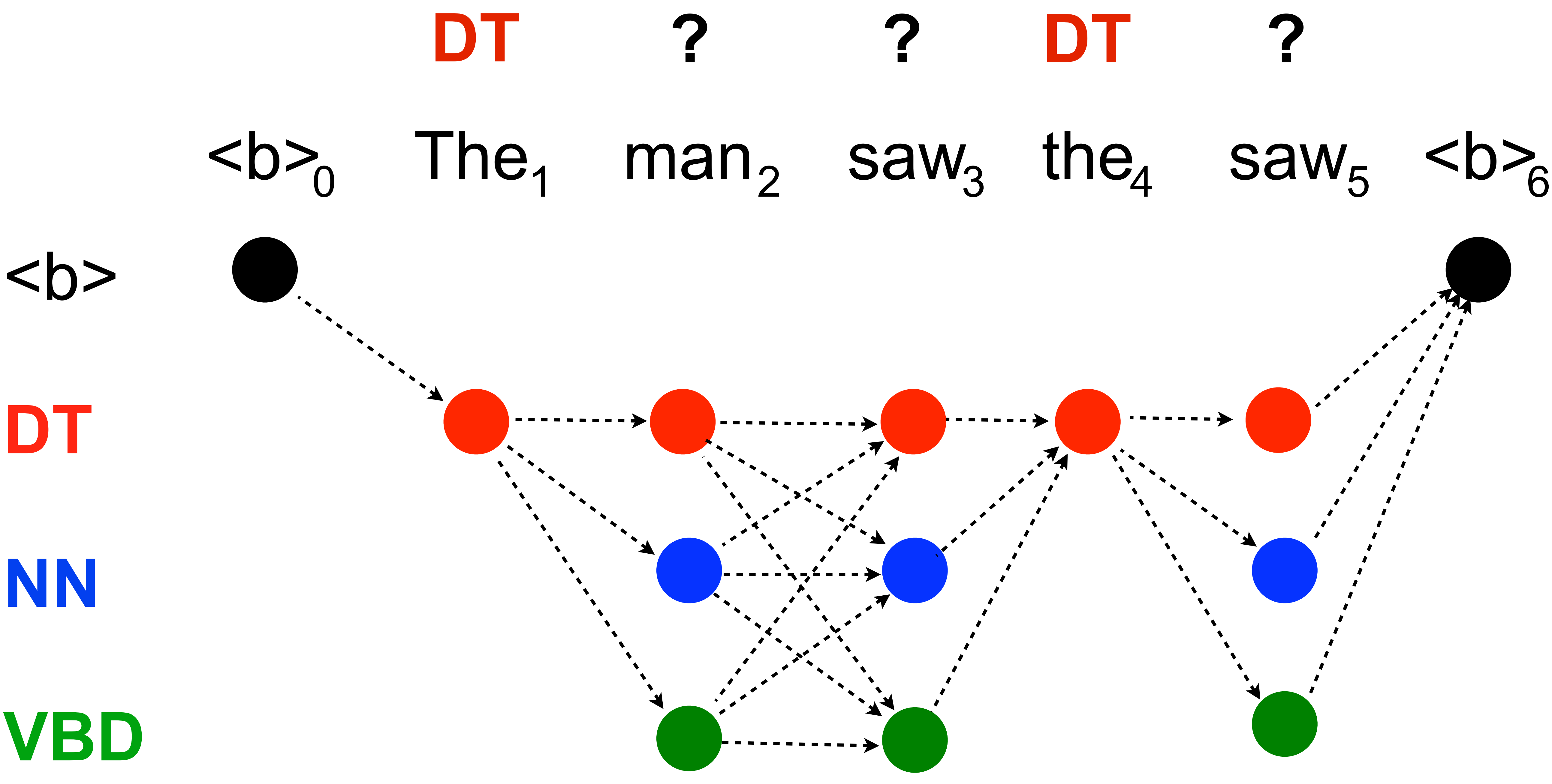
Model Minimization



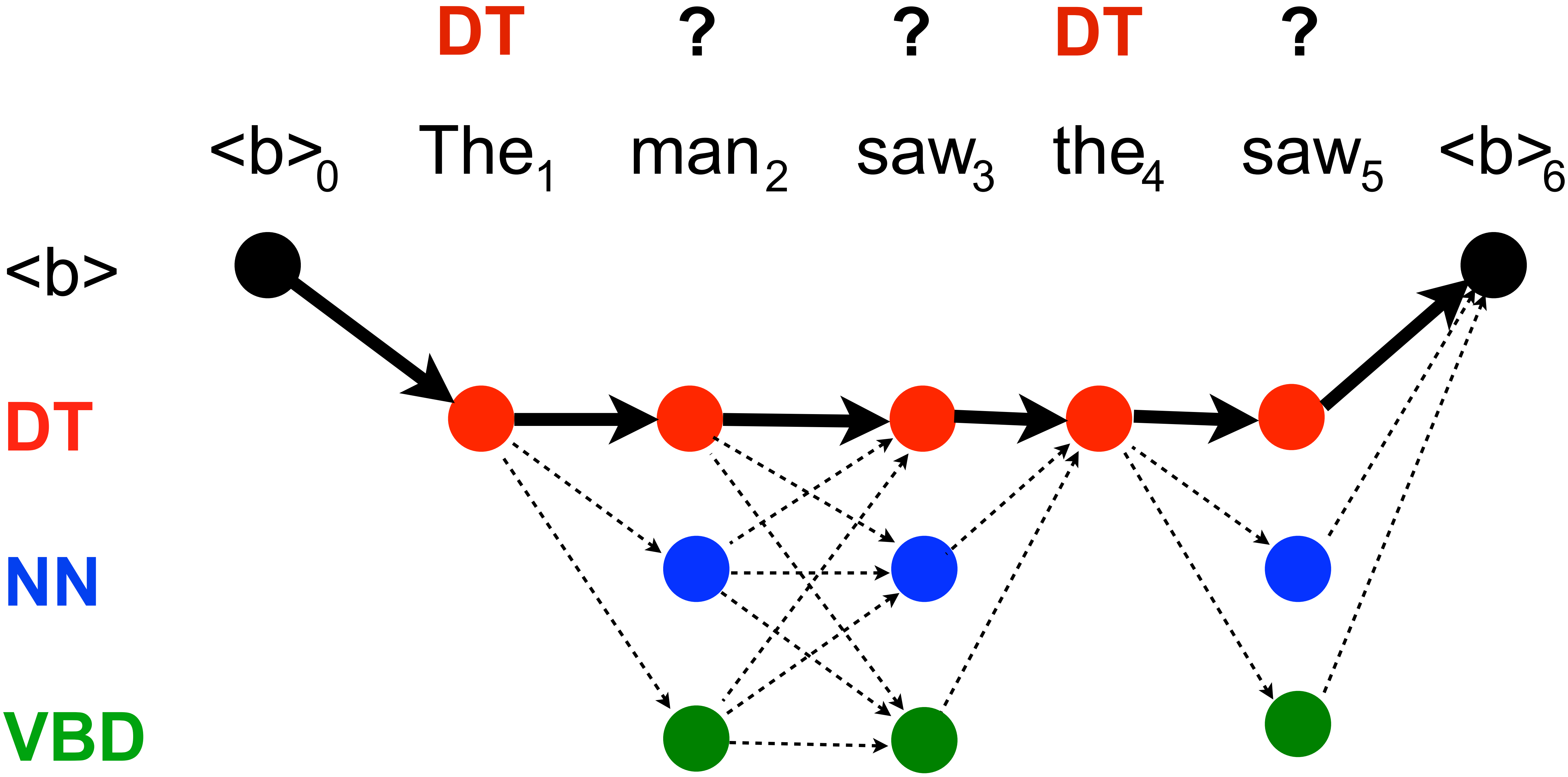
Model Minimization



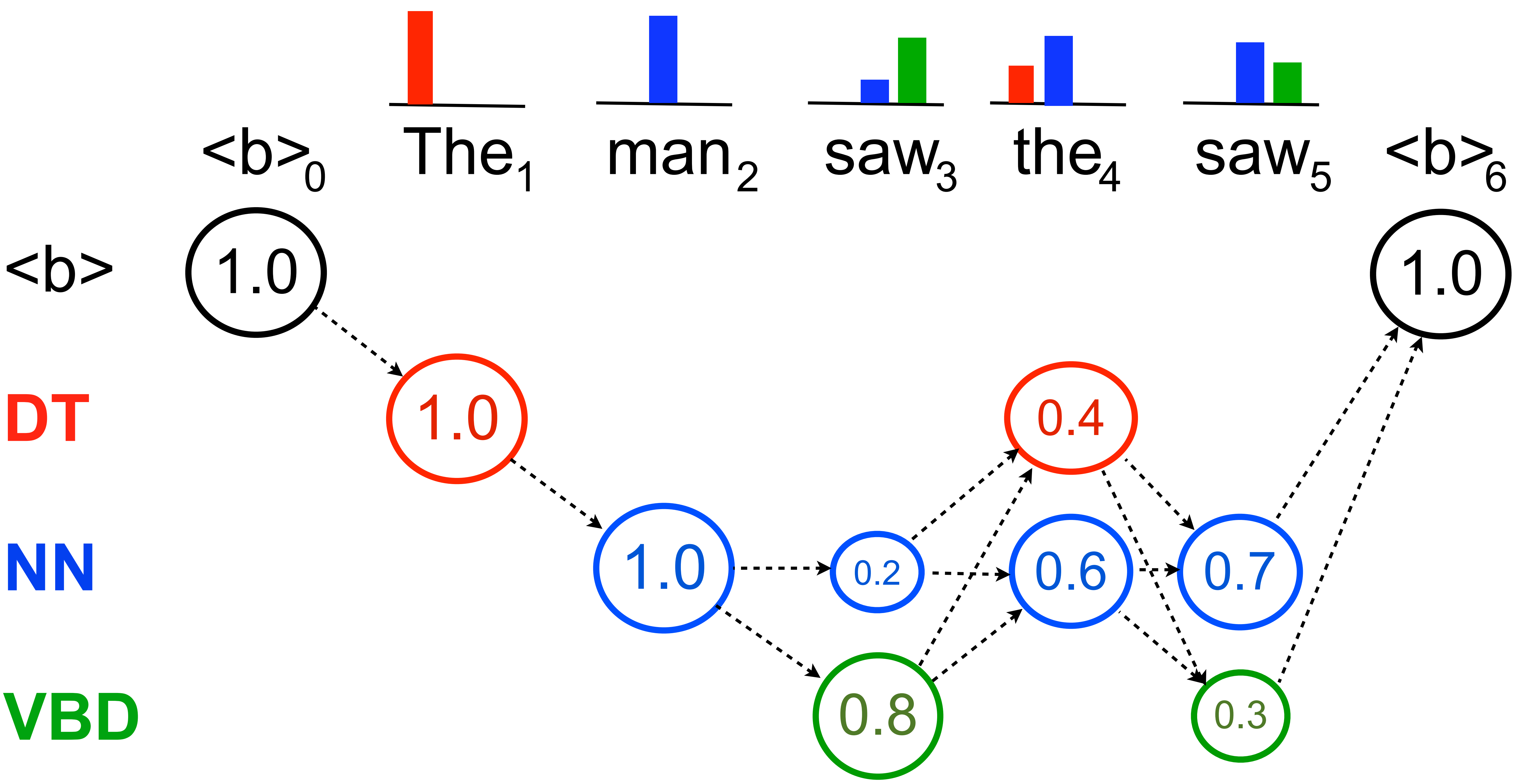
Model Minimization



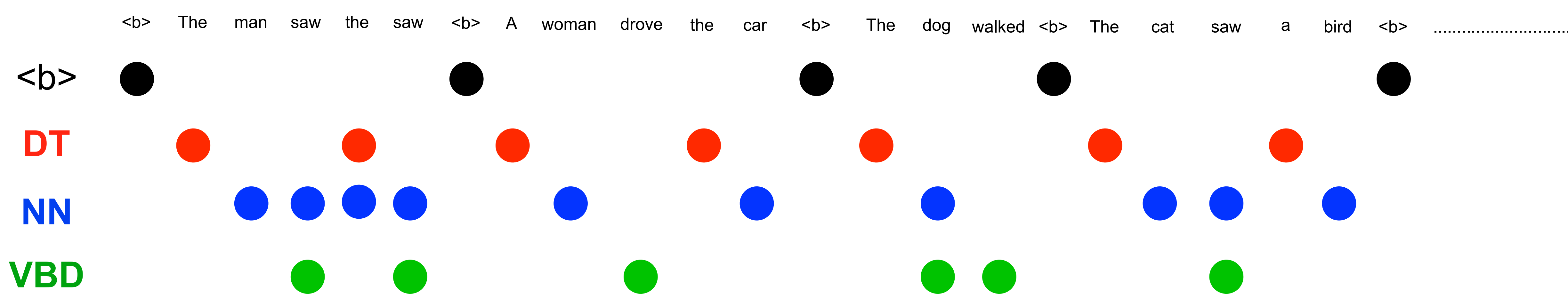
Model Minimization



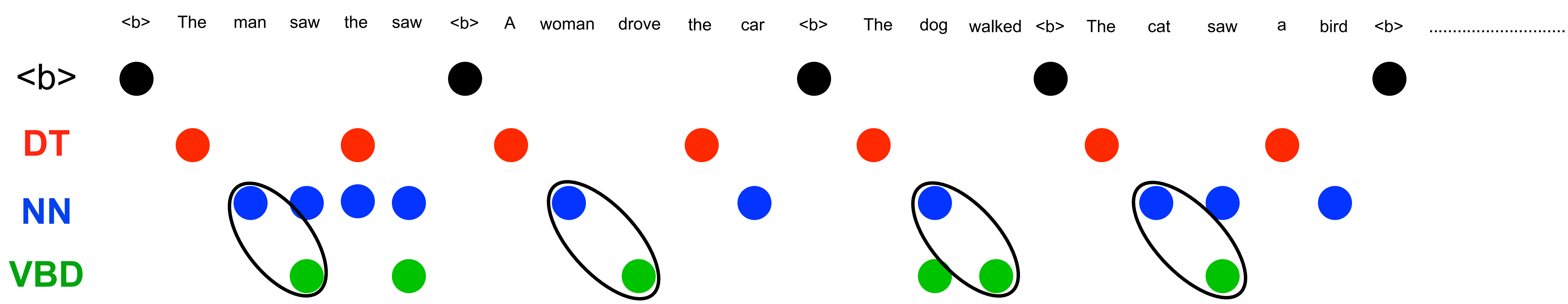
Model Minimization



Model Minimization

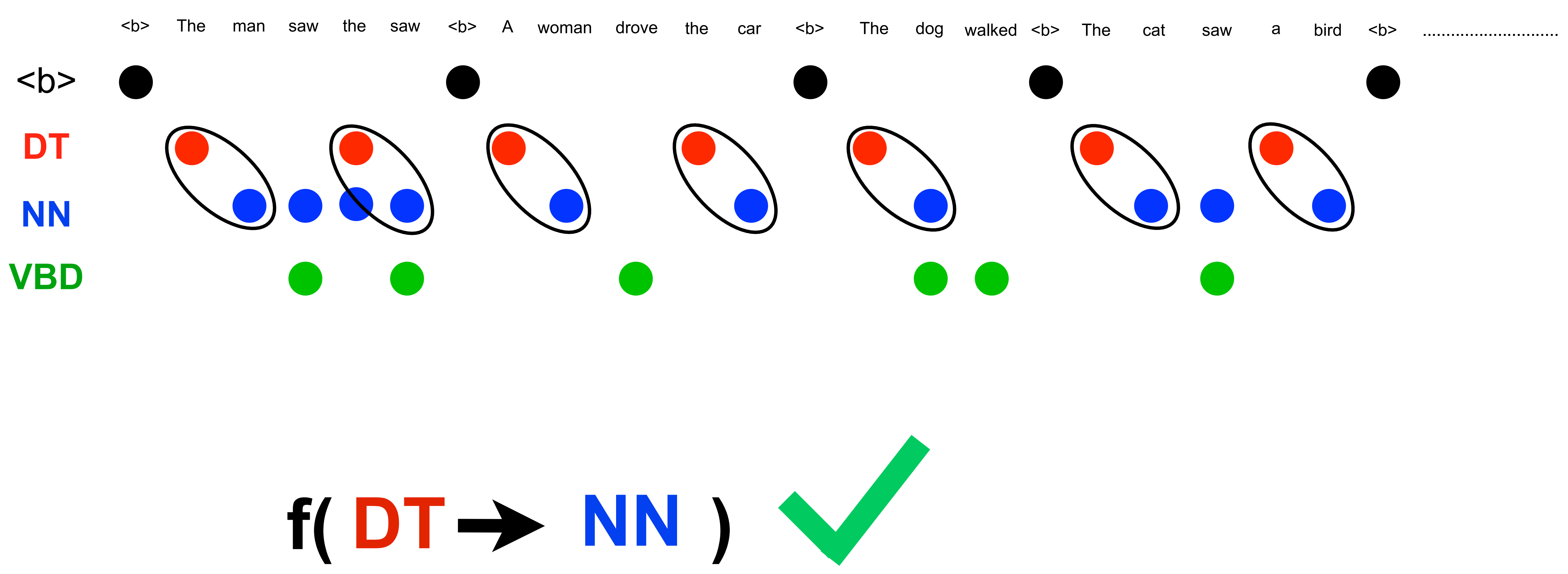


Model Minimization

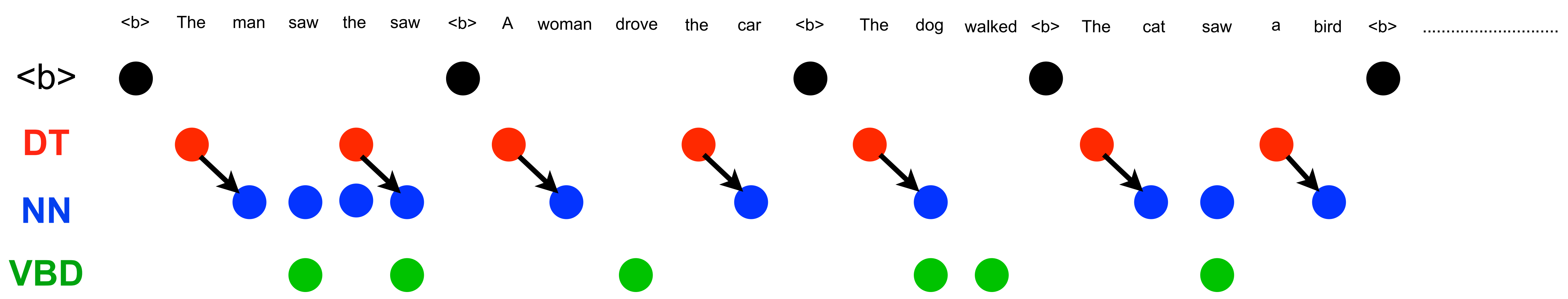


$$f(\mathrm{NN} \rightarrow \mathrm{VBD}) \begin{cases} \text{tag bigram occurrences} \\ \text{weights on their nodes} \end{cases}$$

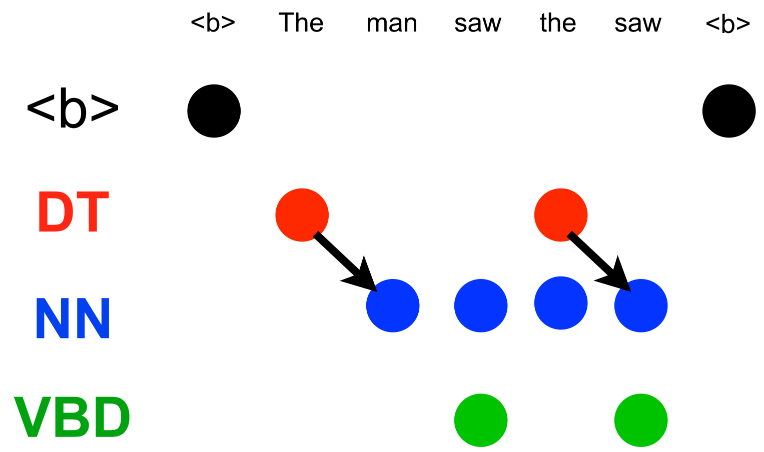
Model Minimization



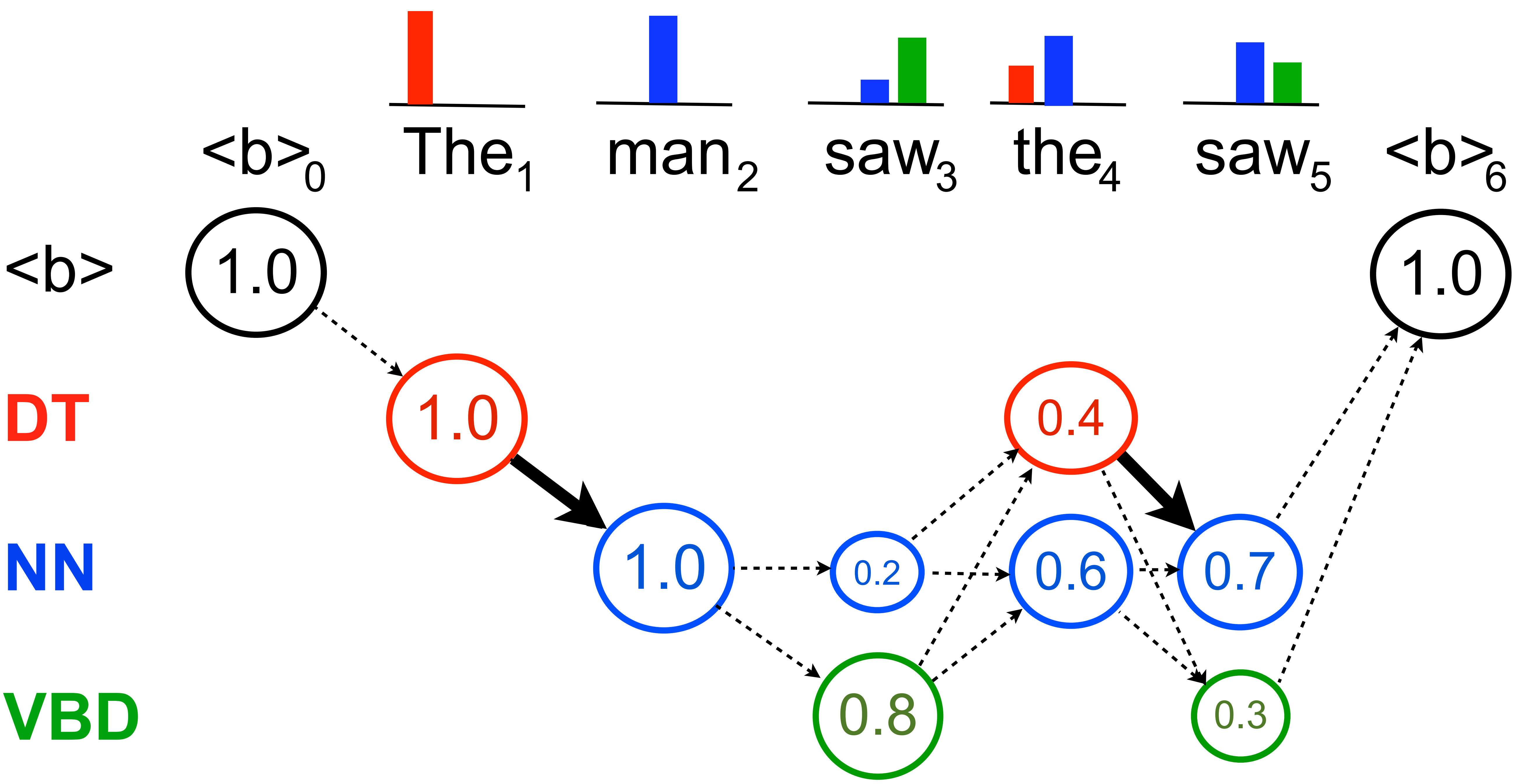
Model Minimization



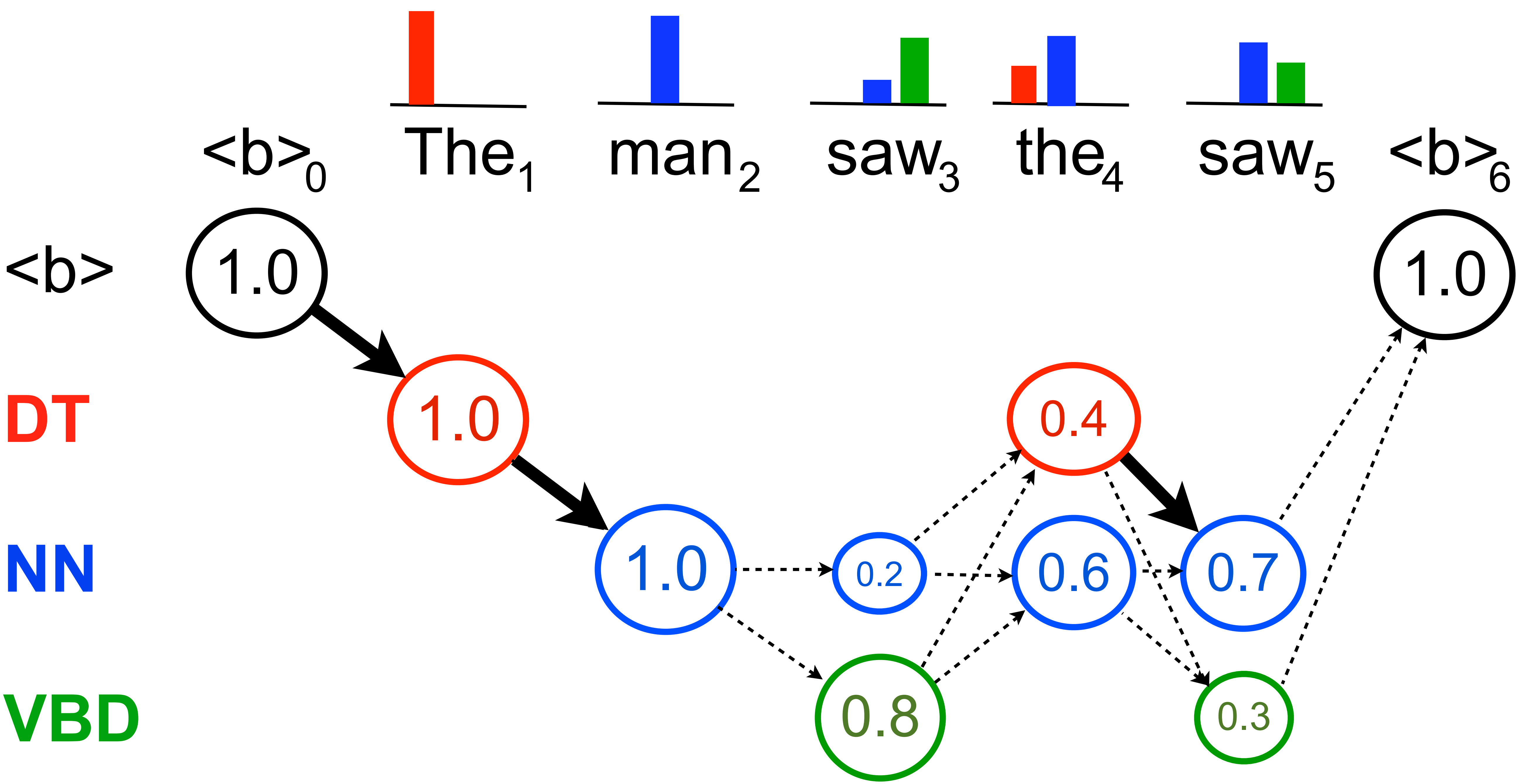
Model Minimization



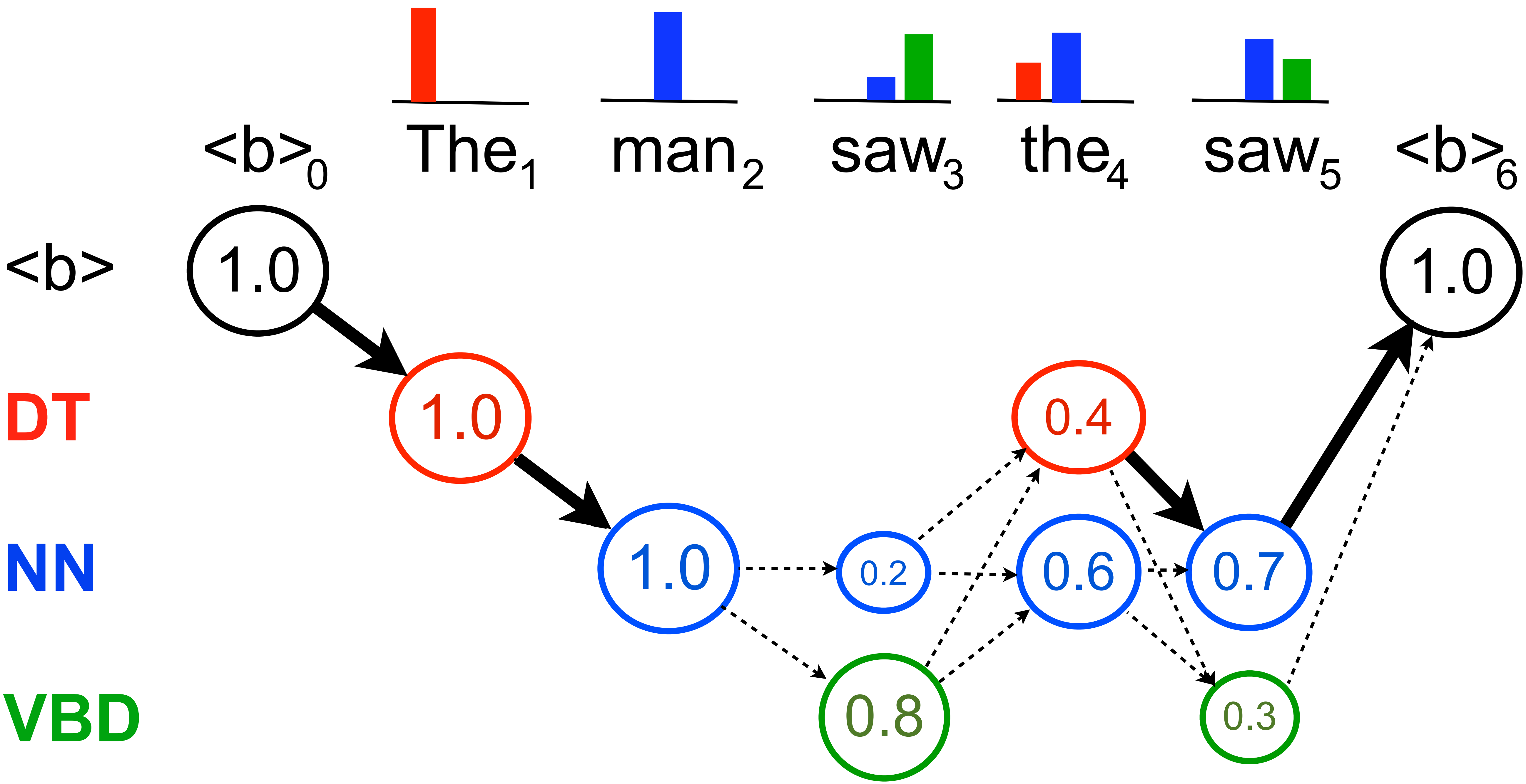
Model Minimization



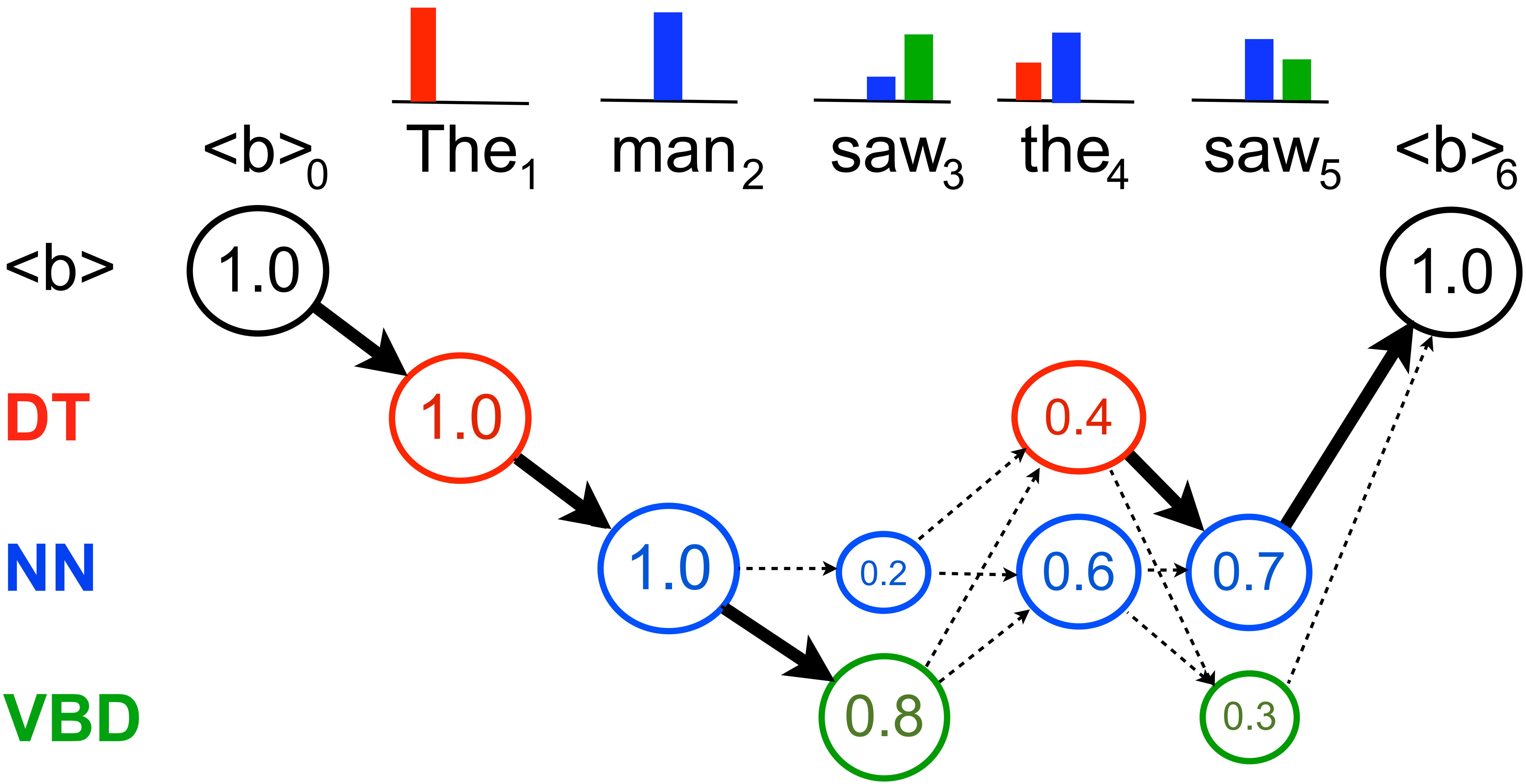
Model Minimization



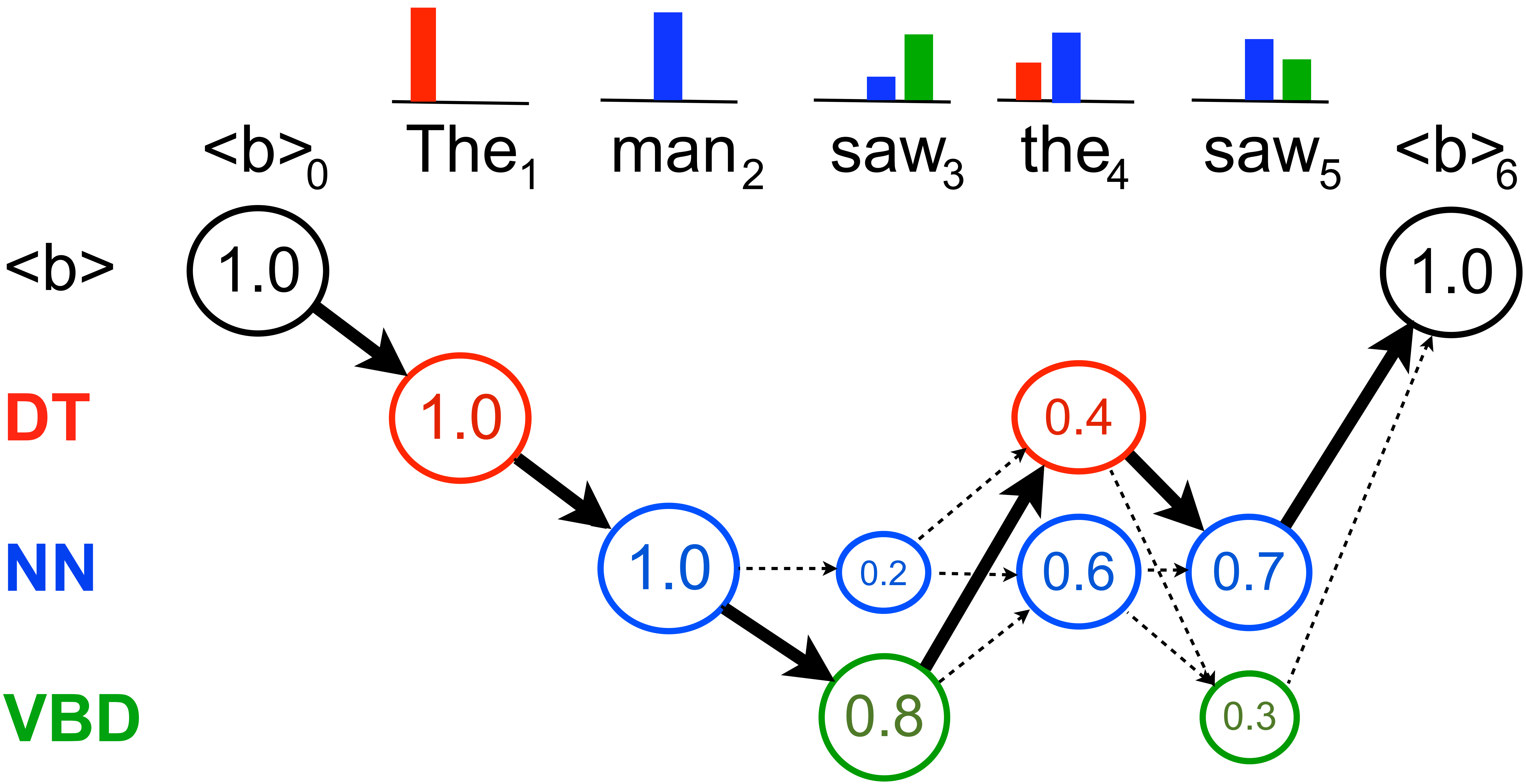
Model Minimization



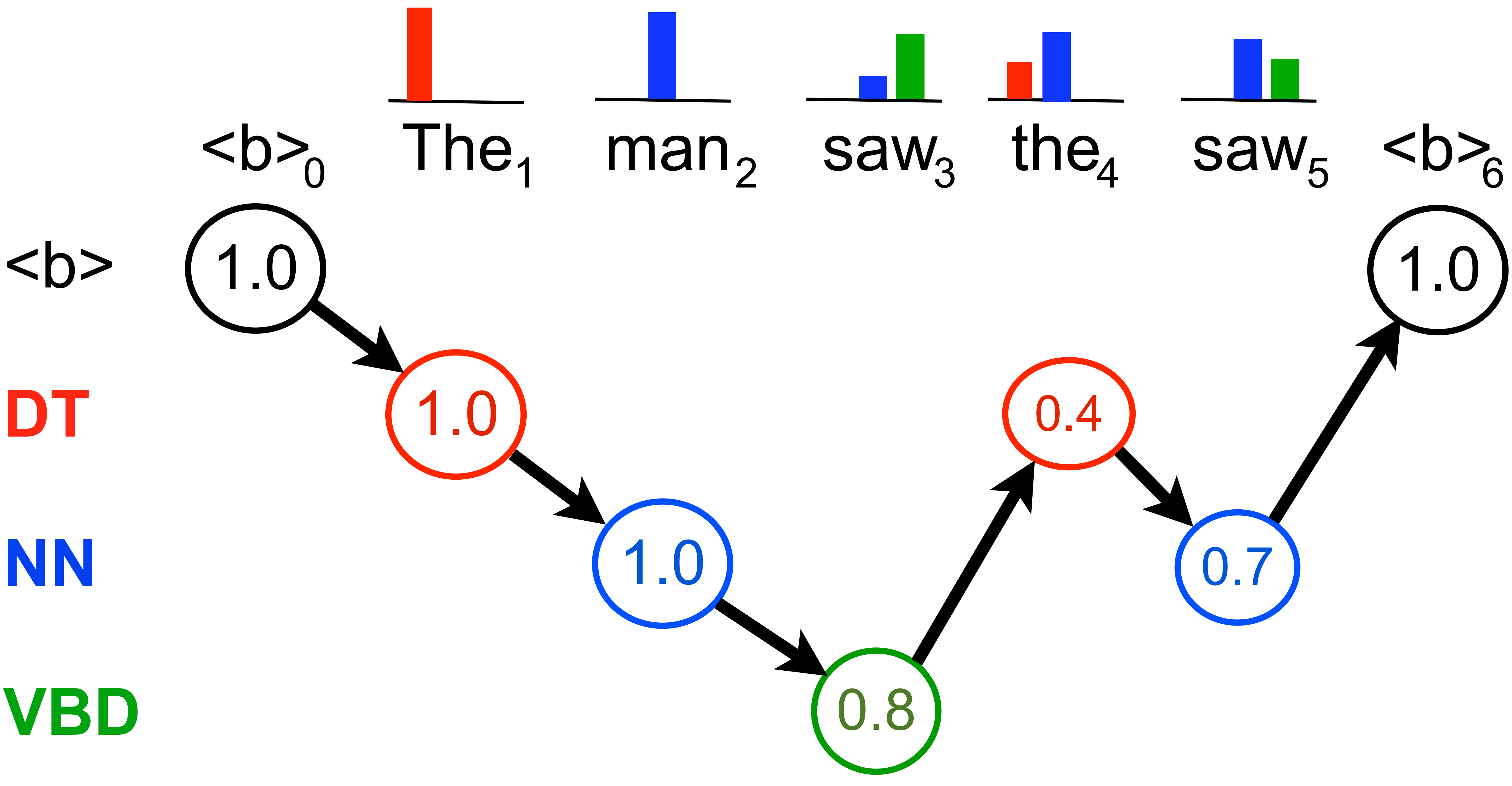
Model Minimization



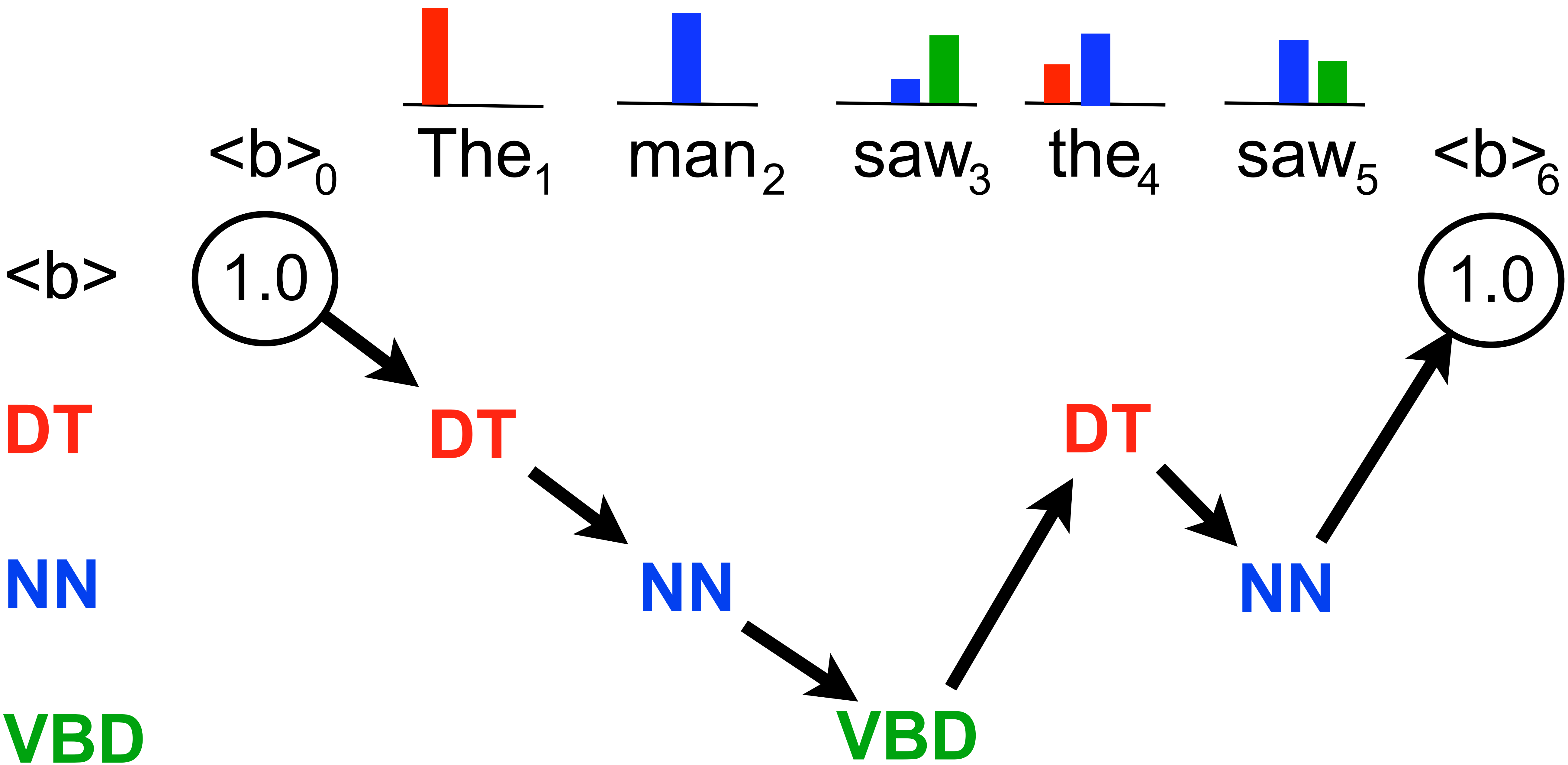
Model Minimization



Model Minimization



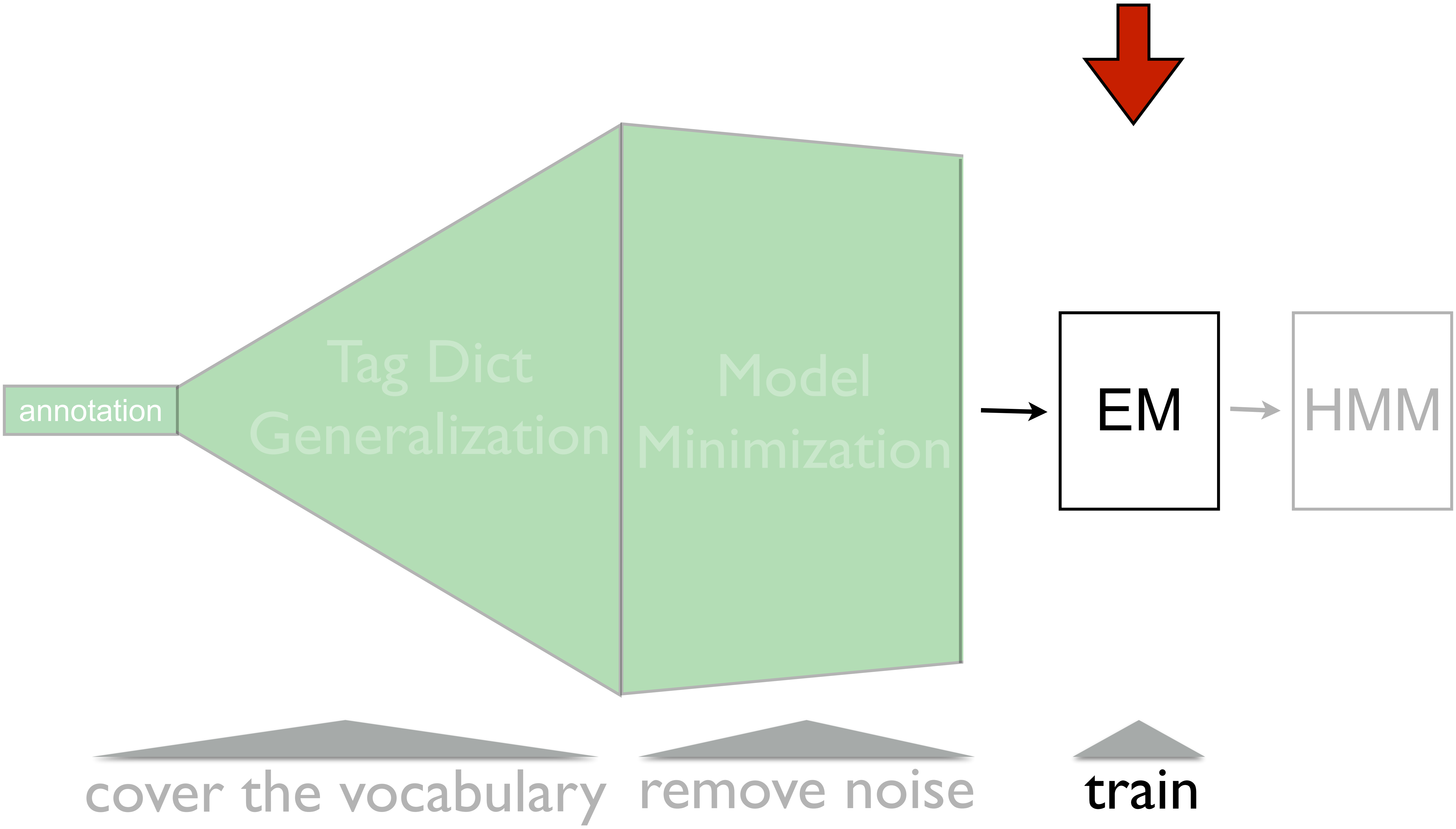
Model Minimization



Model Minimization

<b>$_0$	The$_1$	man$_2$	saw$_3$	the$_4$	saw$_5$	<b>$_6$
	DT	**NN**	**VBD**	**DT**	**NN**	

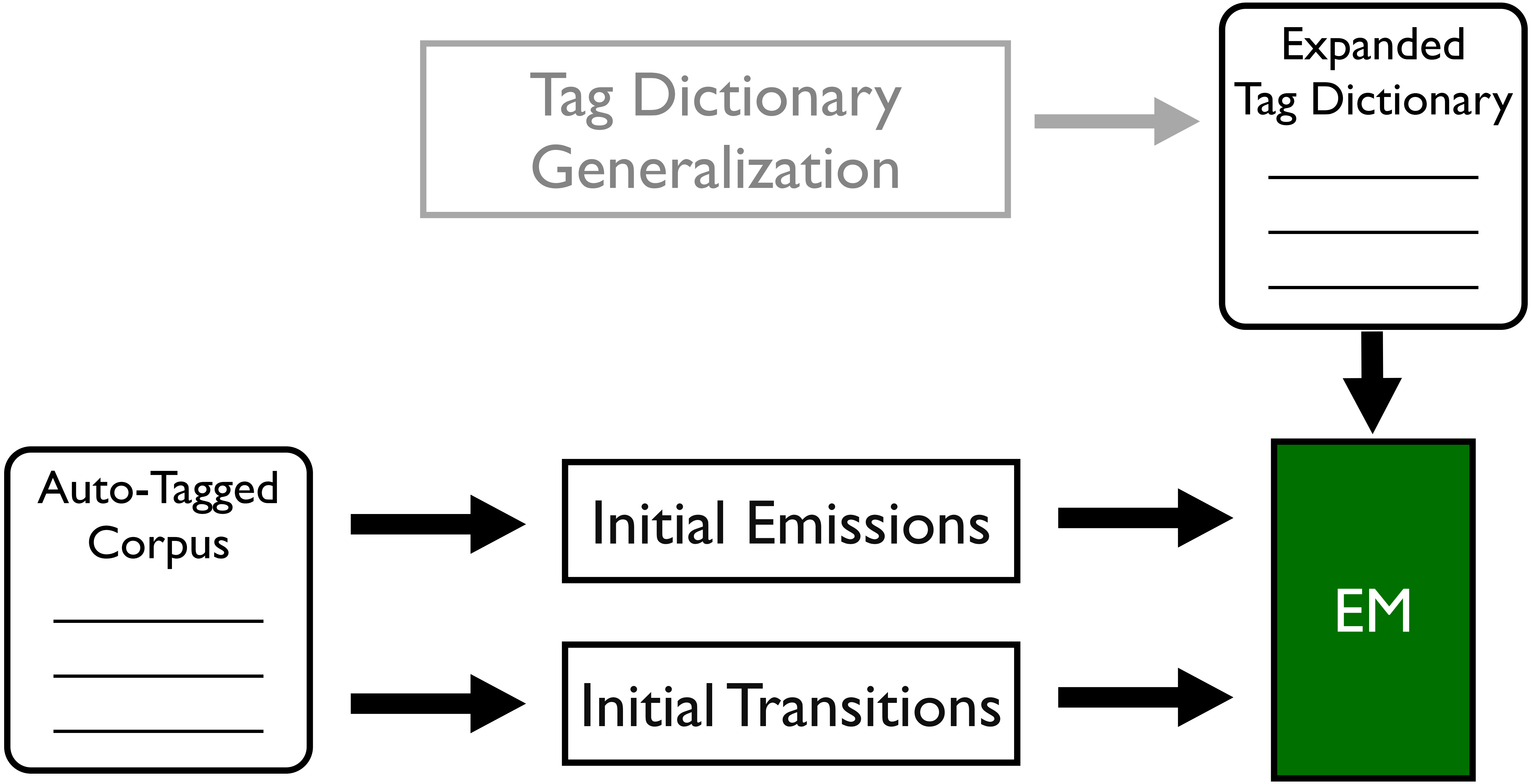
Our Approach



EM Training

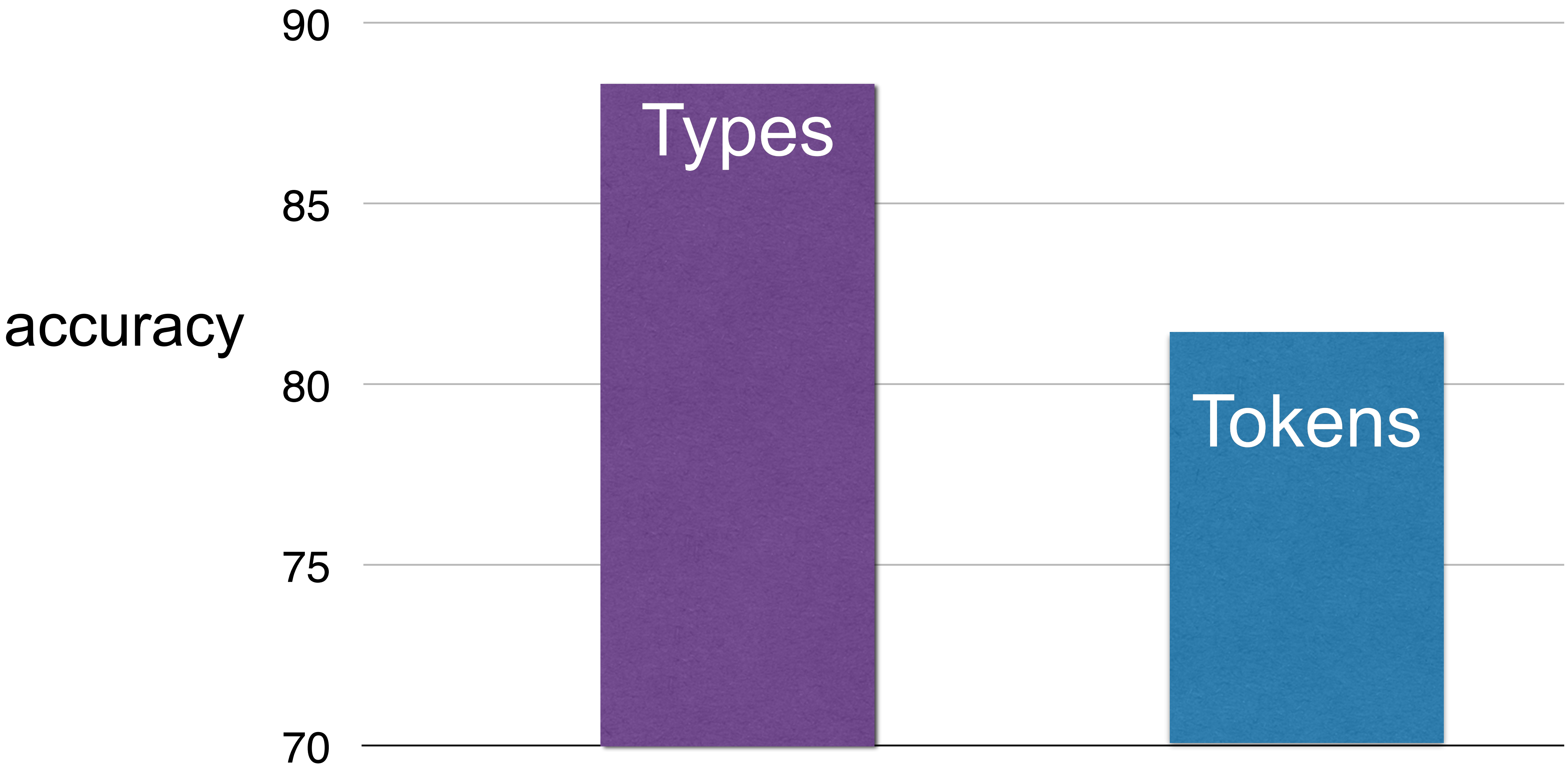
<b>$_0$	The$_1$	man$_2$	saw$_3$	the$_4$	saw$_5$	<b>$_6$
	DT	**NN**	**VBD**	**DT**	**NN**	

EM Training


Tag Dictionary Generalization
Expanded Tag Dictionary
Auto-Tagged Corpus
Initial Emissions
Initial Transitions
EM

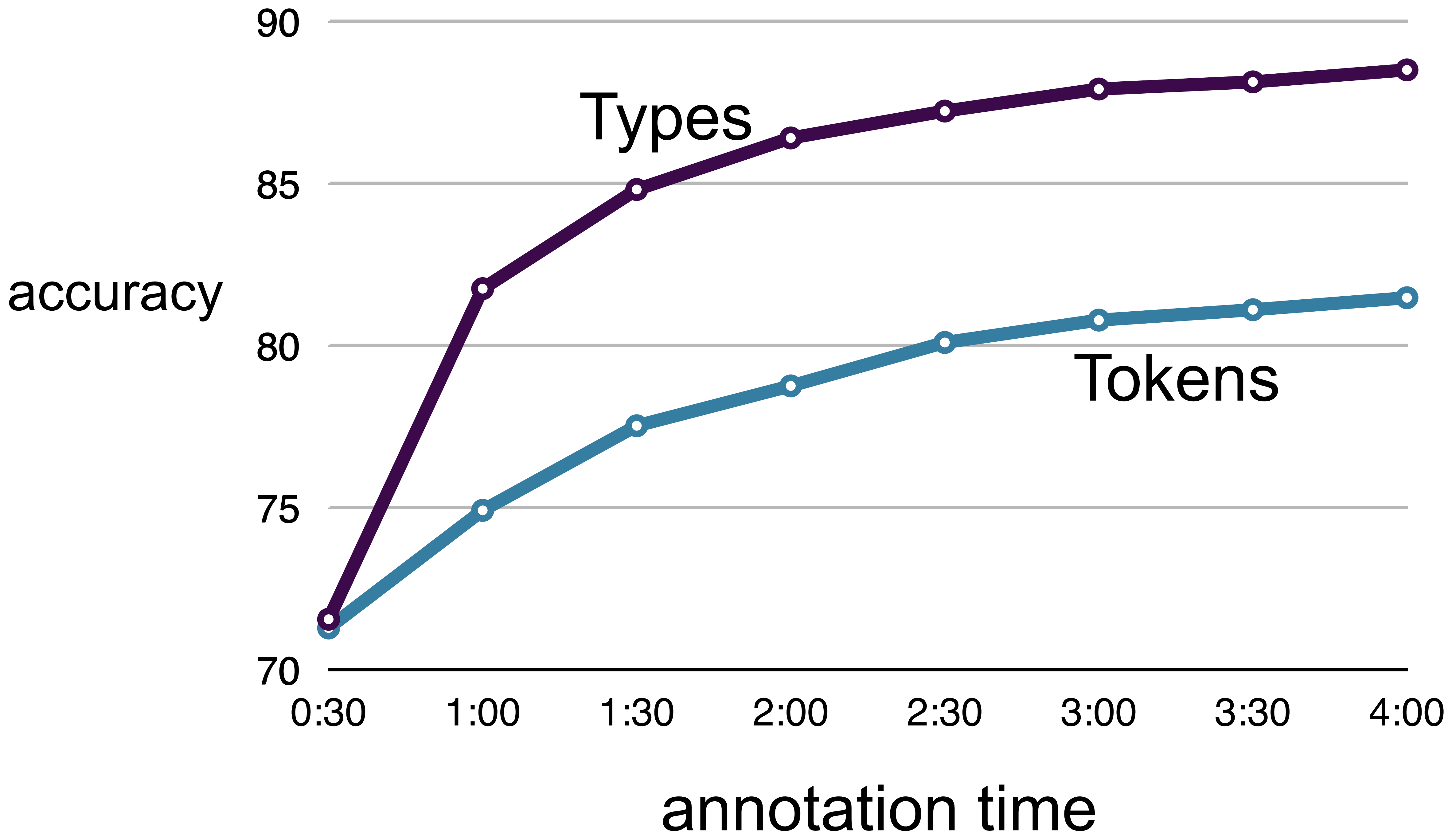
Results

Types vs. Tokens



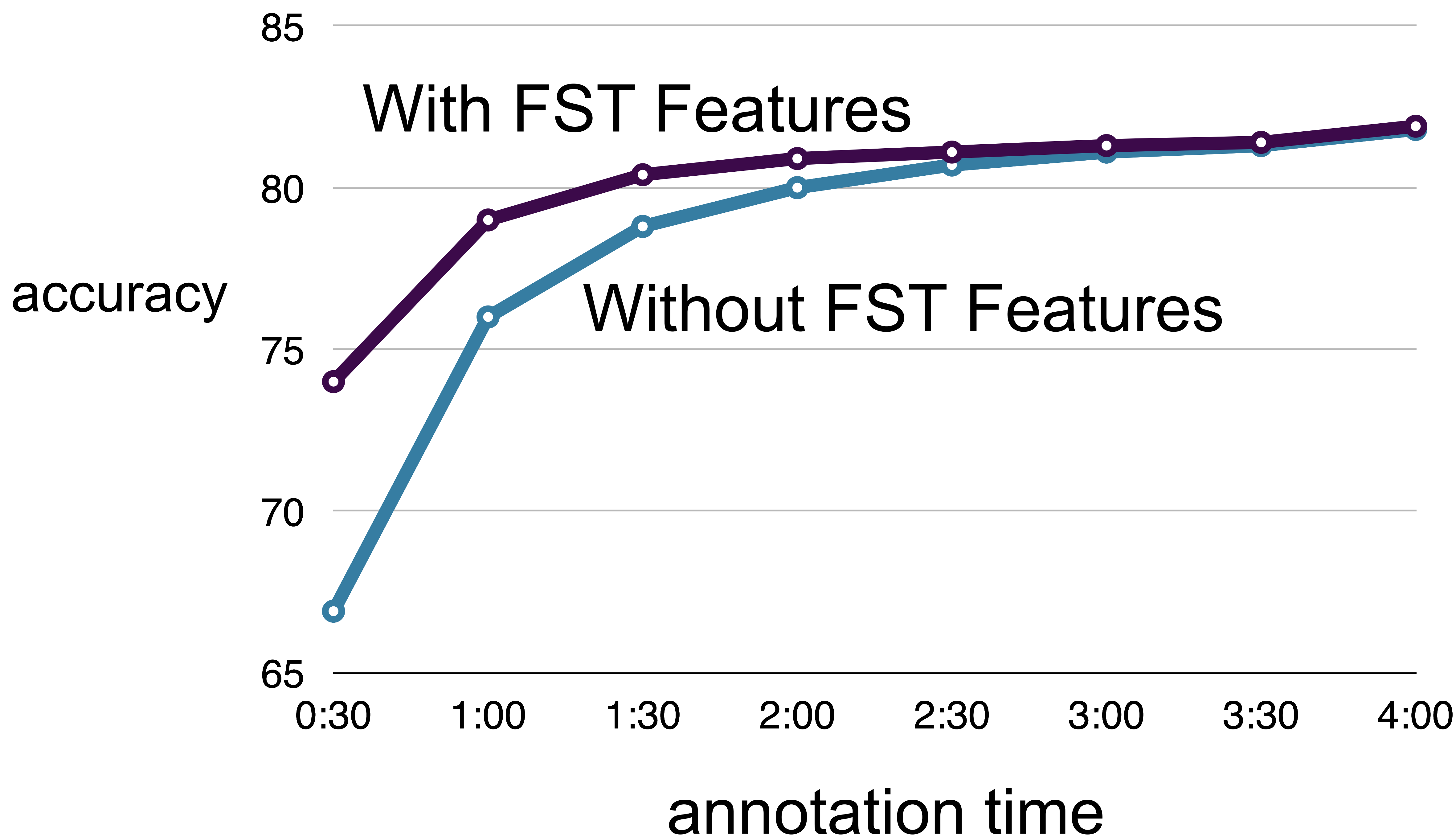
[English]

Types vs. Tokens



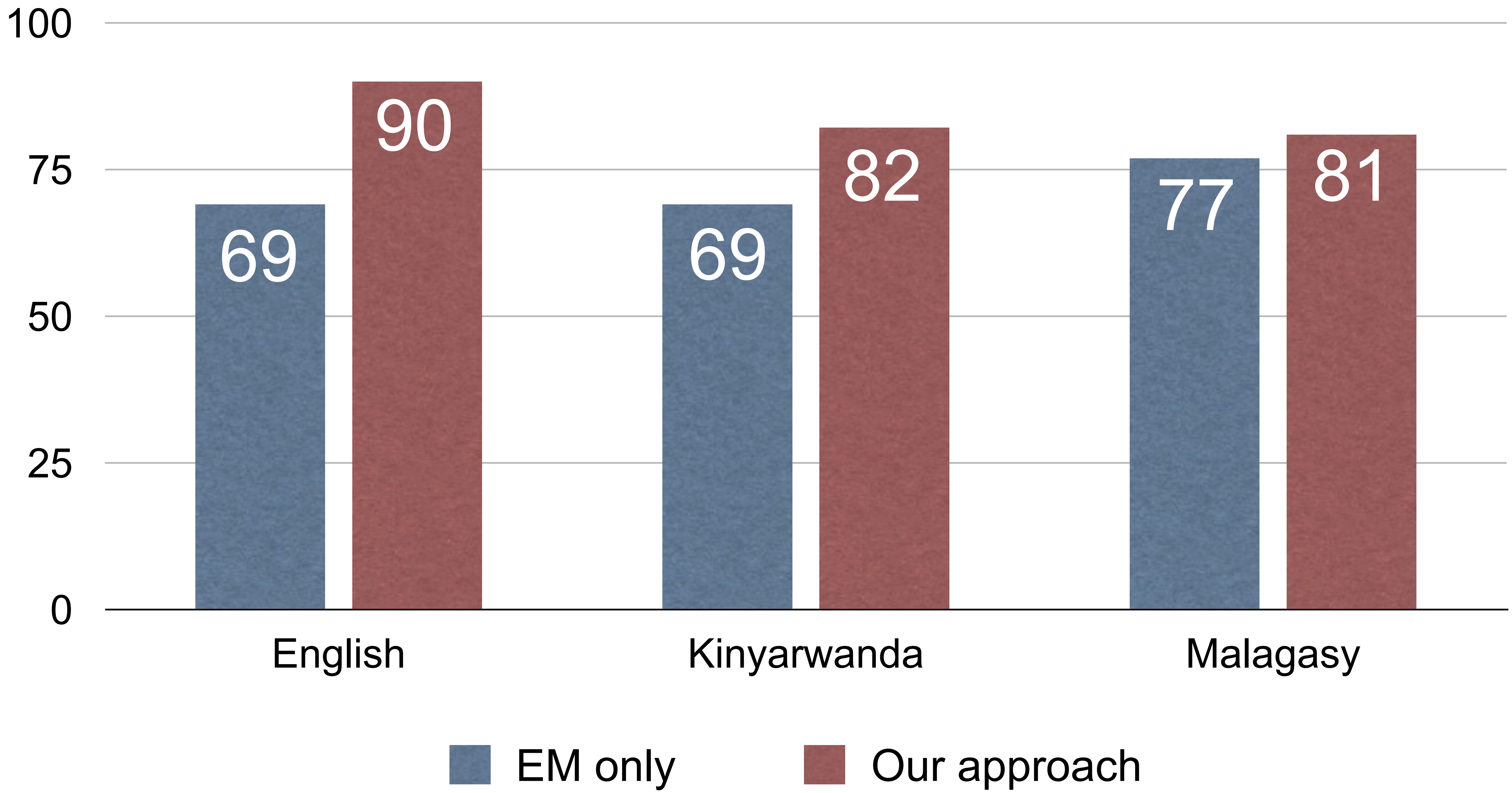
[English]

Morphological Analysis



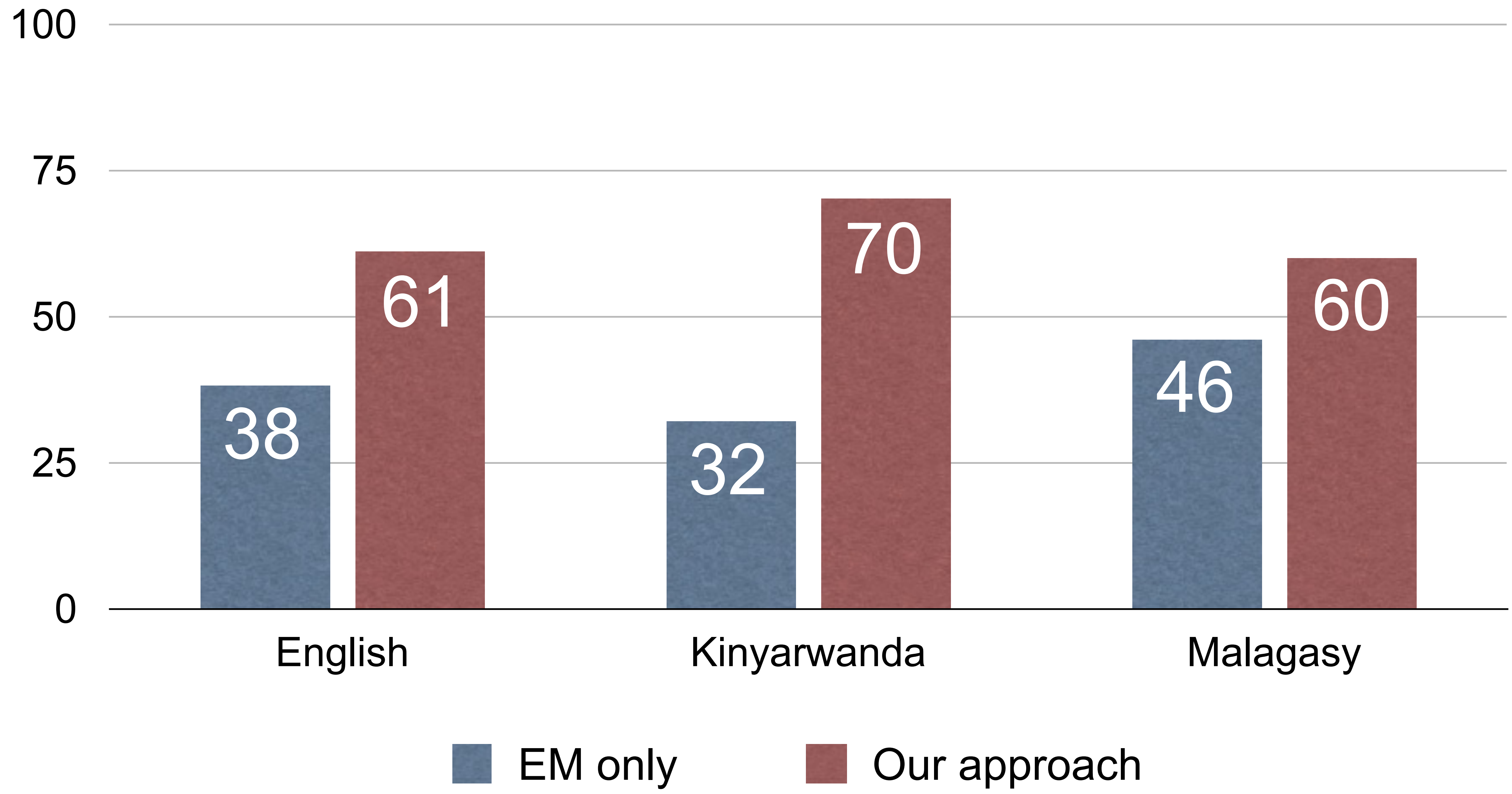
[Kinyarwanda]

Total Accuracy



[4 hours of type annotation]

Unknown Word Accuracy



[2 hours of type annotation]

English Results

12 tags	All of **Wiktionary** (Li et al., 2012)	87%
	Parallel Corpus (Täckström et al., 2013)	89%
45 tags	**4-hours** (Garrette et al., 2013)	**90%**

Rich Morphology

Parallel Corpus (Täckström et al., 2013)

Turkish	65%

4-hours (Garrette et al., 2013)

Kinyarwanda	**82%**

Current Work

- Minimally supervised CCG supertagging and parsing
- Human-provided GFL annotations

Conclusion

- Our approach is able to achieve results better that or comparable to others, but given significantly less input.
- Our annotations are available to others.
- Software available as well.