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.

Low-Resource Languages

Supervised training is not an option.

We do semi-supervised training.

Low-Resource Languages

Supervised training is not an option.

We do semi-supervised training.

➡ Annotate some data by hand

Low-Resource Languages

Supervised training is not an option.

We do semi-supervised training.

→ Annotate some data by hand

... cheaply

Semi-Supervised Training

[Kupiec, 1992]
[Merialdo, 1994]

Semi-Supervised Training

HMM with Expectation-Maximization (EM)

[Kupiec, 1992]
[Merialdo, 1994]

Semi-Supervised Training

HMM with Expectation-Maximization (EM)

Need:

[Kupiec, 1992]
[Merialdo, 1994]

Semi-Supervised Training

HMM with Expectation-Maximization (EM)

Need:

Large **raw** corpus

[Kupiec, 1992]
[Merialdo, 1994]

Semi-Supervised Training

HMM with Expectation-Maximization (EM)

Need:

Large **raw** corpus

Tag dictionary

[Kupiec, 1992]
[Merialdo, 1994]

Semi-Supervised Training

HMM with Expectation-Maximization (EM)

Need:

- Large **raw** corpus ← know how to get this
- Tag dictionary

[Kupiec, 1992]
[Merialdo, 1994]

Semi-Supervised Training

HMM with Expectation-Maximization (EM)

Need:

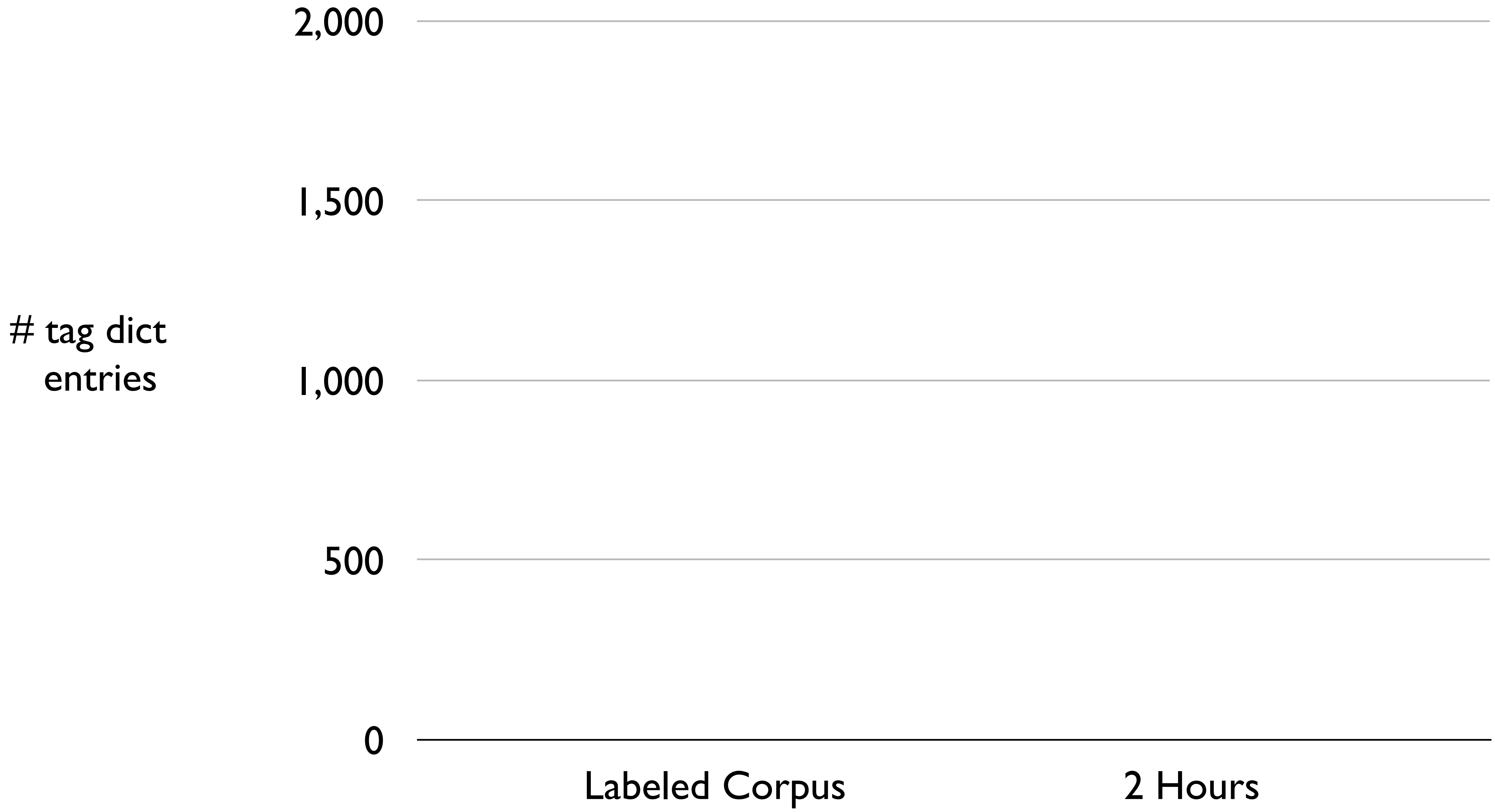
- Large **raw** corpus ← know how to get this
- Tag dictionary ← where is this from?

[Kupiec, 1992]
[Merialdo, 1994]

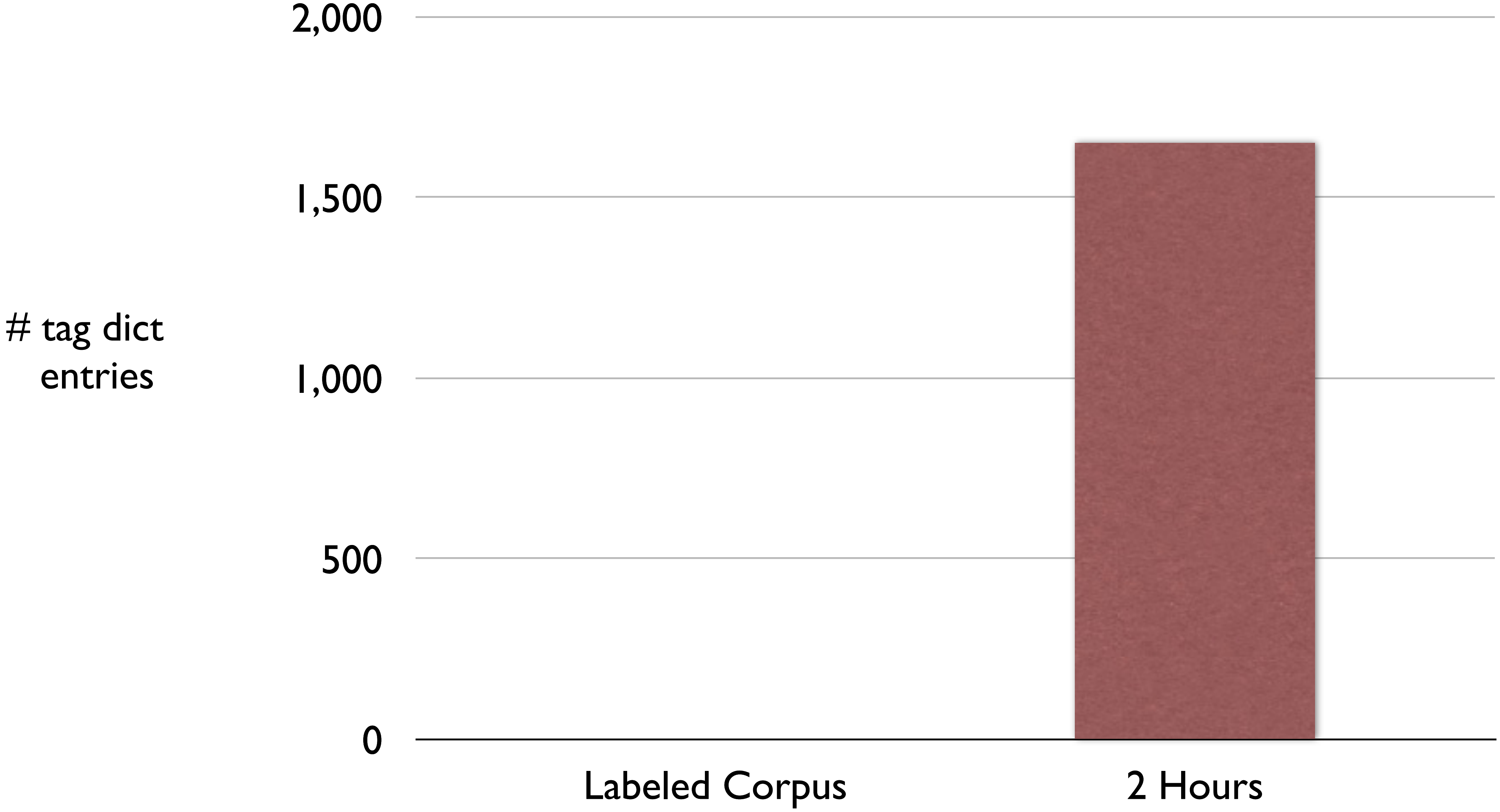
A **Real** Tag Dictionary

# tag dict
entries

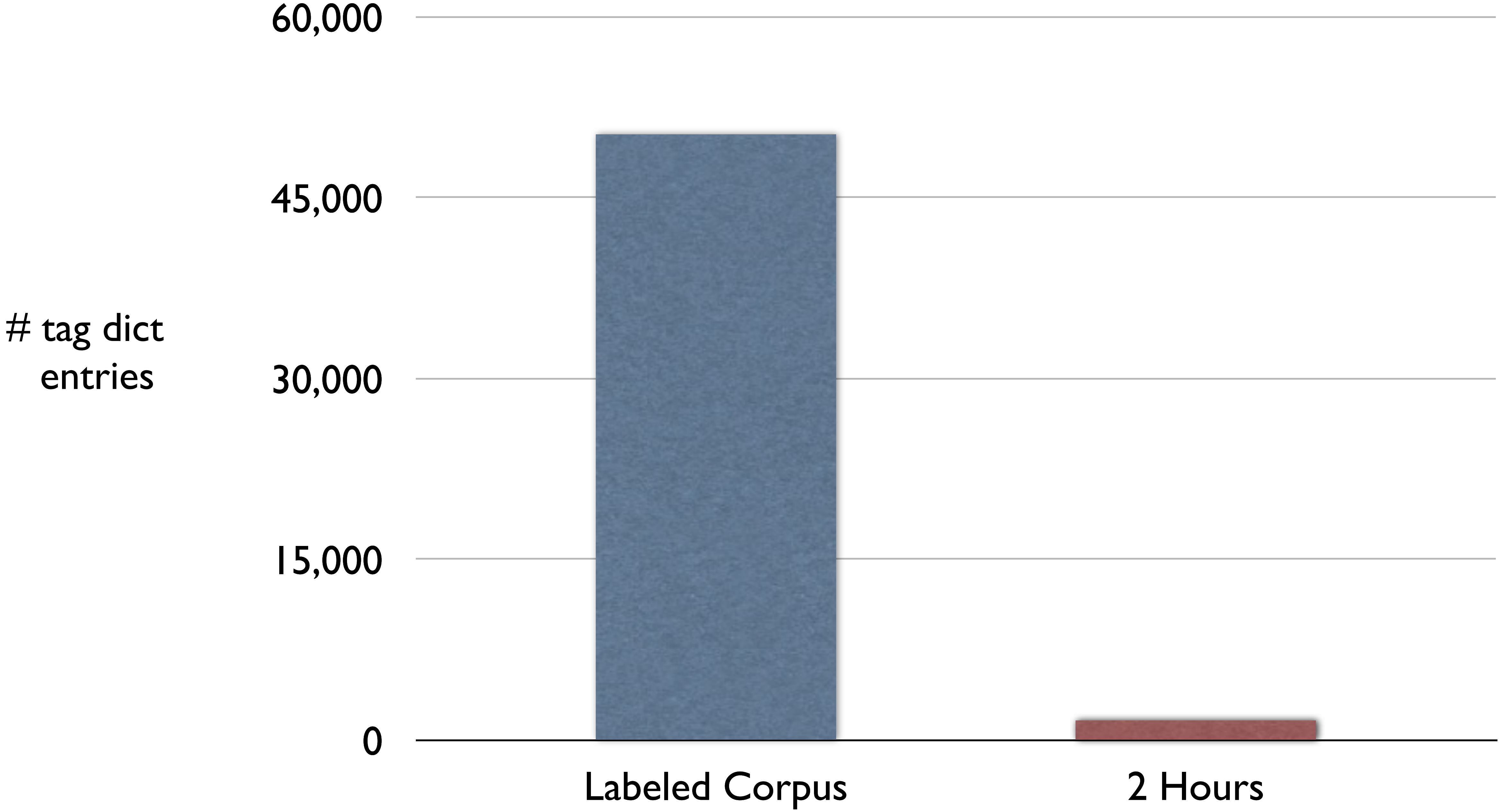
A **Real** Tag Dictionary



A **Real** Tag Dictionary



A **Real** Tag Dictionary



A **Real** Tag Dictionary

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

A **Real** Tag Dictionary

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

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

Our Approach

annotation

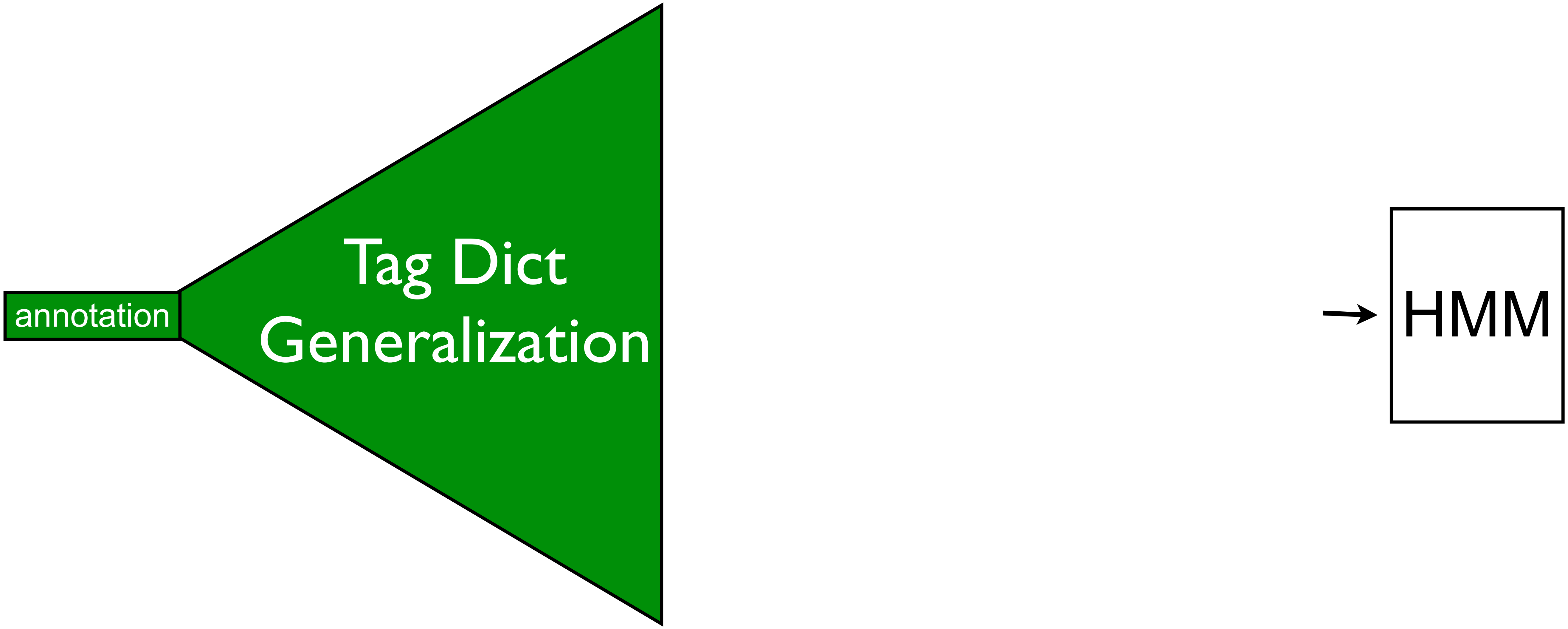
HMM

Our Approach

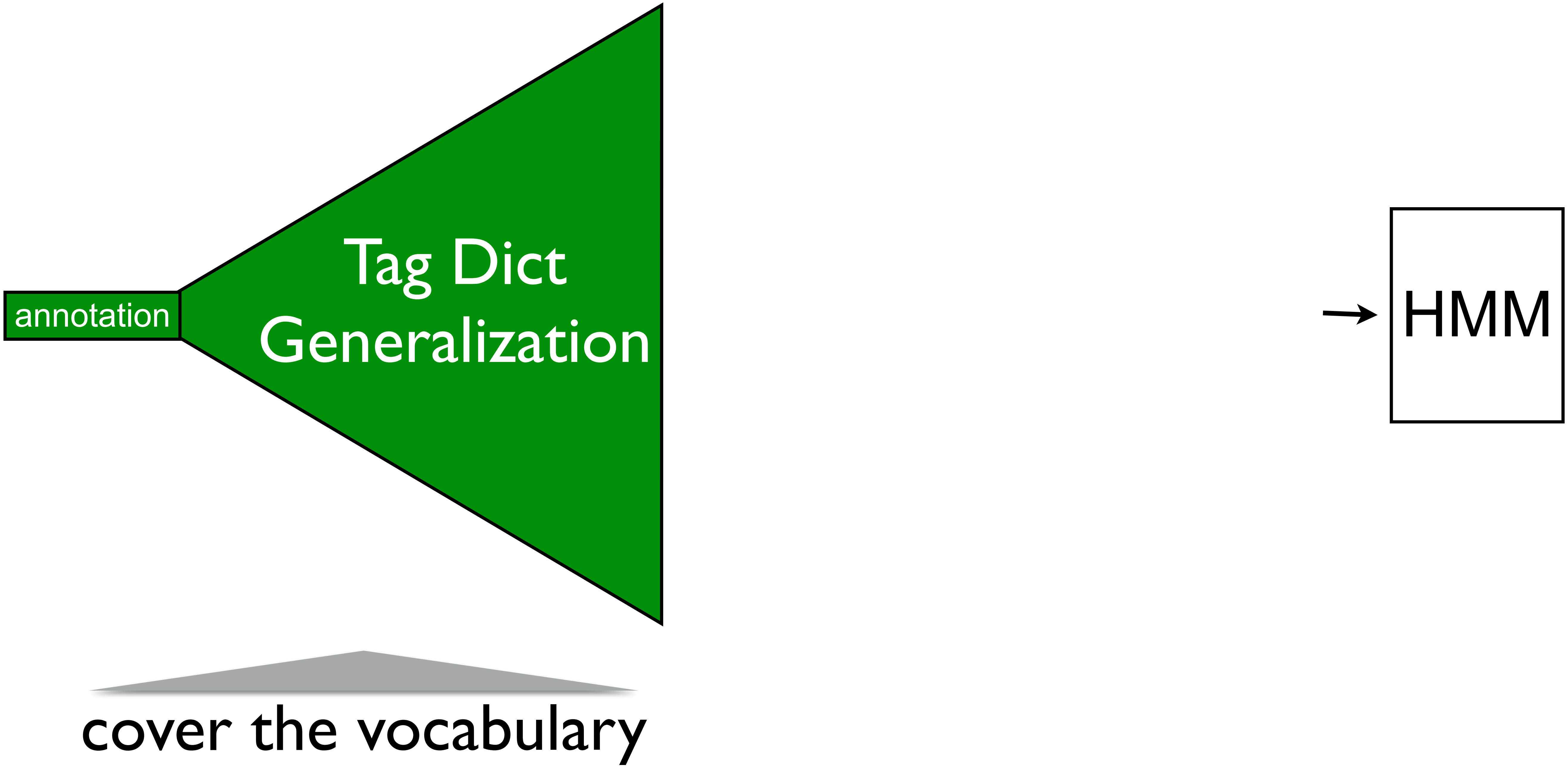
annotation

HMM

Our Approach



Our Approach



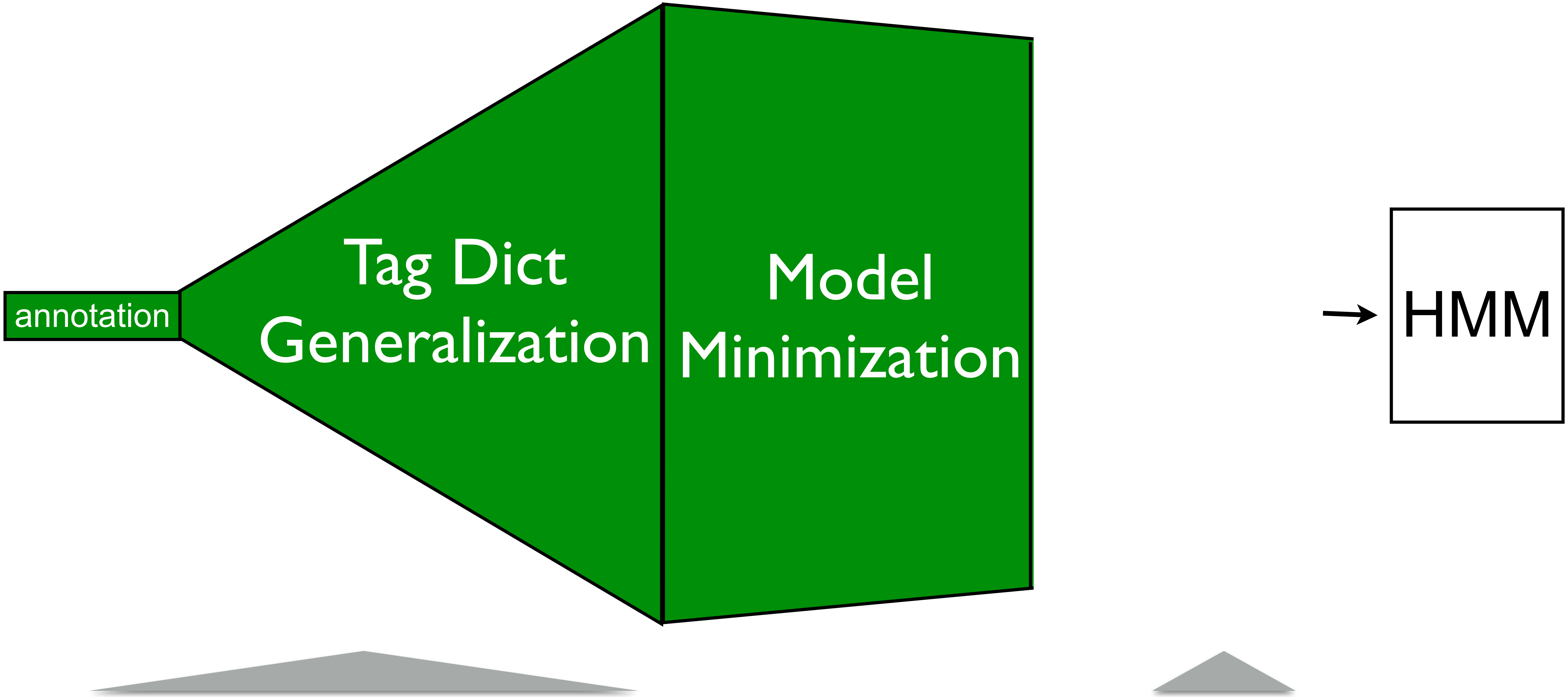
annotation

Tag Dict Generalization

HMM

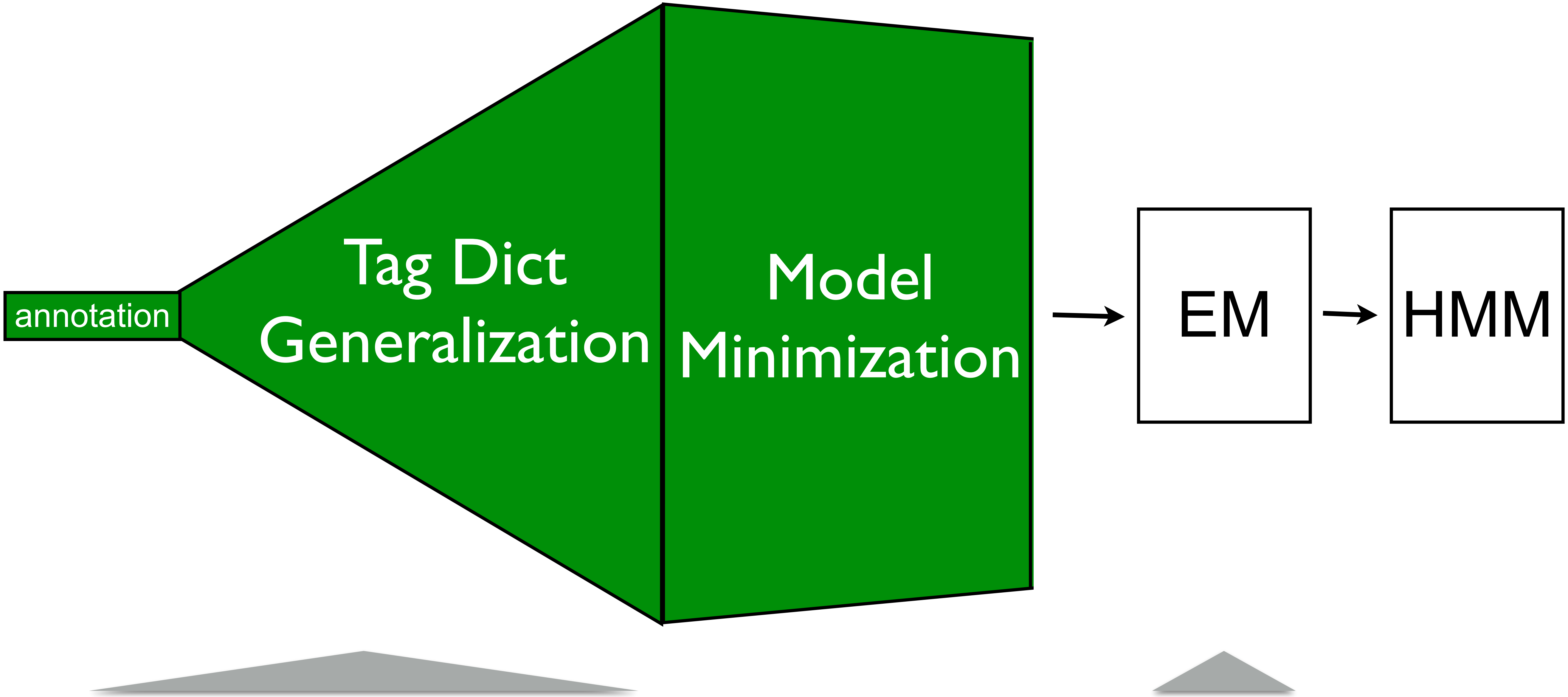
cover the vocabulary

Our Approach



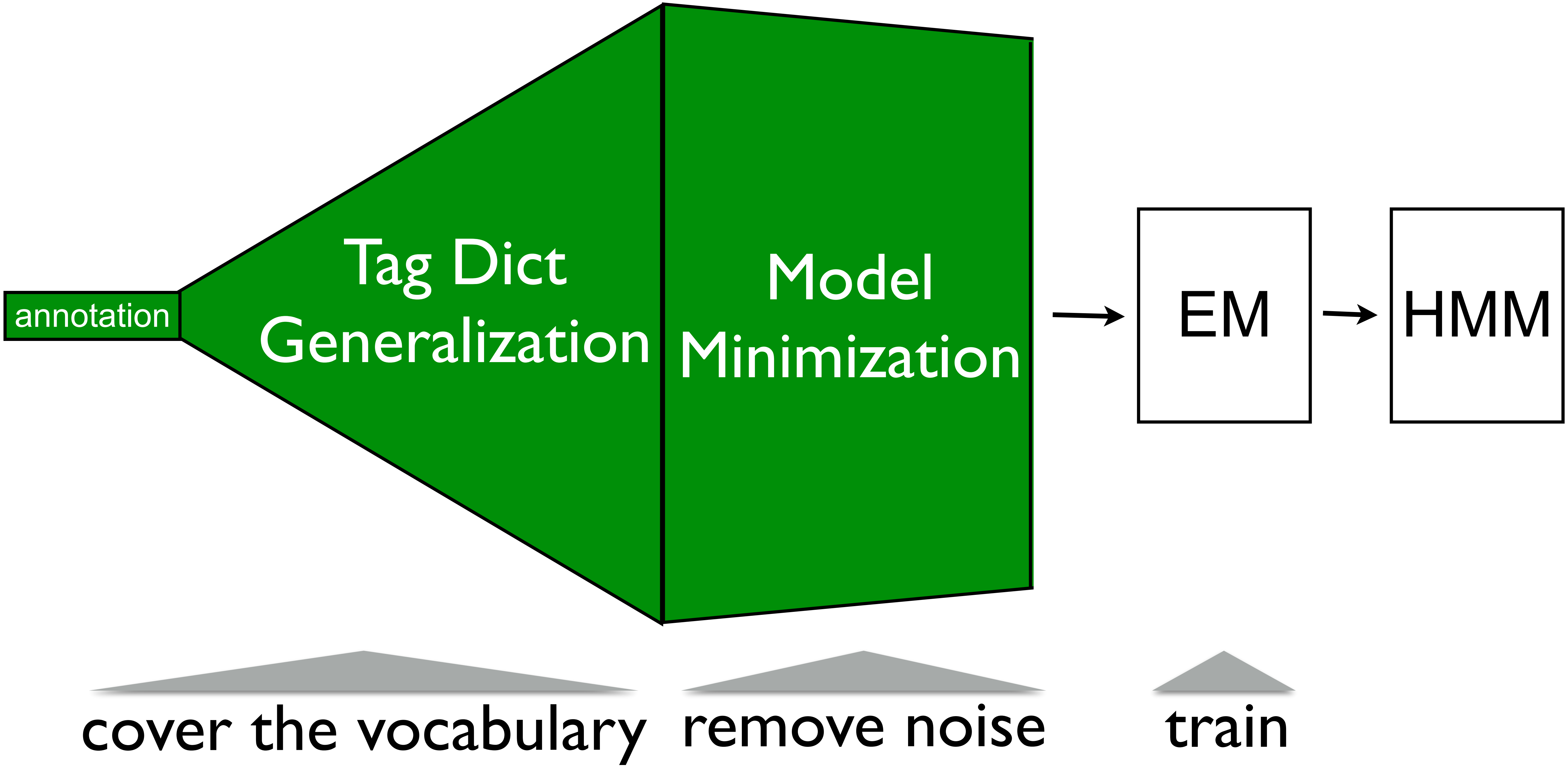
cover the vocabulary remove noise

Our Approach

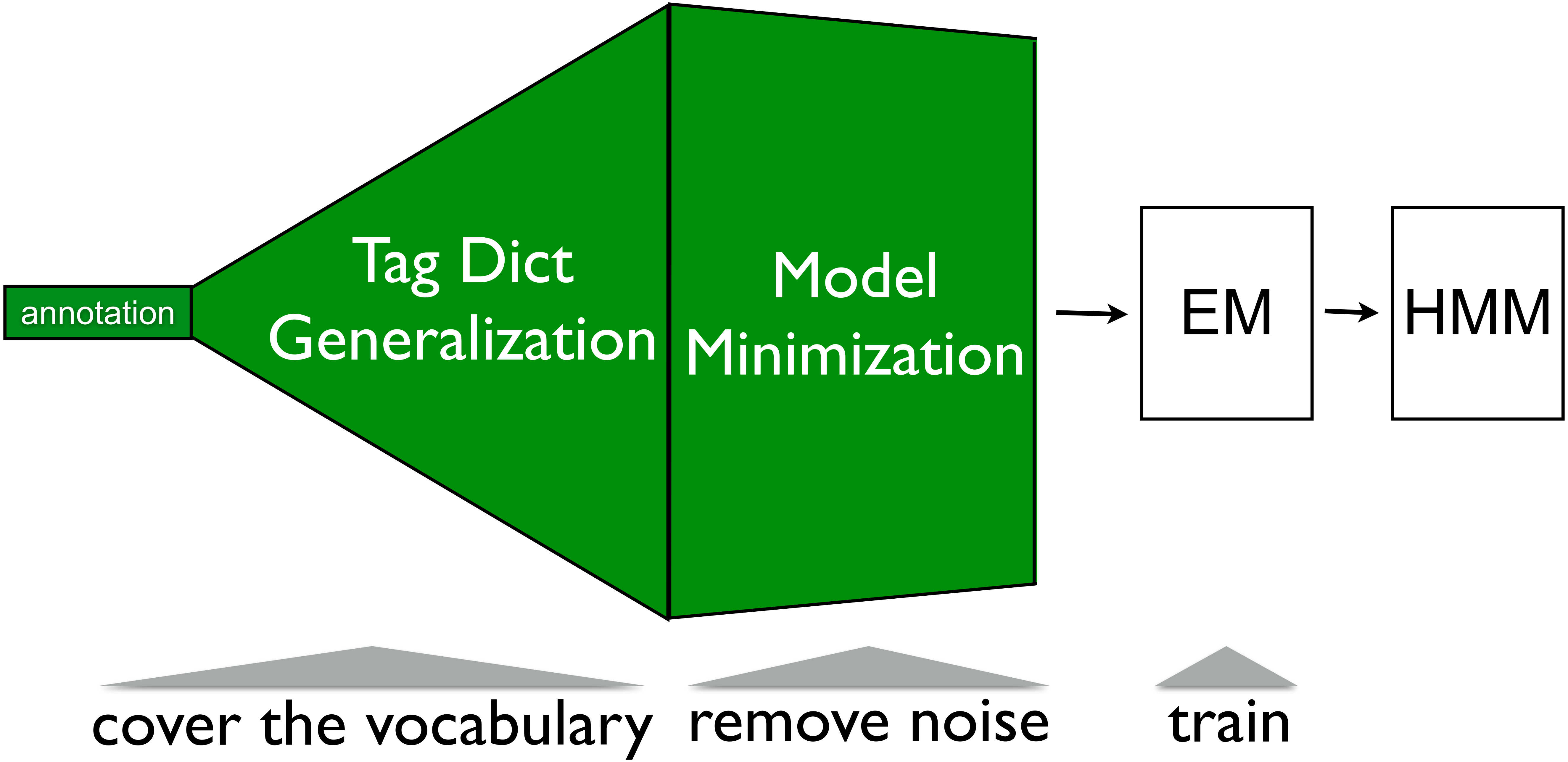


cover the vocabulary remove noise

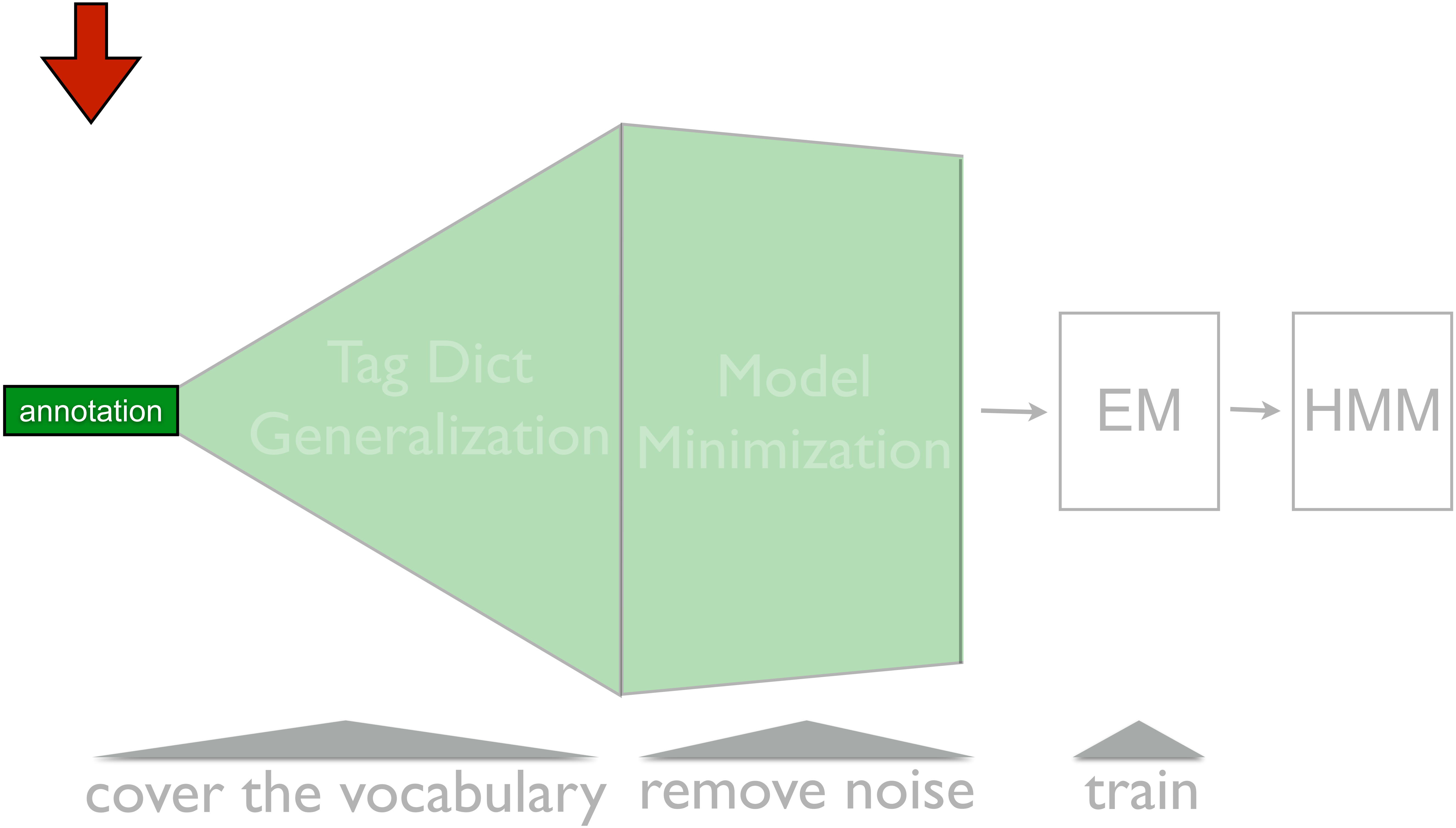
Our Approach



Our Approach



Our Approach



Collecting Annotations

Task #1

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

Collecting Annotations

Task #1

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

,

the

.

of

to

a

and

⋮

only

can

York

into

after

president

⋮

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**

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 .

Mr. Vinken is chairman of Elsevier N.V. , the Dutch publishing group .

⋮

Collecting Annotations

Task #2

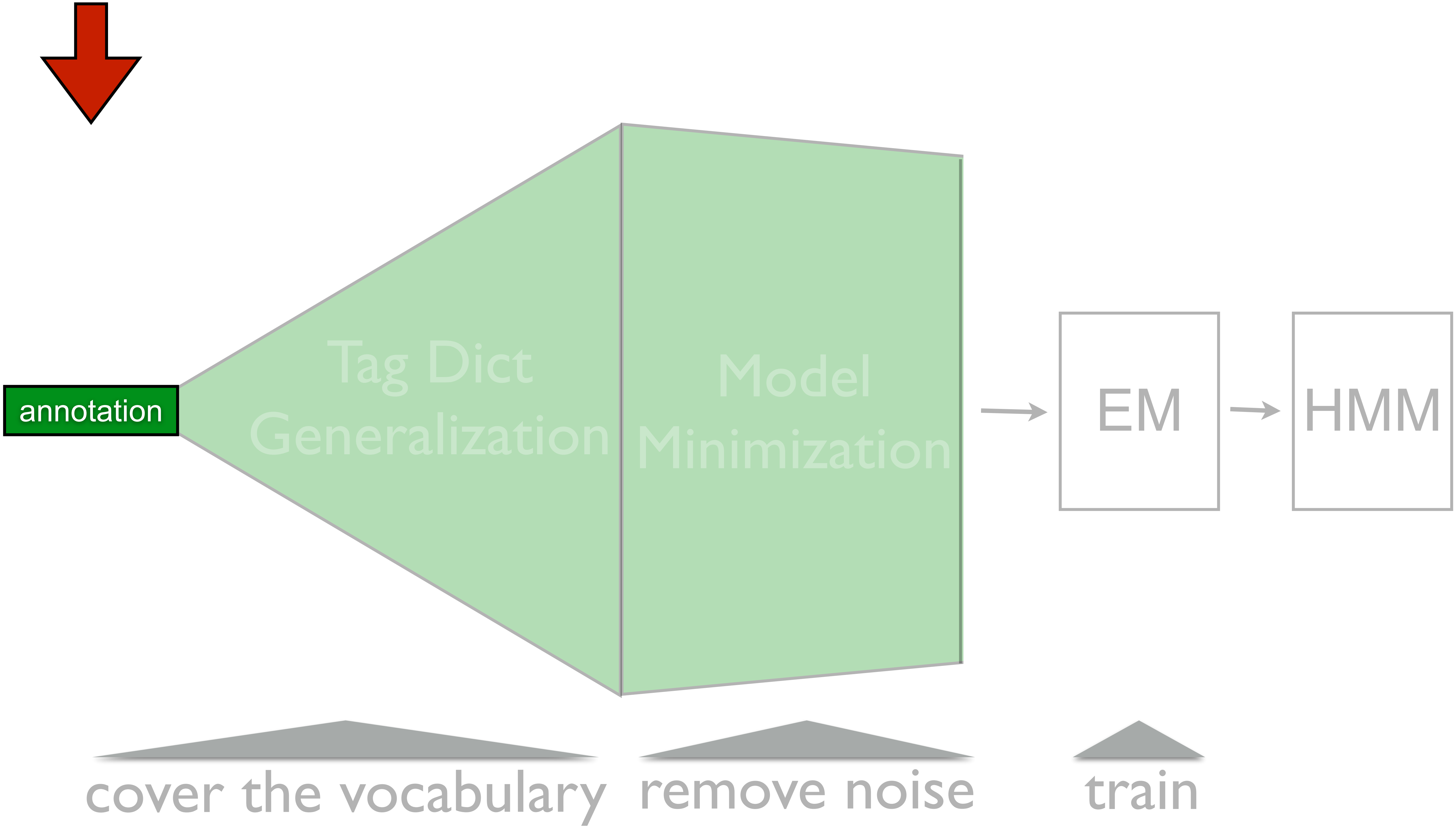
Up to 4 hours to annotate **full sentences**

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	.

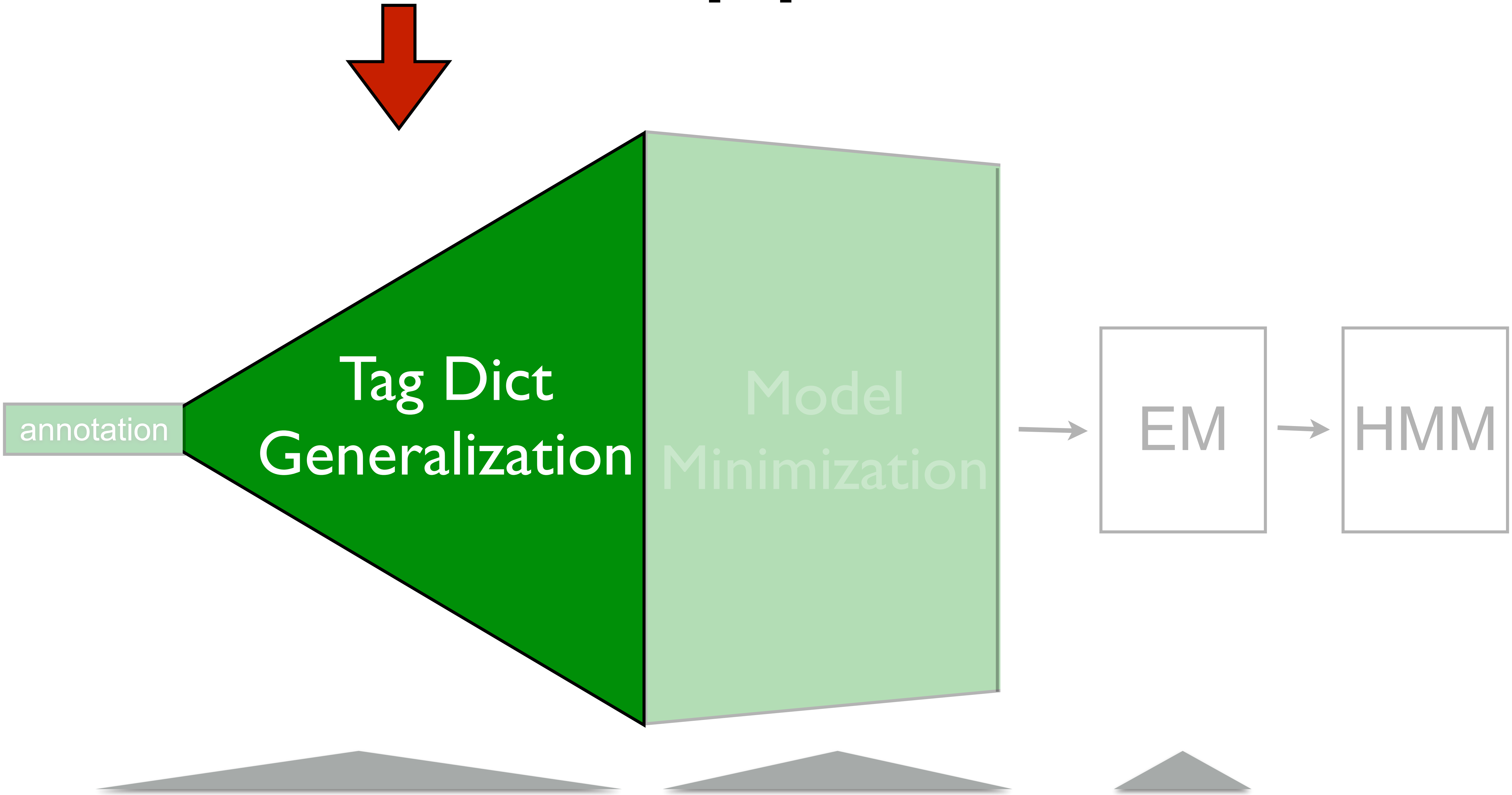
Mr.	Vinken	is	chairman	of	Elsevier	N.V.	,	the	Dutch	publishing	group	.
NNP	NNP	VB	NN	IN	NNP	NNP	,	DT	JJ	JJ	NN	.

⋮

Our Approach



Our Approach



annotation

Tag Dict Generalization

Model Minimization

EM

HMM

cover the vocabulary remove noise train

Tag Dict Generalization

These annotations are too sparse!

Tag Dict Generalization

These annotations are too sparse!

➡ Generalize to the entire vocabulary

Tag Dict Generalization

Our strategy: Label Propagation

[Talukdar and Crammer. 2009]

Tag Dict Generalization

Our strategy: Label Propagation

- **Connect** annotations to raw corpus tokens

[Talukdar and Crammer. 2009]

Tag Dict Generalization

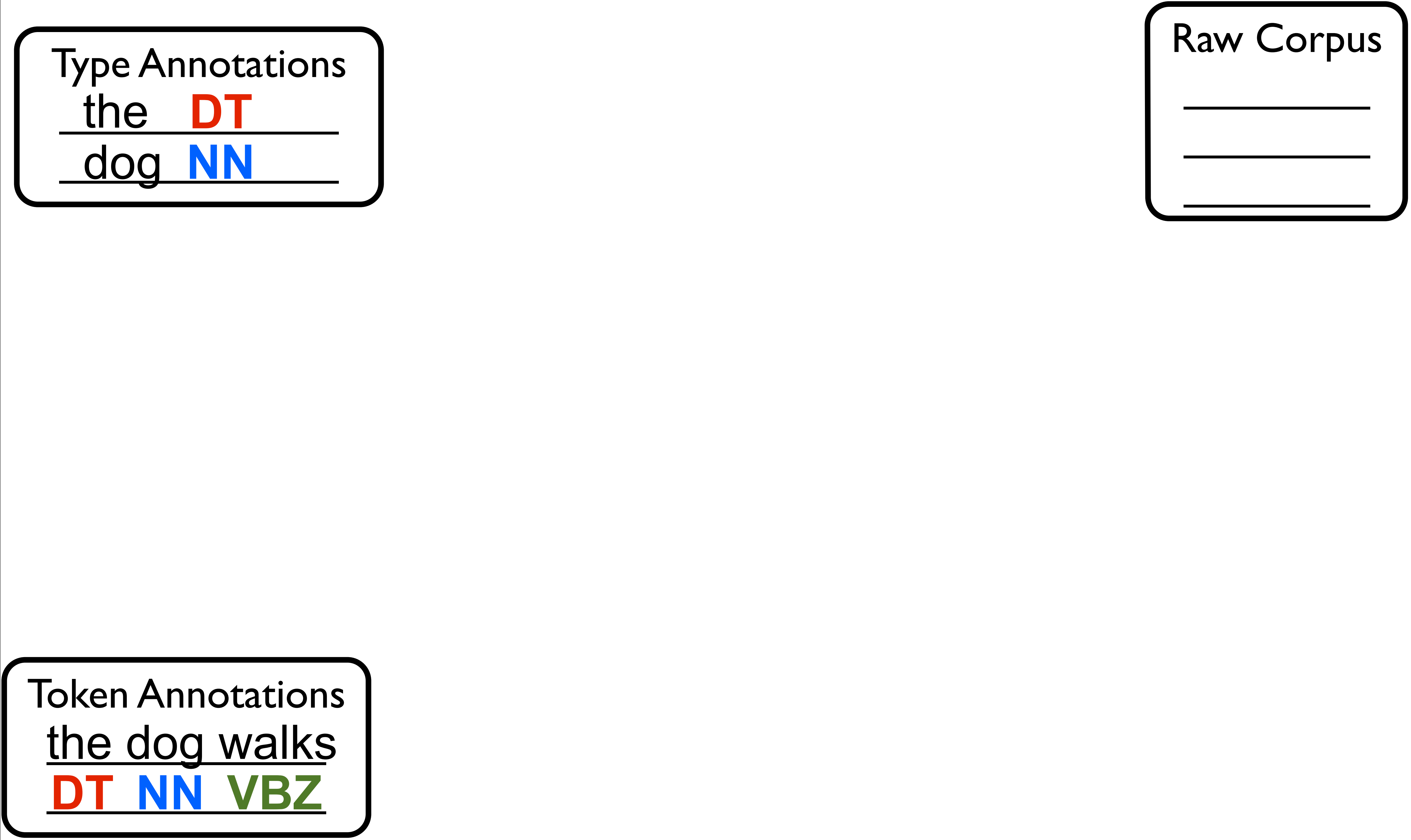
Our strategy: Label Propagation

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

[Talukdar and Crammer. 2009]

Tag Dict Generalization

Tag Dict Generalization



Type Annotations

the DT

dog NN

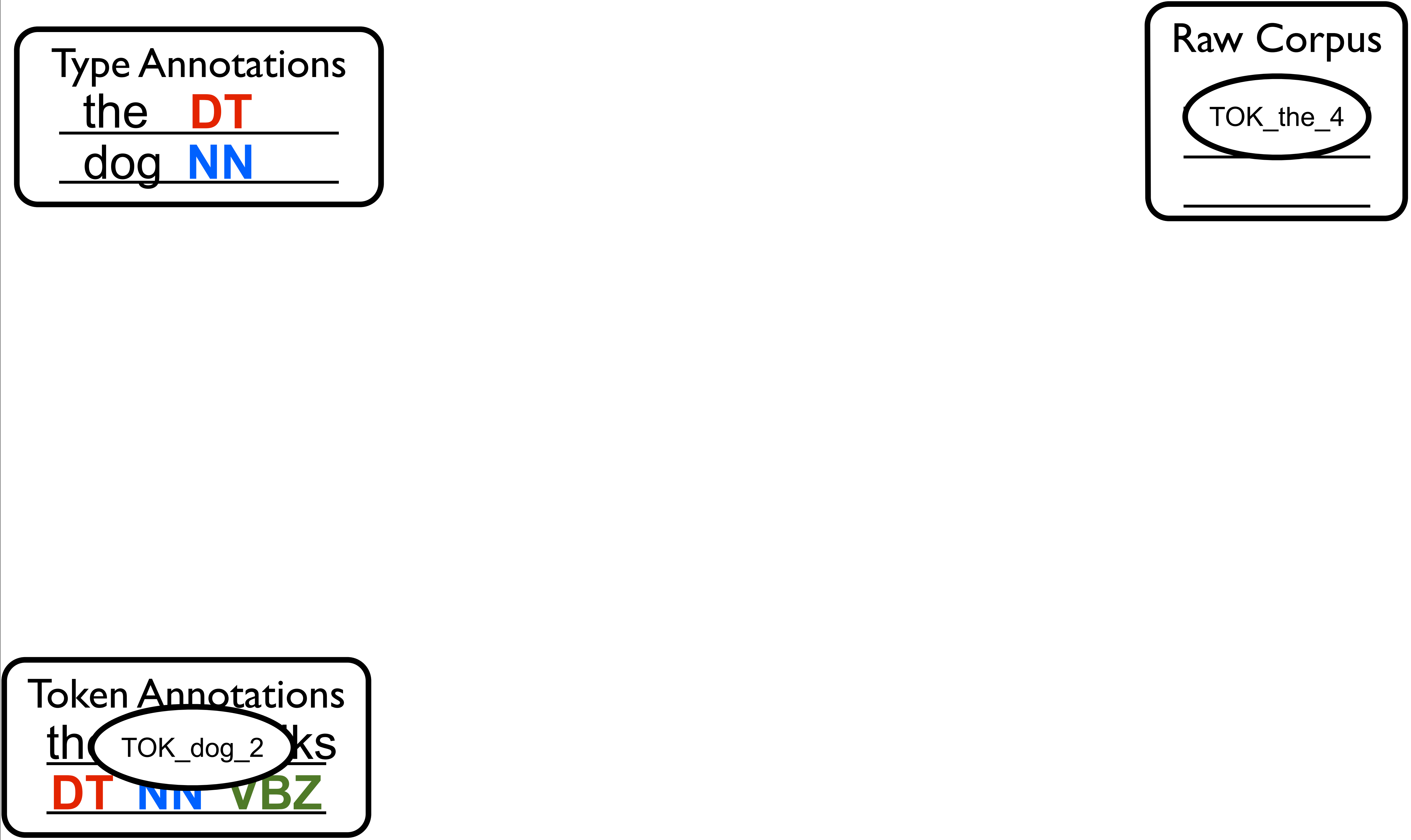
Raw Corpus

Token Annotations

the dog walks

DT NN VBZ

Tag Dict Generalization



Type Annotations

the	DT
dog	NN

Raw Corpus

TOK_the_4

Token Annotations

th TOK_dog_2 ks

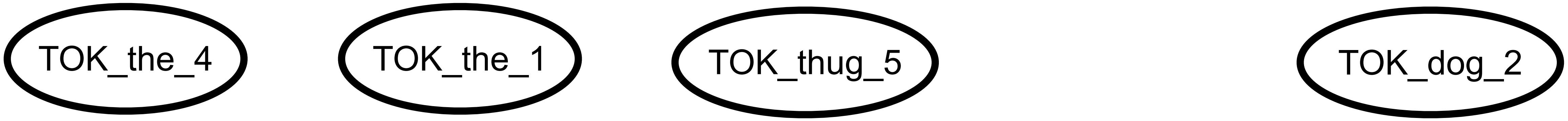
DT NN VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

Raw Corpus



TOK_the_4

TOK_the_1

TOK_thug_5

TOK_dog_2

Token Annotations

the dog walks

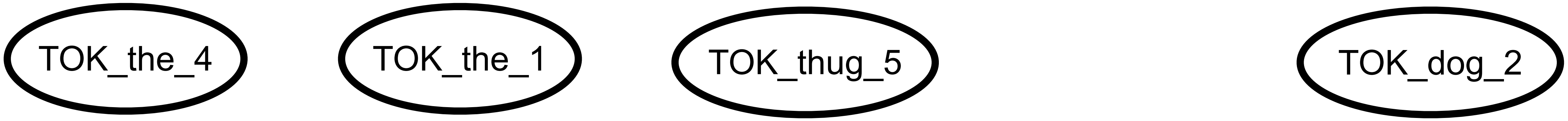
DT NN VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

Raw Corpus


TOK_the_4

TOK_the_1

TOK_thug_5

TOK_dog_2

Token Annotations

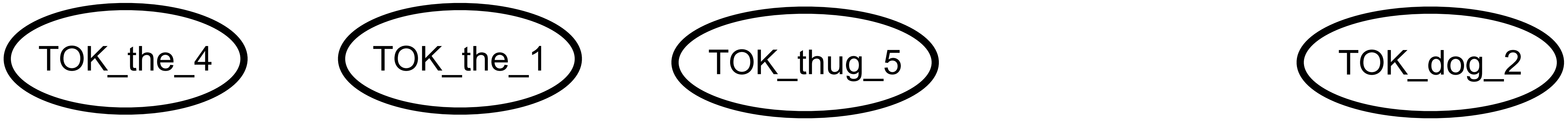
the	dog	walks
DT	NN	VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

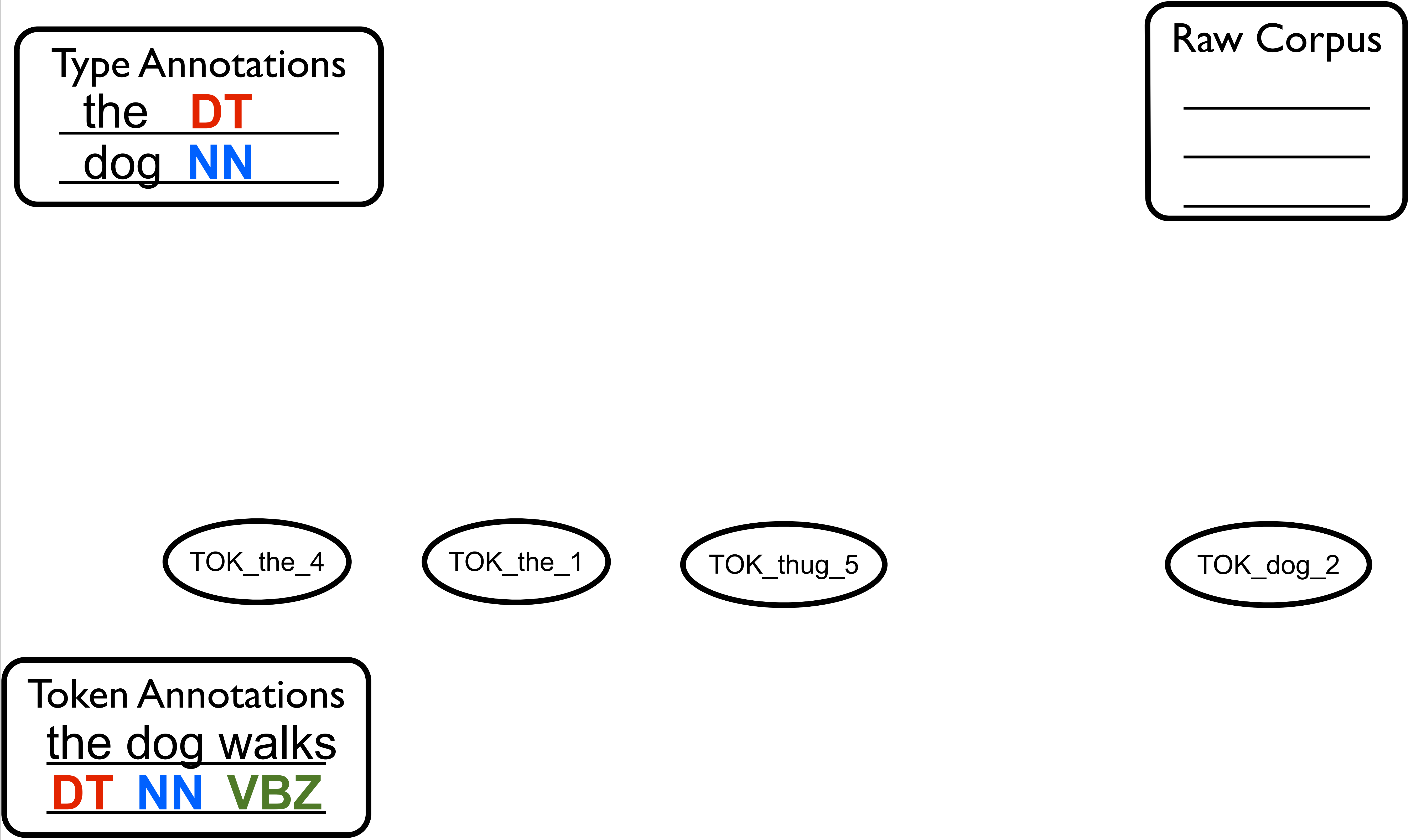
Raw Corpus



Token Annotations

the dog walks
DT NN VBZ

Tag Dict Generalization



Type Annotations

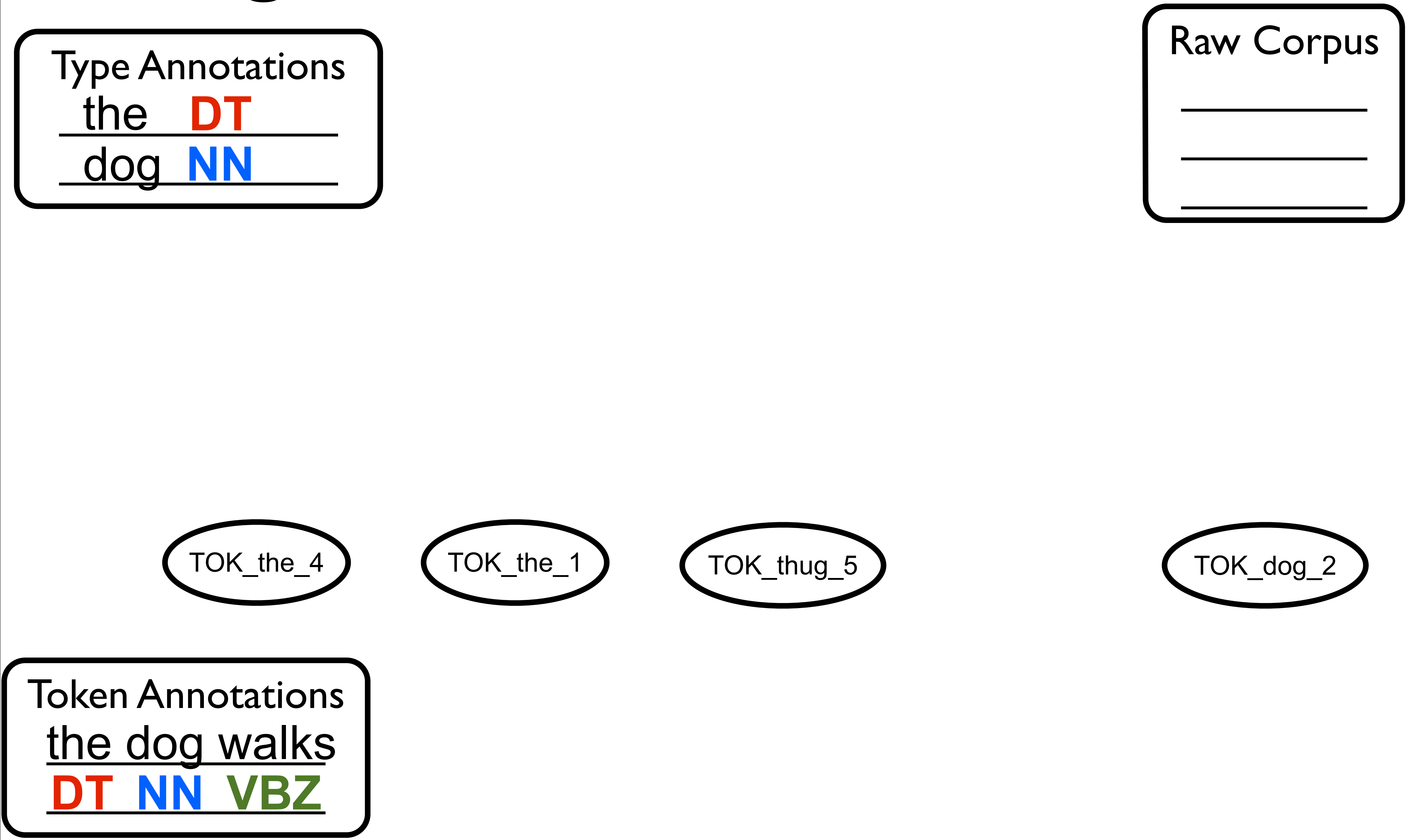
the	DT
dog	NN

Raw Corpus

Token Annotations

the	dog	walks
DT	NN	VBZ

Tag Dict Generalization



Type Annotations

the	DT
dog	NN

Raw Corpus

TOK_the_4

TOK_the_1

TOK_thug_5

TOK_dog_2

Token Annotations

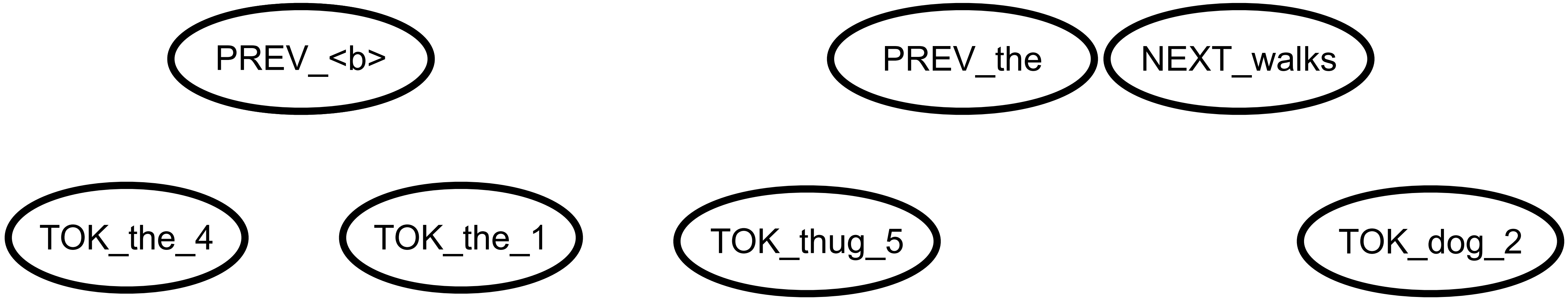
the	dog	walks
DT	NN	VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

Raw Corpus



Token Annotations

the	dog	walks
DT	NN	VBZ

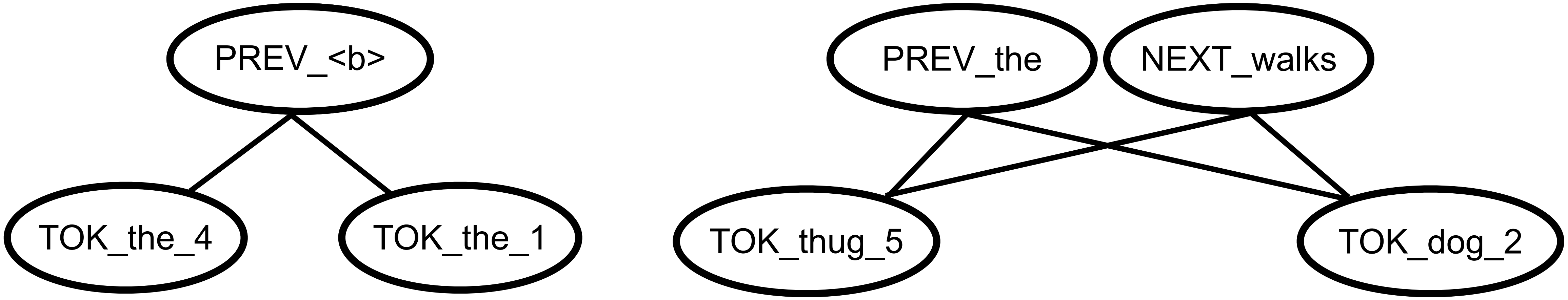
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

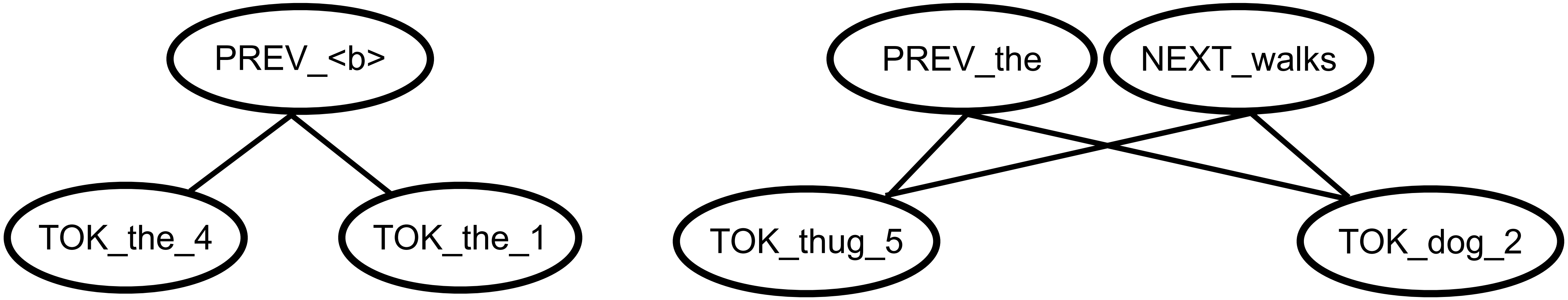
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

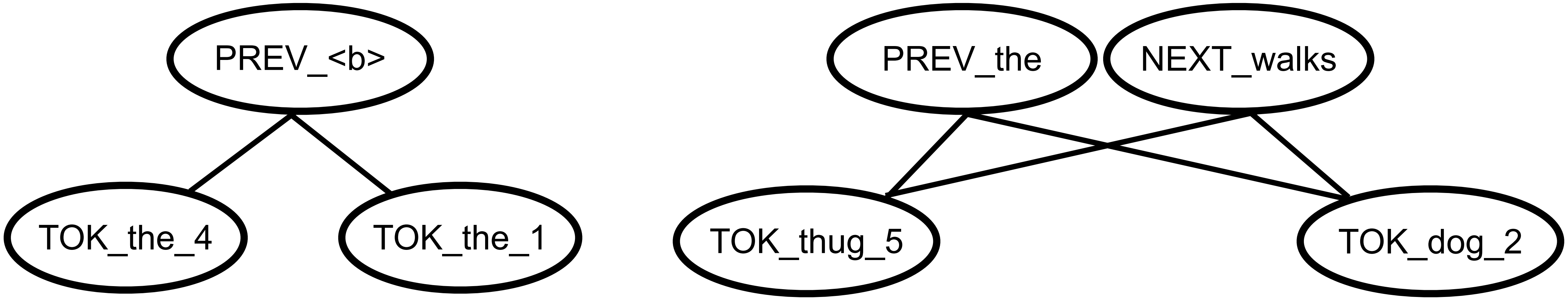
Tag Dict Generalization

Type Annotations

th TYPE_dog

dog NN

Raw Corpus



Token Annotations

the dog walks

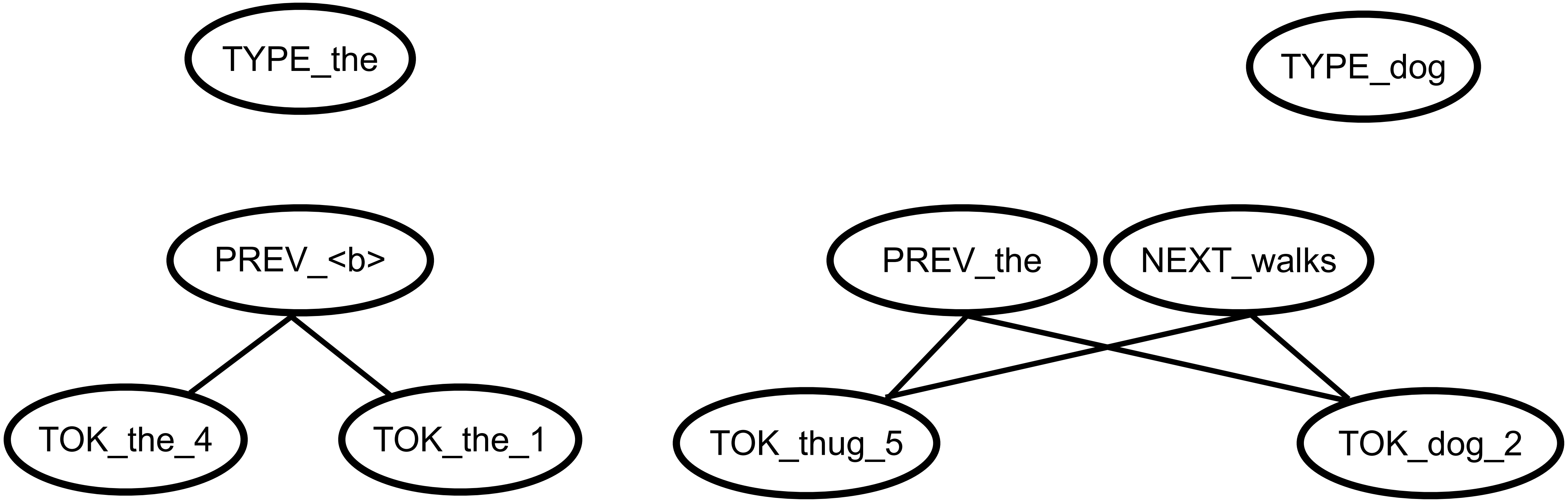
DT NN VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

Raw Corpus



TYPE_the

TYPE_dog

PREV_<b>

PREV_the

NEXT_walks

TOK_the_4

TOK_the_1

TOK_thug_5

TOK_dog_2

Token Annotations

the dog walks

DT NN VBZ

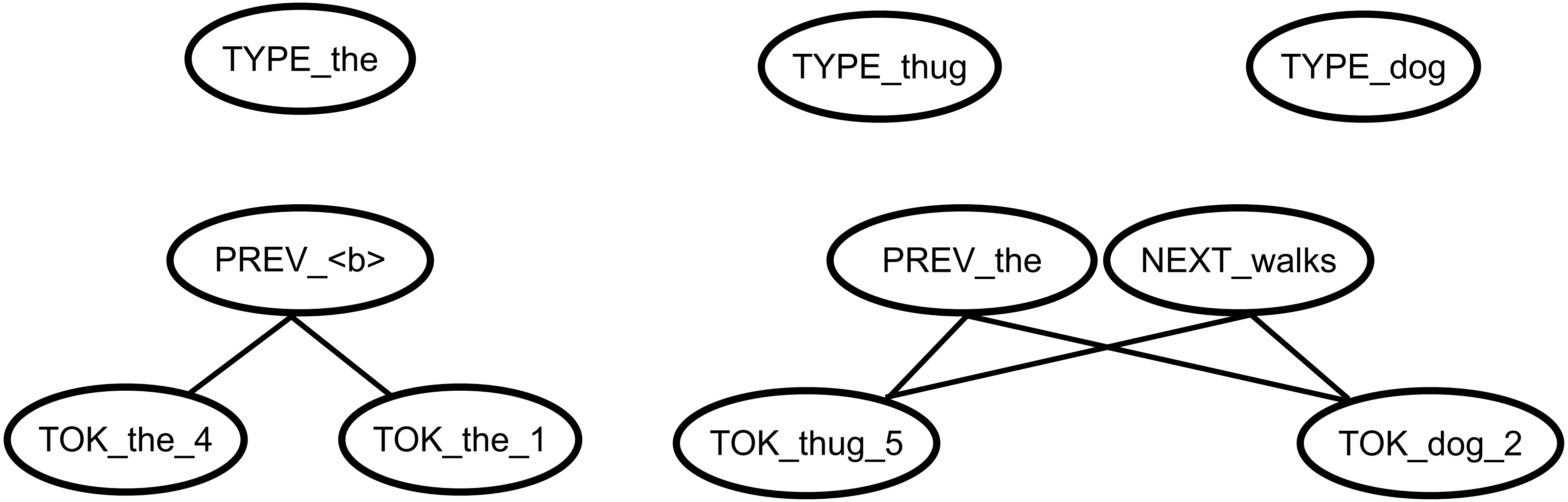
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

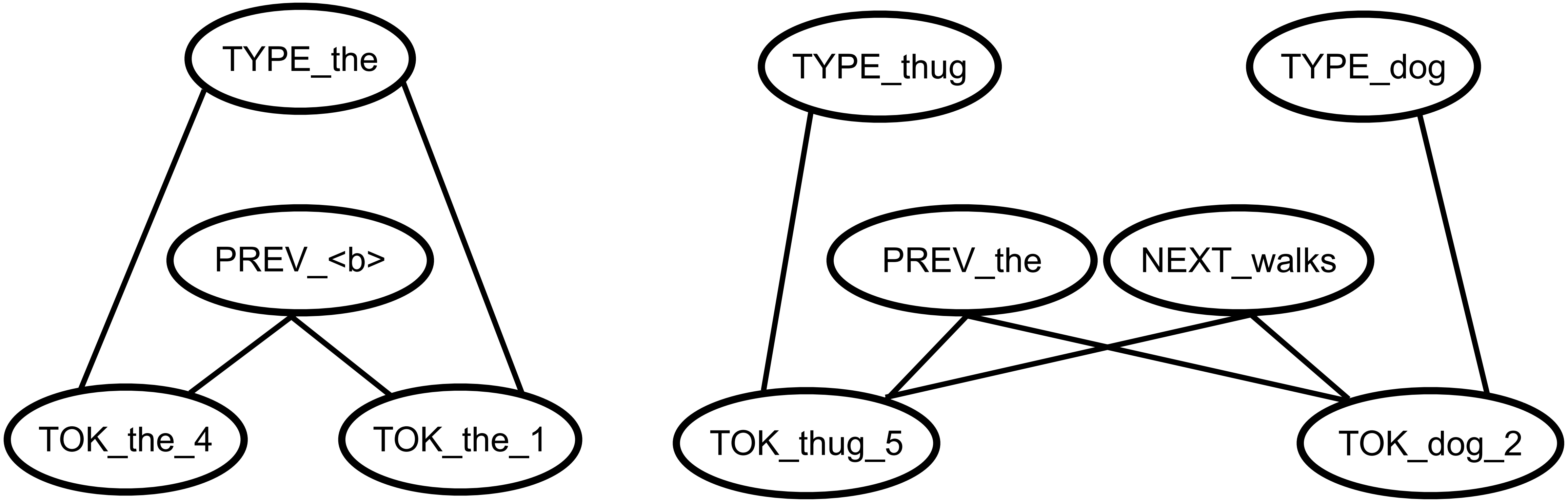
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



TYPE_the

PREV_<b>

TOK_the_4

TOK_the_1

TYPE_thug

PREV_the

NEXT_walks

TYPE_dog

TOK_thug_5

TOK_dog_2

Token Annotations

the dog walks

DT NN VBZ

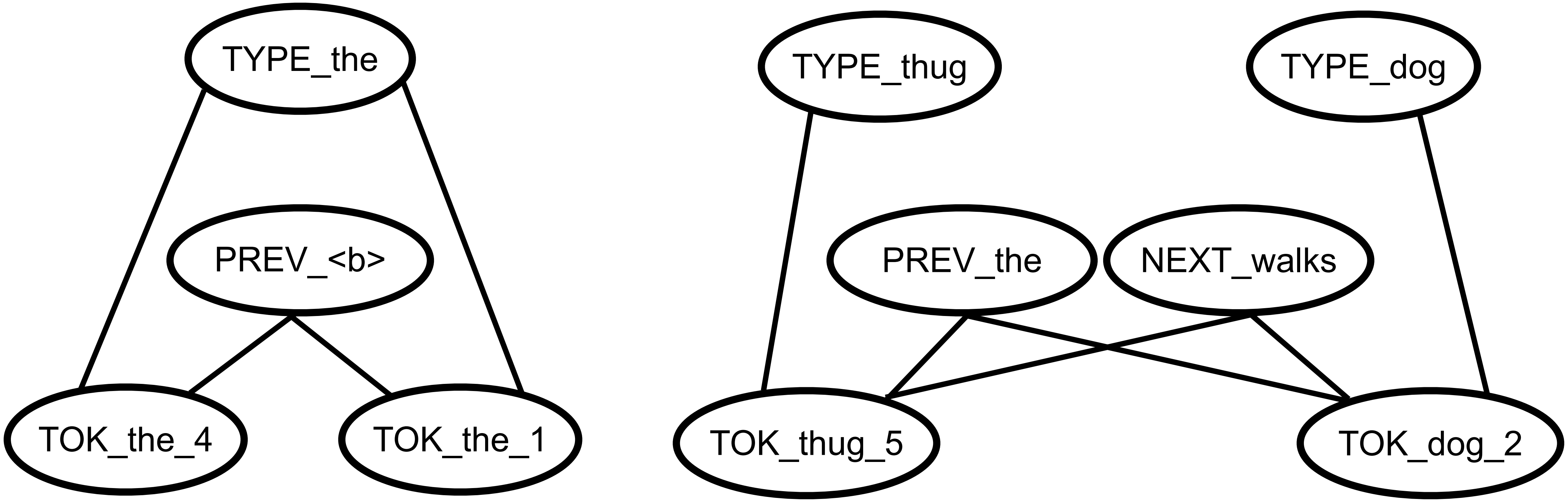
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

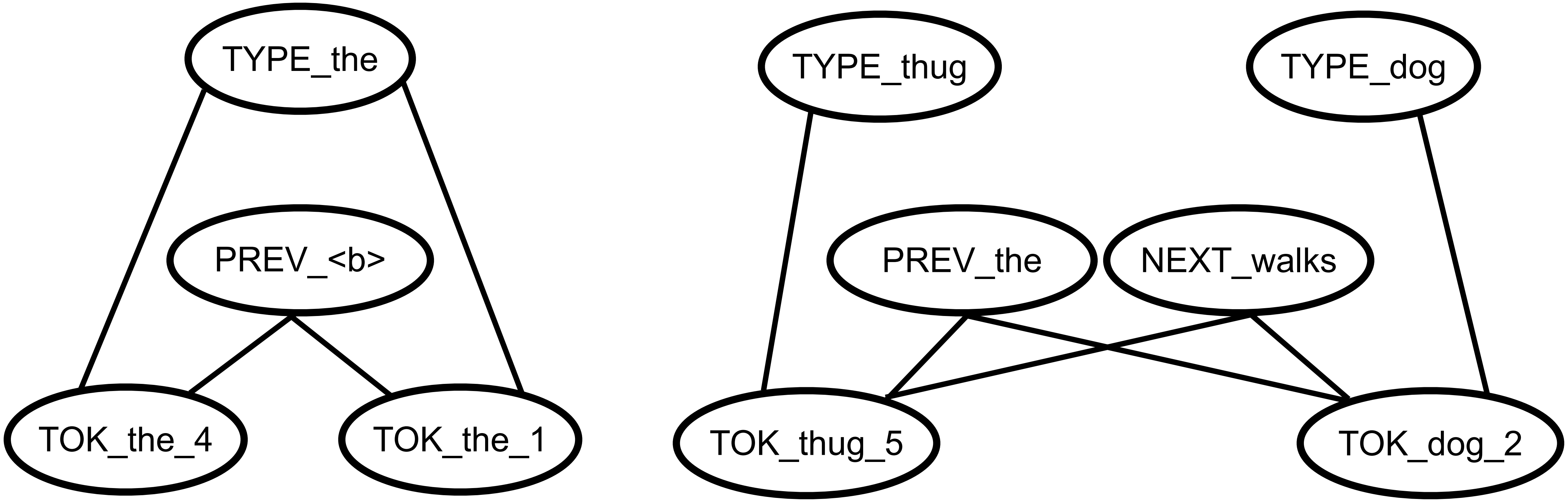
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



Token Annotations

the dog walks

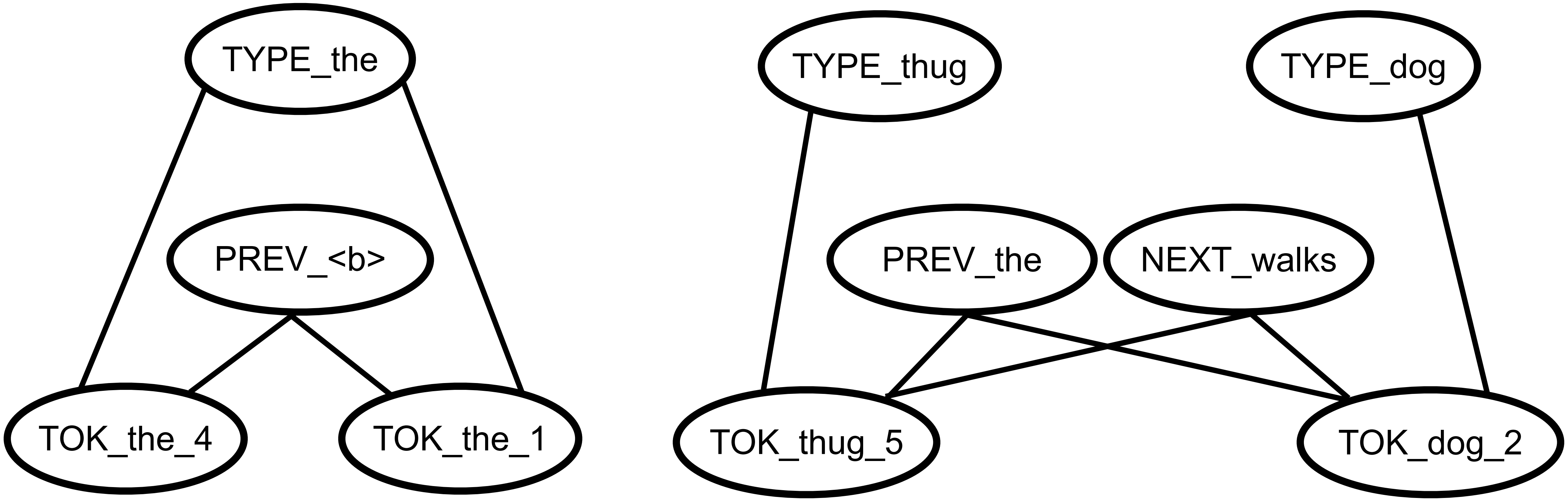
DT NN VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

Raw Corpus



Token Annotations

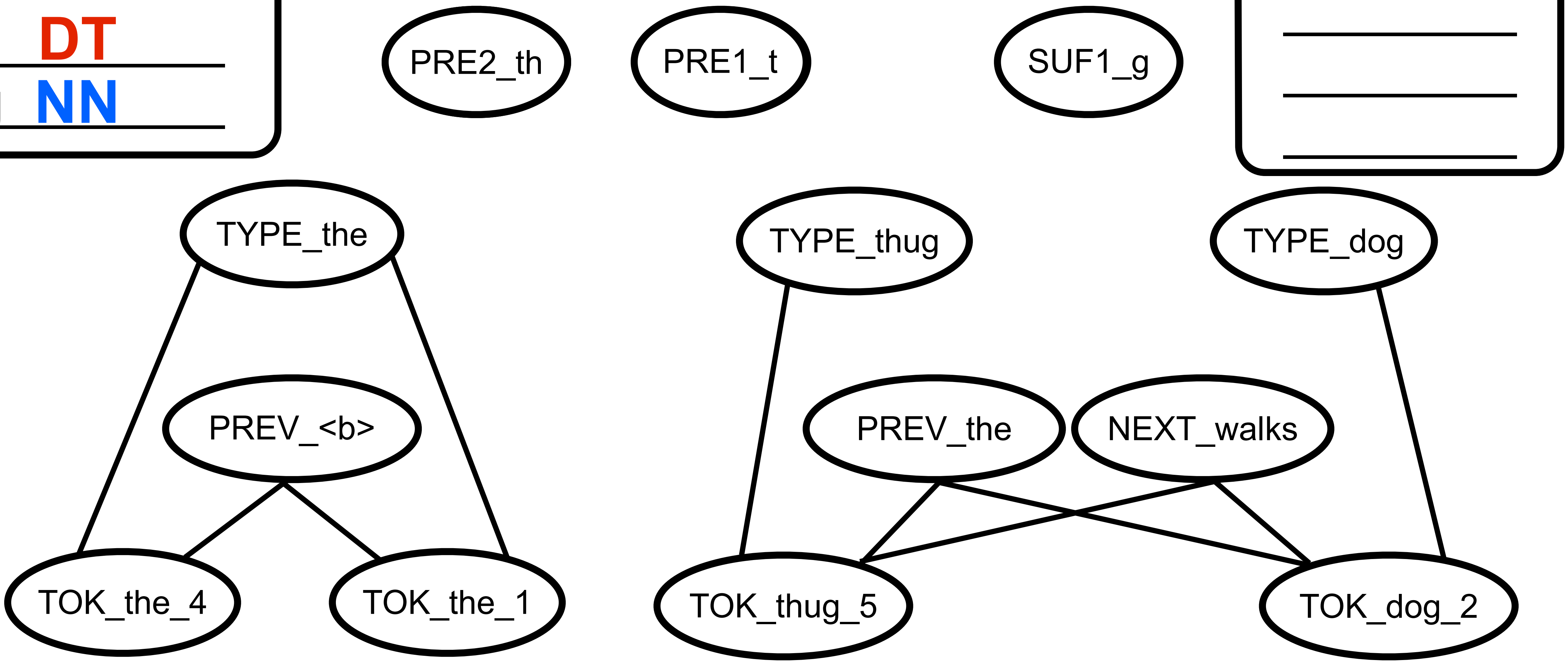
the dog walks
DT NN VBZ

Tag Dict Generalization

Type Annotations

the DT

dog NN



Raw Corpus

Token Annotations

the dog walks

DT NN VBZ

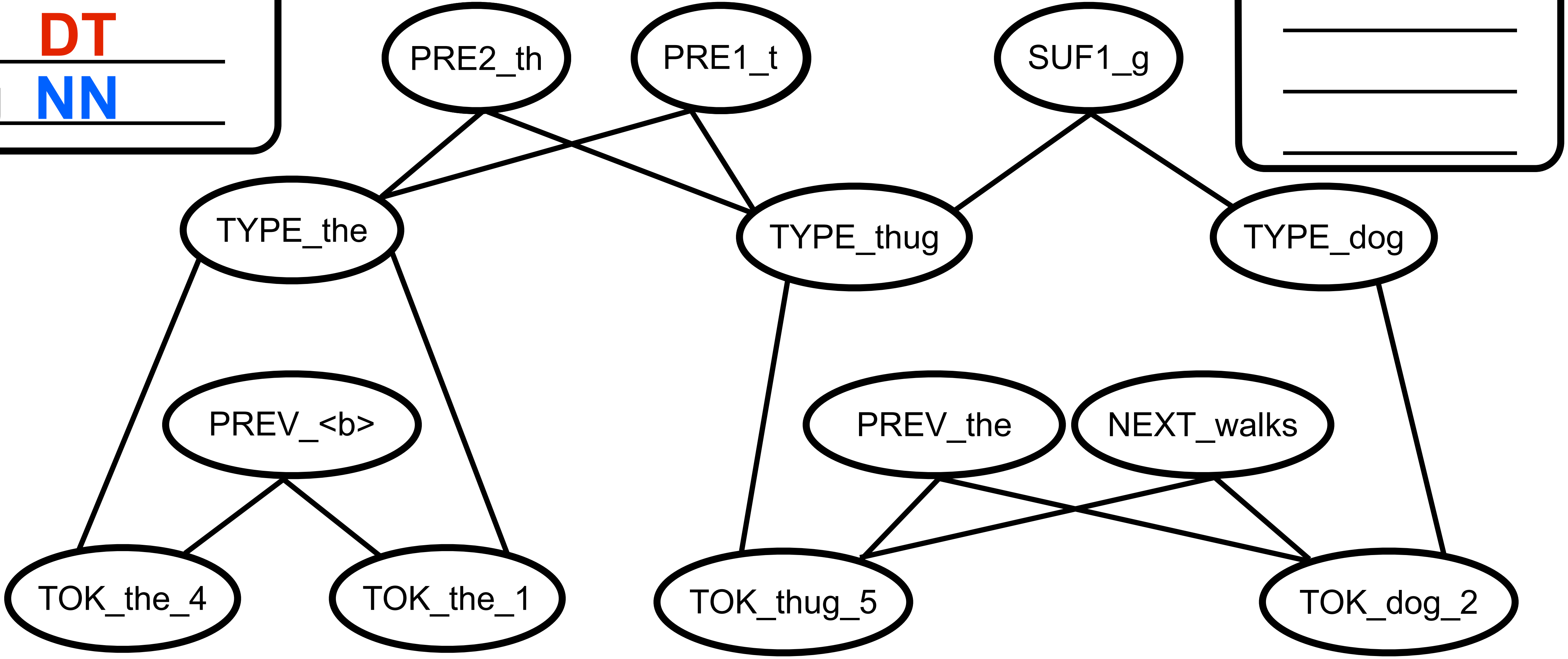
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus

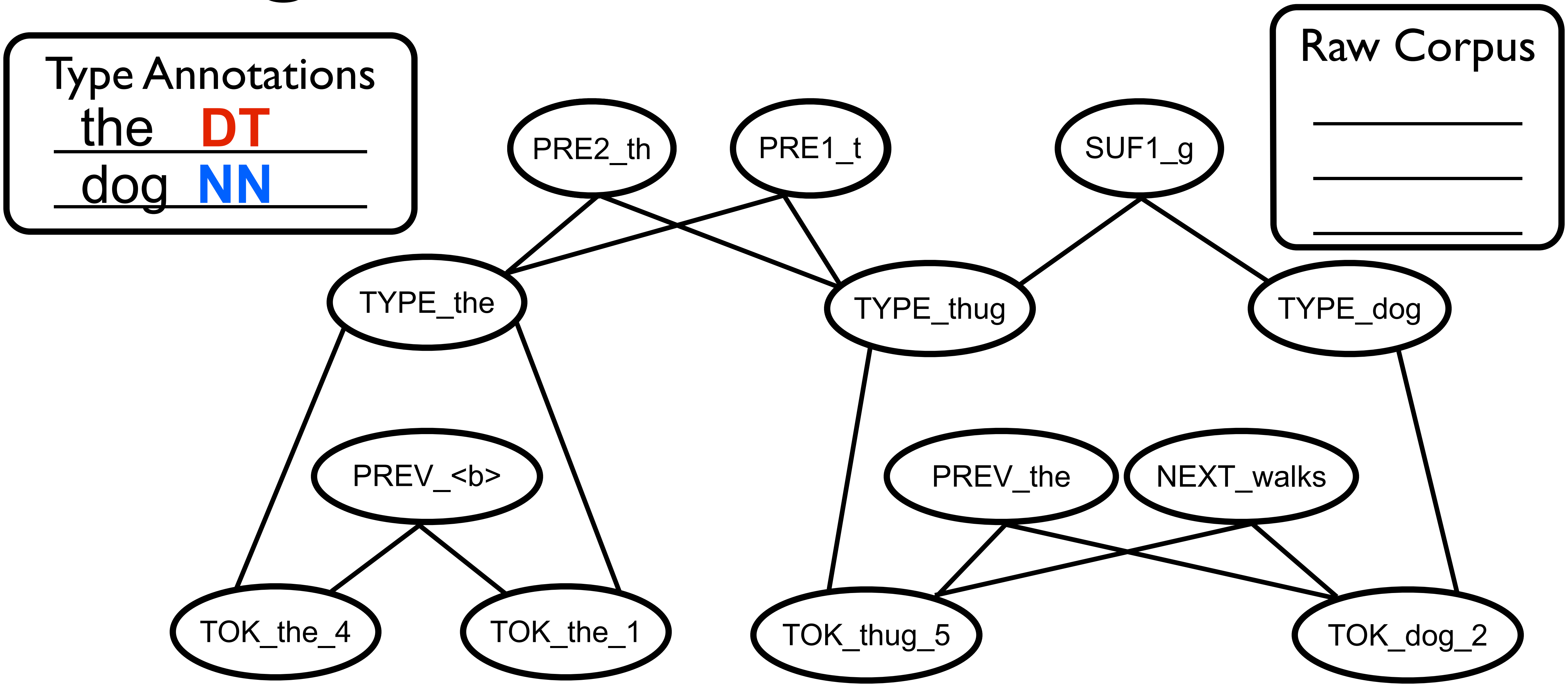


Token Annotations

the dog walks

DT NN VBZ

Tag Dict Generalization



Token Annotations

the dog walks

DT NN VBZ

Any arbitrary features could be used

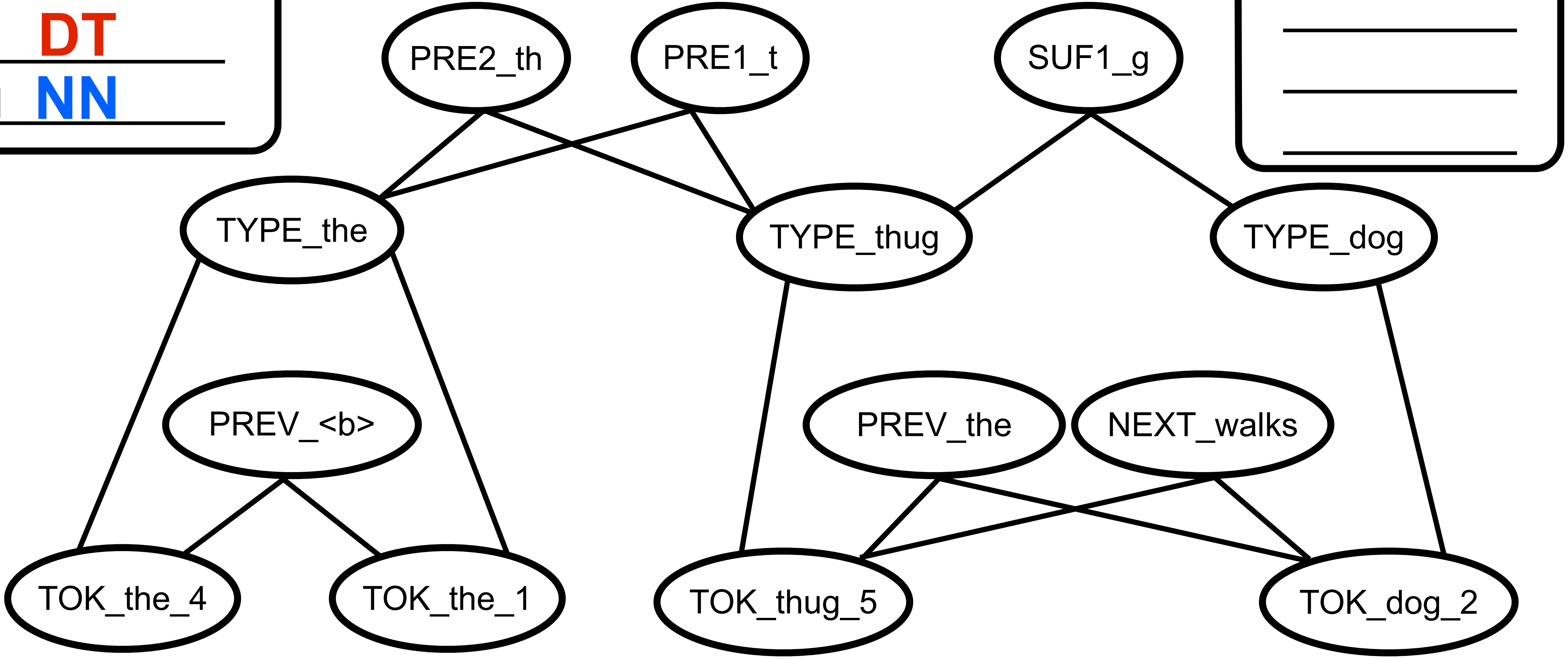
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



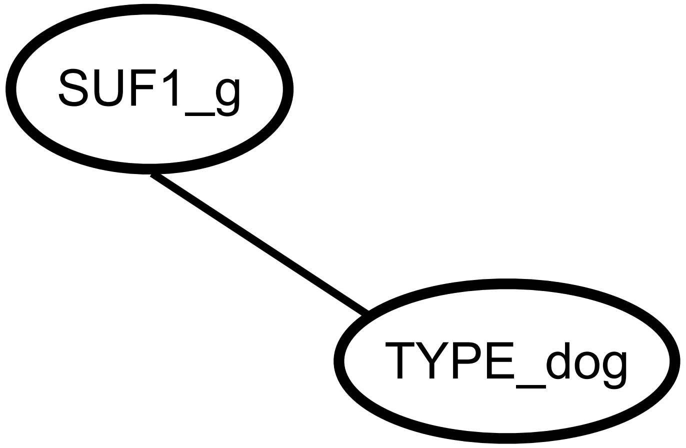
Token Annotations

the dog walks

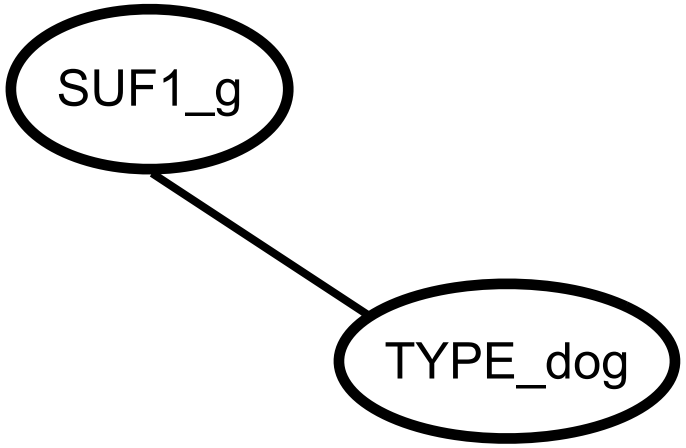
DT NN VBZ

Any arbitrary features could be used

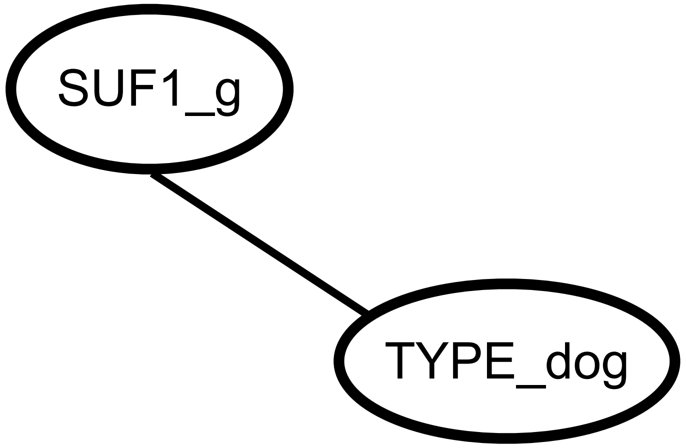
Tag Dict Generalization



Tag Dict Generalization



Tag Dict Generalization



Tag Dict Generalization

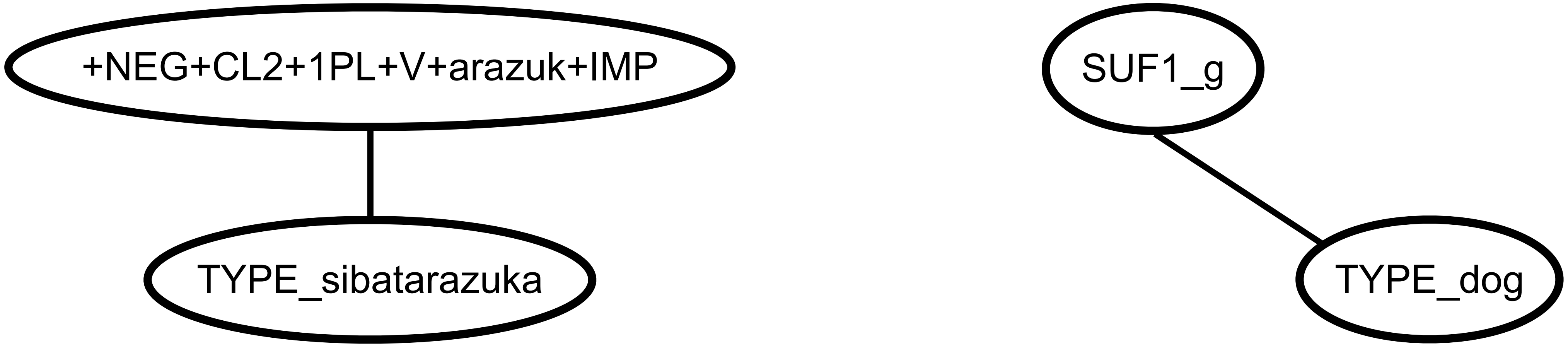
SUF1_g
TYPE_sibatarazuka
TYPE_dog

Tag Dict Generalization

SUF1_g
TYPE_sibatarazuka
TYPE_dog

Finite-State Transducer (FST)

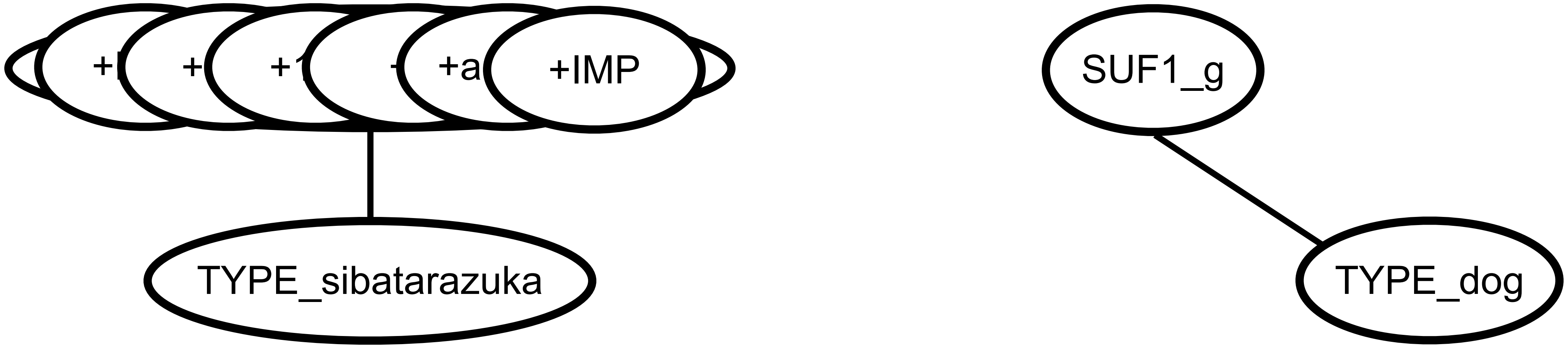
Tag Dict Generalization



Finite-State Transducer (FST)

- Generates morphological analysis

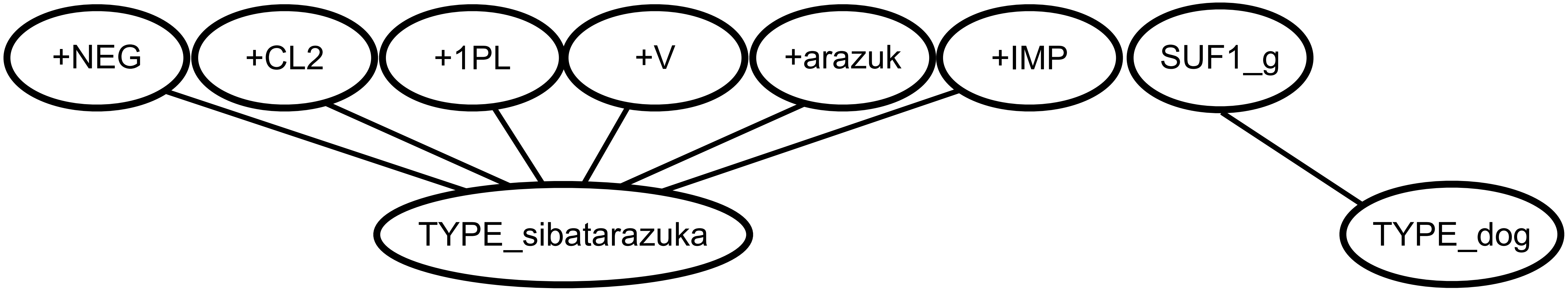
Tag Dict Generalization



Finite-State Transducer (FST)

- Generates morphological analysis

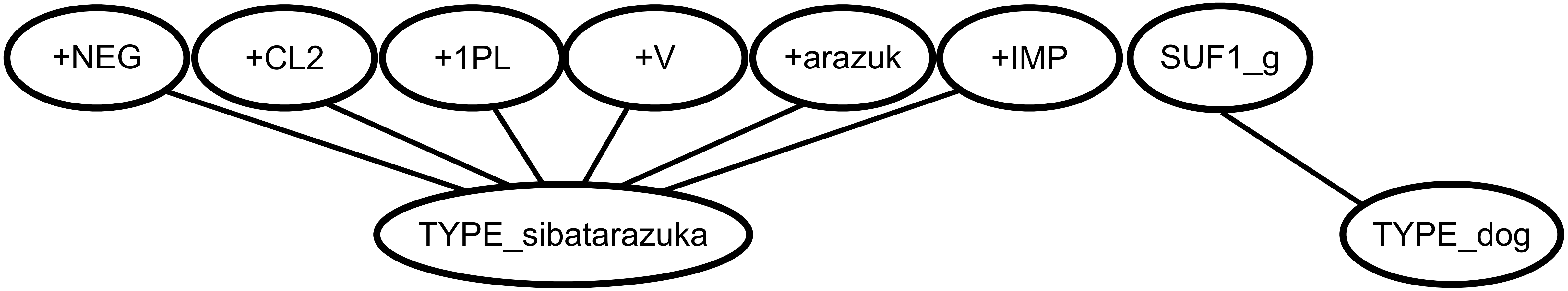
Tag Dict Generalization



Finite-State Transducer (FST)

- Generates morphological analysis

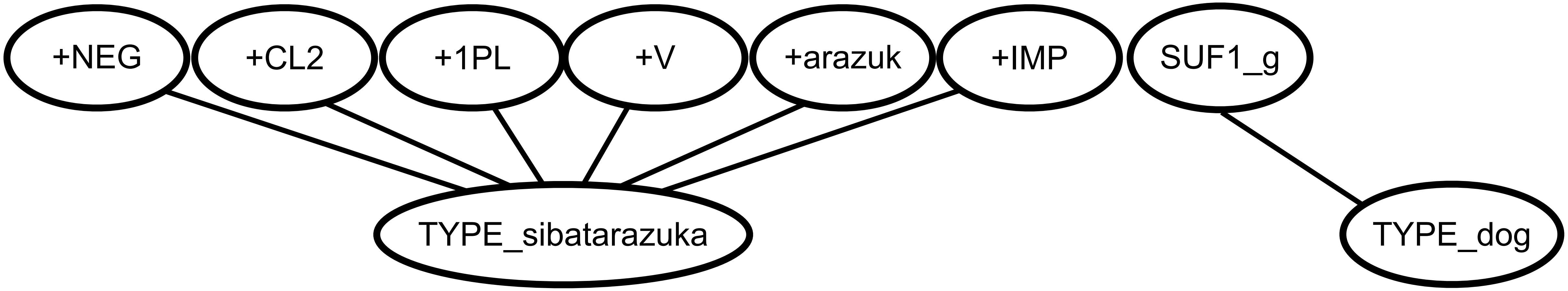
Tag Dict Generalization



Finite-State Transducer (FST)

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

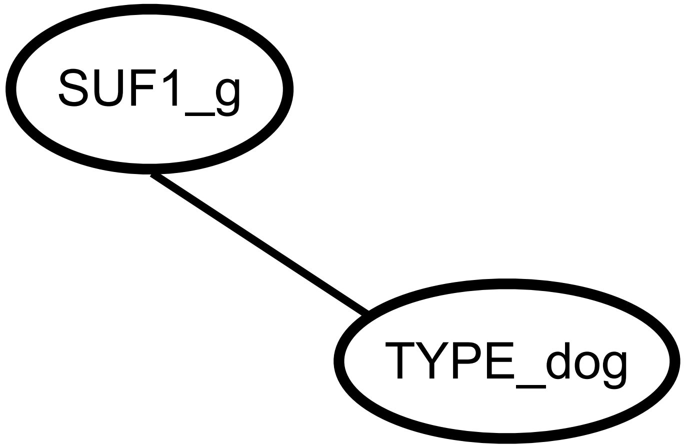
Tag Dict Generalization



Finite-State Transducer (FST)

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

Tag Dict Generalization



Tag Dict Generalization

Type Annotations

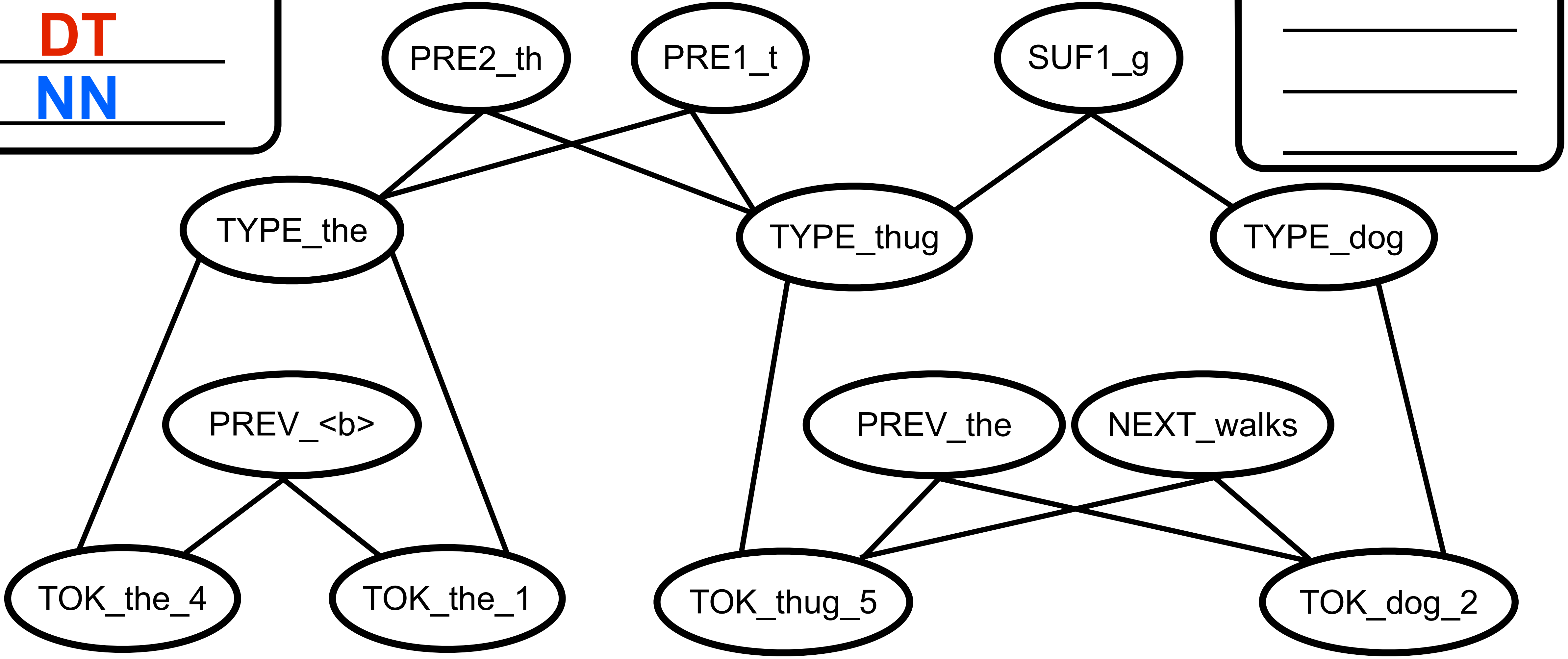
the	DT
dog	NN

Raw Corpus

____

____

____



Token Annotations

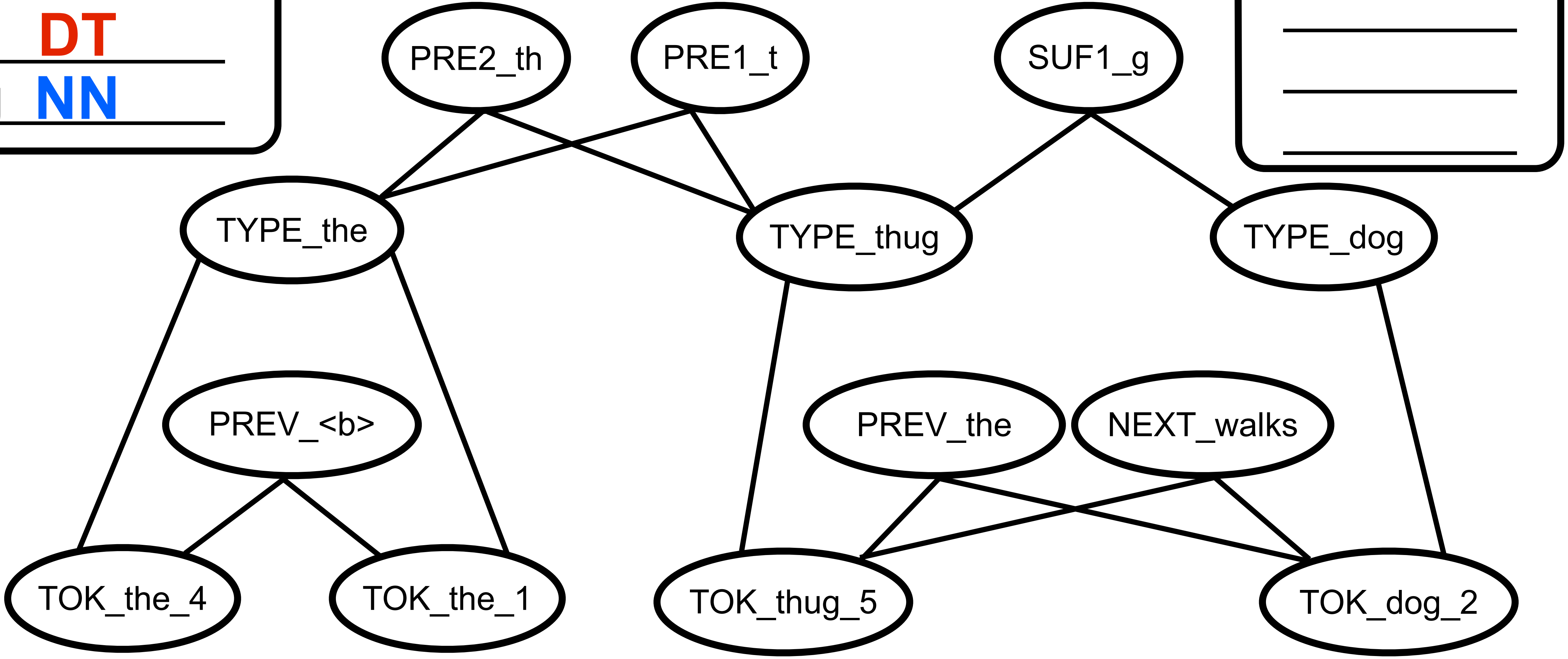
the	dog	walks
DT	NN	VBZ

Tag Dict Generalization

Type Annotations

the	DT
dog	NN

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

Tag Dict Generalization

Type Annotations

the DT

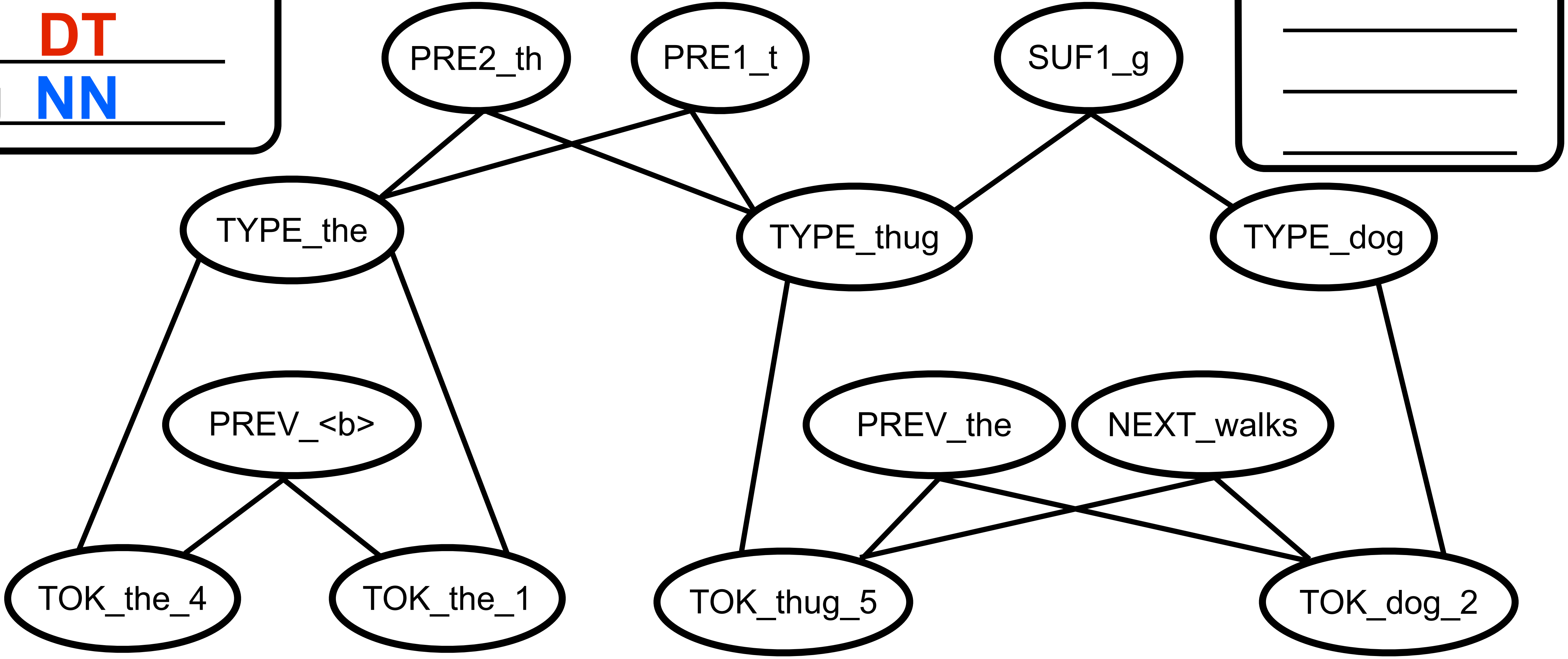
dog NN

Raw Corpus

___

___

___



Token Annotations

the dog walks

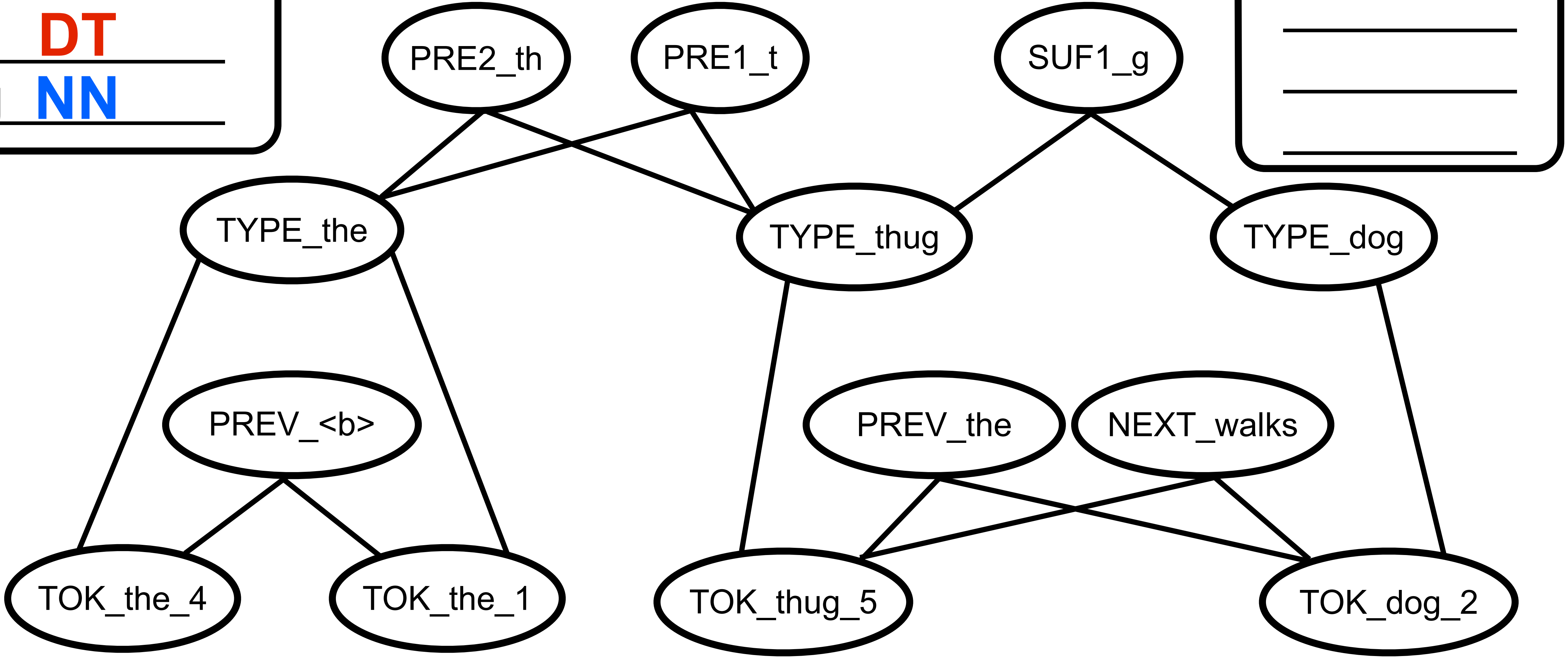
DT NN VBZ

Tag Dict Generalization

Type Annotations

the DT

dog NN



PRE2_th

PRE1_t

SUF1_g

TYPE_the

TYPE_thug

TYPE_dog

PREV_<b>

PREV_the

NEXT_walks

TOK_the_4

TOK_the_1

TOK_thug_5

TOK_dog_2

Raw Corpus

Token Annotations

the dog walks

DT NN VBZ

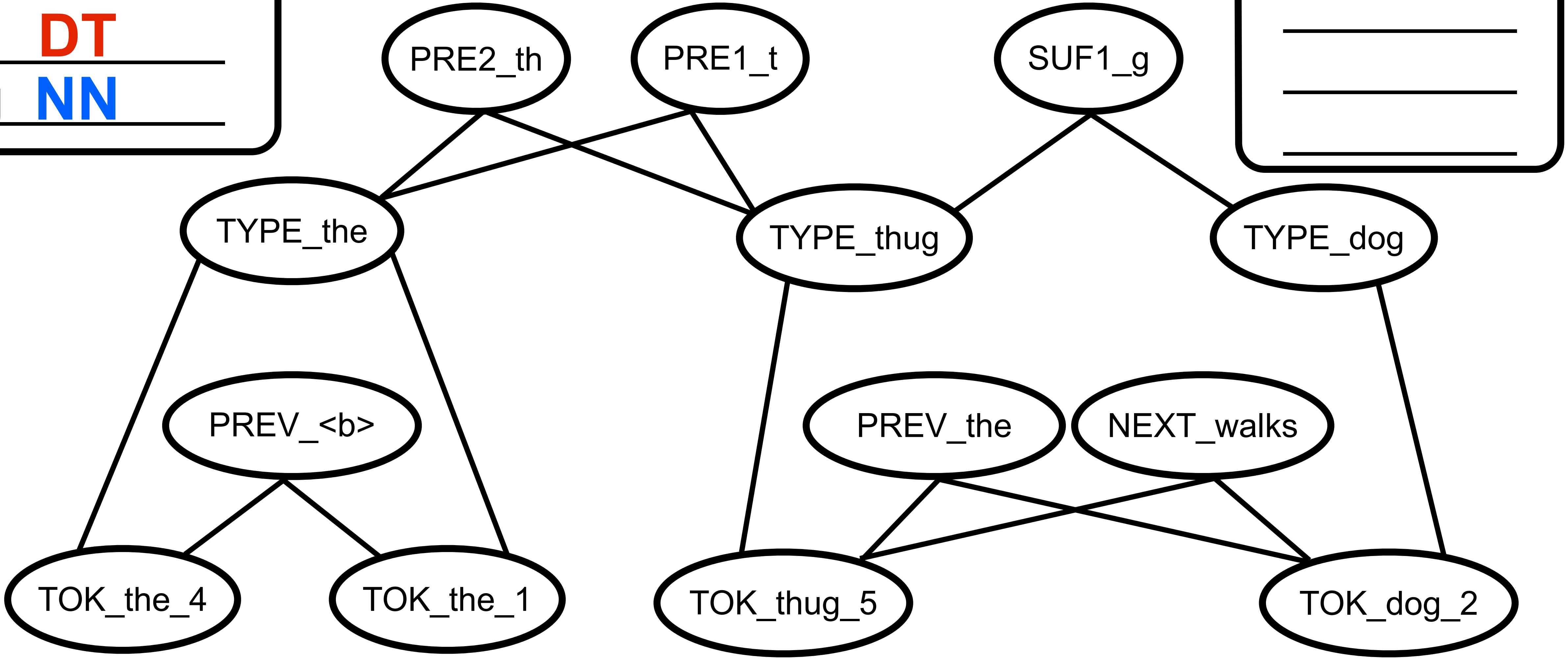
Tag Dict Generalization

Type Annotations

the DT

dog NN

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

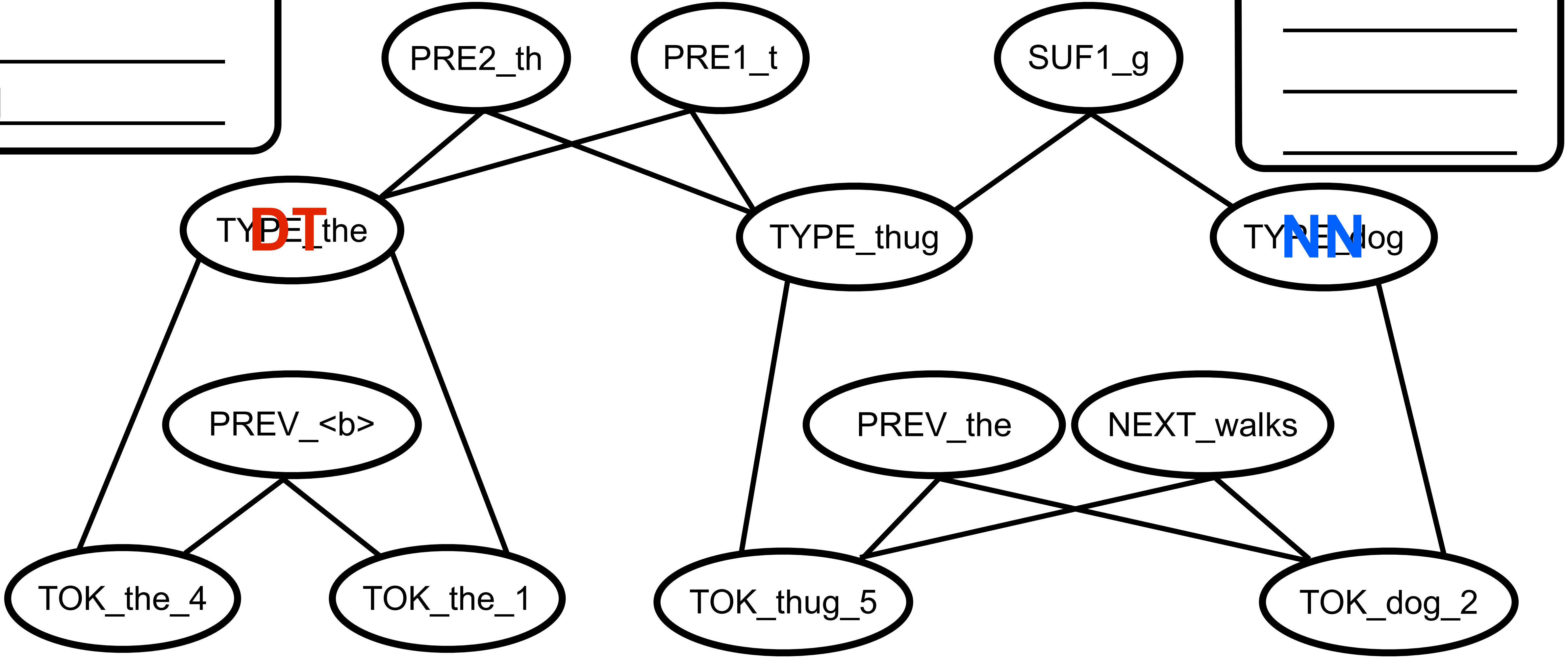
Tag Dict Generalization

Type Annotations

the

dog

Raw Corpus



Token Annotations

the dog walks

DT NN VBZ

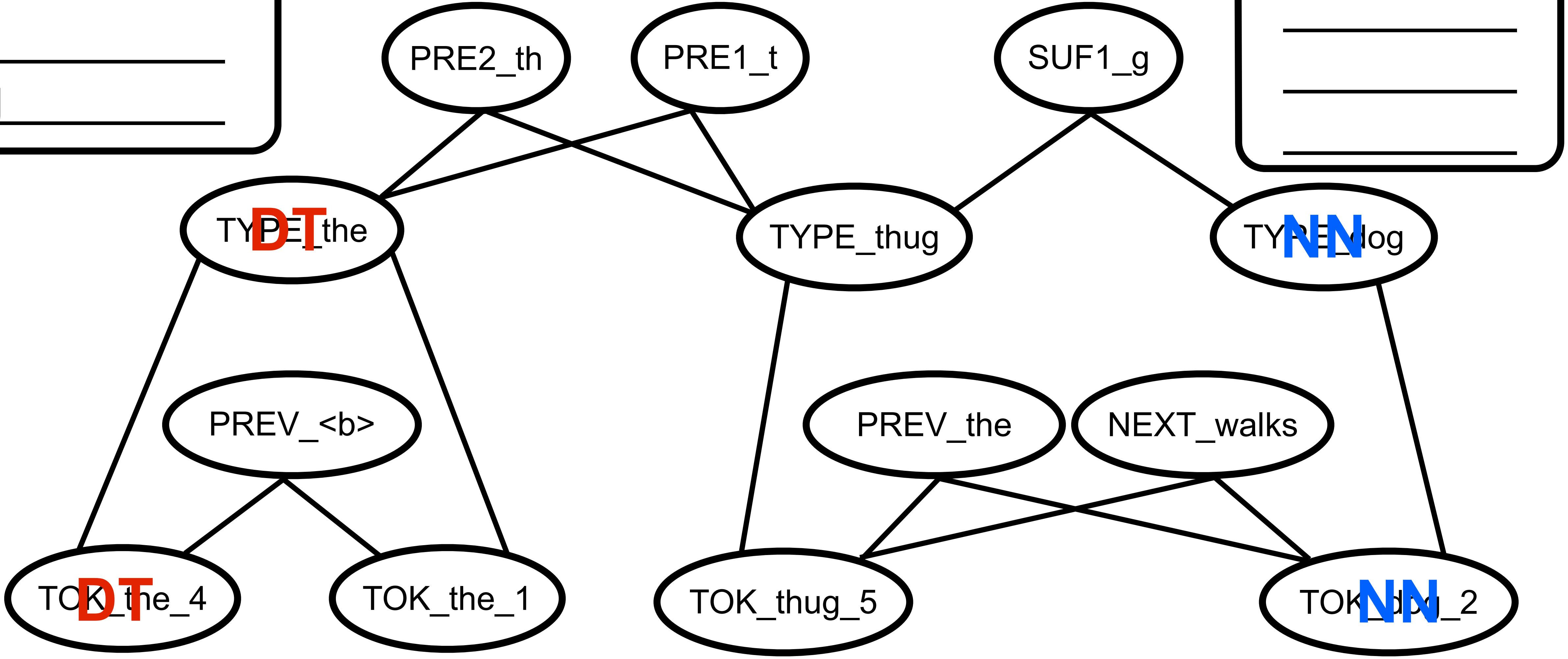
Tag Dict Generalization

Type Annotations

the

dog

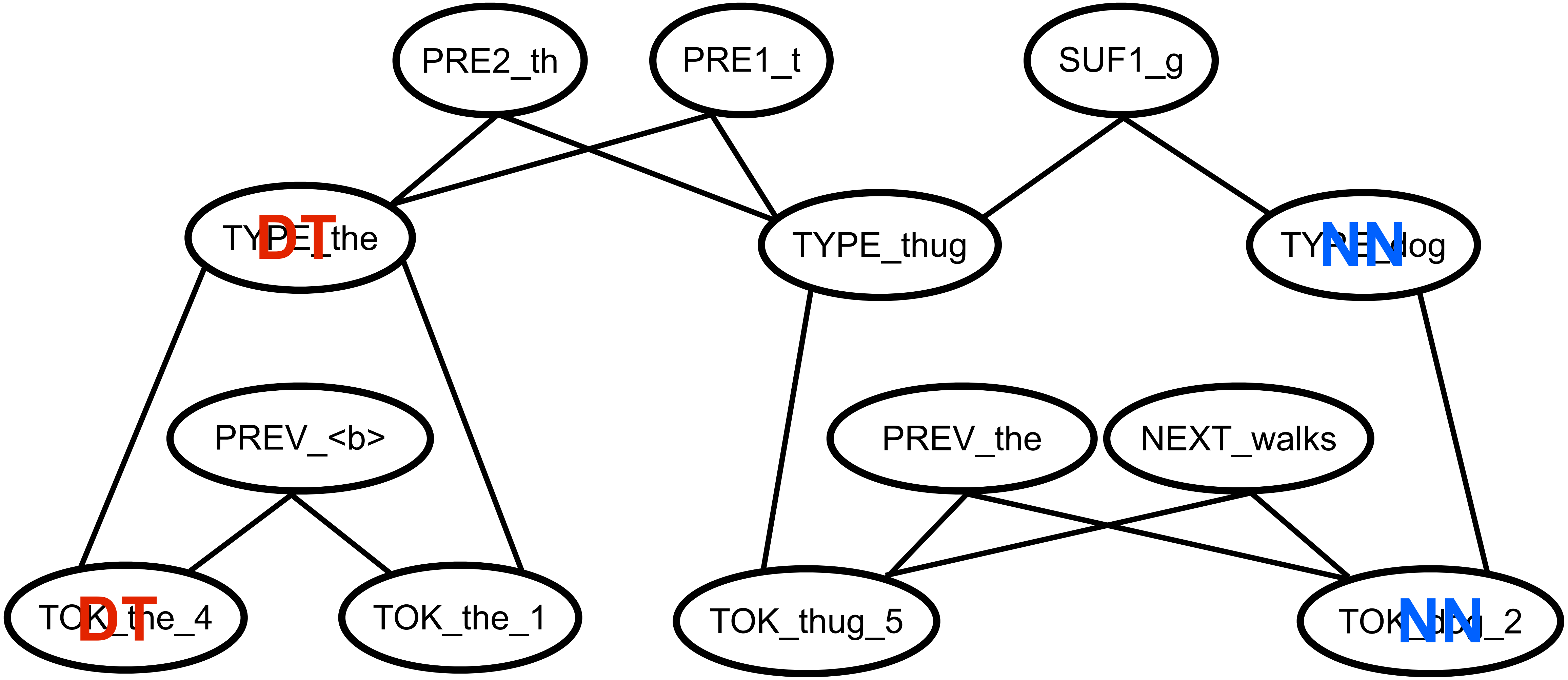
Raw Corpus



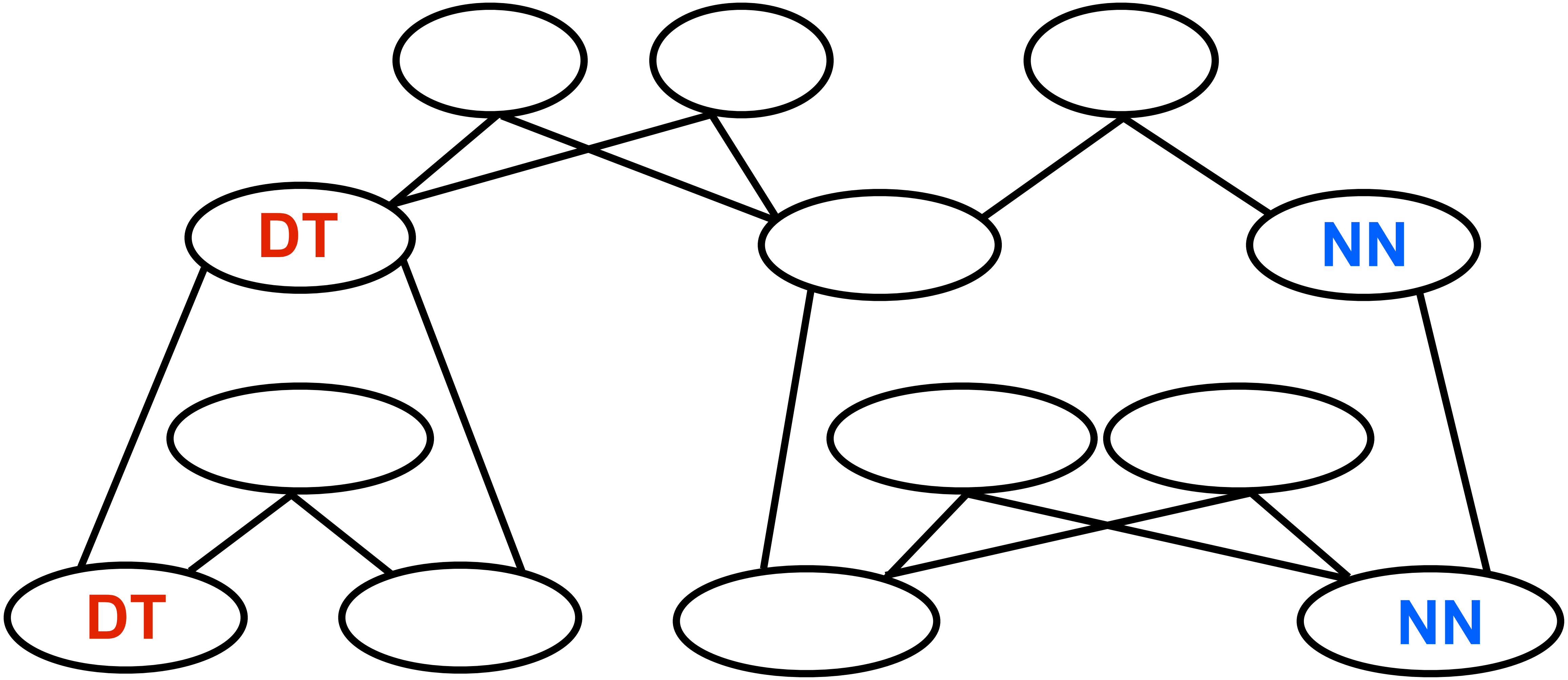
Token Annotations

the dog walks

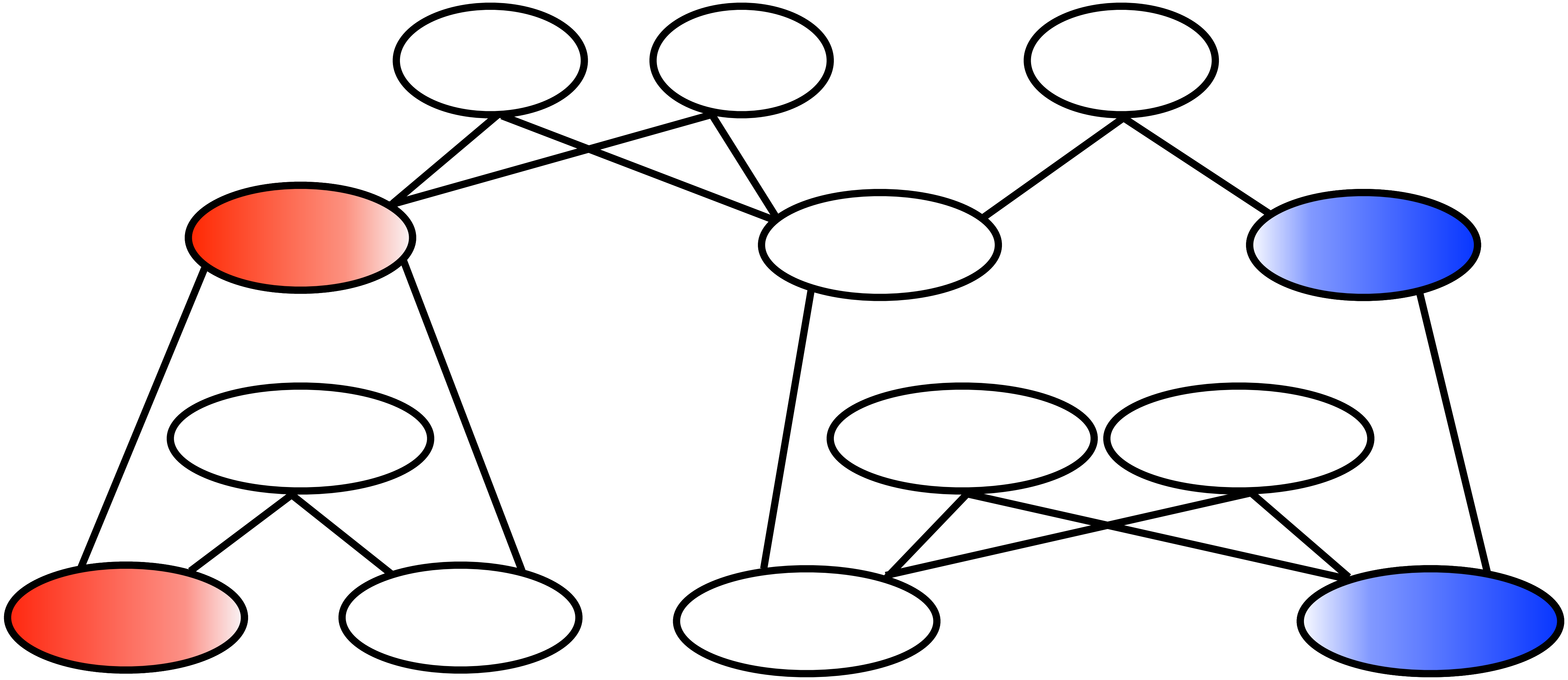
Tag Dict Generalization



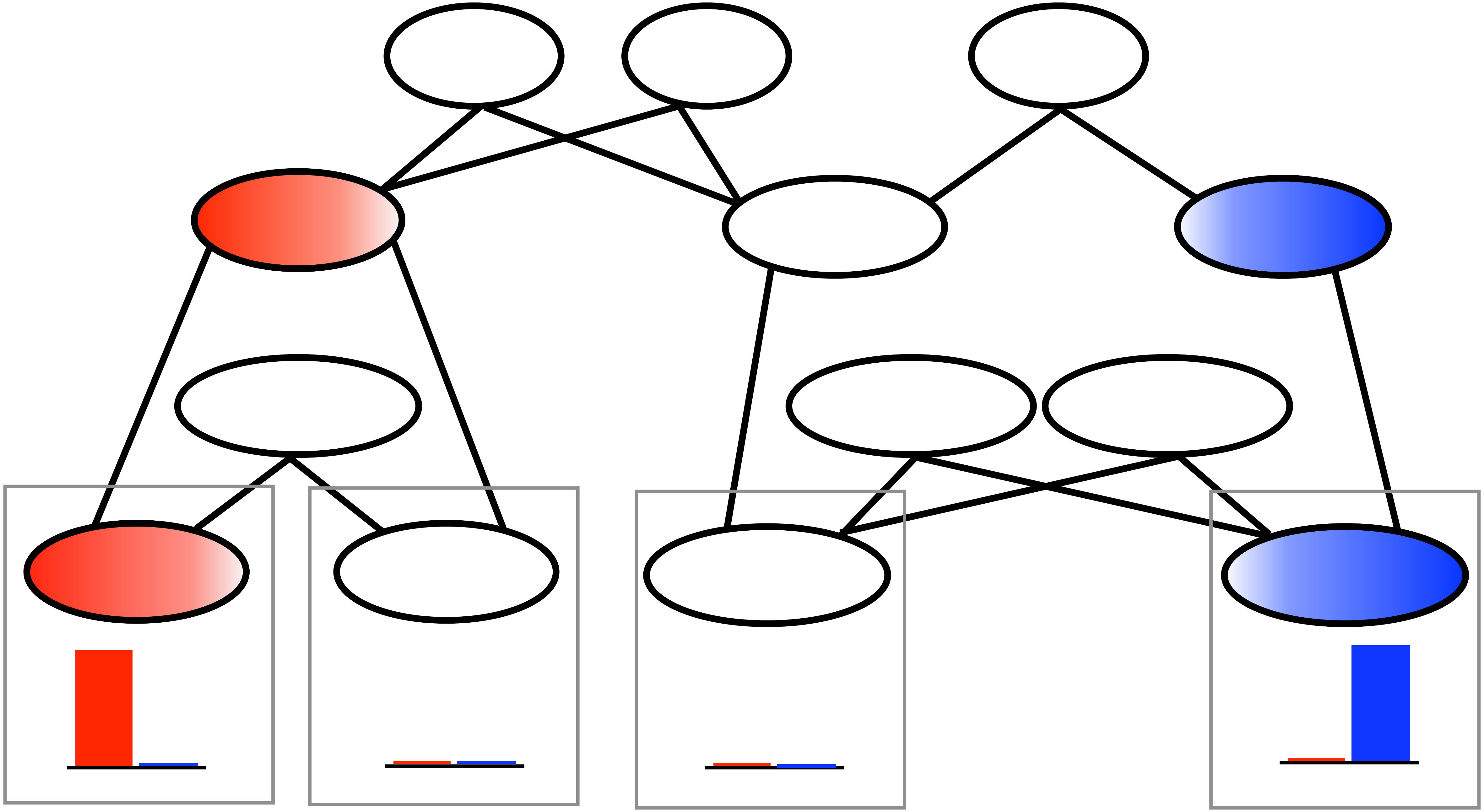
Tag Dict Generalization



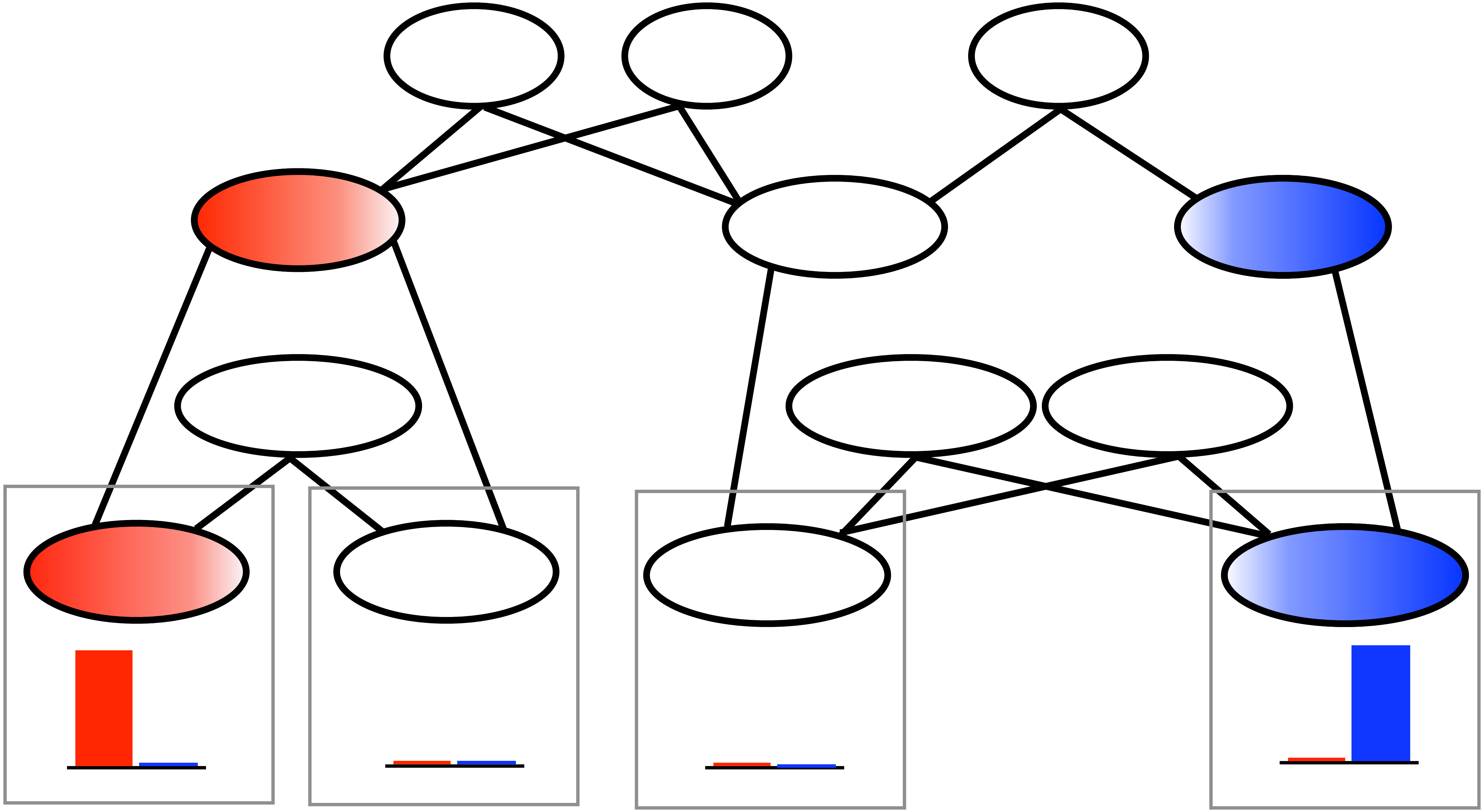
Tag Dict Generalization



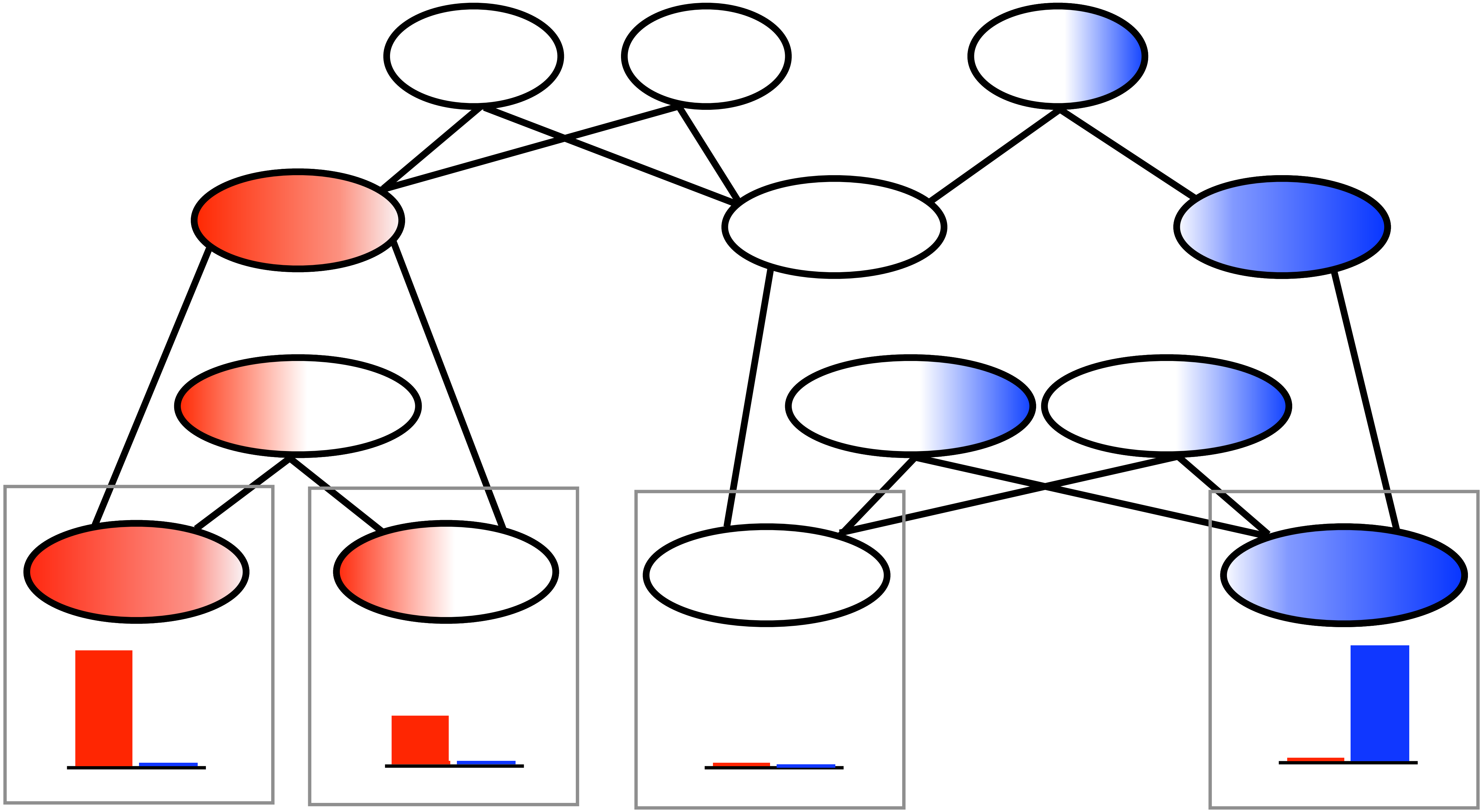
Tag Dict Generalization



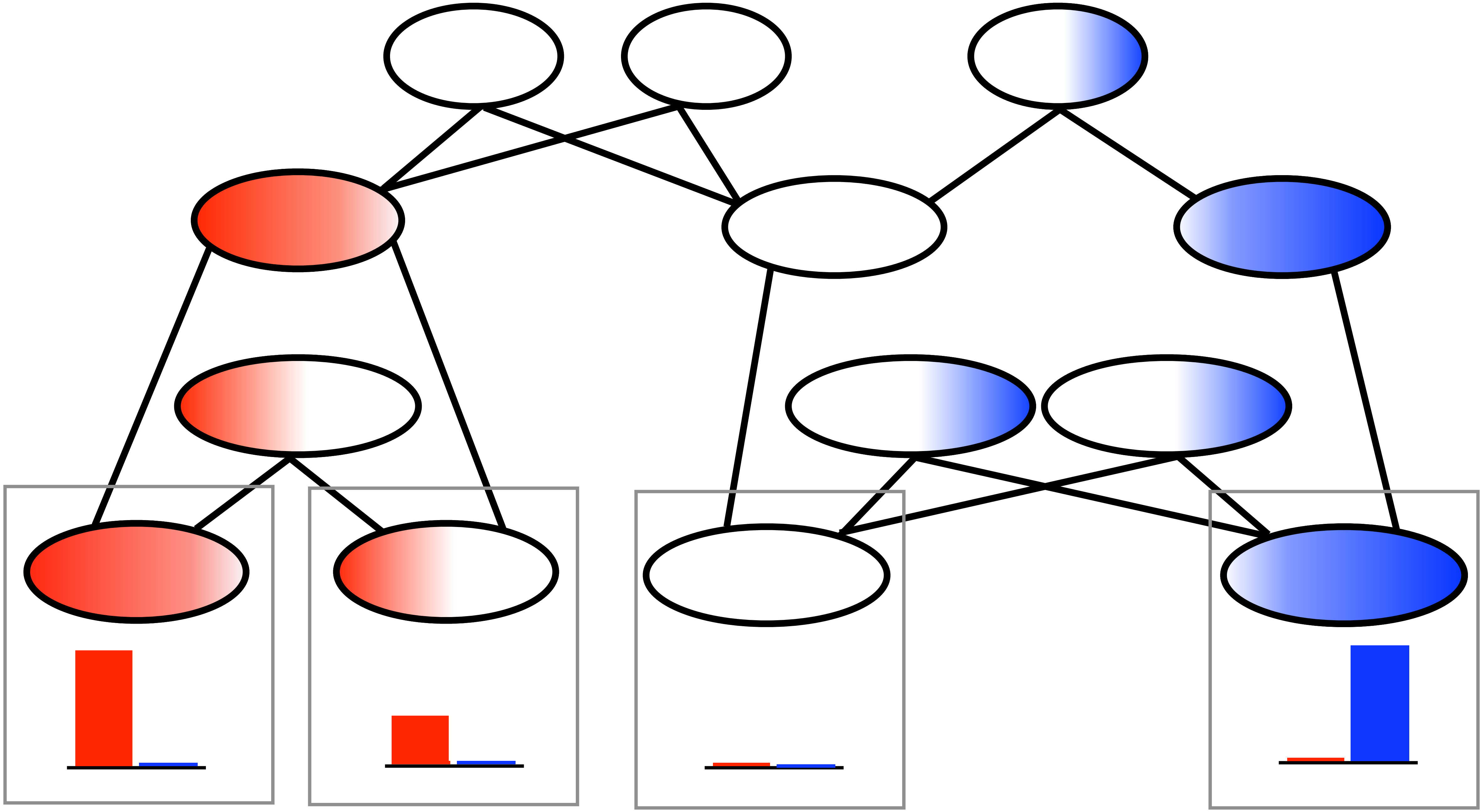
Tag Dict Generalization



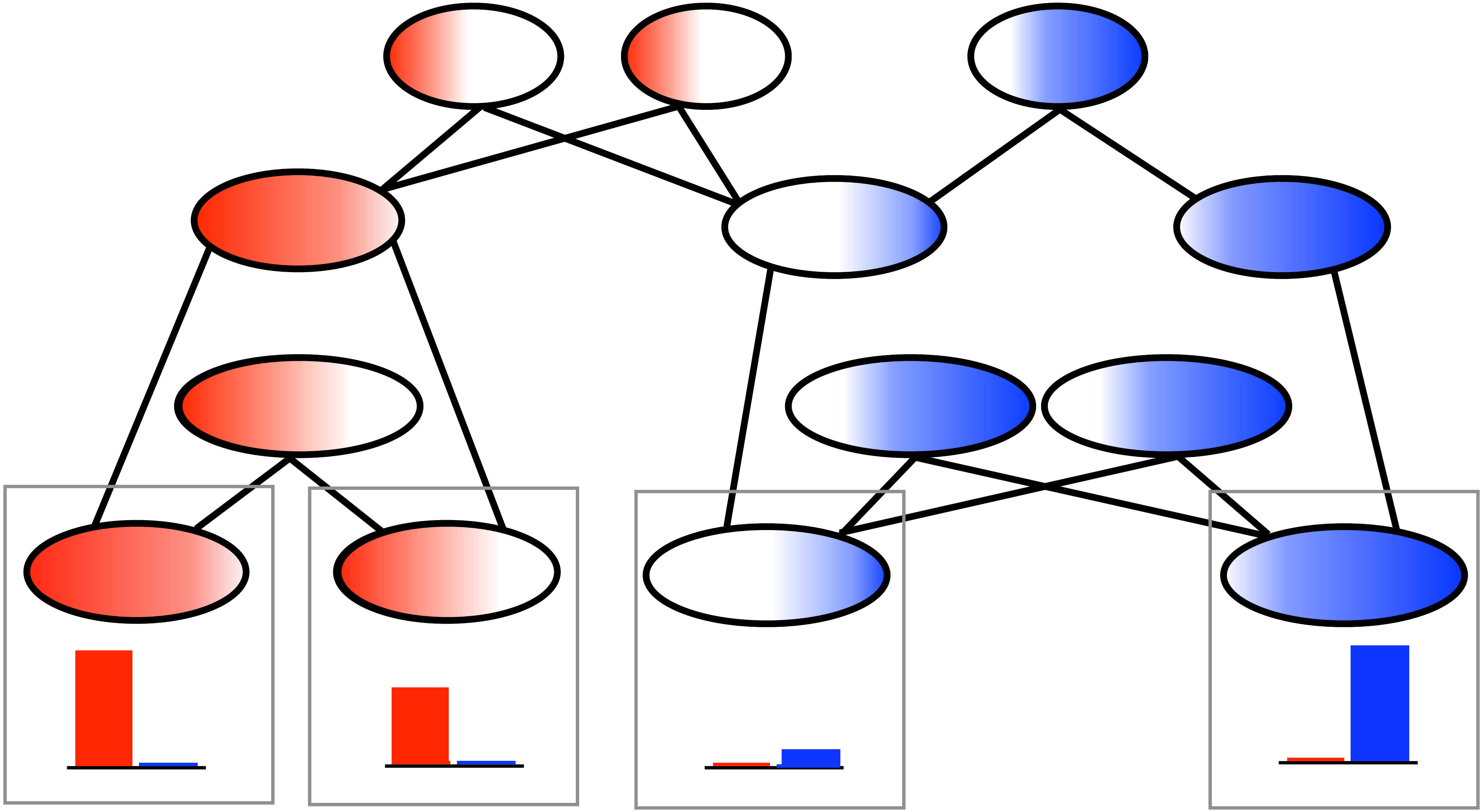
Tag Dict Generalization



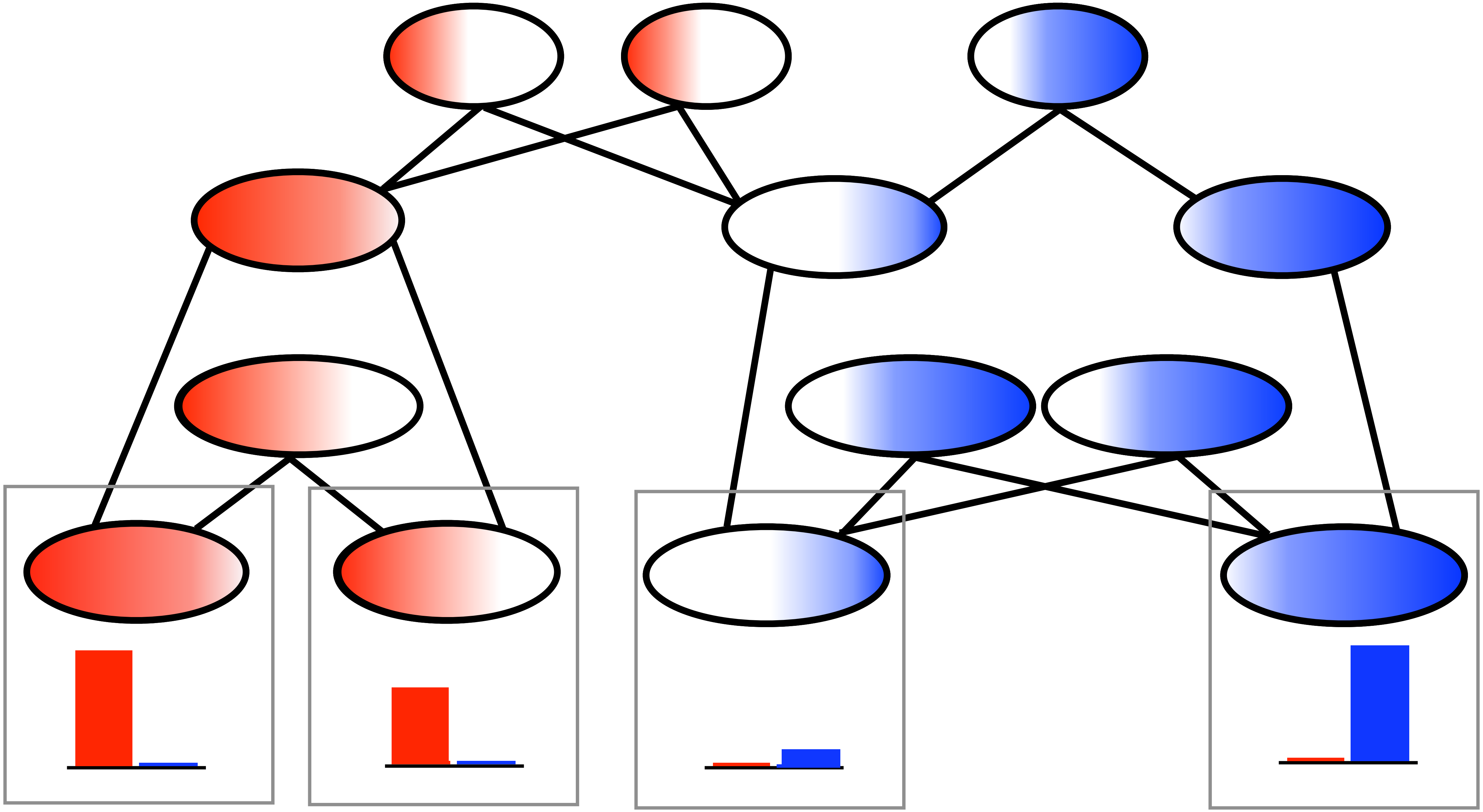
Tag Dict Generalization



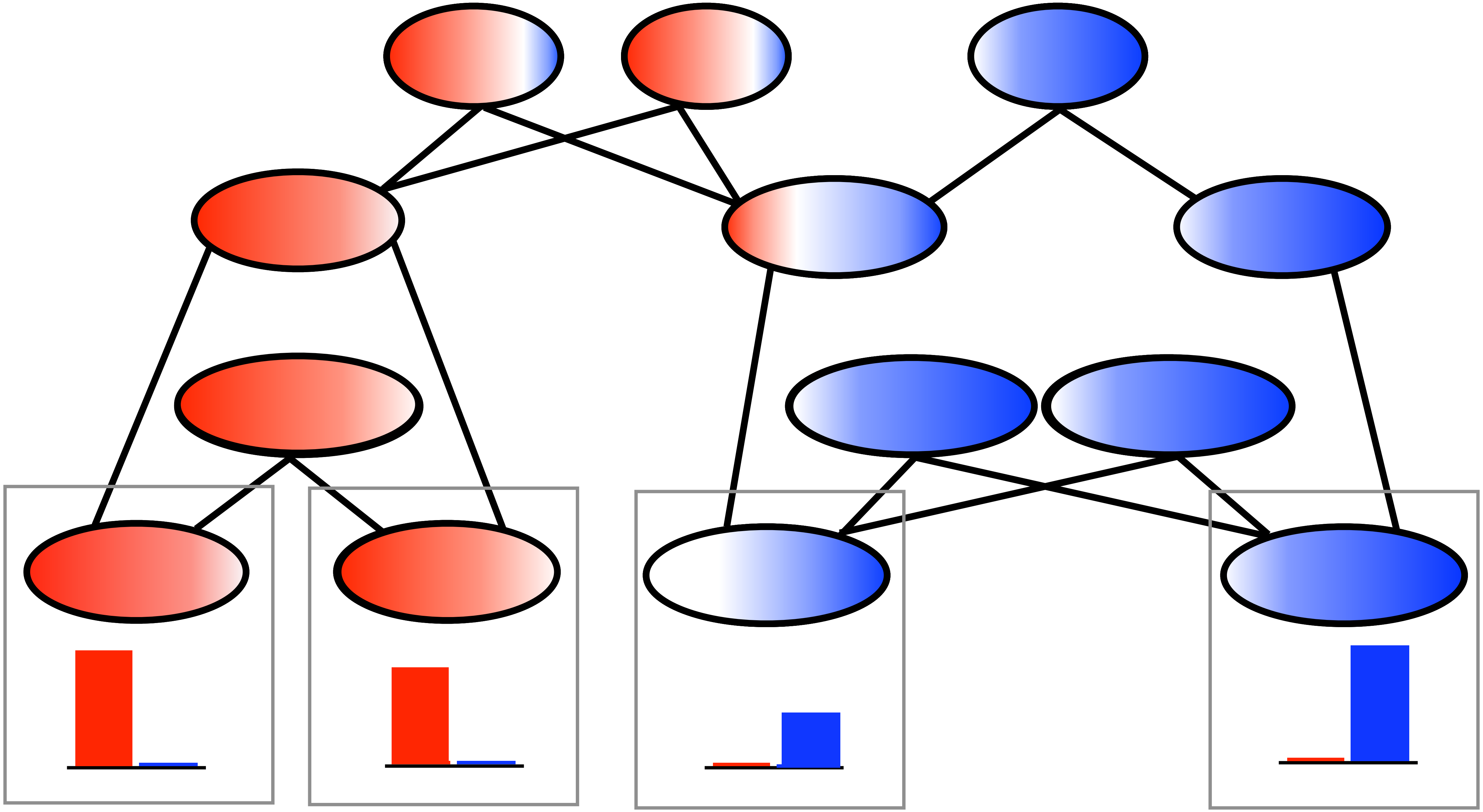
Tag Dict Generalization



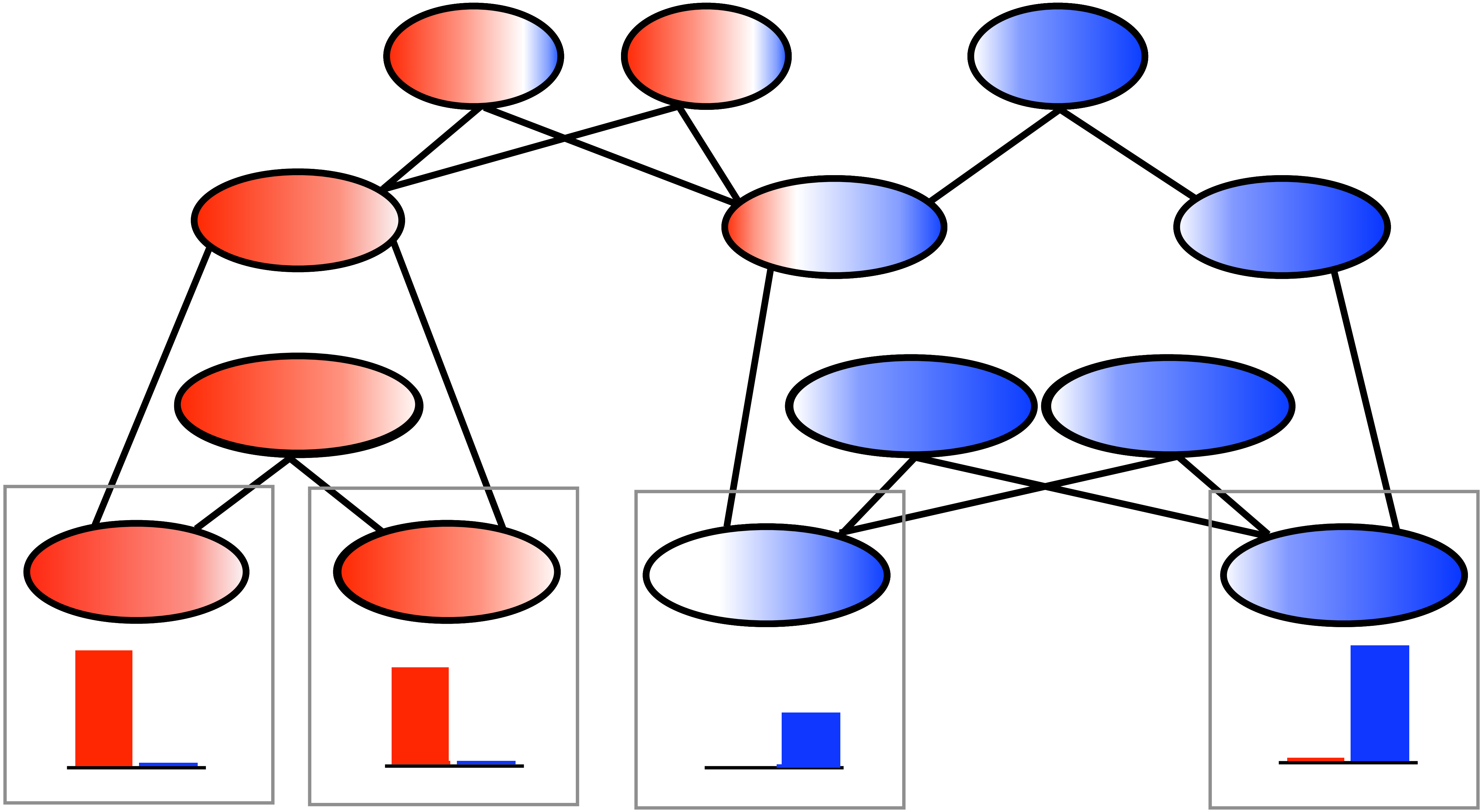
Tag Dict Generalization



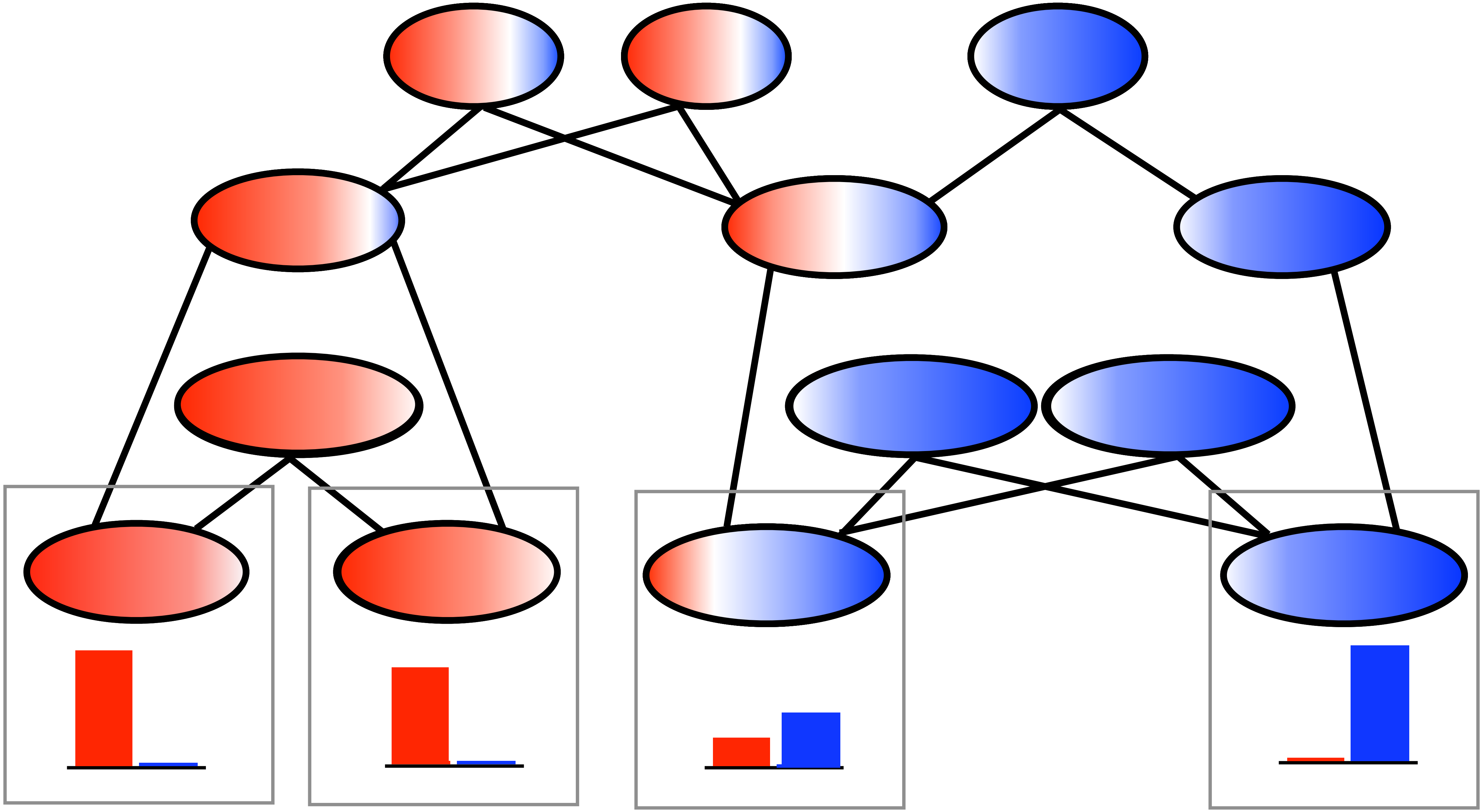
Tag Dict Generalization



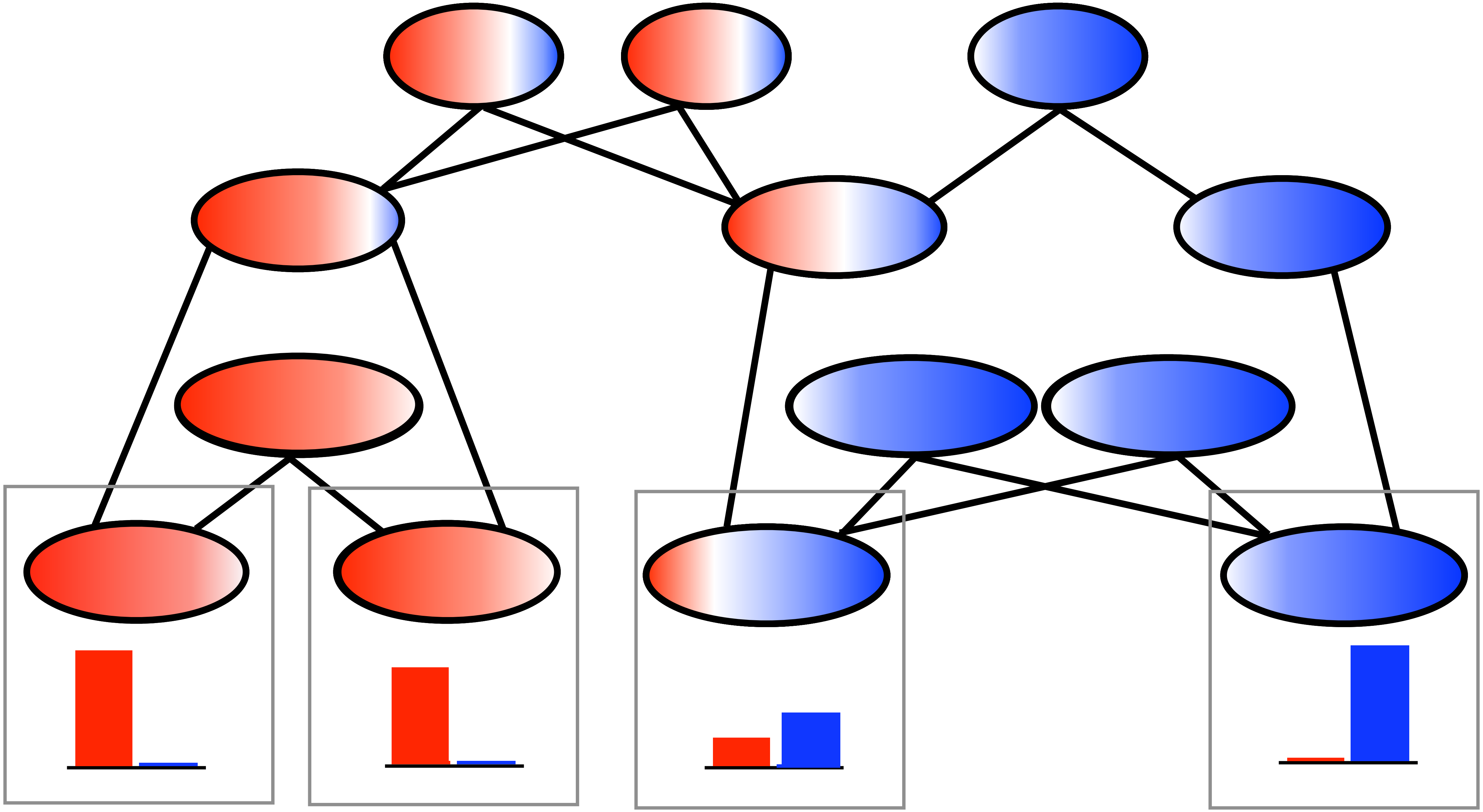
Tag Dict Generalization



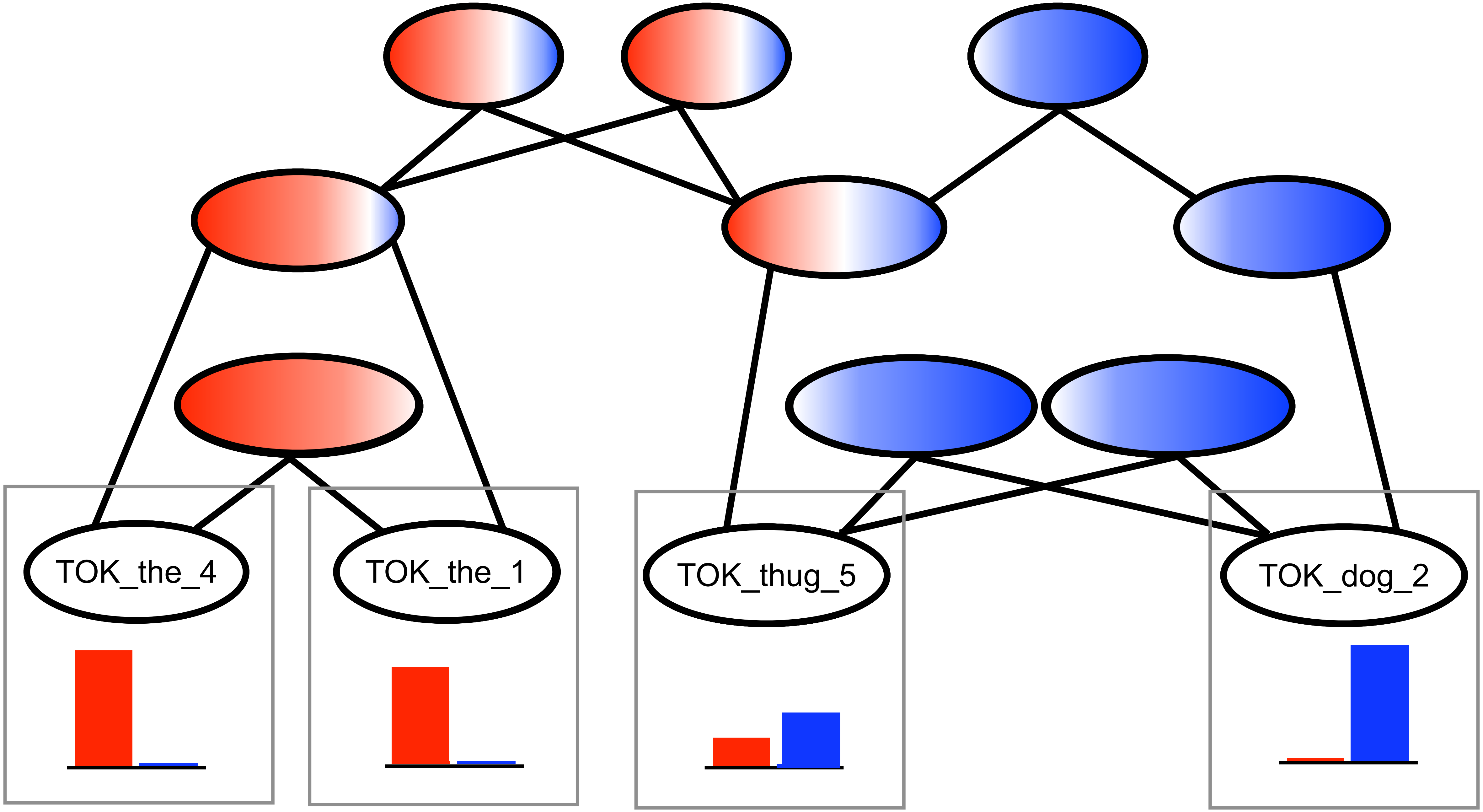
Tag Dict Generalization



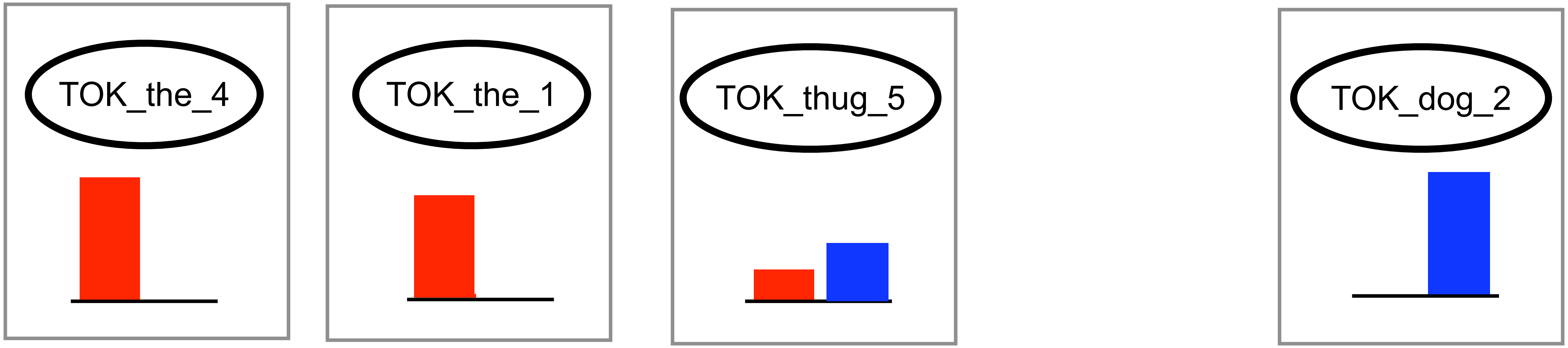
Tag Dict Generalization



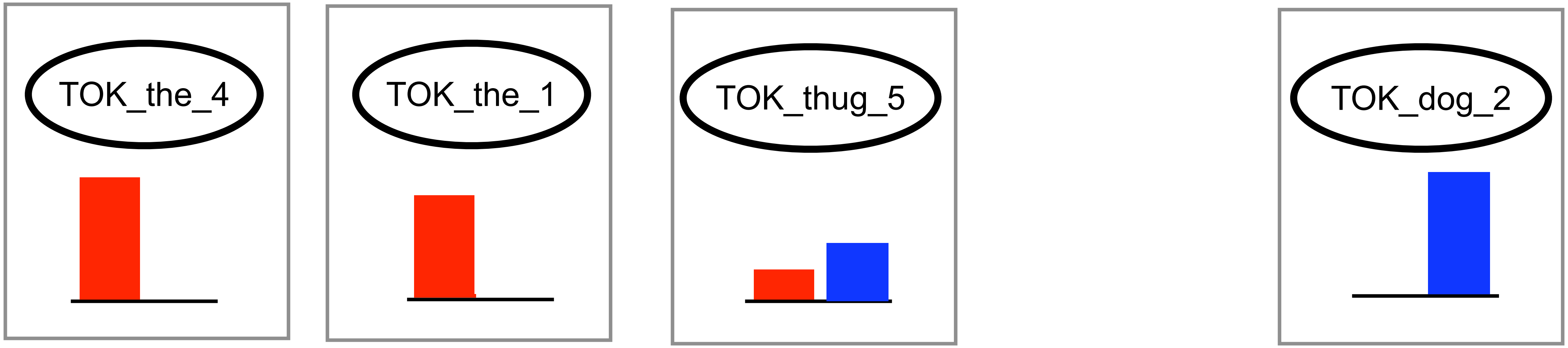
Tag Dict Generalization



Tag Dict Generalization



Tag Dict Generalization



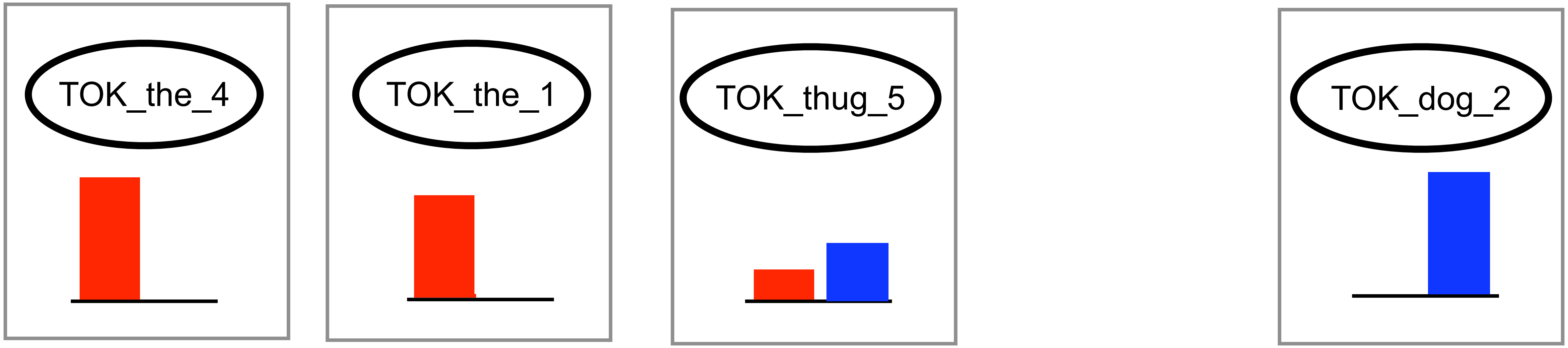
TOK_the_4

TOK_the_1

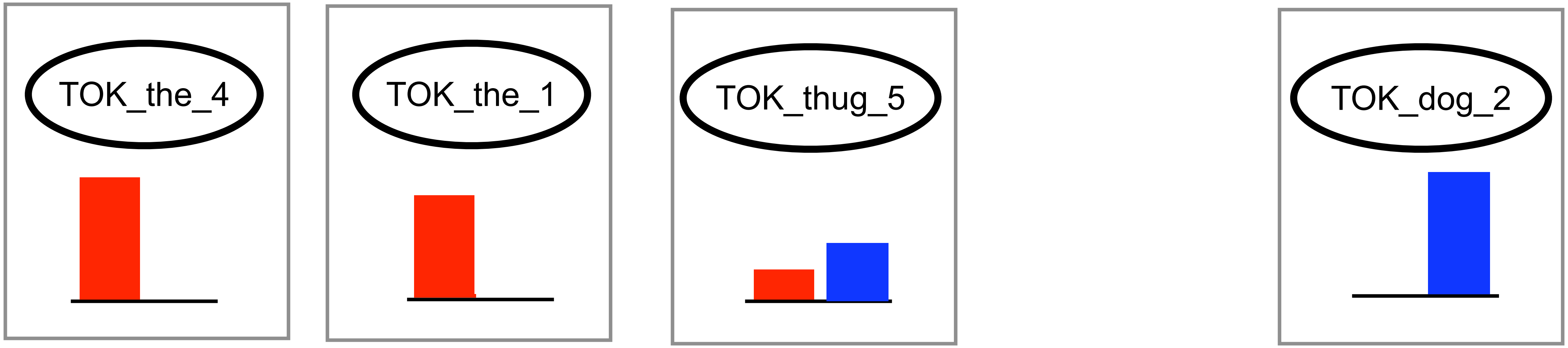
TOK_thug_5

TOK_dog_2

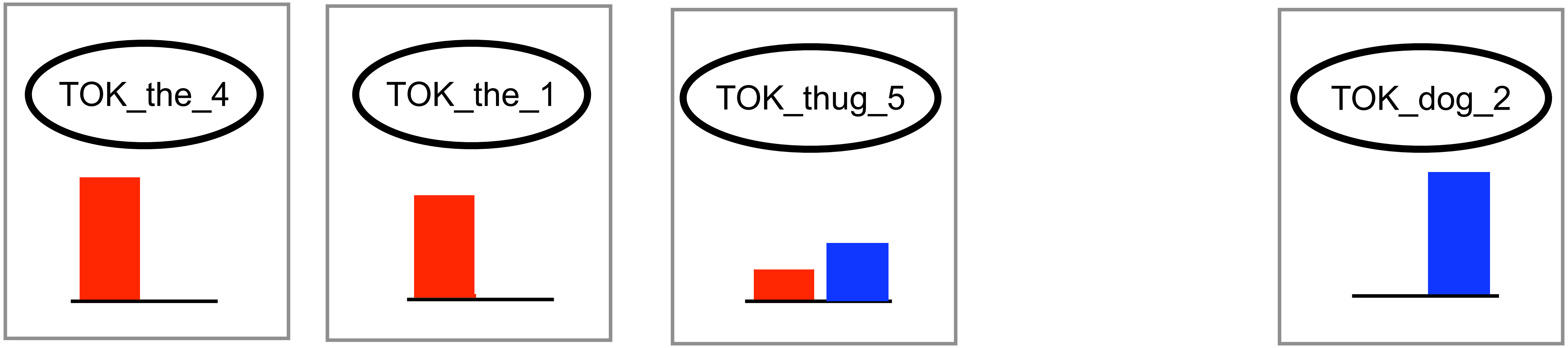
Tag Dict Generalization



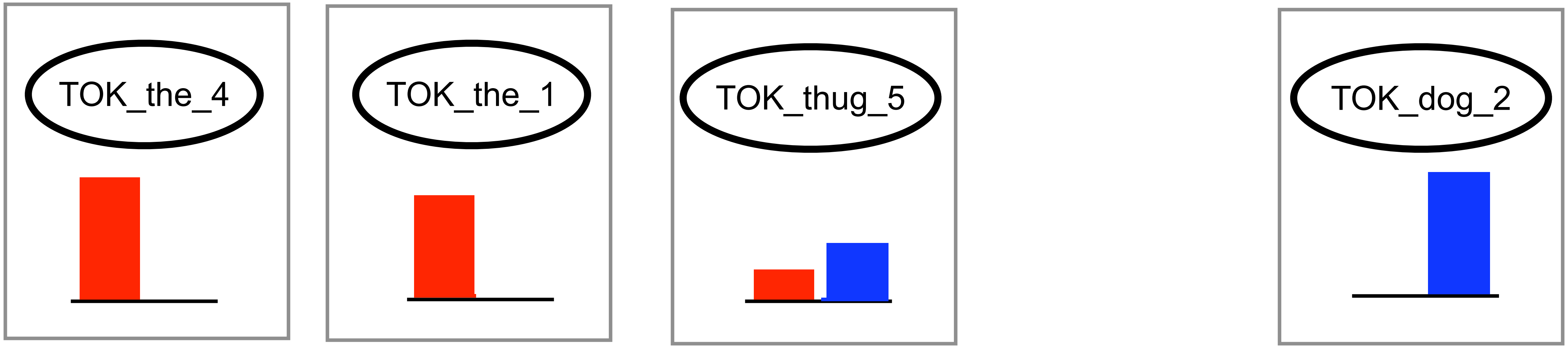
Tag Dict Generalization



Tag Dict Generalization

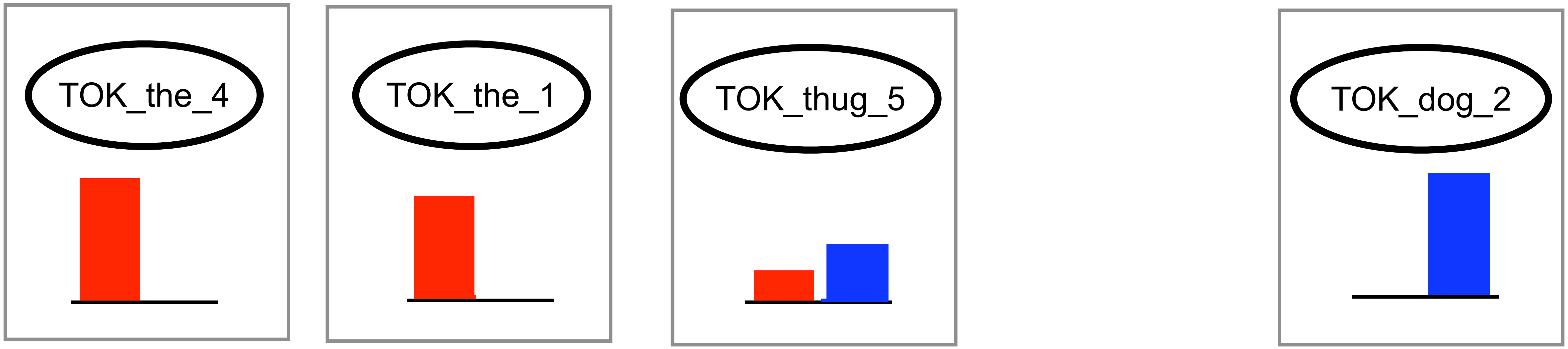


Tag Dict Generalization



Tag Dict Generalization

Result:



TOK_the_4

TOK_the_1

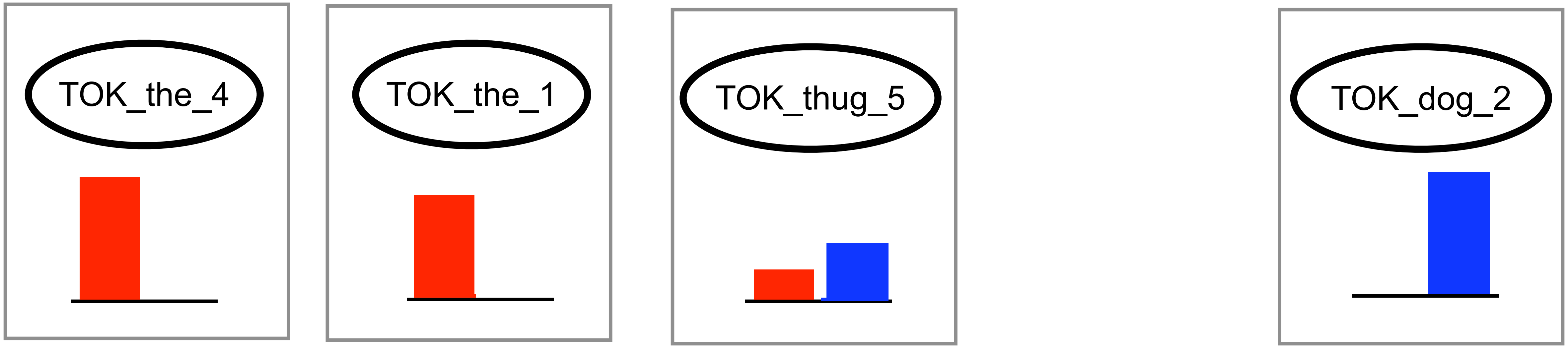
TOK_thug_5

TOK_dog_2

Tag Dict Generalization

Result:

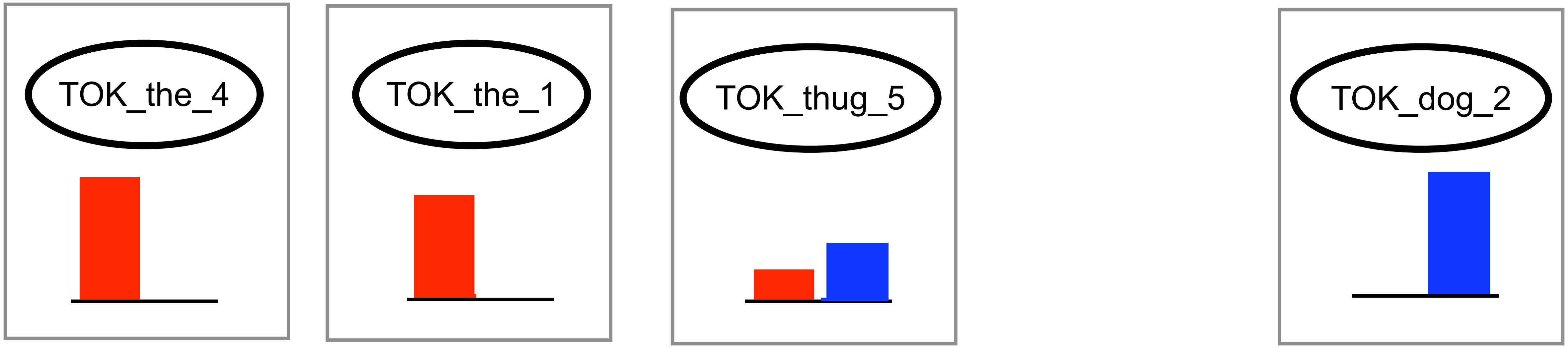
- a tag distribution on every token (soft tagging)



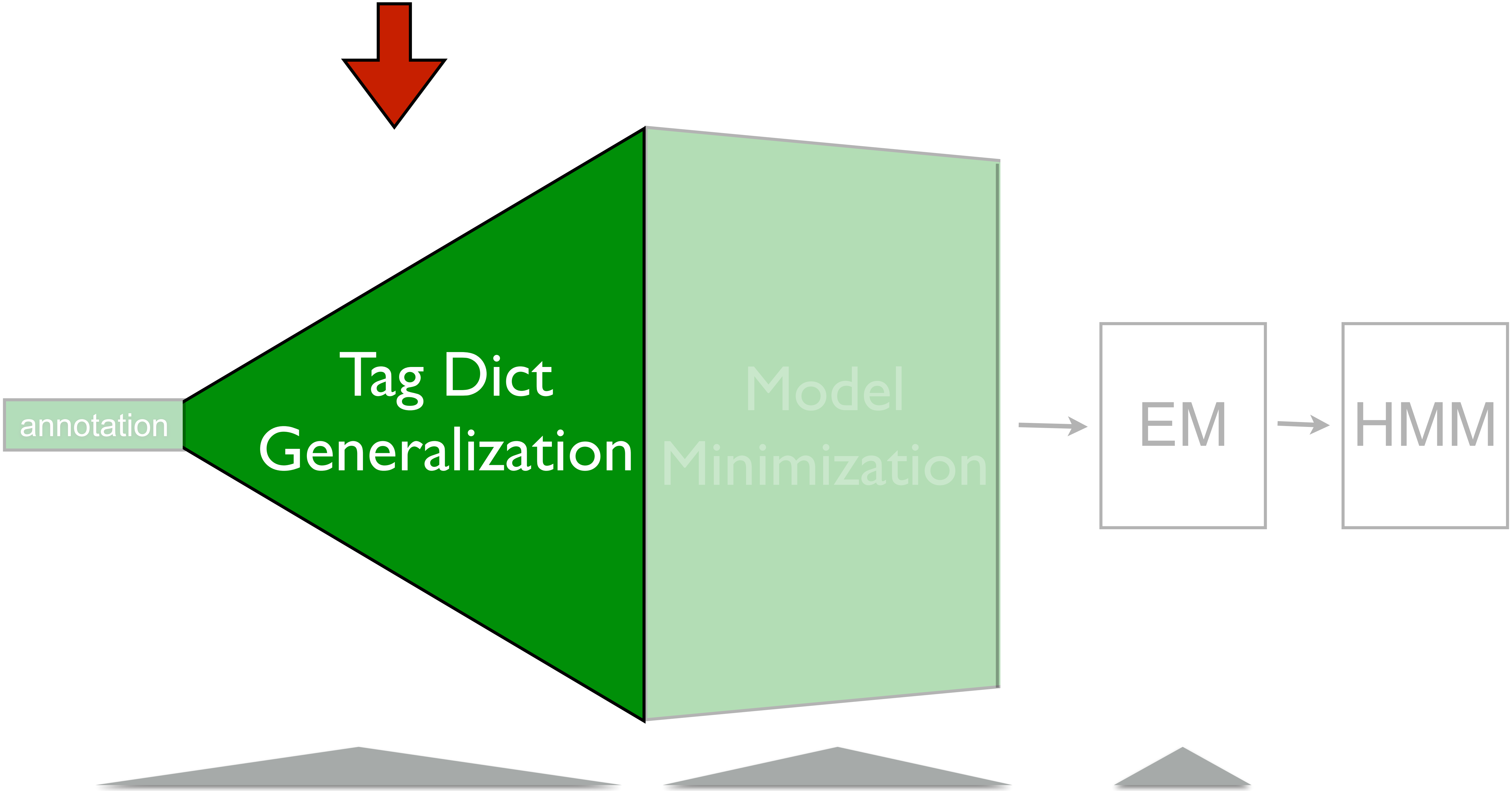
Tag Dict Generalization

Result:

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

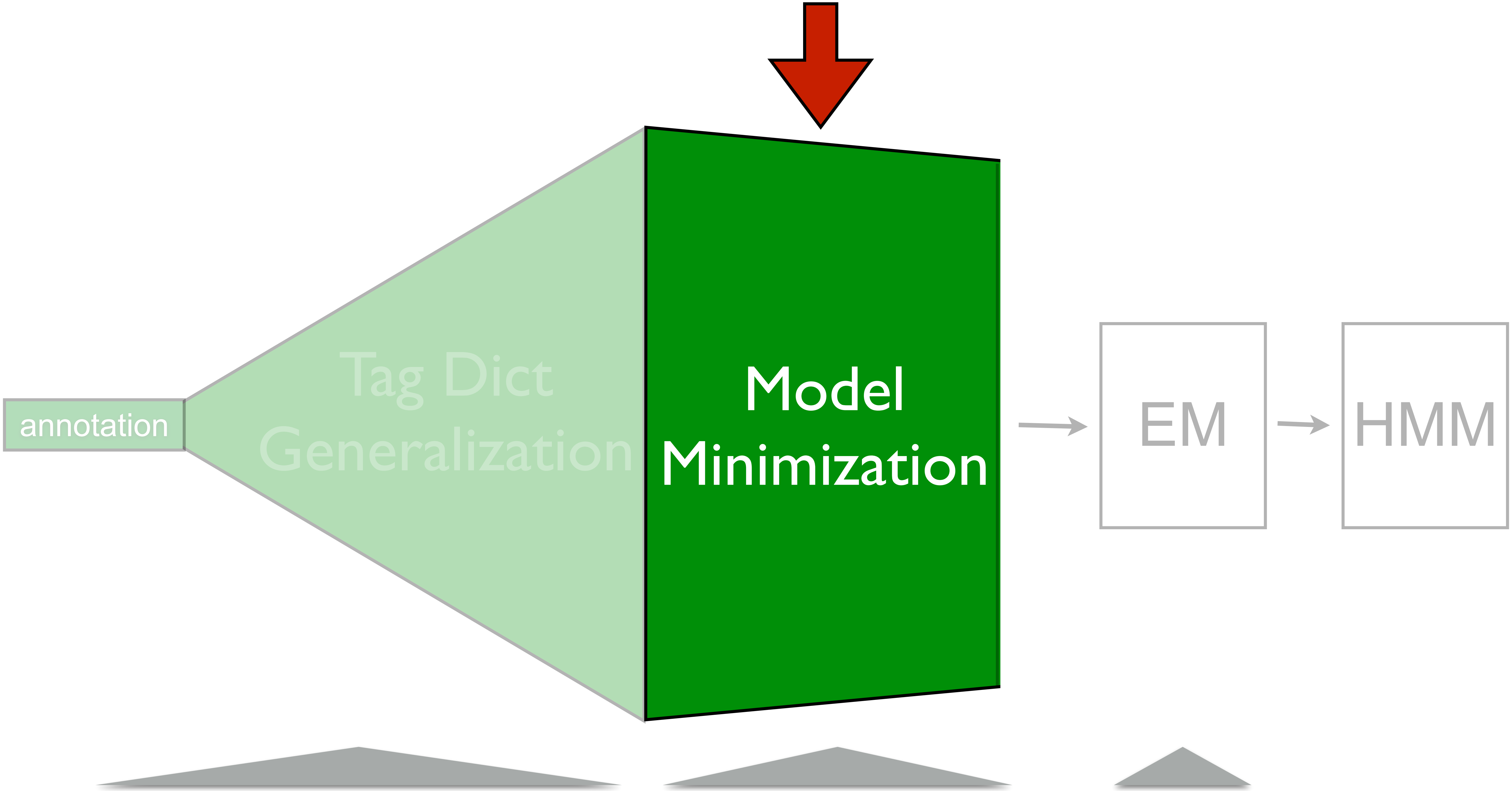


Our Approach



cover the vocabulary remove noise train

Our Approach



cover the vocabulary

remove noise

train

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

<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

Model Minimization

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

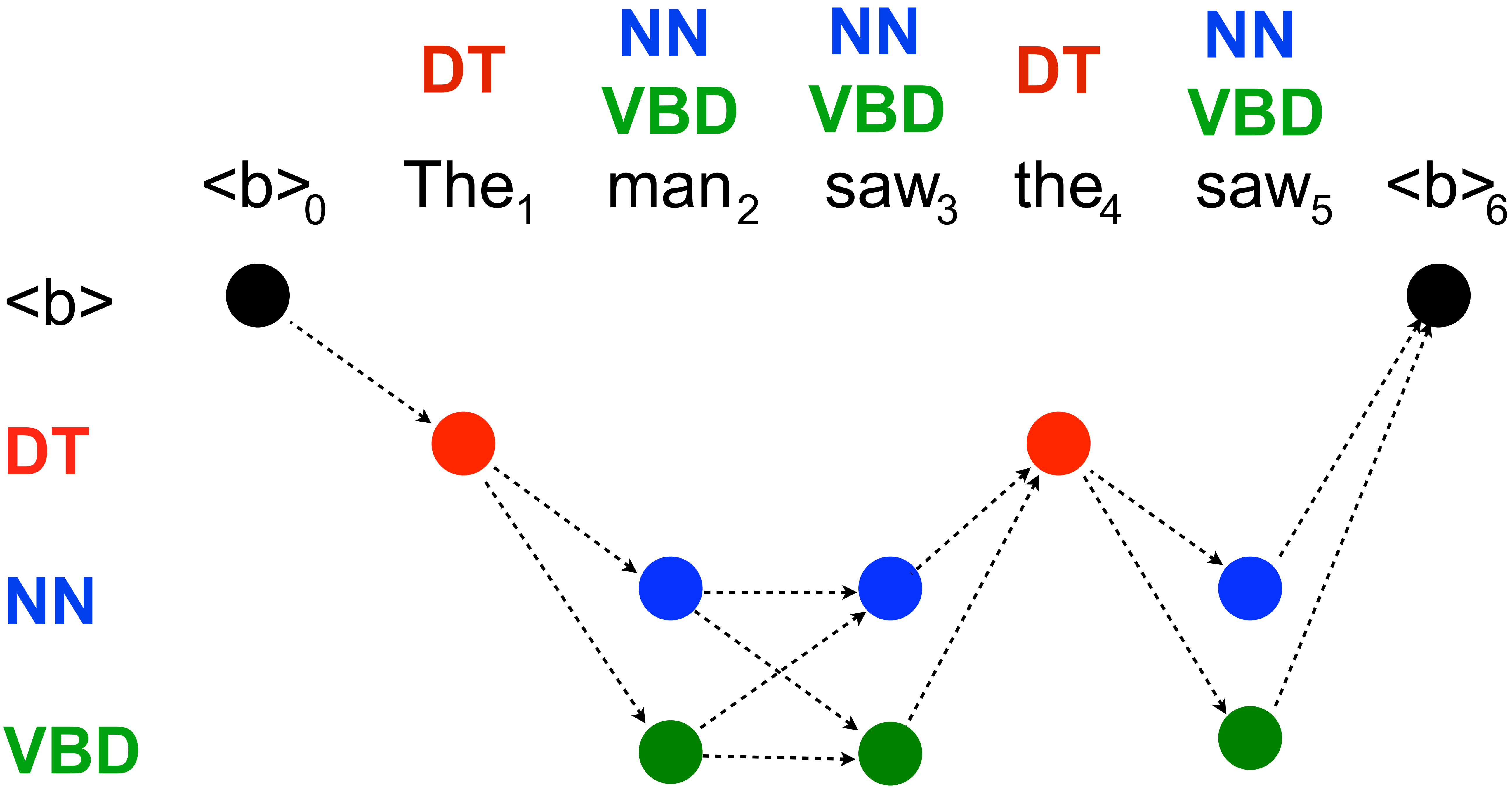
Model Minimization

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

Model Minimization

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

Model Minimization



DT NN VBD NN VBD DT NN VBD

<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

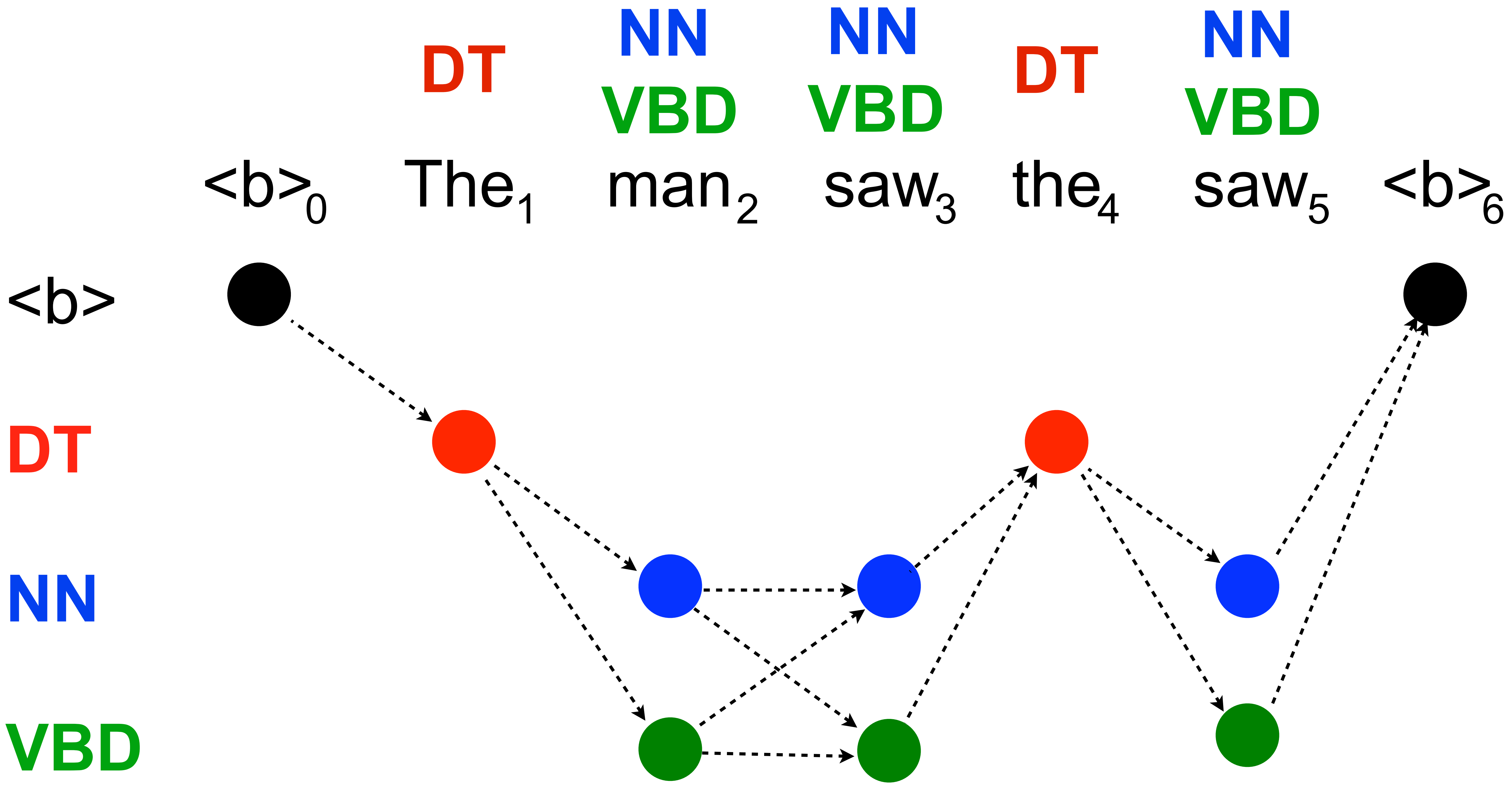
<b>

DT

NN

VBD

Model Minimization



DT NN VBD NN VBD DT NN VBD

<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

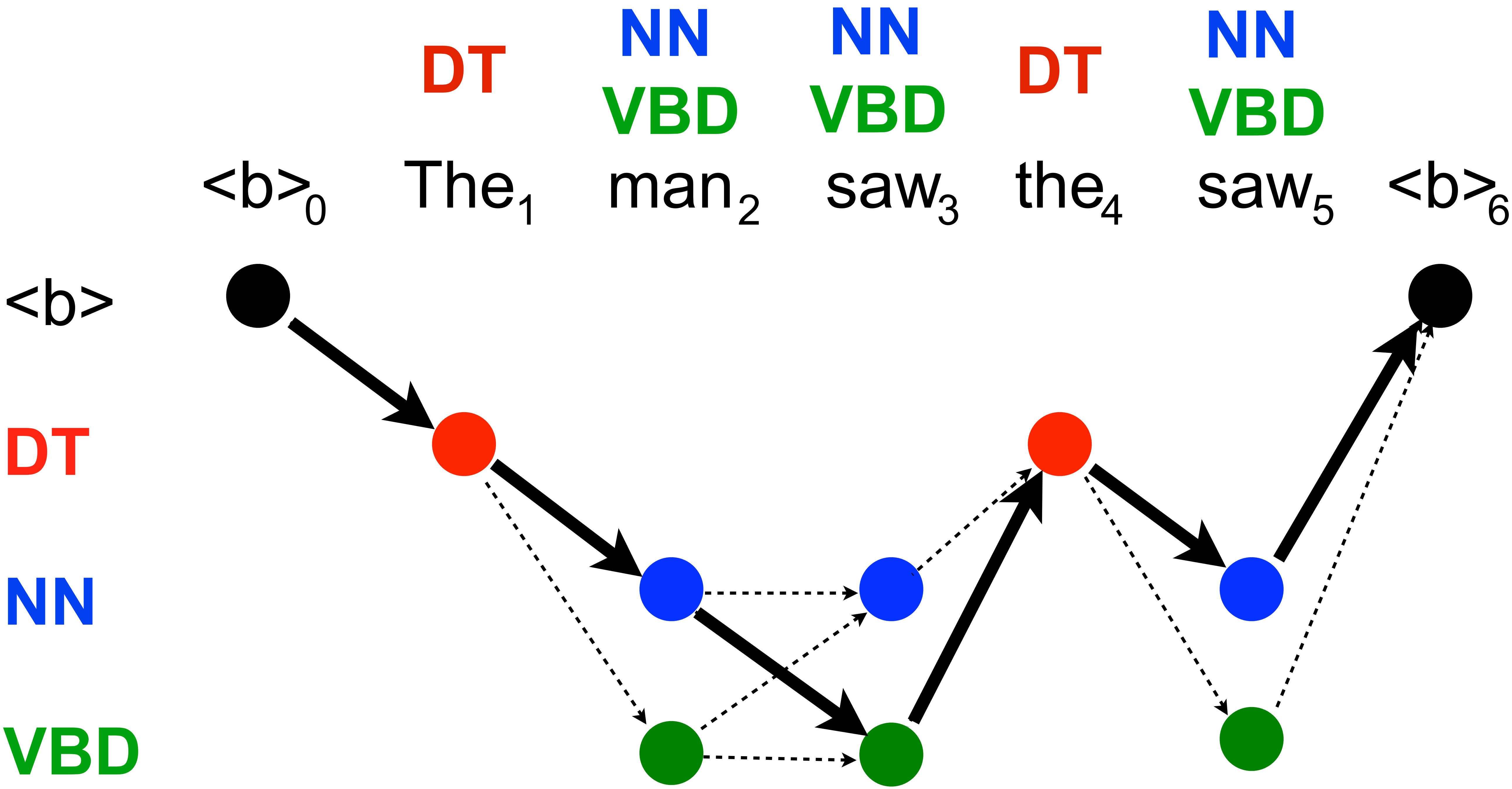
<b>

DT

NN

VBD

Model Minimization



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

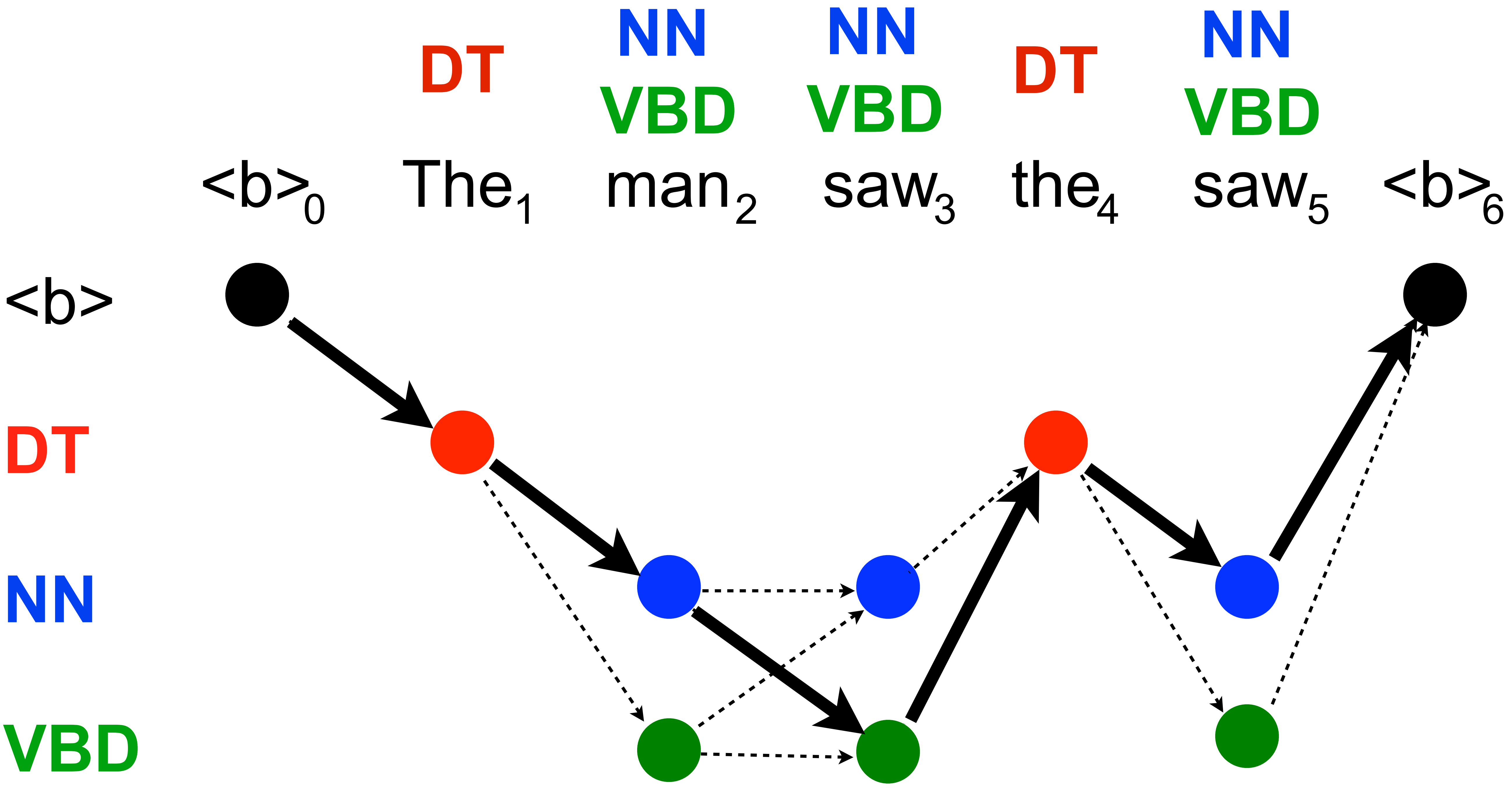
<b>

DT

NN

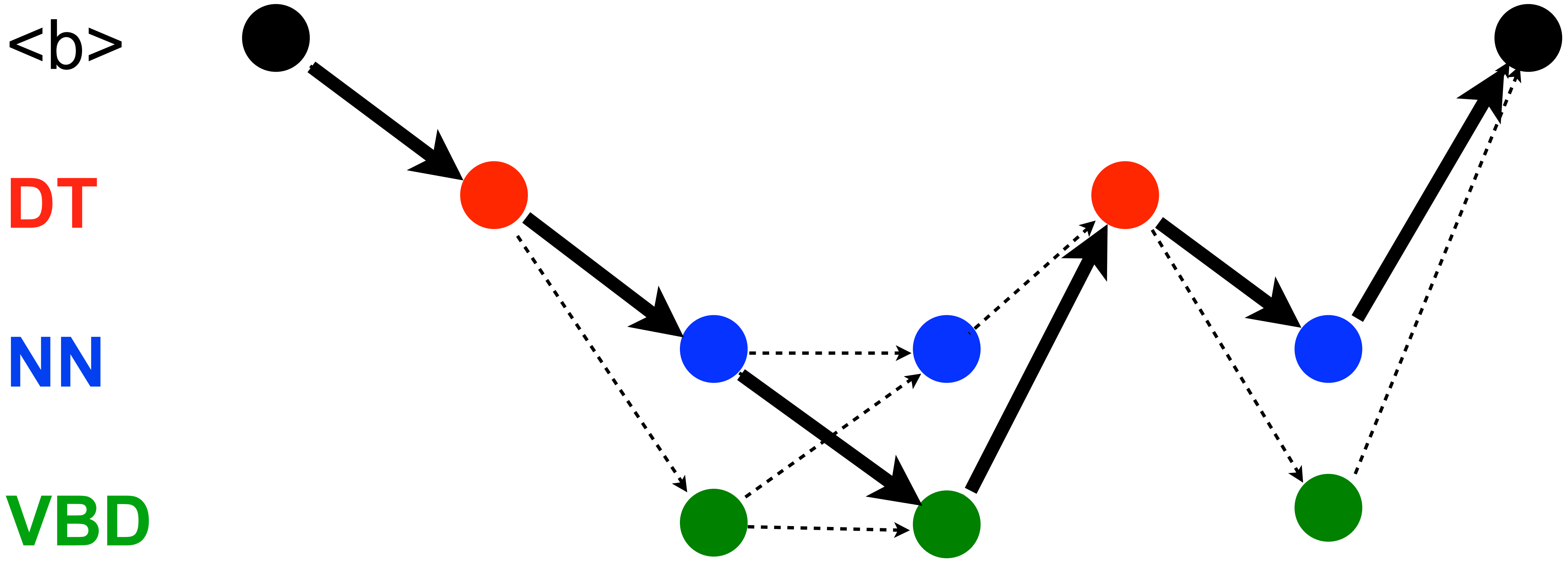
VBD

Model Minimization



Model Minimization

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



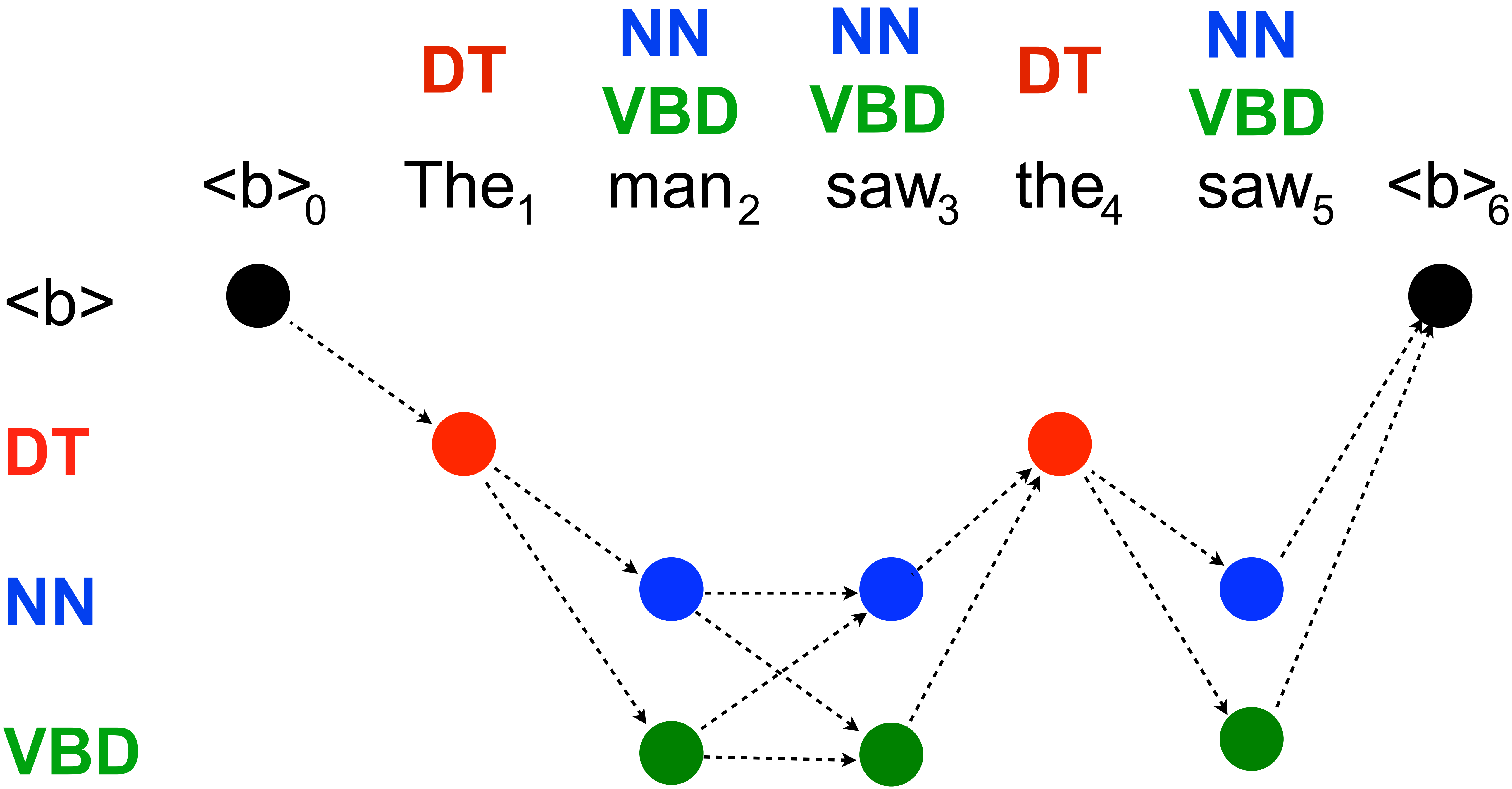
<b>

DT

NN

VBD

Model Minimization



DT NN VBD NN VBD DT NN VBD

<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

<b>

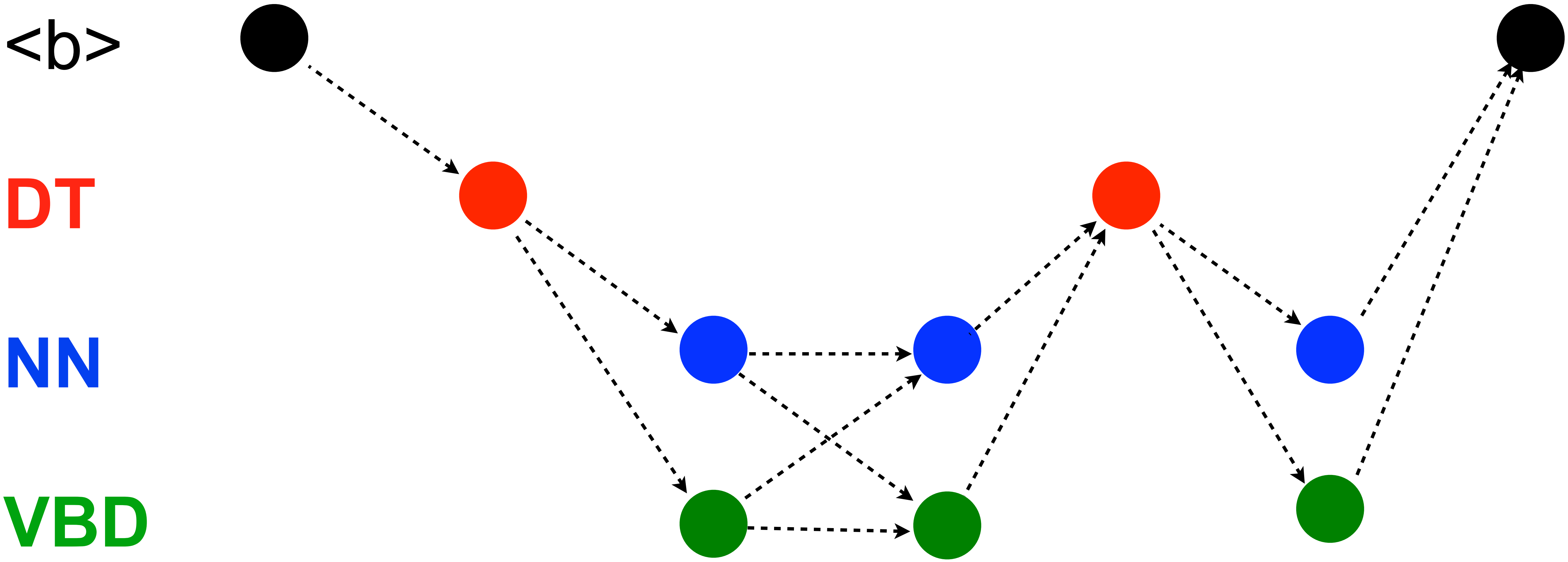
DT

NN

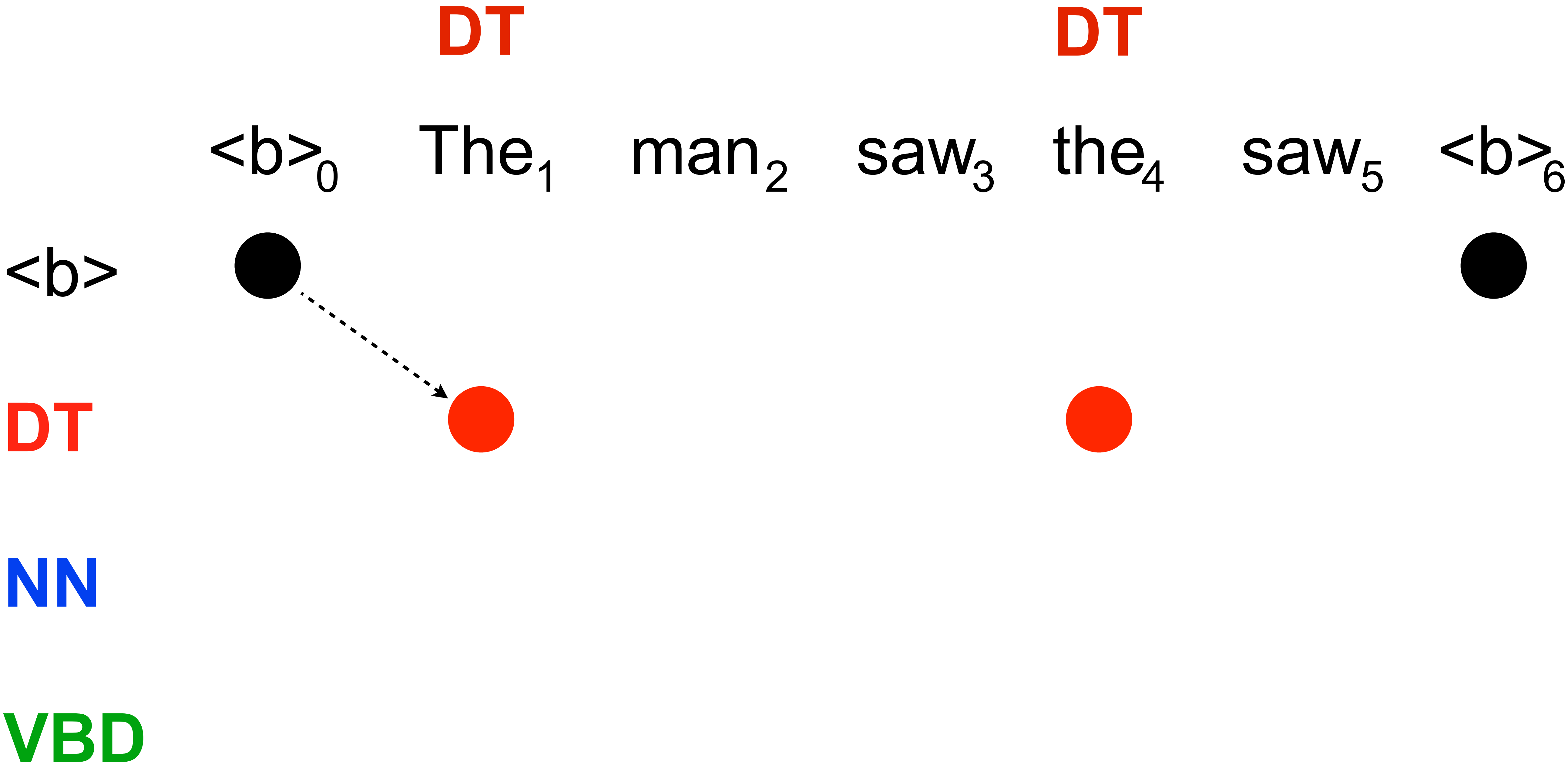
VBD

Model Minimization

	DT	NN VBD	NN VBD	DT	NN VBD	
<b>_0	The_1	man_2	saw_3	the_4	saw_5	<b>_6



Model Minimization



DT DT

<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

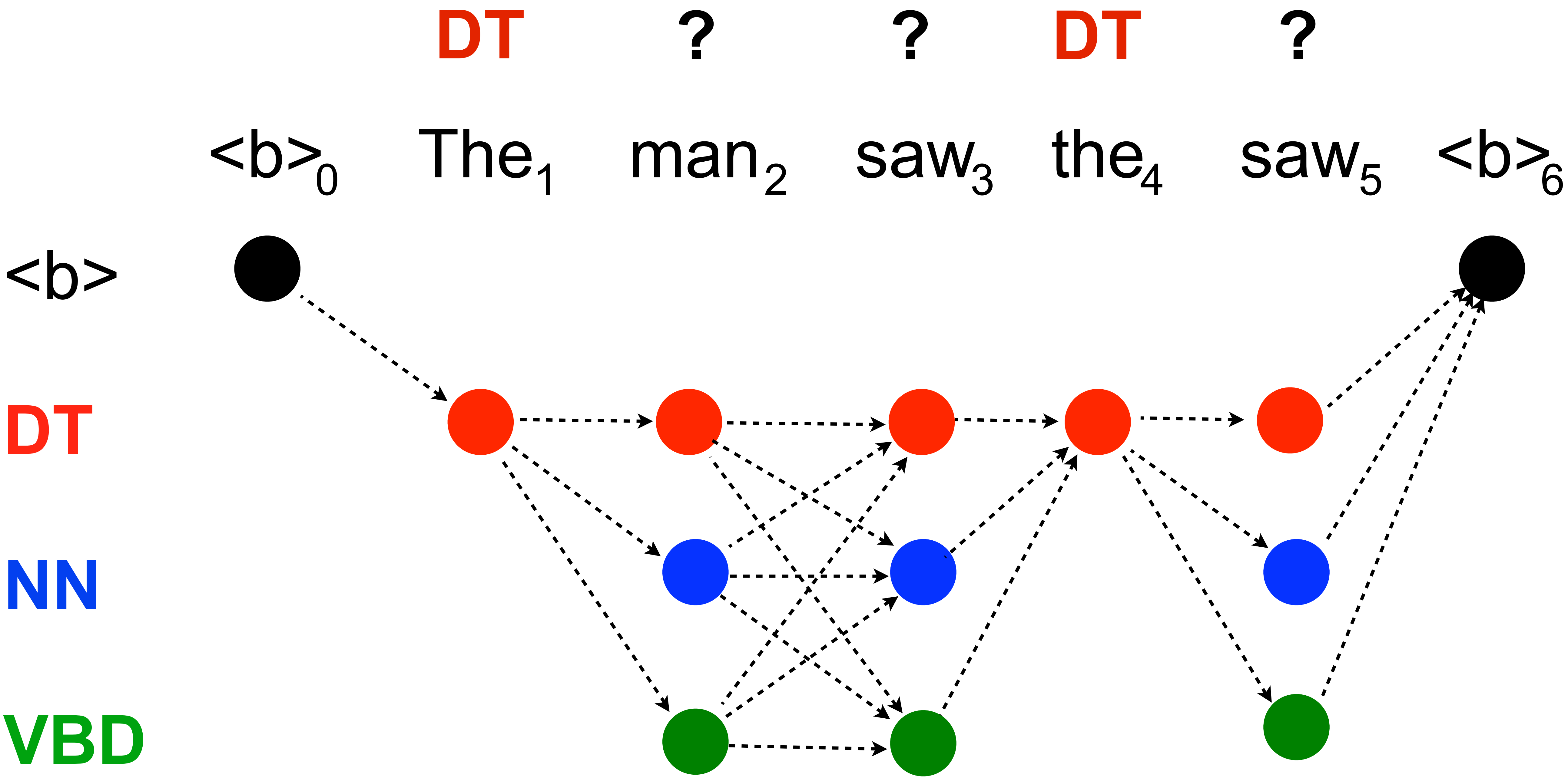
<b>

DT

NN

VBD

Model Minimization



DT ? ? DT ?

<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

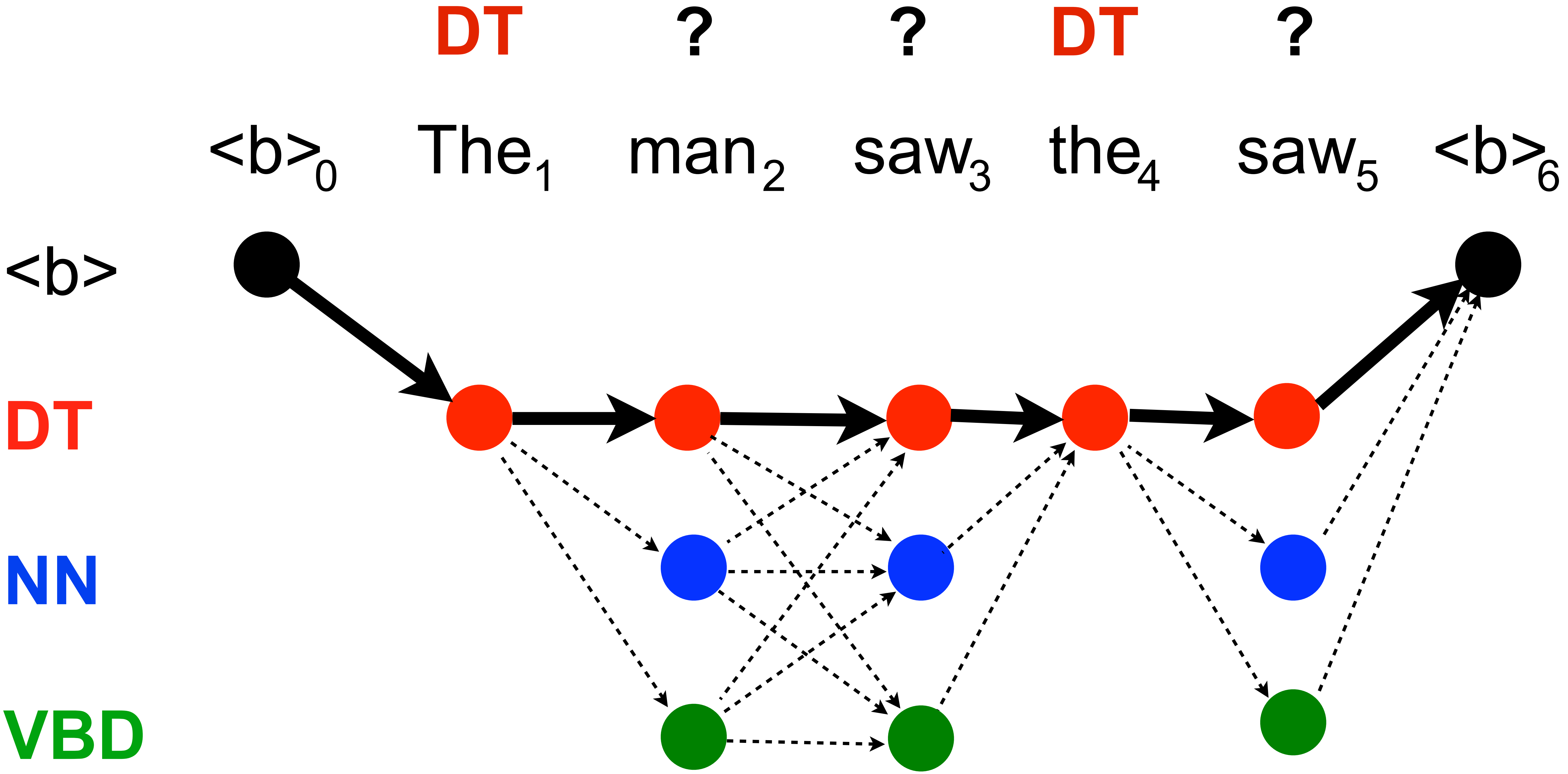
<b>

DT

NN

VBD

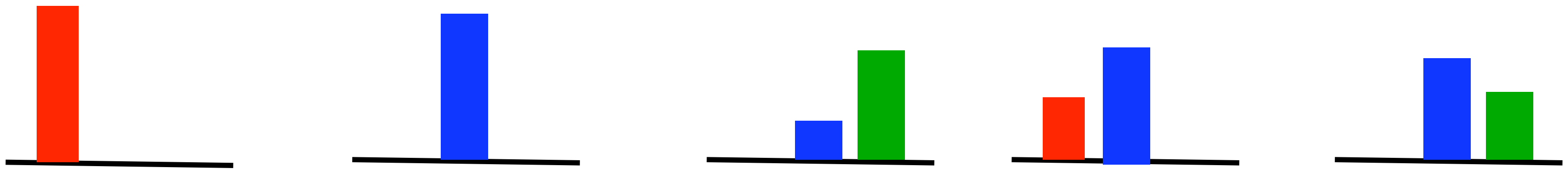
Model Minimization



Model Minimization

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

Model Minimization



<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

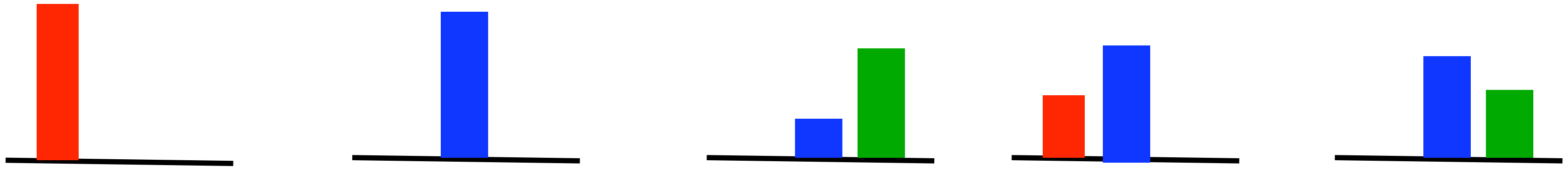
<b>

DT

NN

VBD

Model Minimization



<b>$_0$ The$_1$ man$_2$ saw$_3$ the$_4$ saw$_5$ <b>$_6$

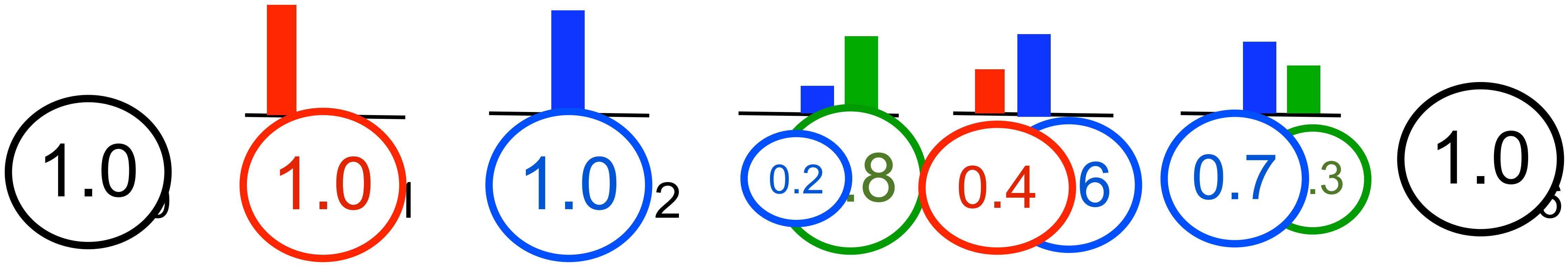
<b>

DT

NN

VBD

Model Minimization



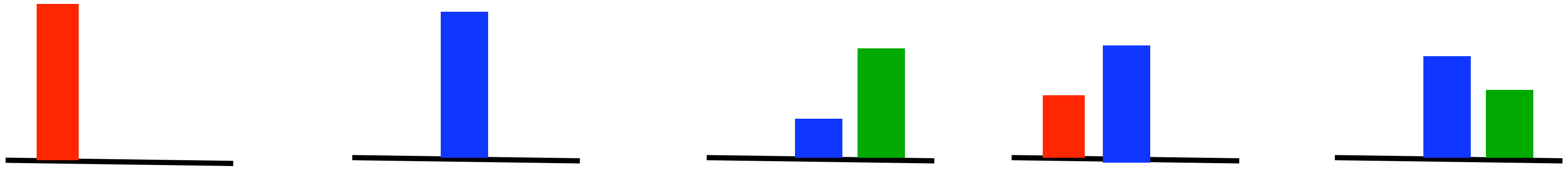
<b>

DT

NN

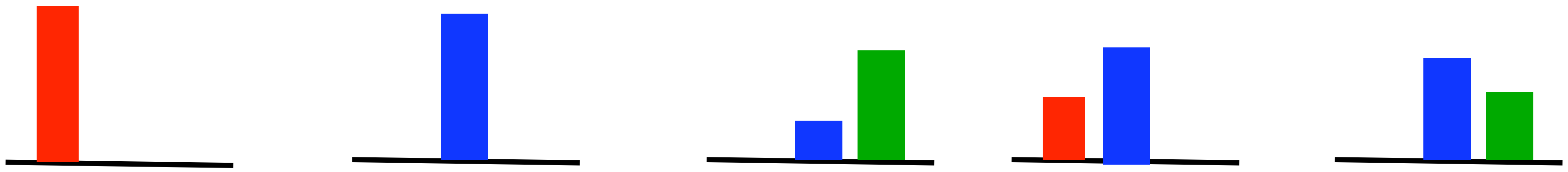
VBD

Model Minimization



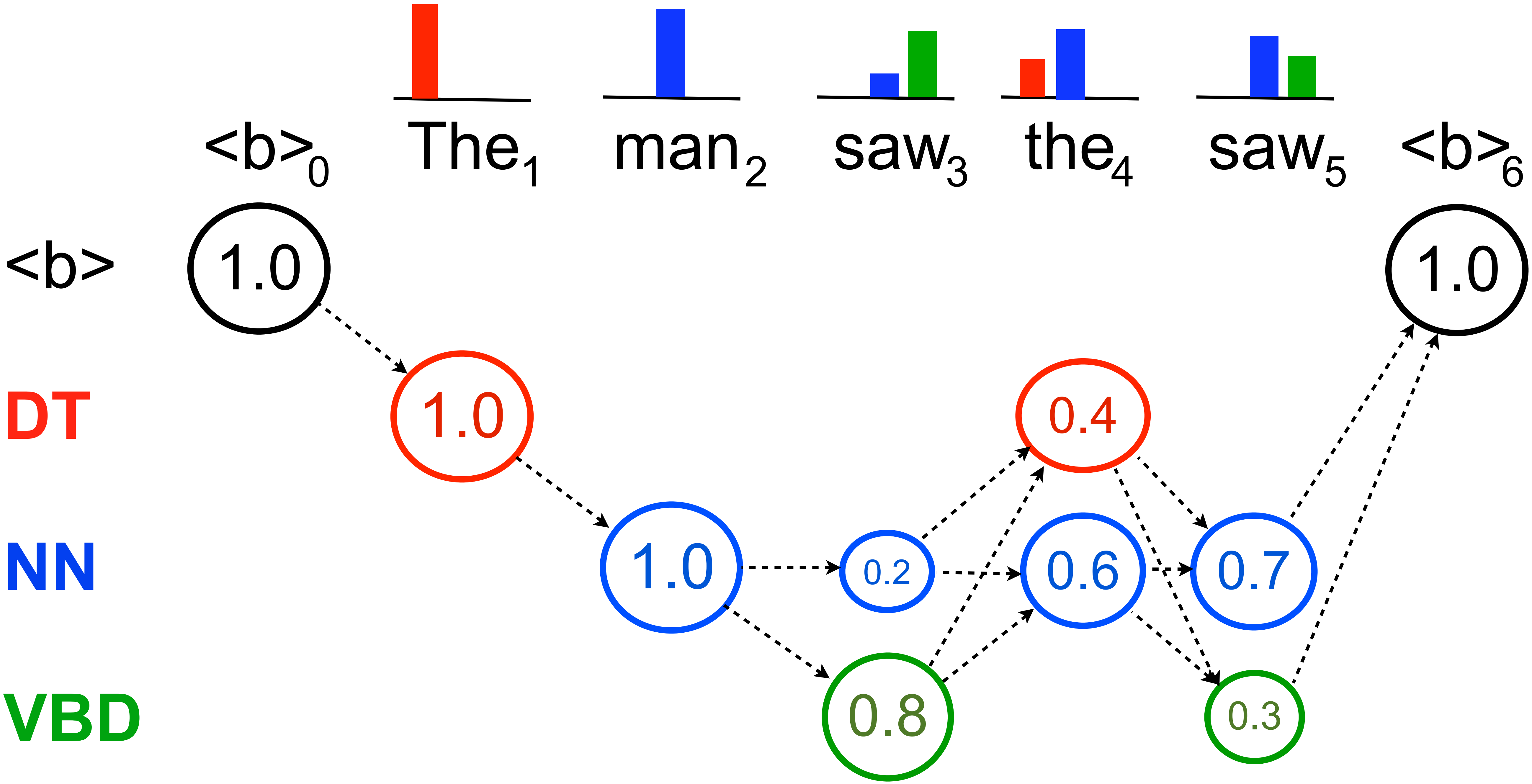
	<b>$_0$	The$_1$	man$_2$	saw$_3$	the$_4$	saw$_5$	<b>$_6$
<b>	1.0						1.0
DT		1.0			0.4		
NN			1.0	0.2	0.6	0.7	
VBD				0.8		0.3	

Model Minimization

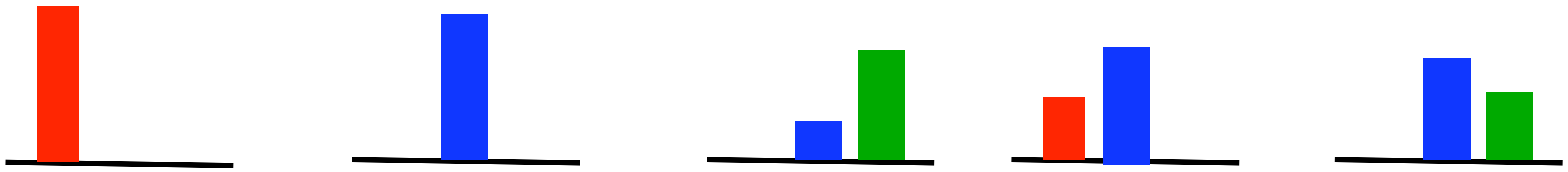


	<b>$_0$	The$_1$	man$_2$	saw$_3$	the$_4$	saw$_5$	<b>$_6$
<b>	1.0						1.0
DT		1.0			0.4		
NN			1.0	0.2	0.6	0.7	
VBD				0.8		0.3	

Model Minimization



Model Minimization



	<b>	The	man	saw	the	saw	<b>
<b>	1.0						1.0
DT		1.0			0.4		
NN			1.0	0.2	0.6	0.7	
VBD				0.8		0.3	

Model Minimization

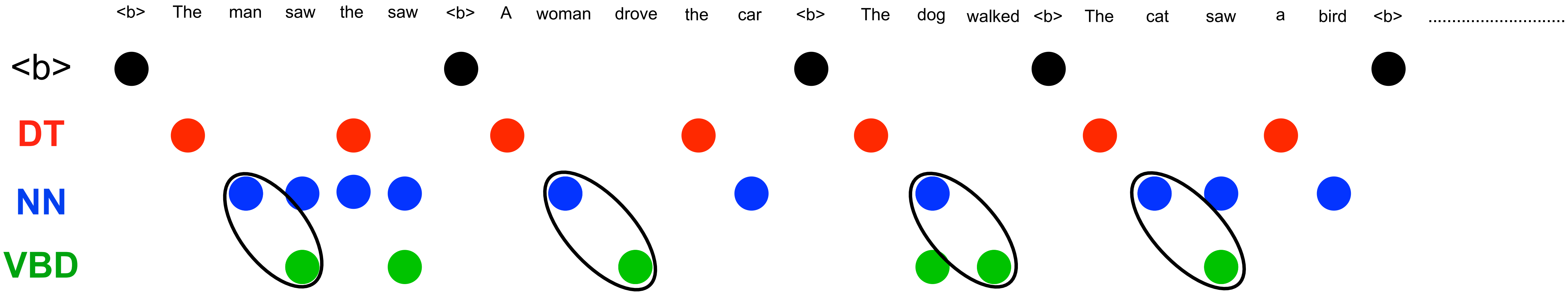
	<b>	The	man	saw	the	saw	<b>	A	woman	drove	the	car	<b>	The	dog	walked	<b>	The	cat	saw	a	bird	<b>	....
<b>	●						●						●				●						●	
DT		●			●			●			●			●				●			●			
NN			●	●	●	●			●			●			●				●	●		●		
VBD				●		●				●					●	●				●				

Model Minimization

	<b>	The	man	saw	the	saw	<b>	A	woman	drove	the	car	<b>	The	dog	walked	<b>	The	cat	saw	a	bird	<b>	....
<b>	●						●						●				●						●	
DT		●			●			●			●			●				●			●			
NN			●	●	●	●			●			●			●				●	●		●		
VBD				●		●				●					●	●				●				

$$f(NN \rightarrow VBD) \Big\langle$$

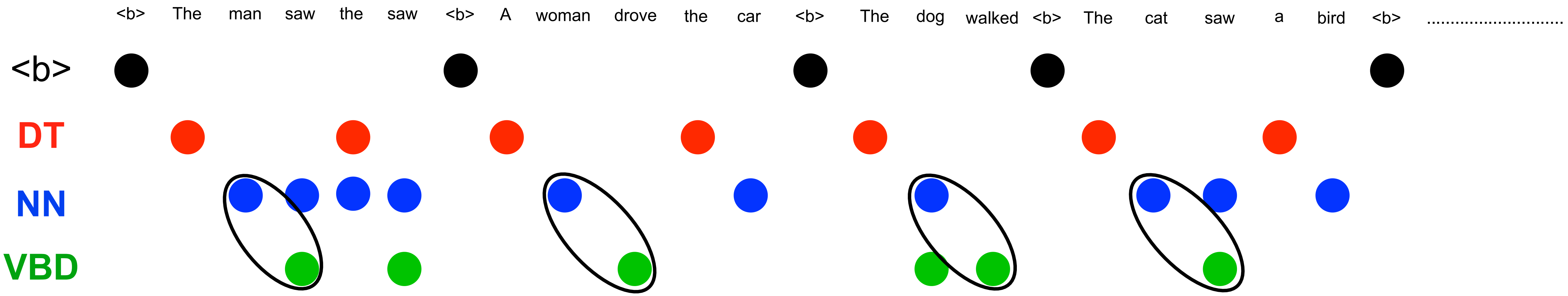
Model Minimization



<b> The man saw the saw <b> A woman drove the car <b> The dog walked <b> The cat saw a bird <b> ....
<b>
DT
NN
VBD

$f(\text{NN} \rightarrow \text{VBD})$ tag bigram occurrences

Model Minimization


<b> The man saw the saw <b> A woman drove the car <b> The dog walked <b> The cat saw a bird <b> ....
<b>
DT
NN
VBD

f(NN → VBD)

tag bigram occurrences

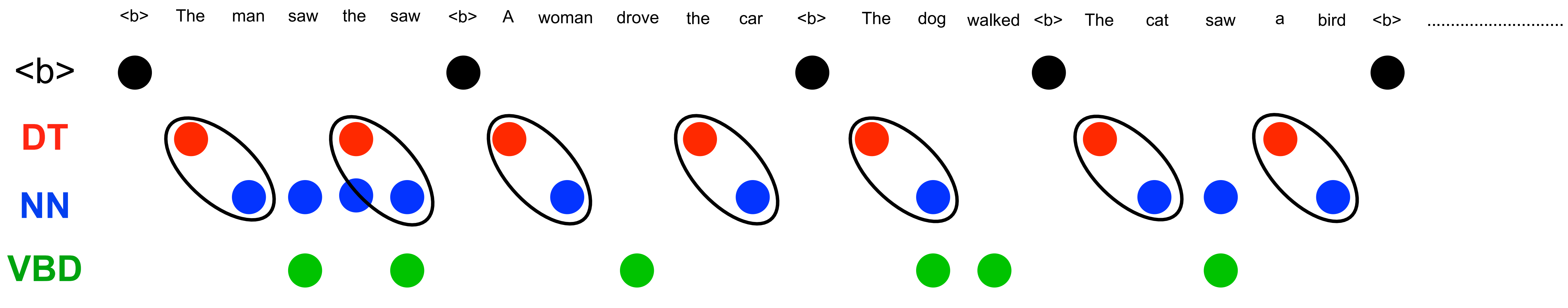
weights on their nodes

Model Minimization

	<b>	The	man	saw	the	saw	<b>	A	woman	drove	the	car	<b>	The	dog	walked	<b>	The	cat	saw	a	bird	<b>	....
<b>	●						●						●				●						●	
DT		●			●			●			●			●				●			●			
NN			●	●	●	●			●			●			●				●	●		●		
VBD				●		●				●					●	●				●				

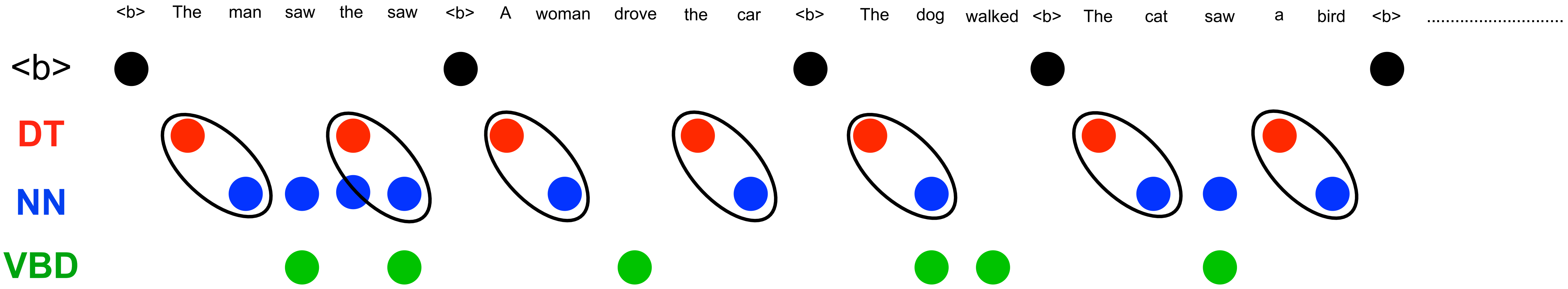
f(DT → NN)

Model Minimization



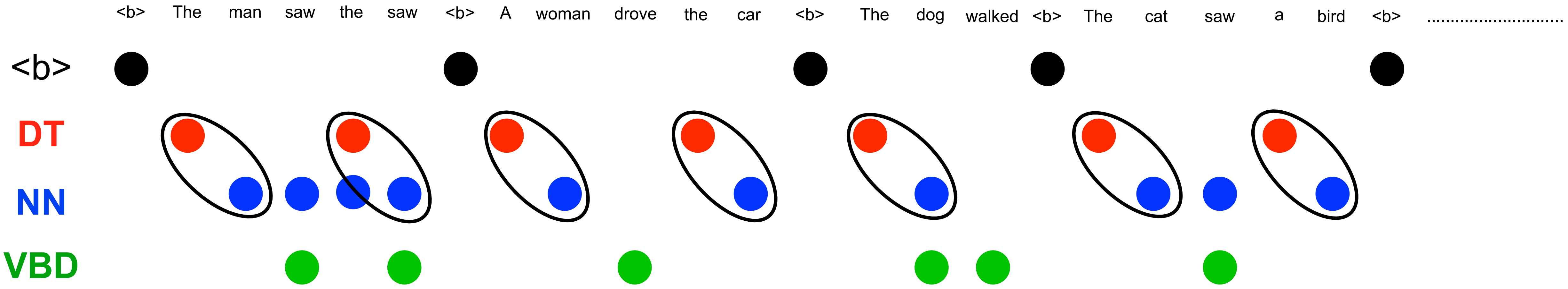
f(DT → NN)

Model Minimization


<b> The man saw the saw <b> A woman drove the car <b> The dog walked <b> The cat saw a bird <b> ....
<b>
DT
NN
VBD

f(DT → NN) ✓

Model Minimization



<b> The man saw the saw <b> A woman drove the car <b> The dog walked <b> The cat saw a bird <b> ....

<b>

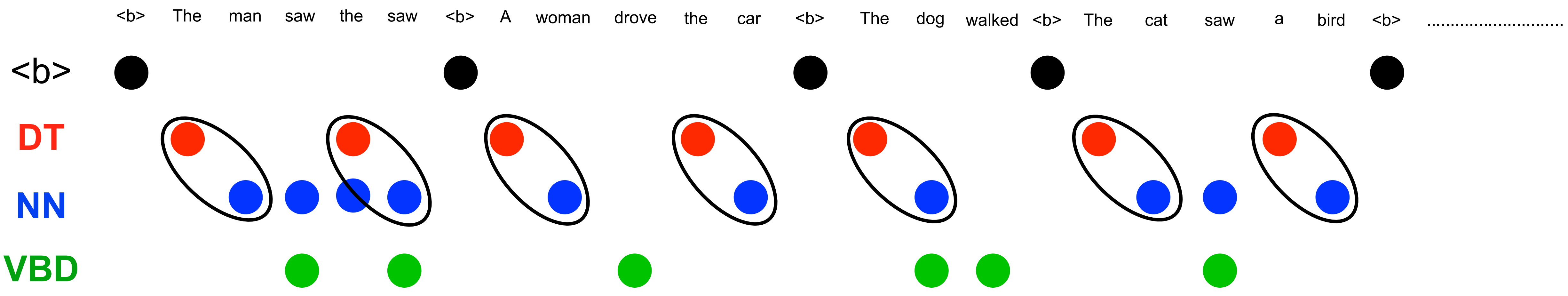
DT

NN

VBD

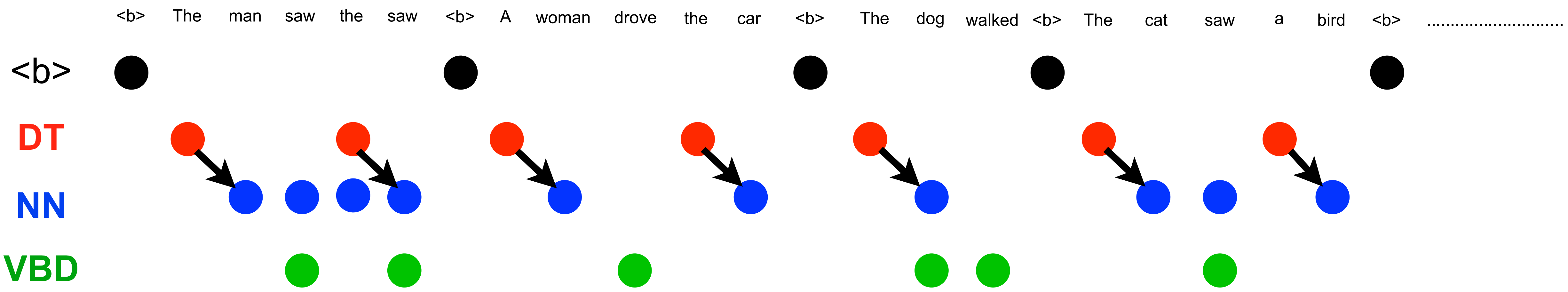
f(DT → NN) ✓

Model Minimization


<b> The man saw the saw <b> A woman drove the car <b> The dog walked <b> The cat saw a bird <b> ....
<b>
DT
NN
VBD

f(DT → NN) ✓

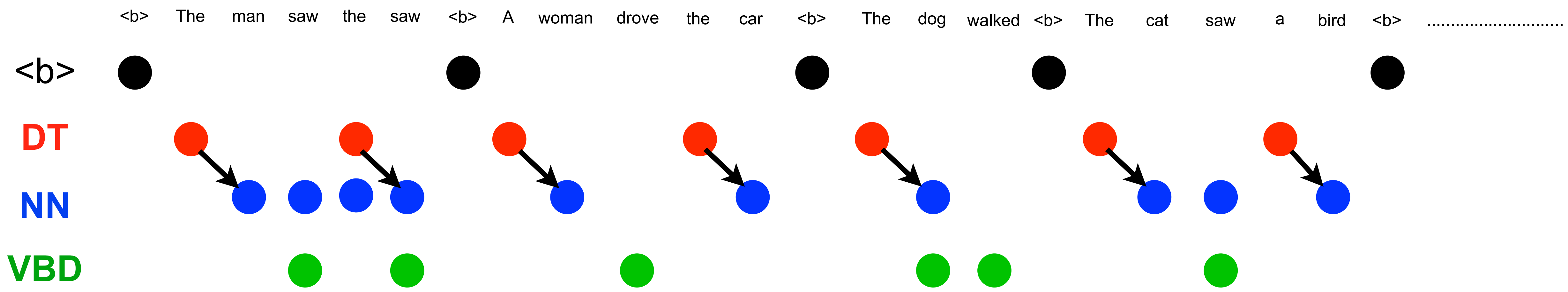
Model Minimization



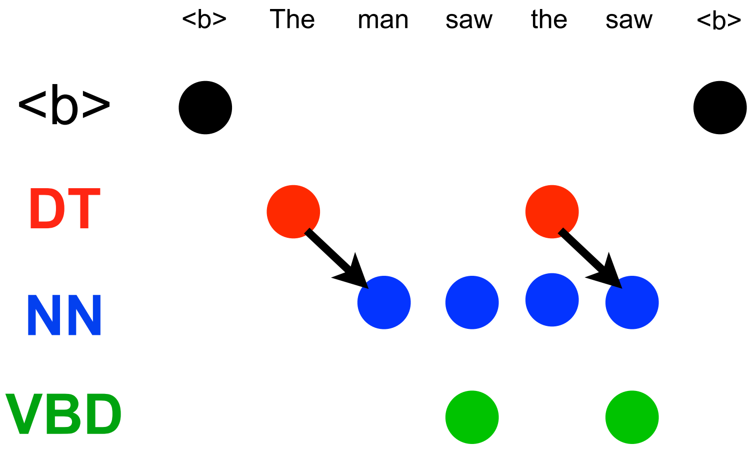
<b> The man saw the saw <b> A woman drove the car <b> The dog walked <b> The cat saw a bird <b> ....
<b>
DT
NN
VBD

f(DT → NN) ✓

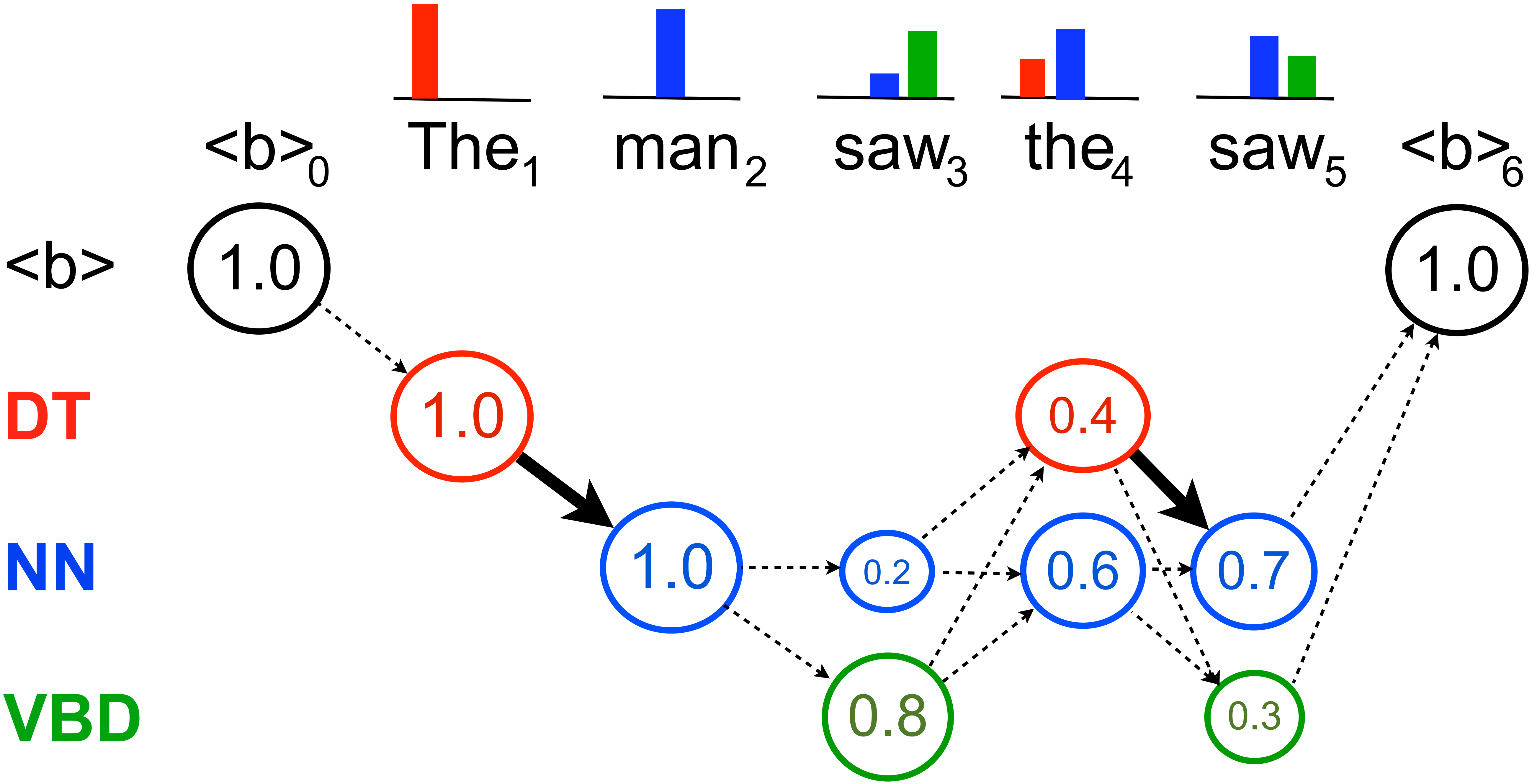
Model Minimization



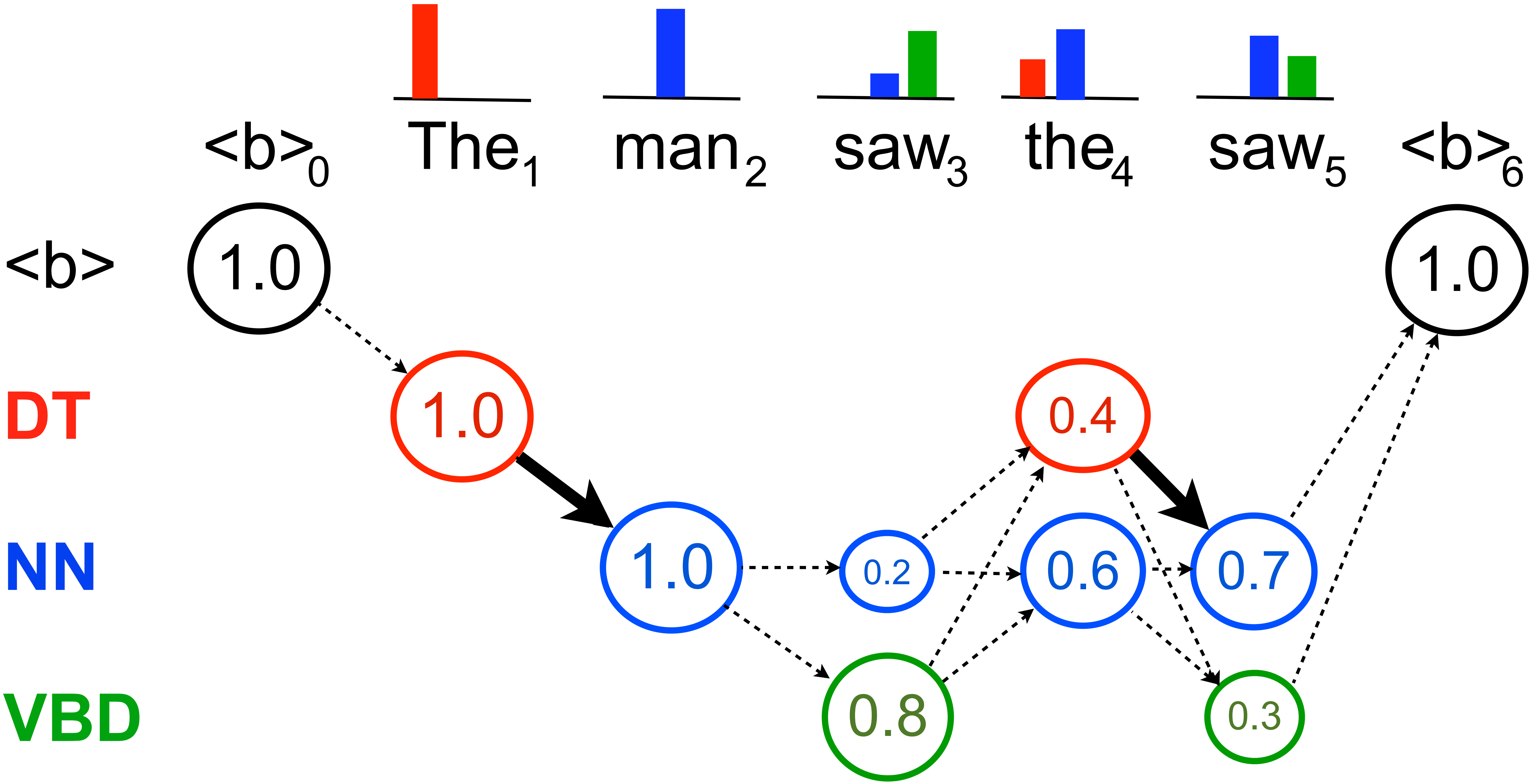
Model Minimization



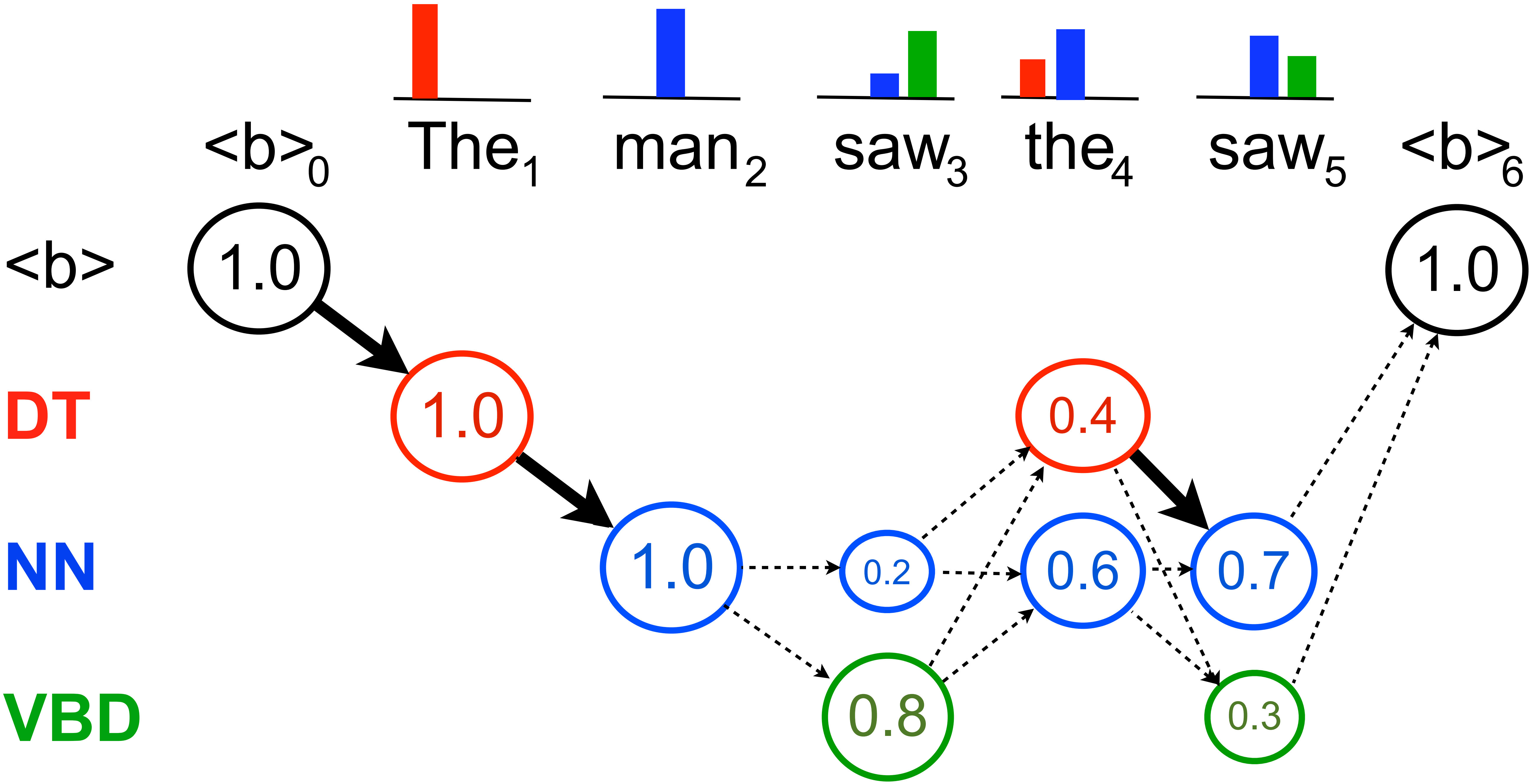
Model Minimization



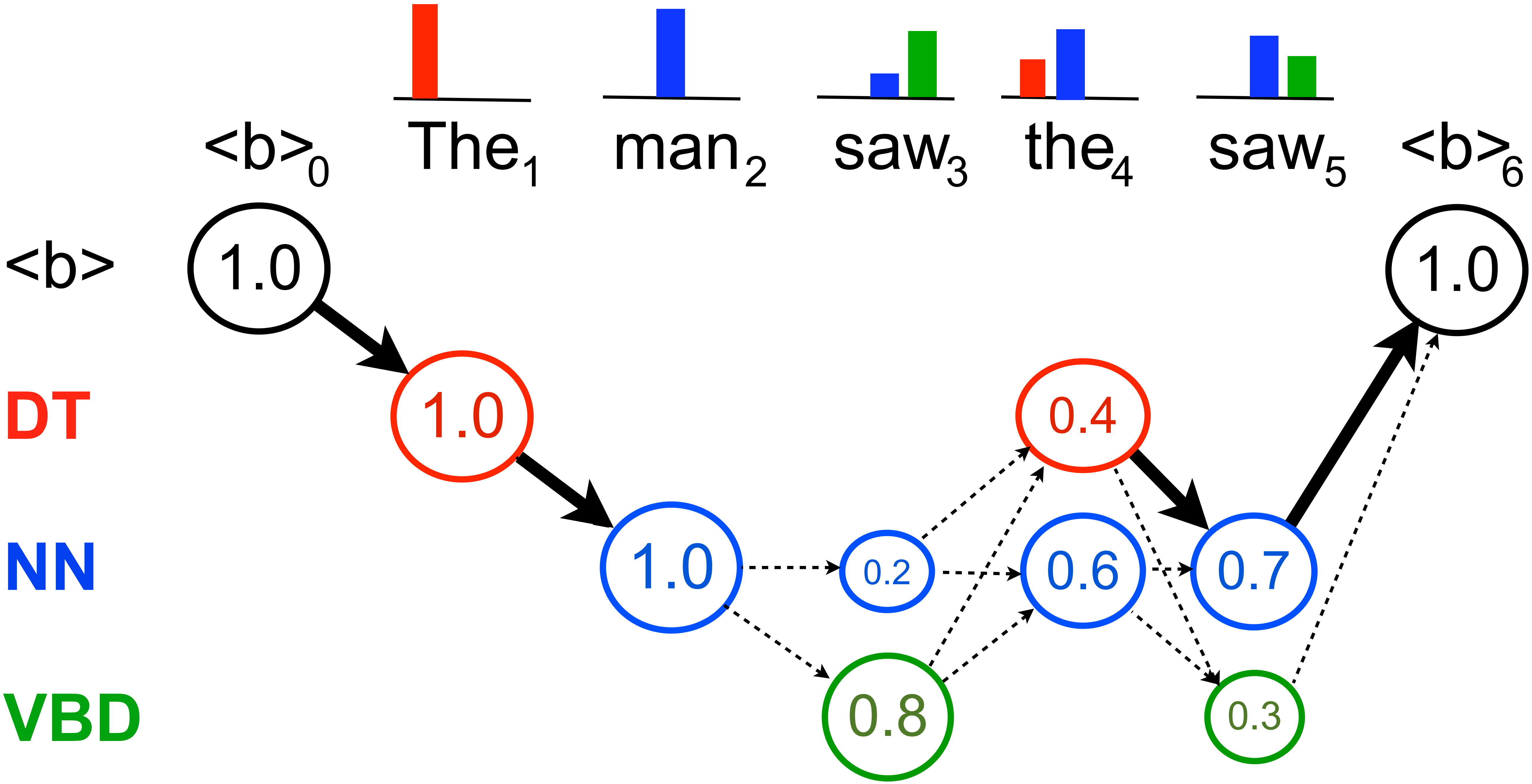
Model Minimization



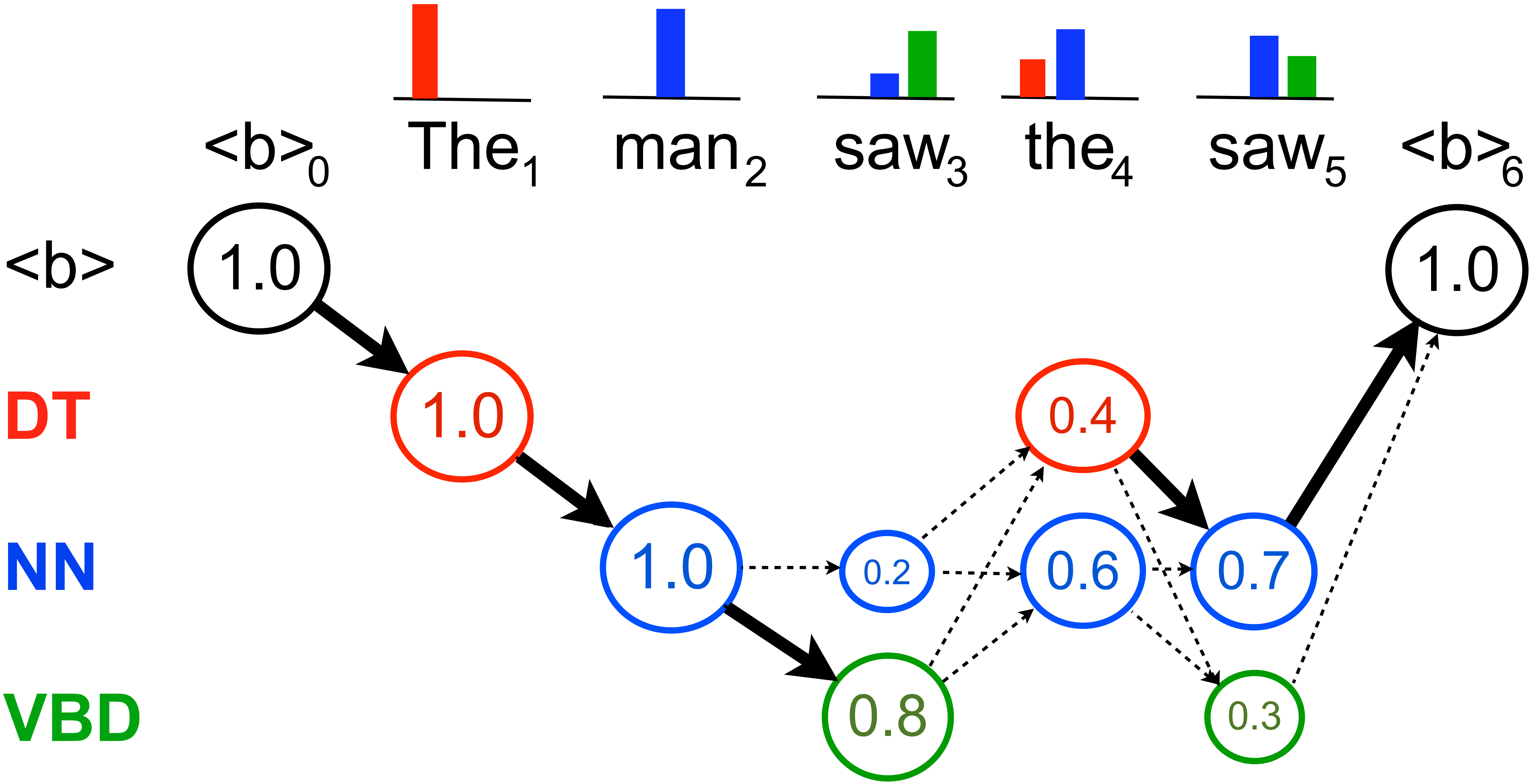
Model Minimization



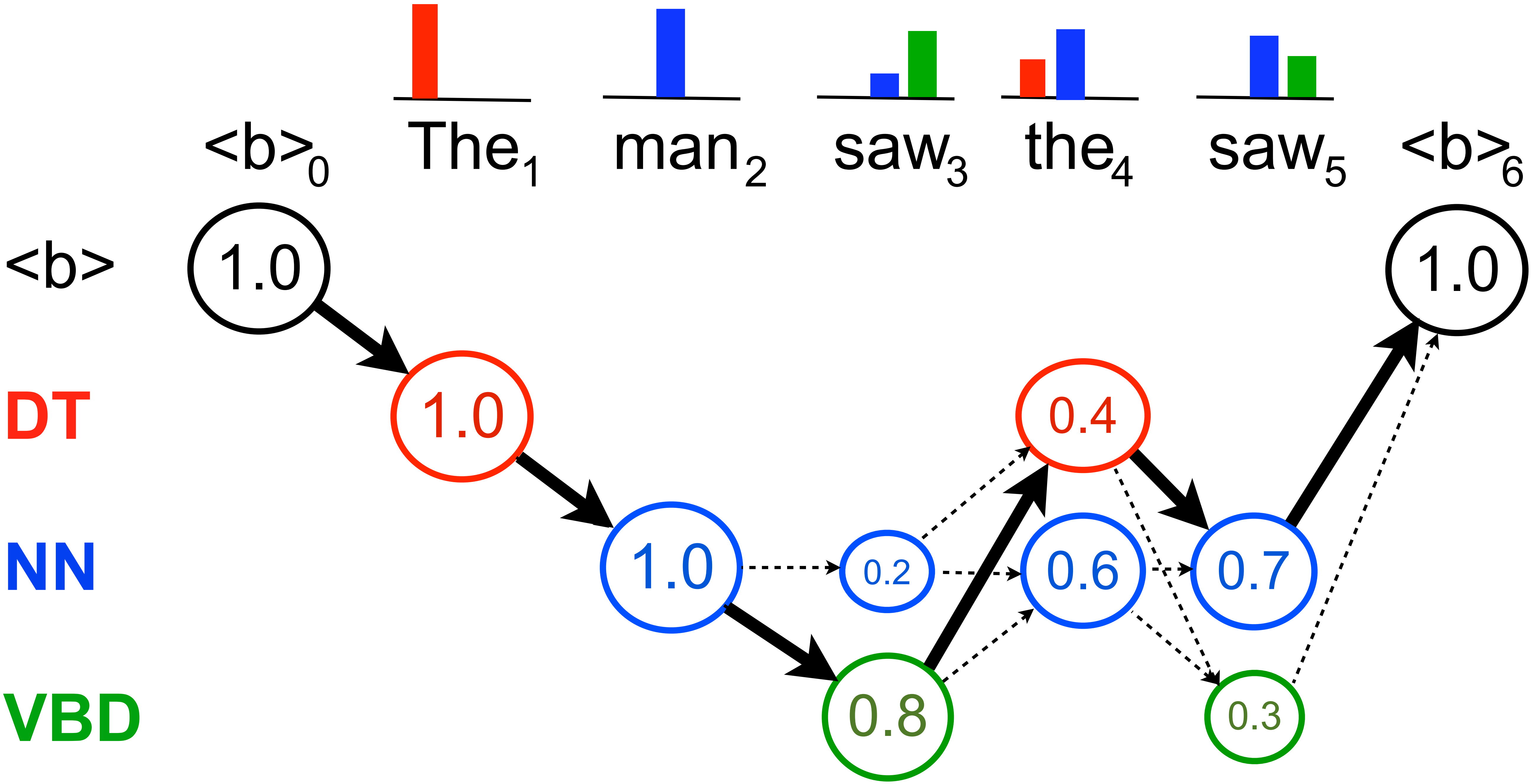
Model Minimization



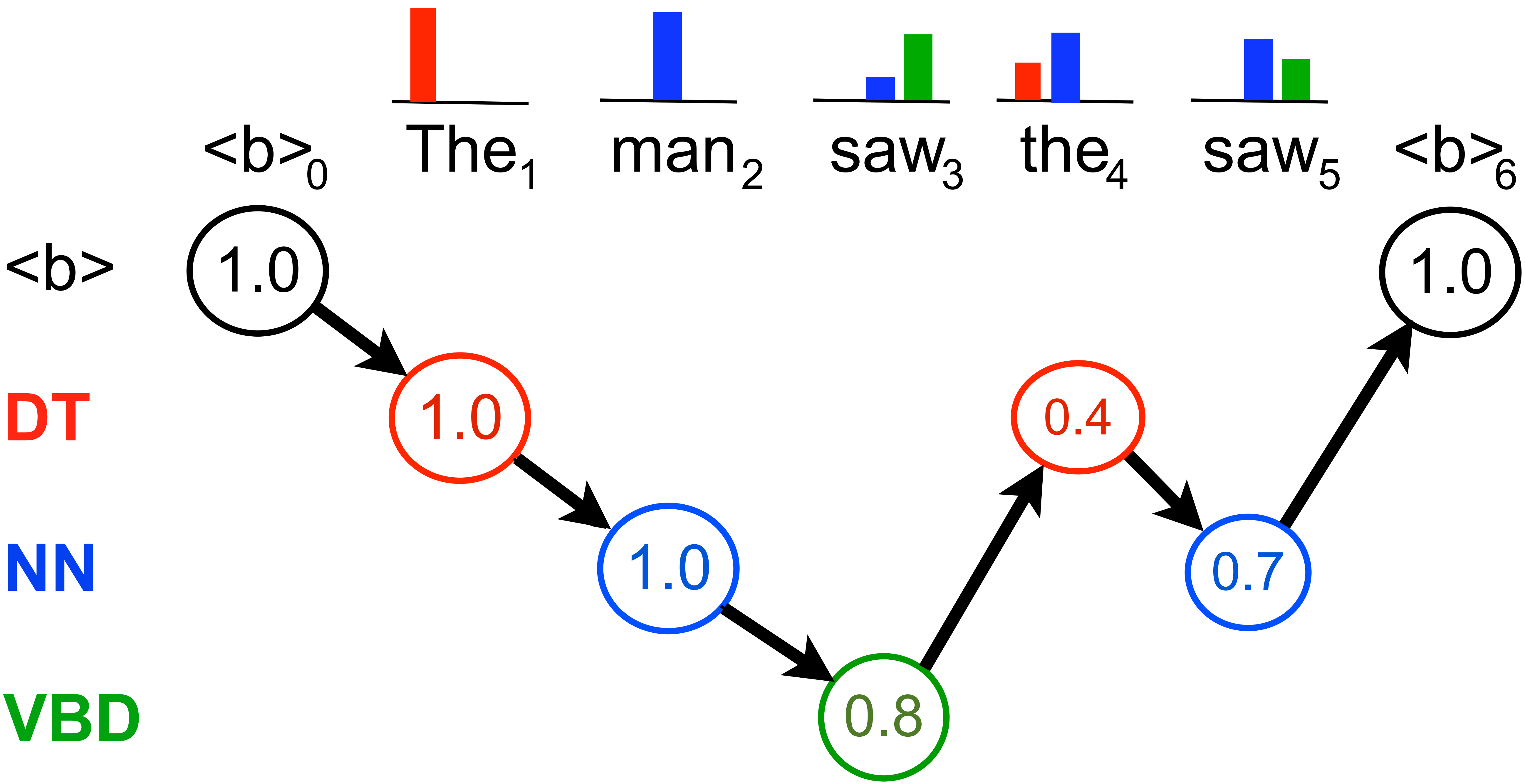
Model Minimization



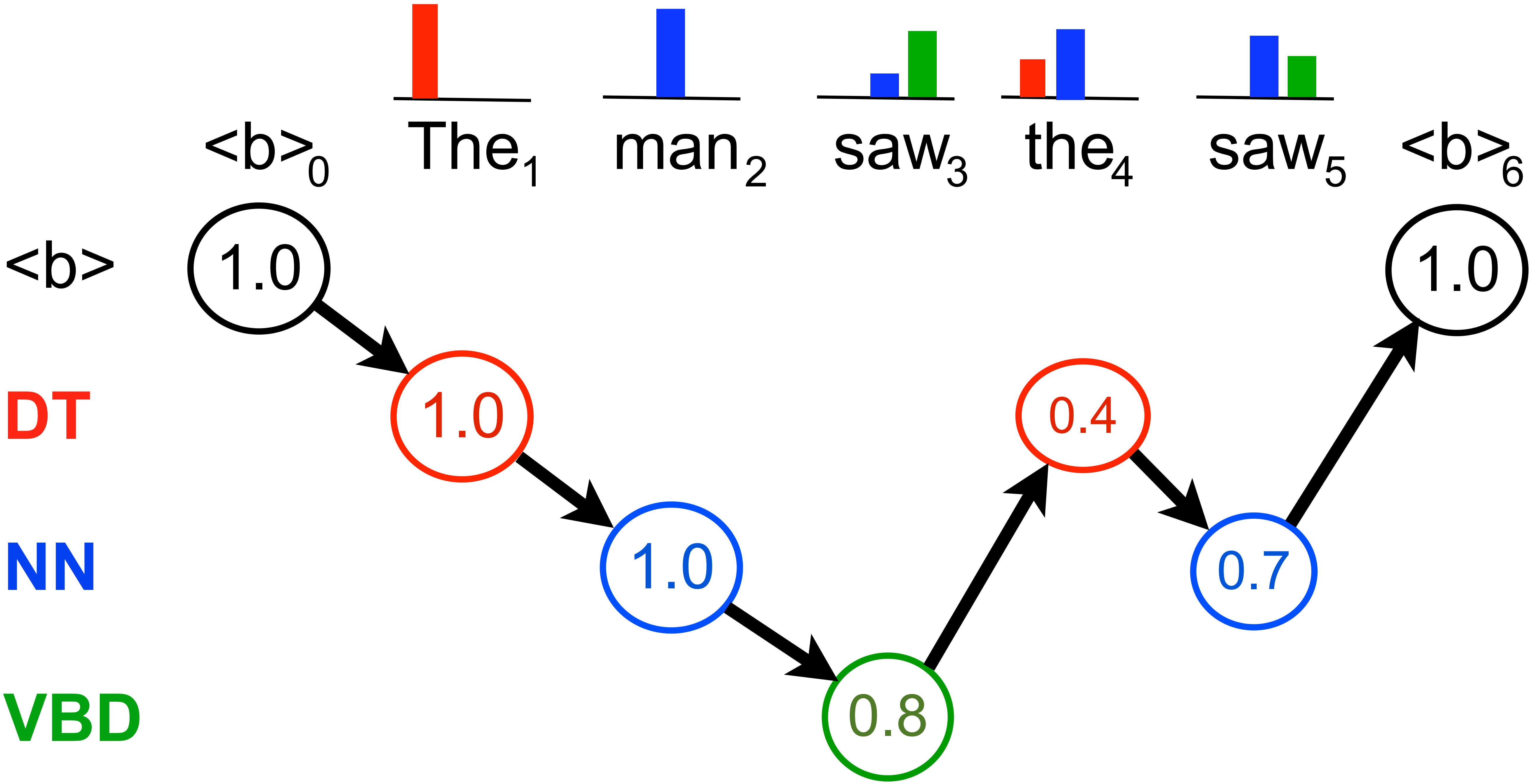
Model Minimization



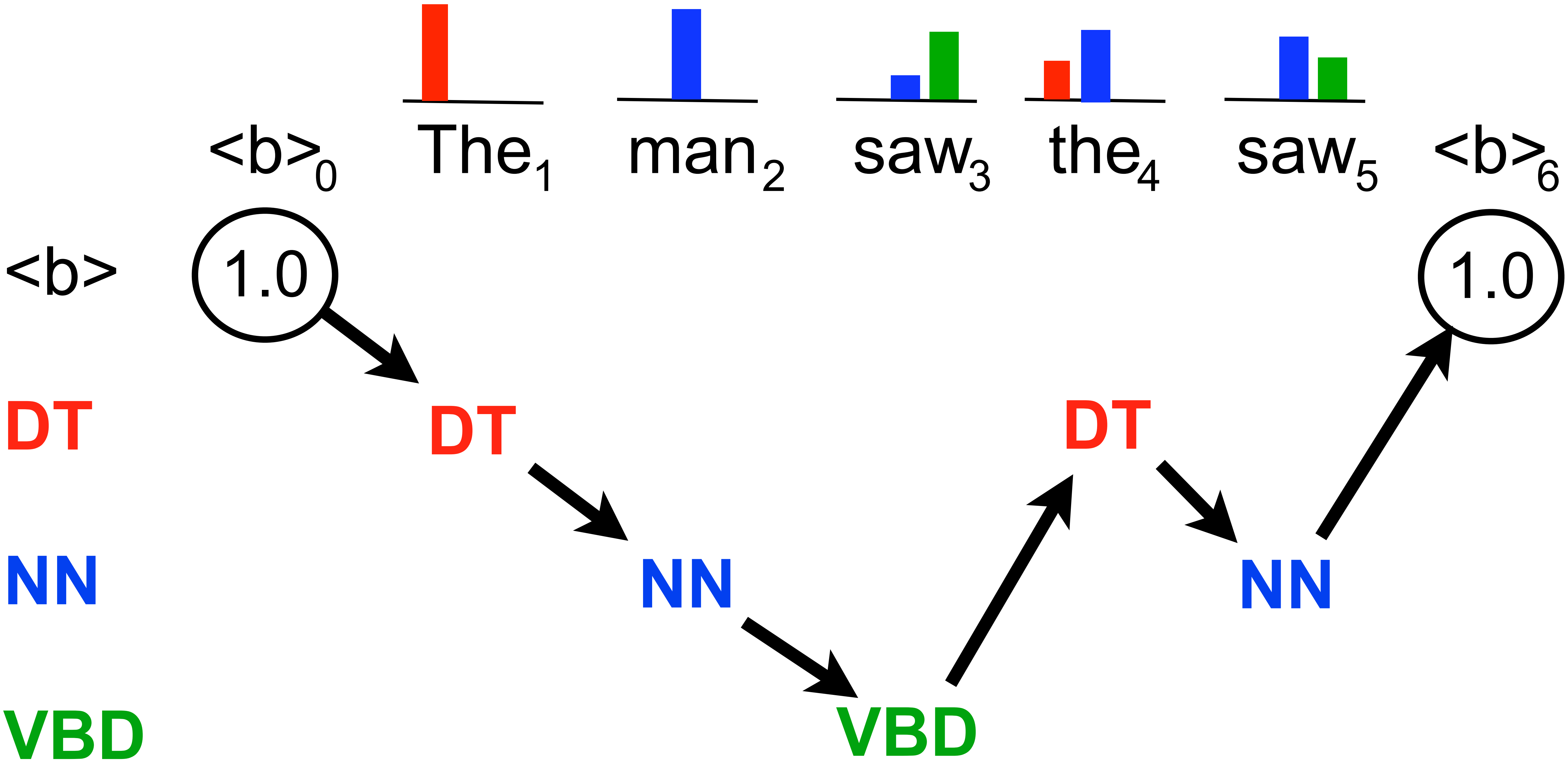
Model Minimization



Model Minimization



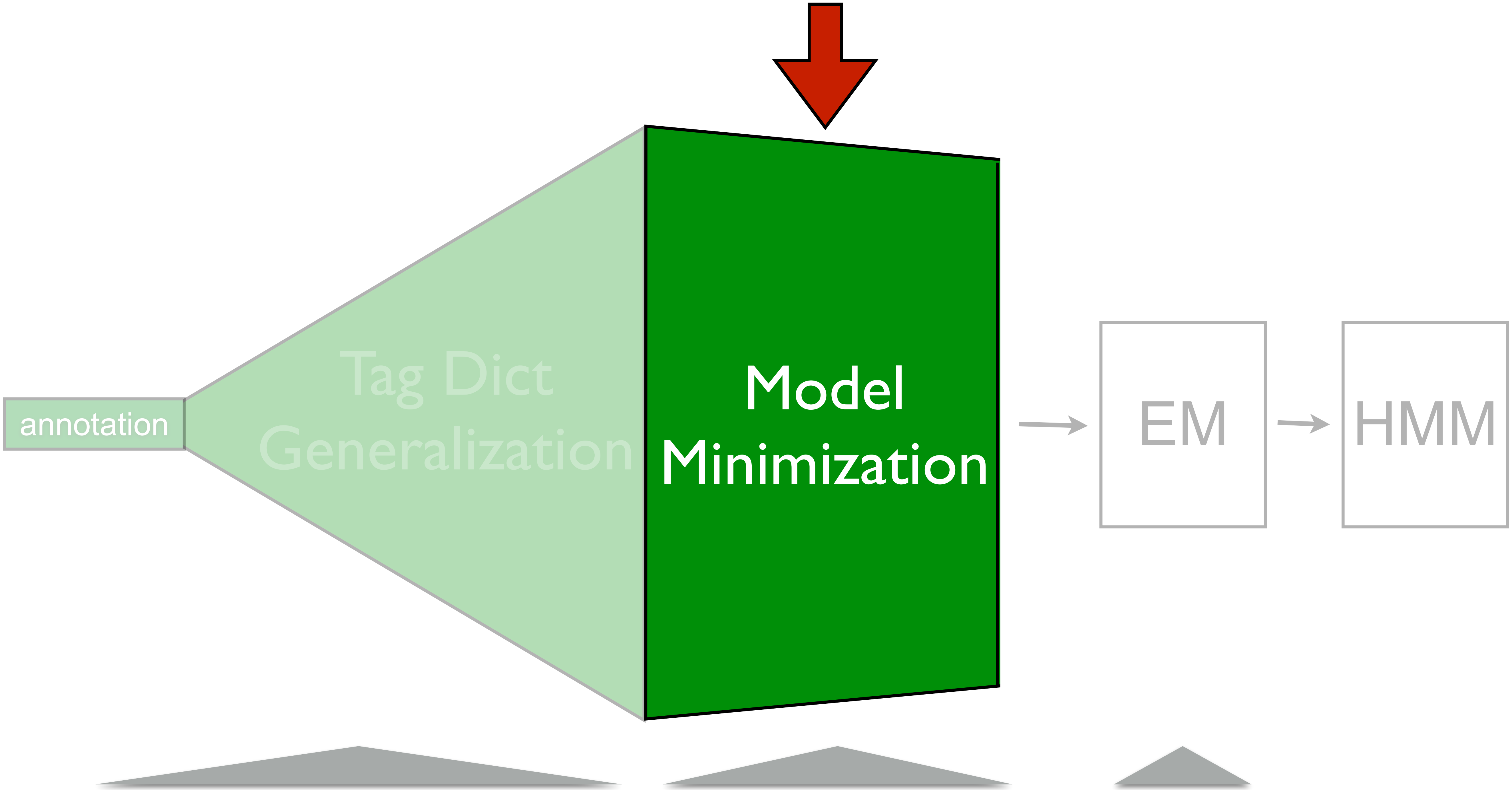
Model Minimization



Model Minimization

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

Our Approach



annotation

Tag Dict Generalization

Model Minimization

EM

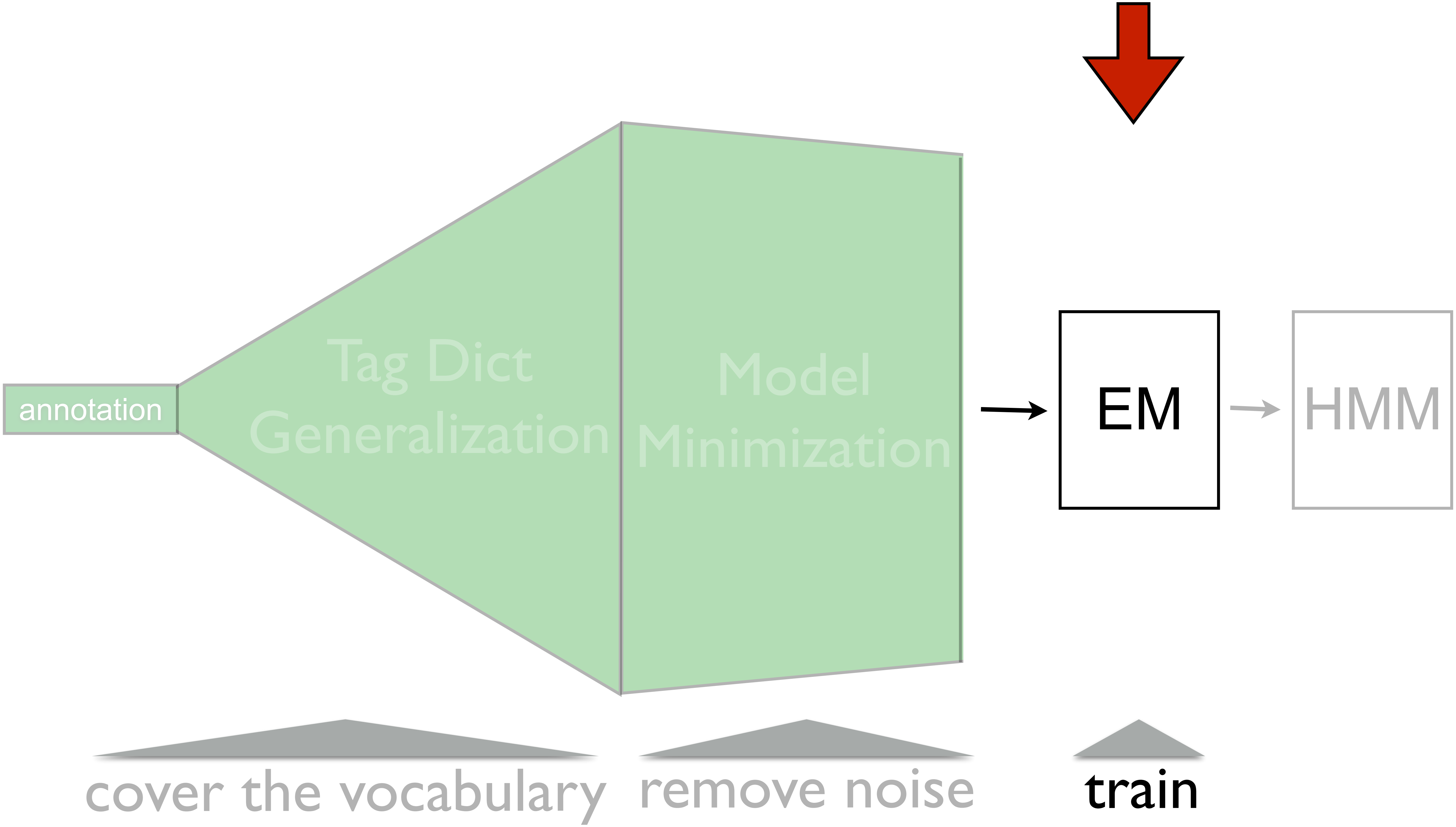
HMM

cover the vocabulary

remove noise

train

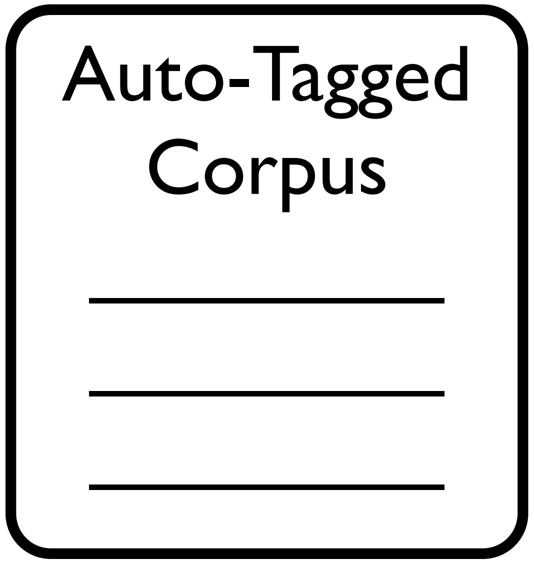
Our Approach



EM Training

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

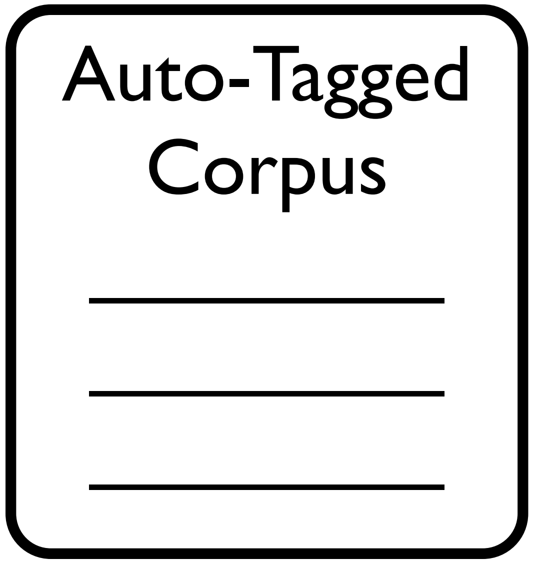
EM Training



Auto-Tagged Corpus

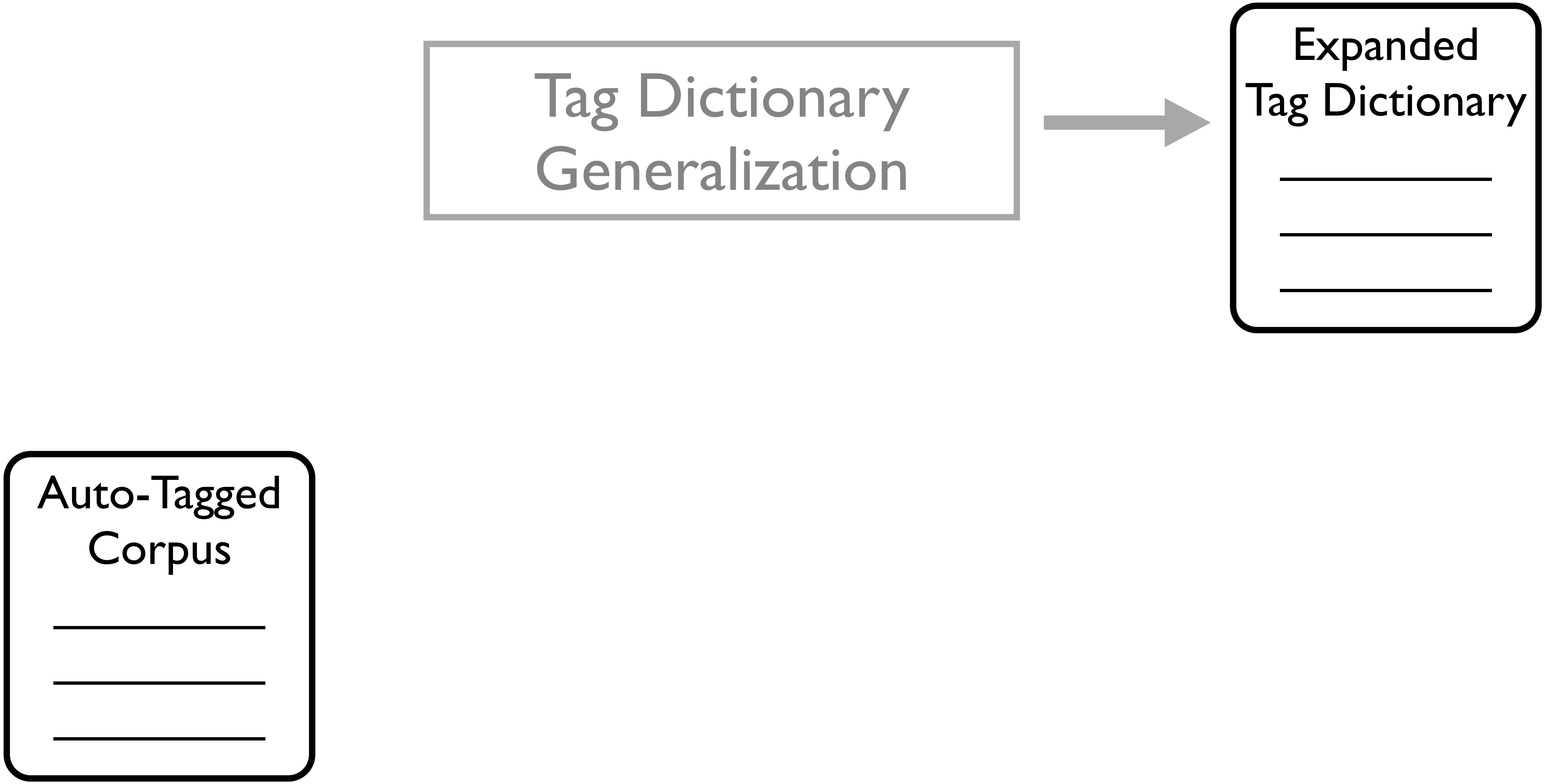
EM Training

Expanded
Tag Dictionary



Auto-Tagged
Corpus

EM Training

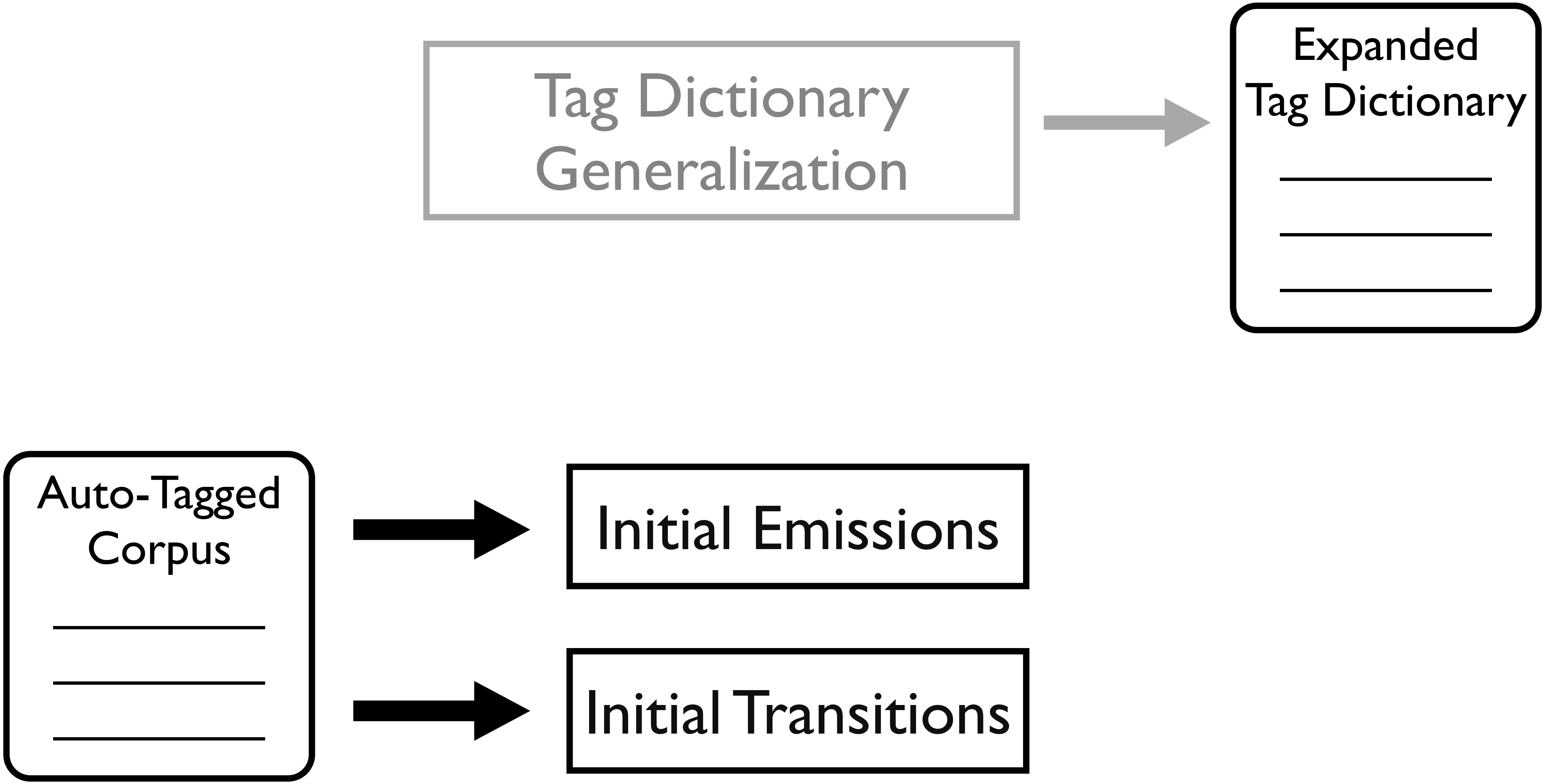


Tag Dictionary Generalization

Expanded Tag Dictionary

Auto-Tagged Corpus

EM Training



Tag Dictionary Generalization

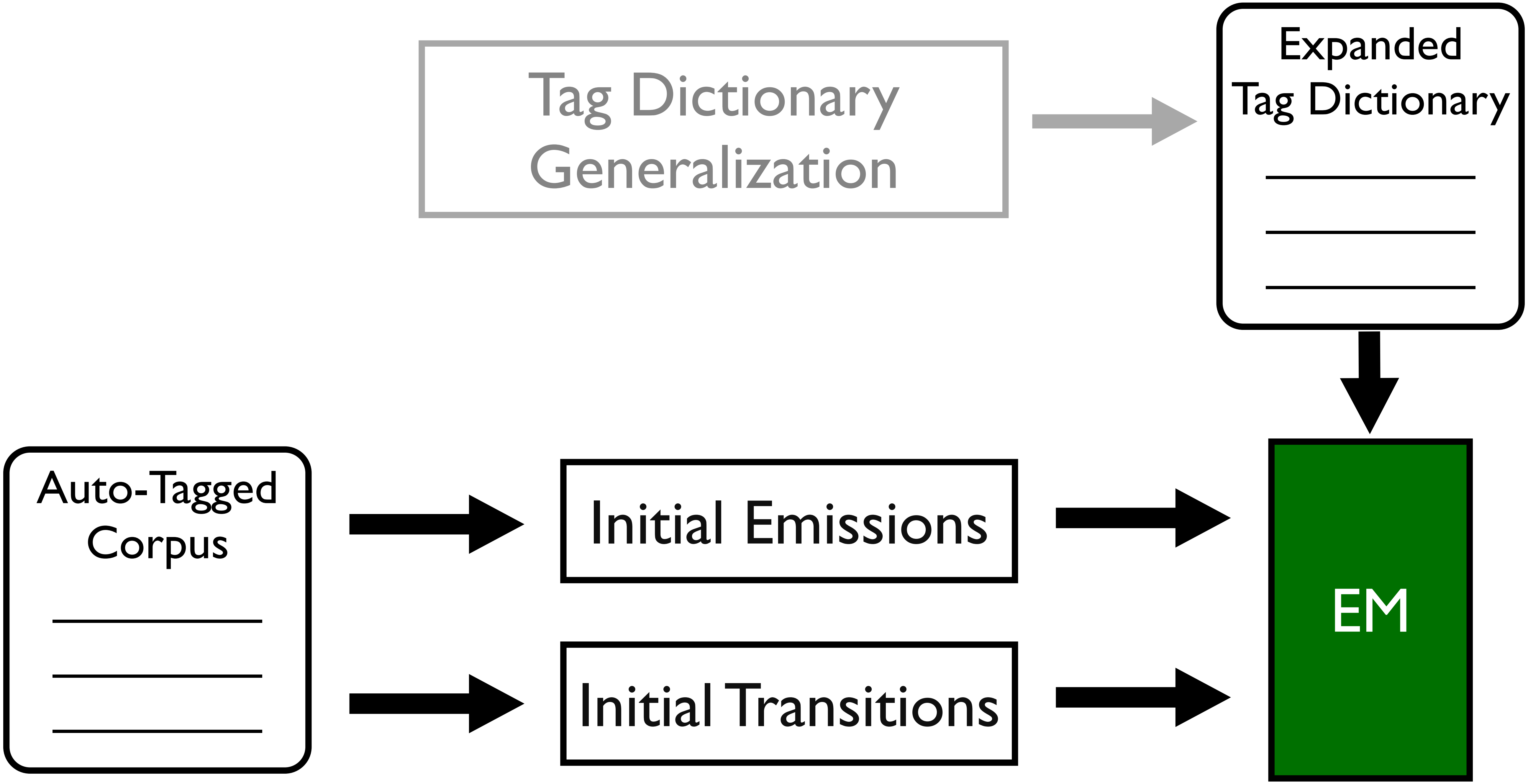
Expanded Tag Dictionary

Auto-Tagged Corpus

Initial Emissions

Initial Transitions

EM Training



Tag Dictionary Generalization

Expanded Tag Dictionary

Auto-Tagged Corpus

Initial Emissions

Initial Transitions

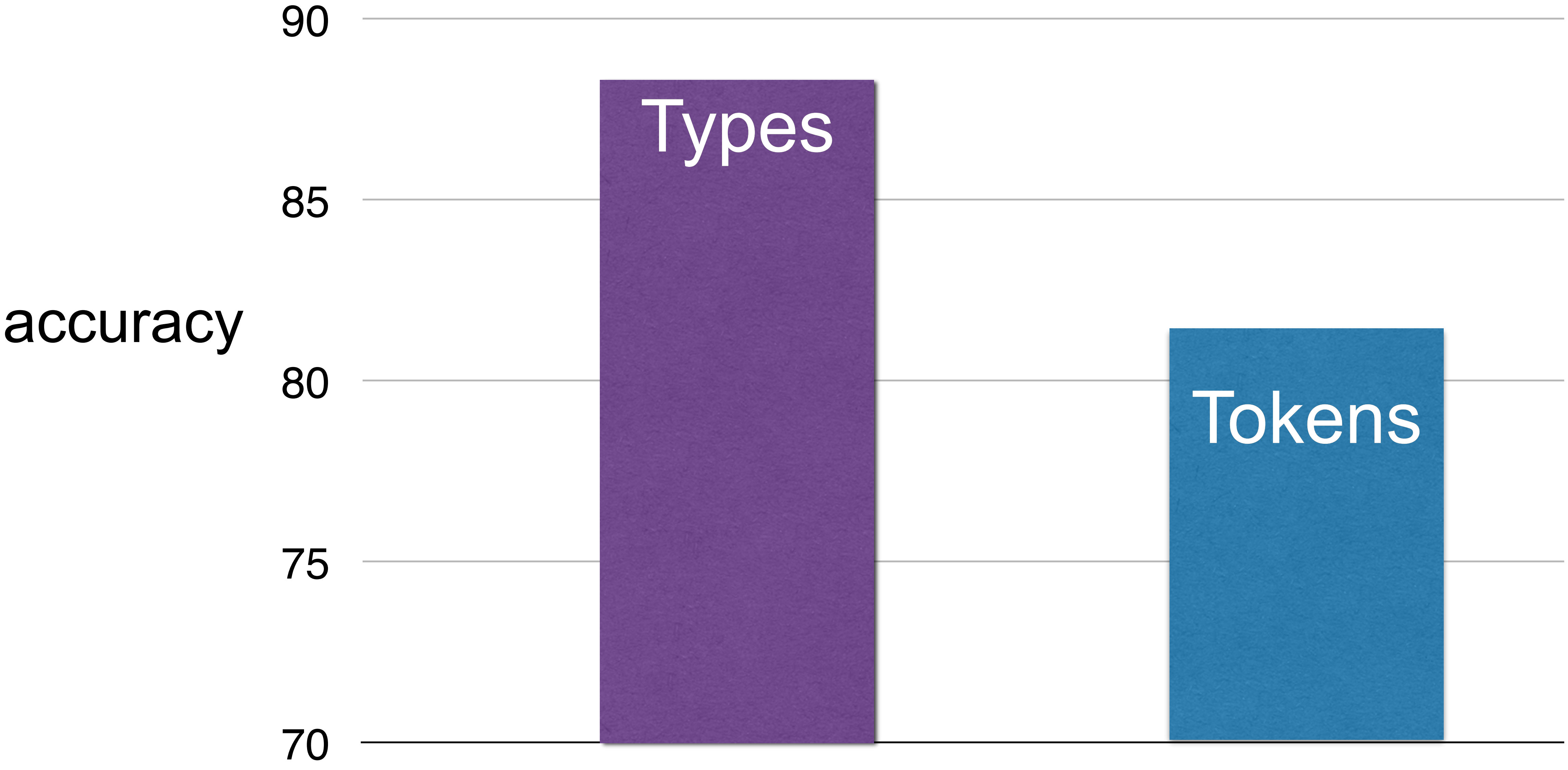
EM

Results

Types vs. Tokens

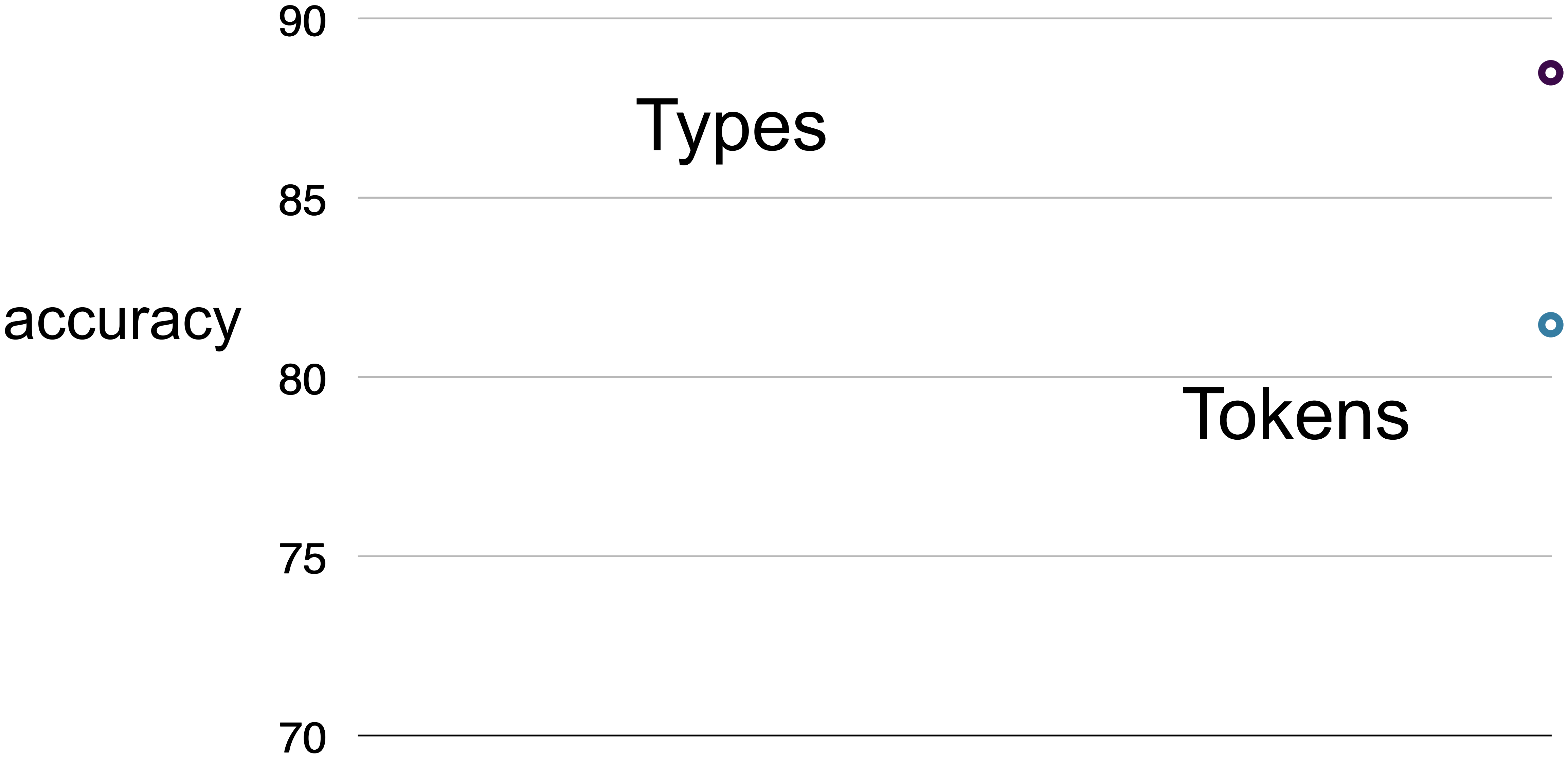
[English]

Types vs. Tokens



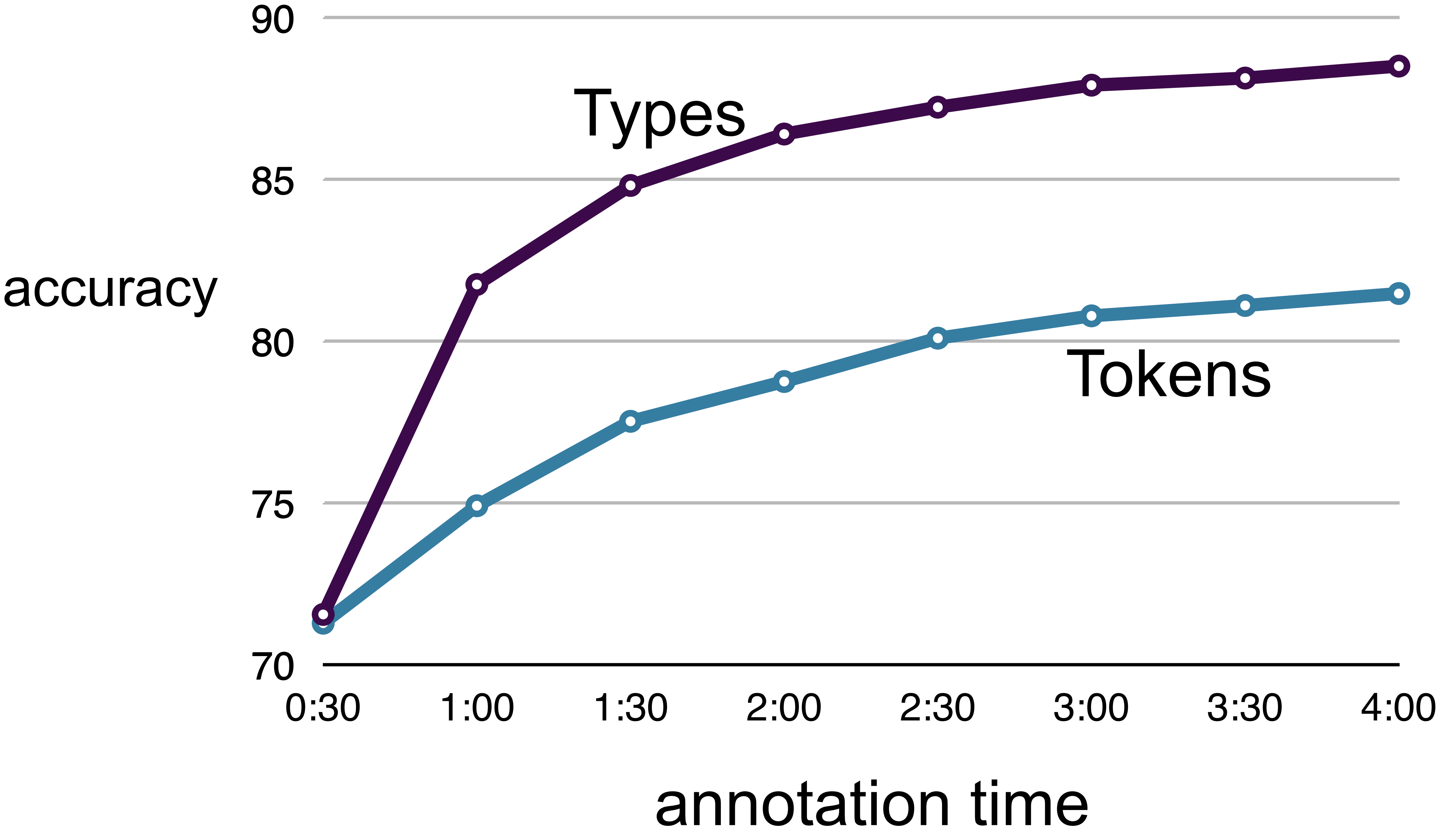
[English]

Types vs. Tokens



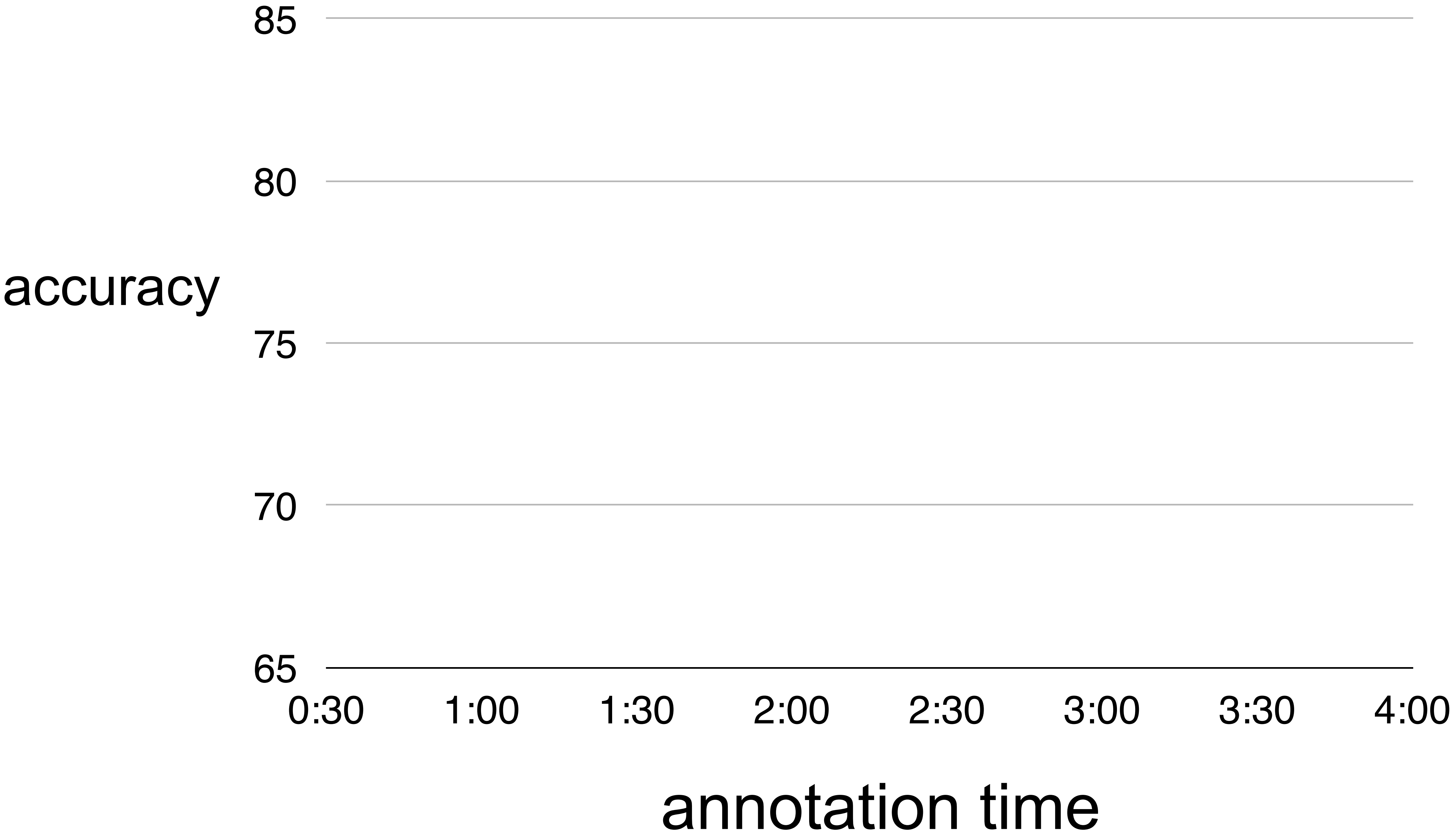
[English]

Types vs. Tokens



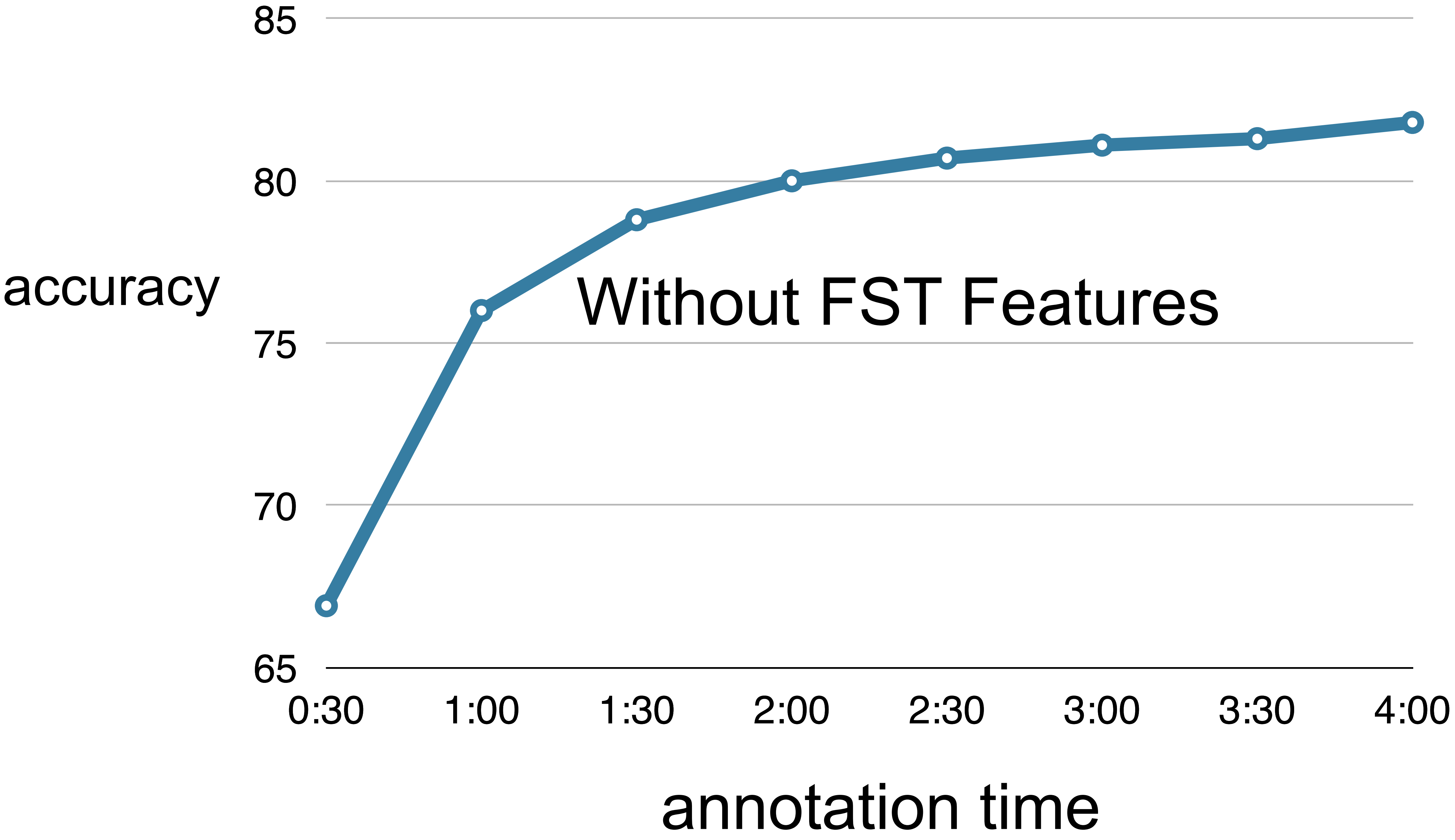
[English]

Morphological Analysis



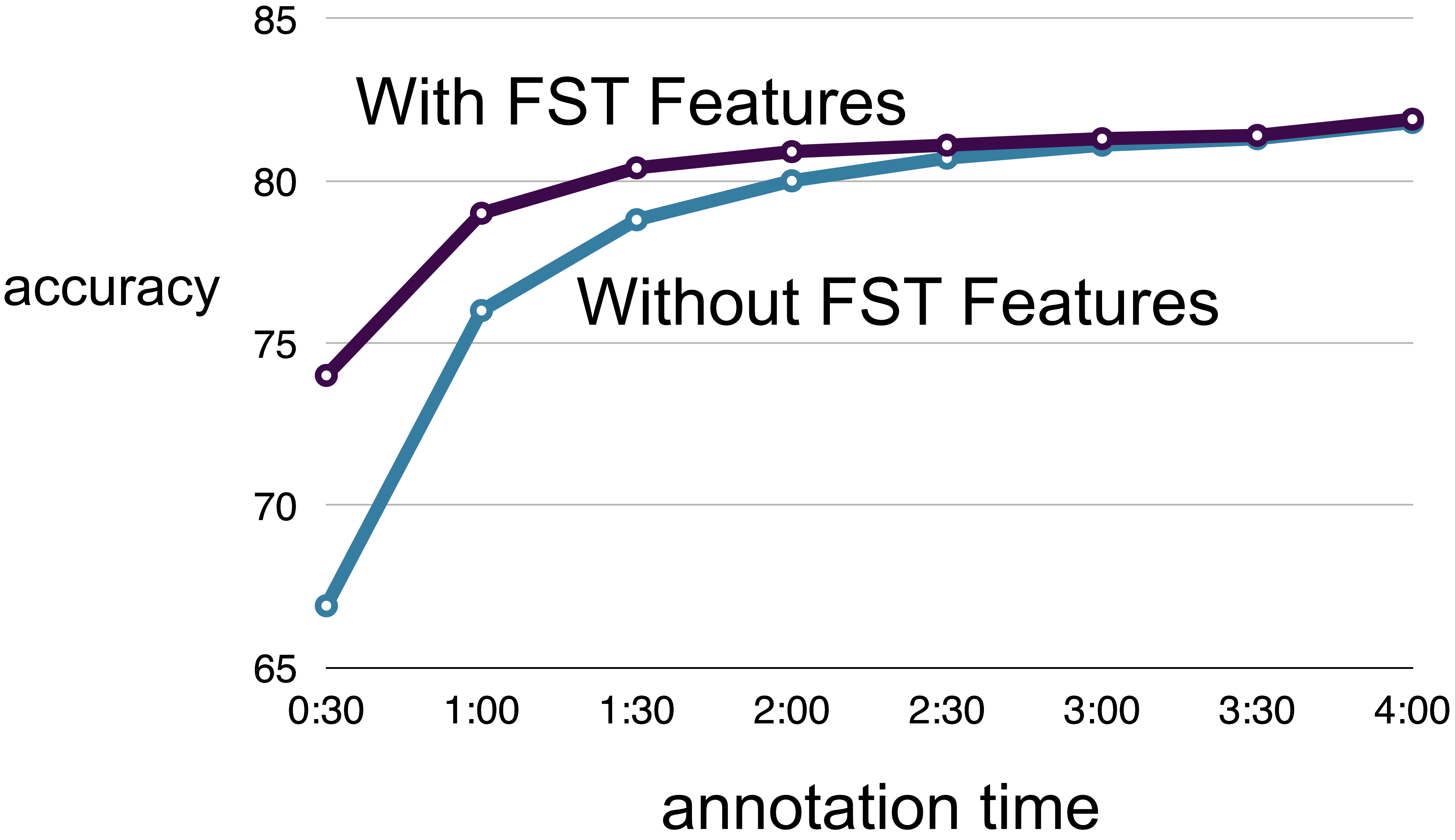
[Kinyarwanda]

Morphological Analysis



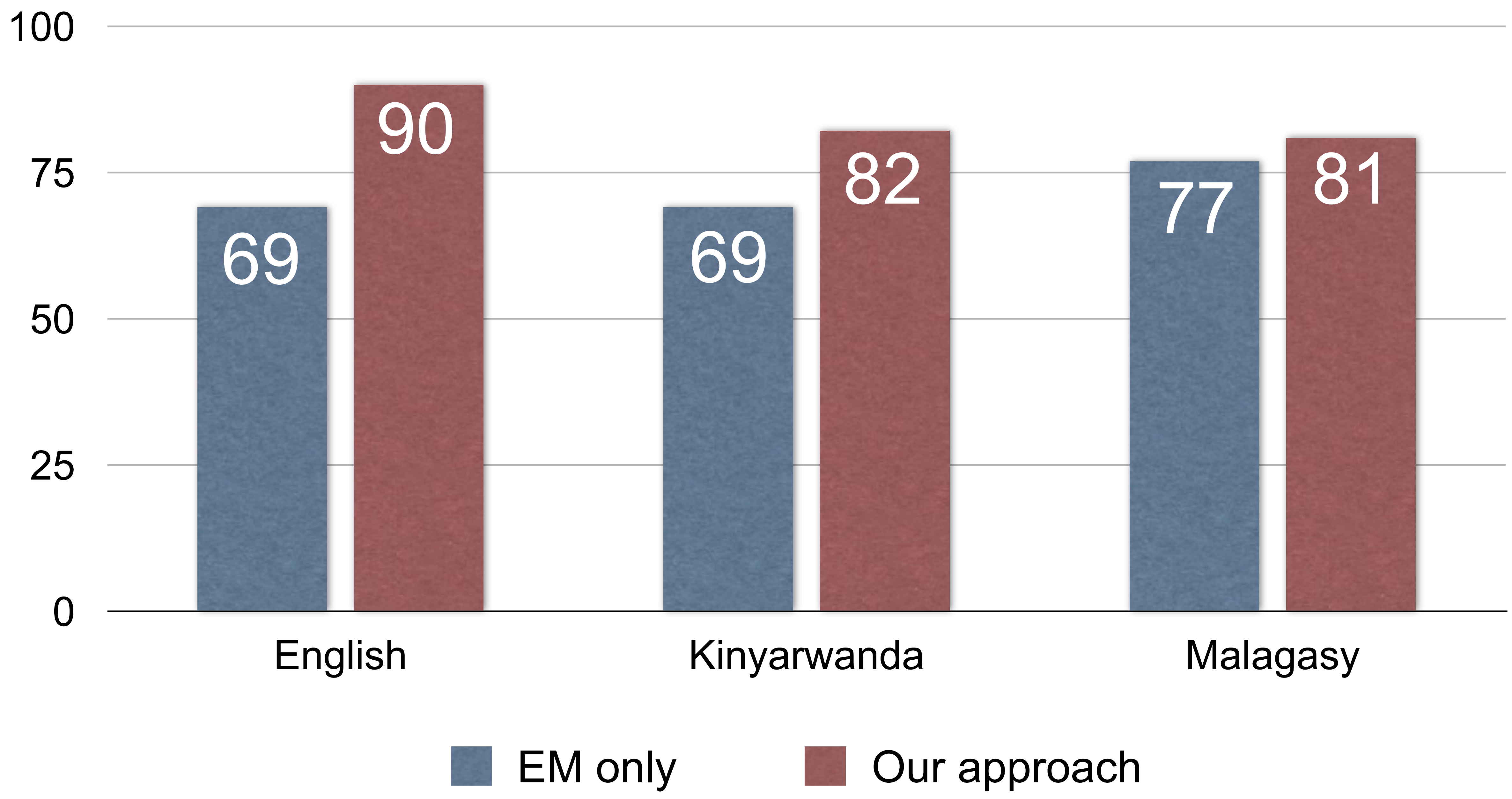
[Kinyarwanda]

Morphological Analysis



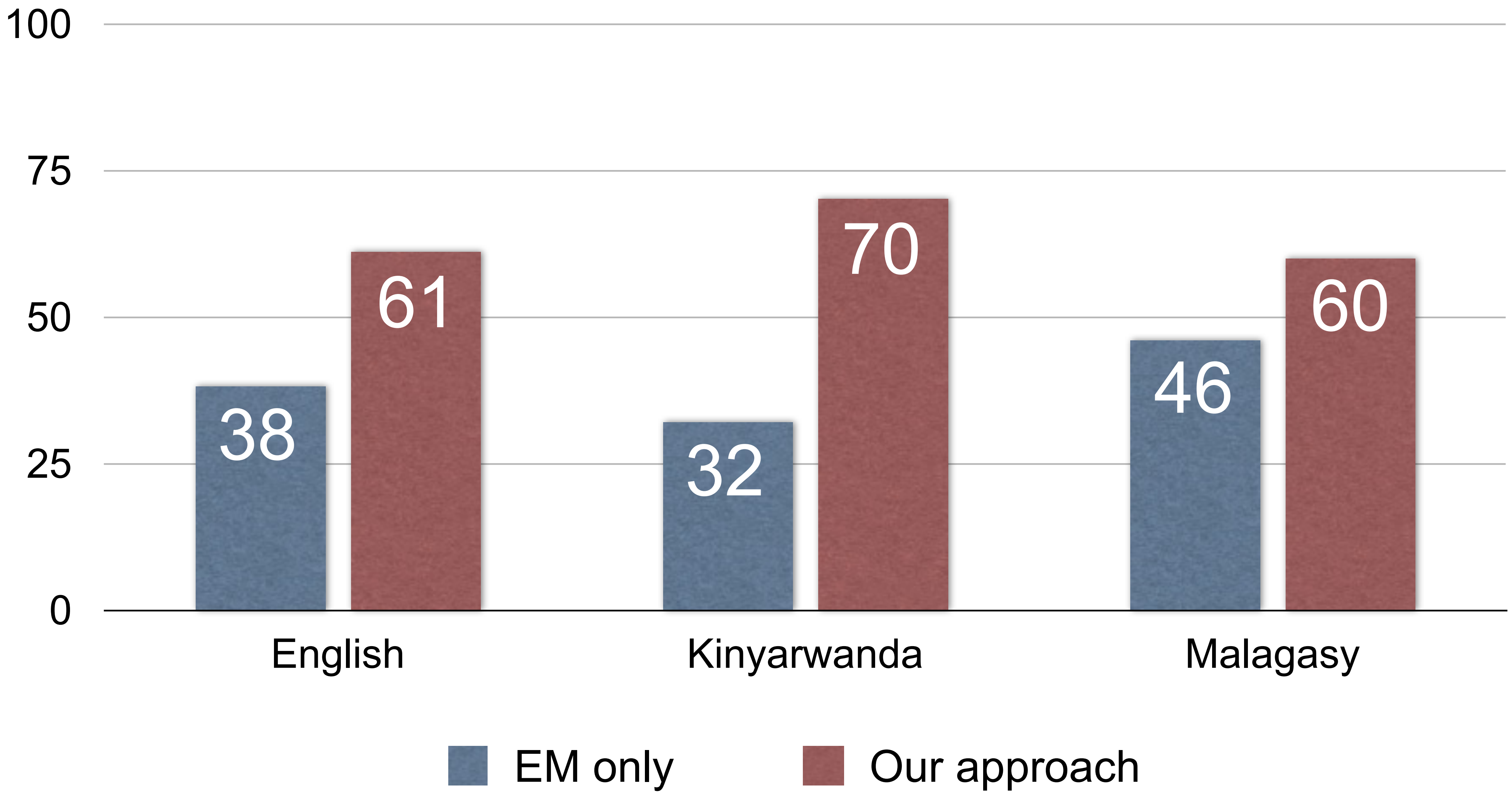
[Kinyarwanda]

Total Accuracy



[4 hours of type annotation]

Unknown Word Accuracy



[2 hours of type annotation]

English Results

English Results

All of **Wiktionary** (Li et al., 2012)

English Results

All of **Wiktionary** (Li et al., 2012) 87%

English Results

All of **Wiktionary** (Li et al., 2012)	87%
Parallel Corpus (Täckström et al., 2013)	

English Results

All of **Wiktionary** (Li et al., 2012)	87%
Parallel Corpus (Täckström et al., 2013)	89%

English Results

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

English Results

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

English Results

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

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

Rich Morphology

Parallel Corpus (Täckström et al., 2013)

Turkish

Rich Morphology

Parallel Corpus (Täckström et al., 2013)

Turkish	65%

Rich Morphology

Parallel Corpus (Täckström et al., 2013)

Turkish 65%

4-hours (Garrette et al., 2013)

Kinyarwanda

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.