Neural CRF Parsing

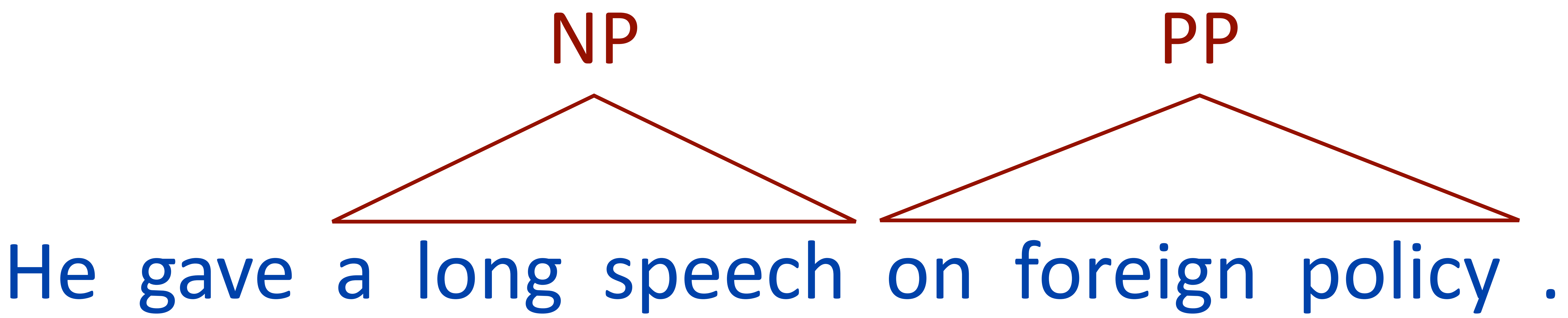
Greg Durrett and Dan Klein
UC Berkeley

Parsing with CKY

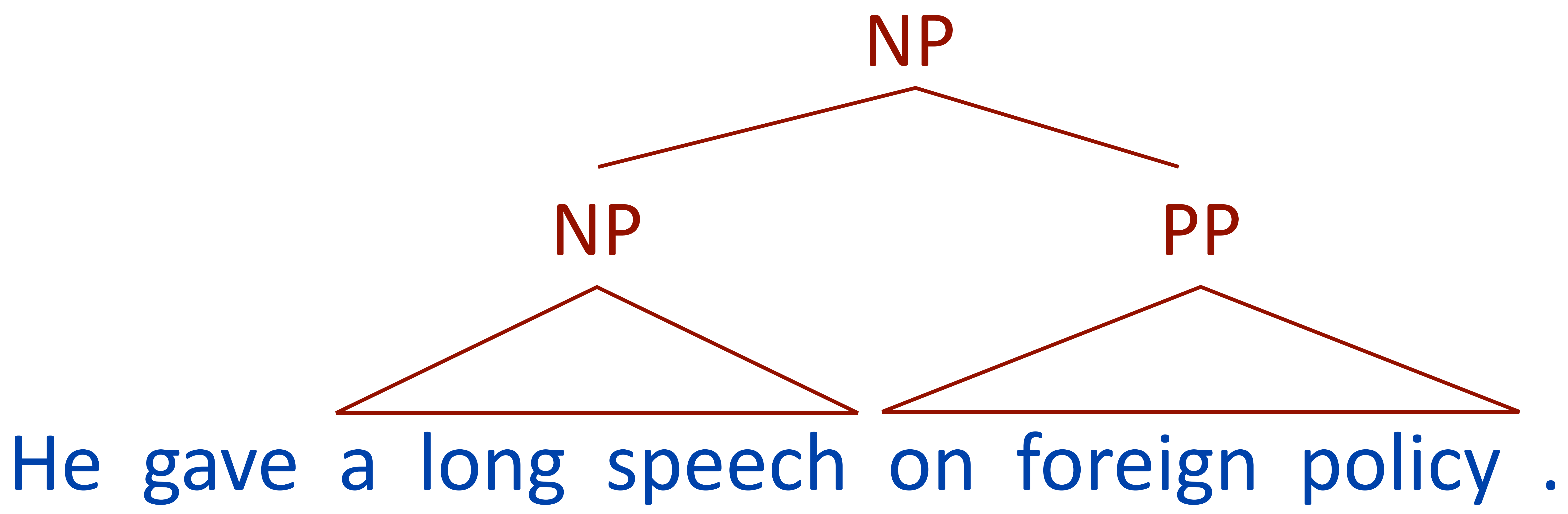
Parsing with CKY

He gave a long speech on foreign policy .

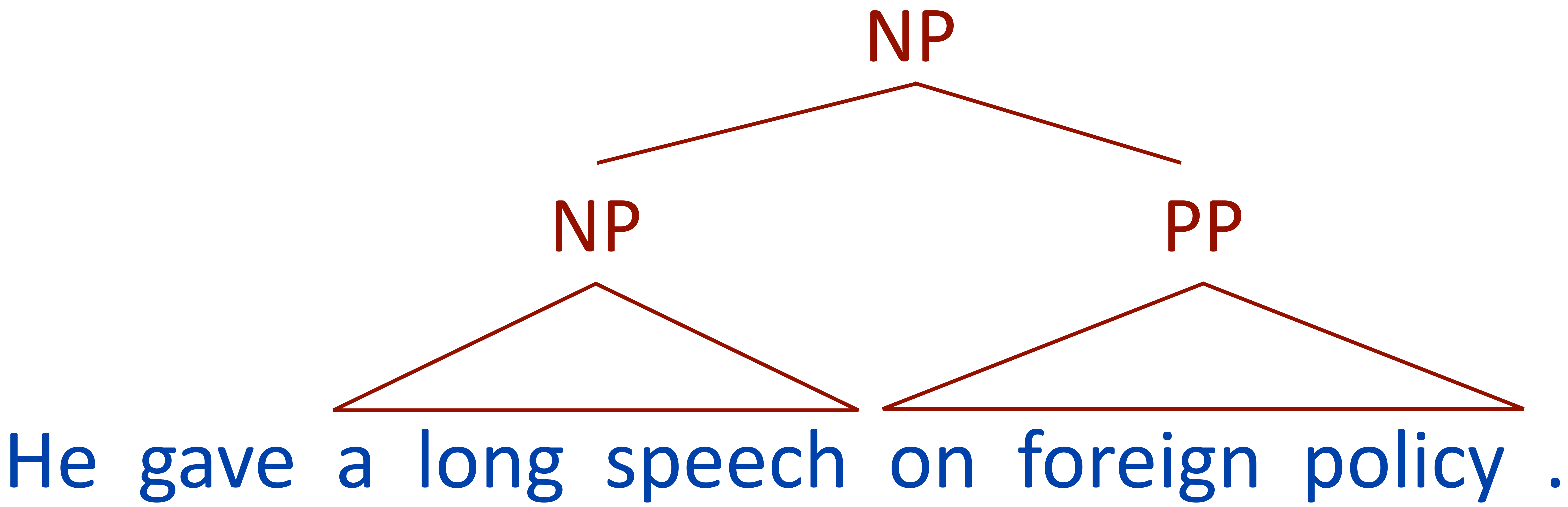
Parsing with CKY



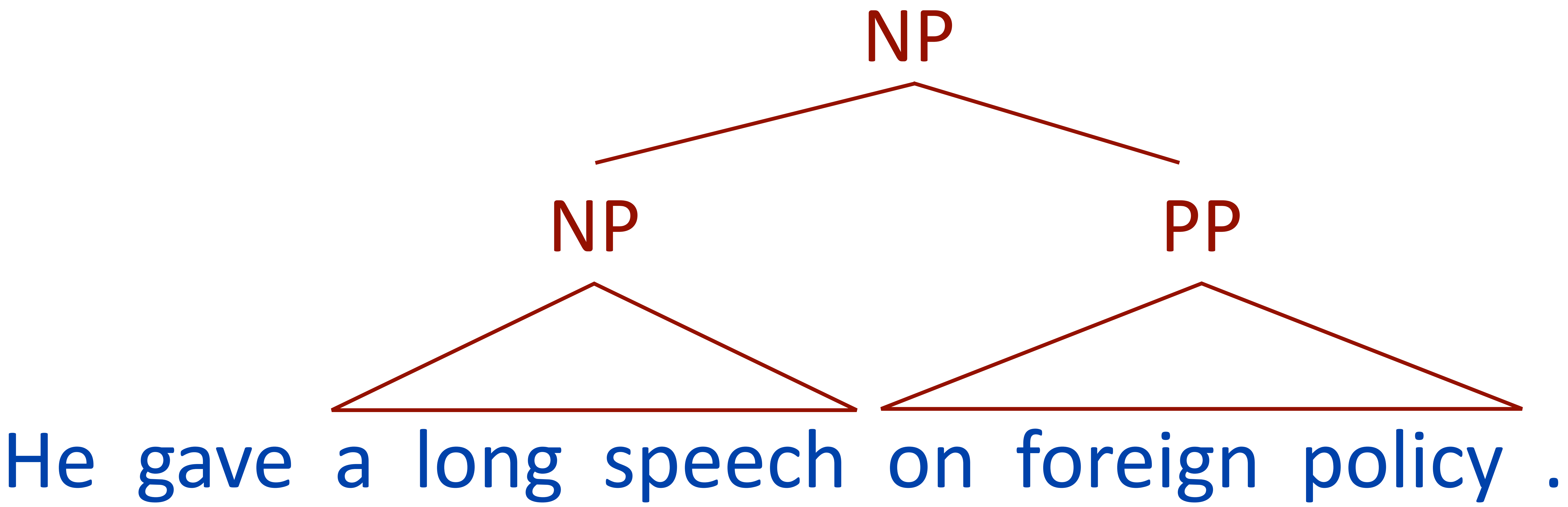
Parsing with CKY



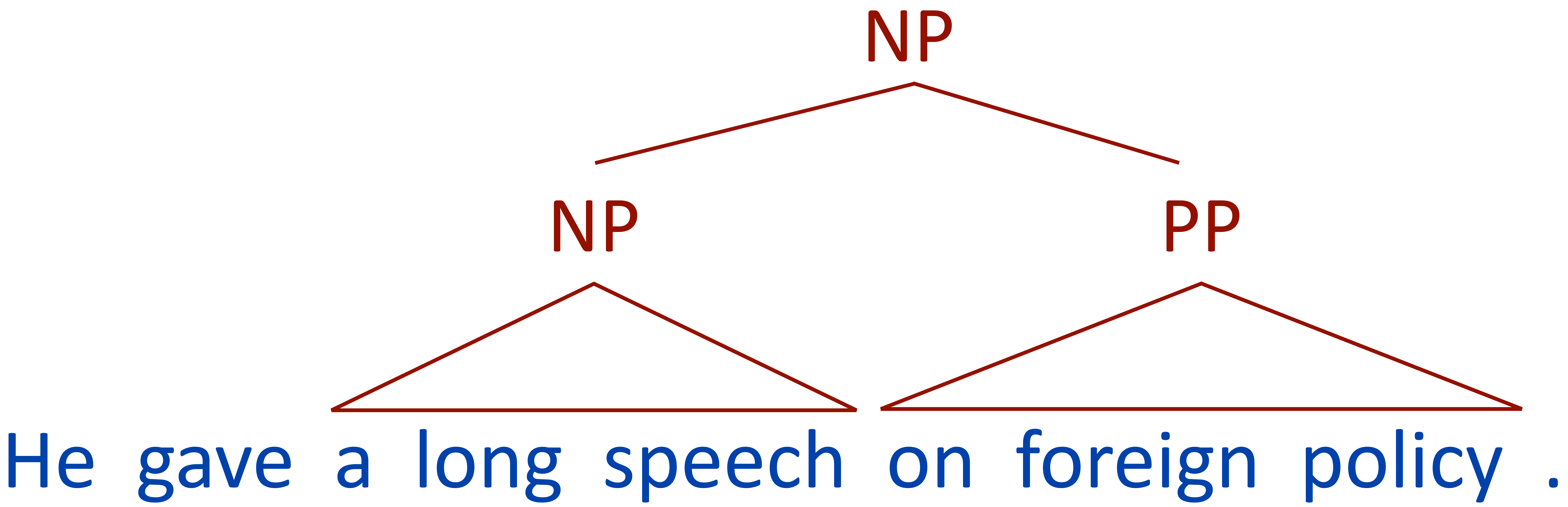
Parsing with CKY



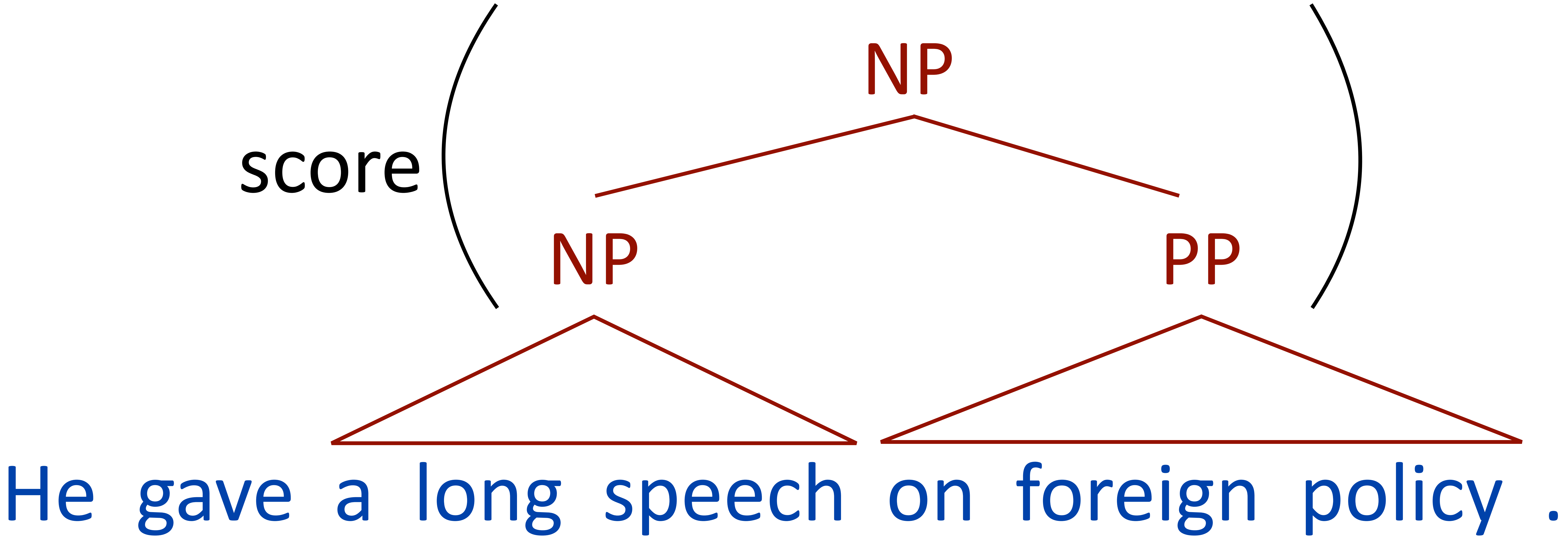
Parsing with CKY



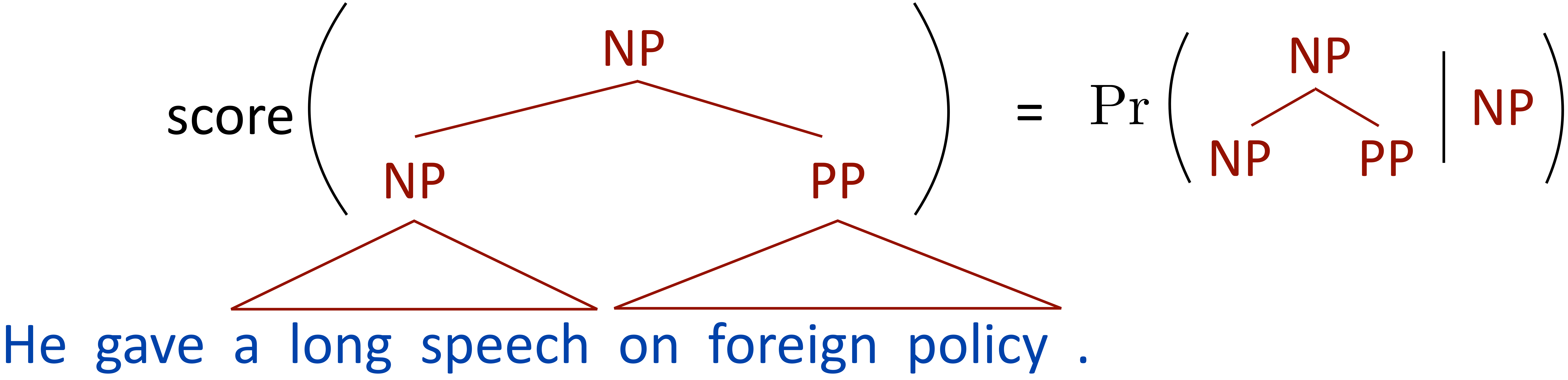
Discrete Structure



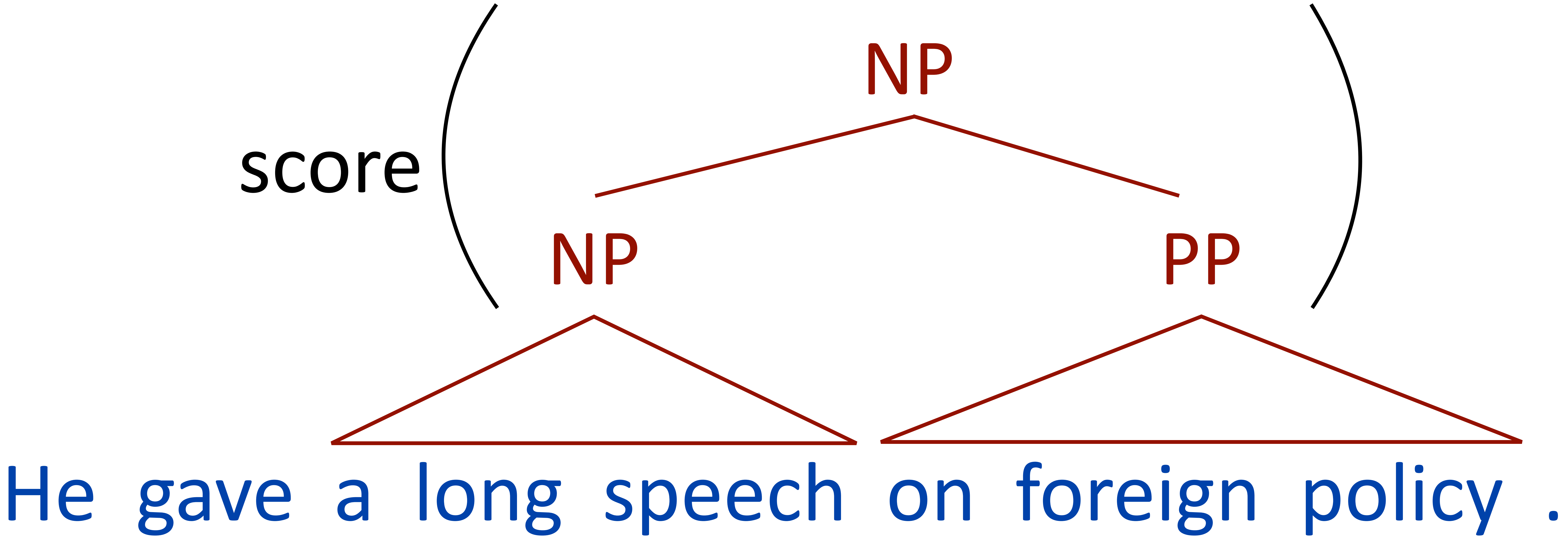
Discrete Structure



Discrete Structure

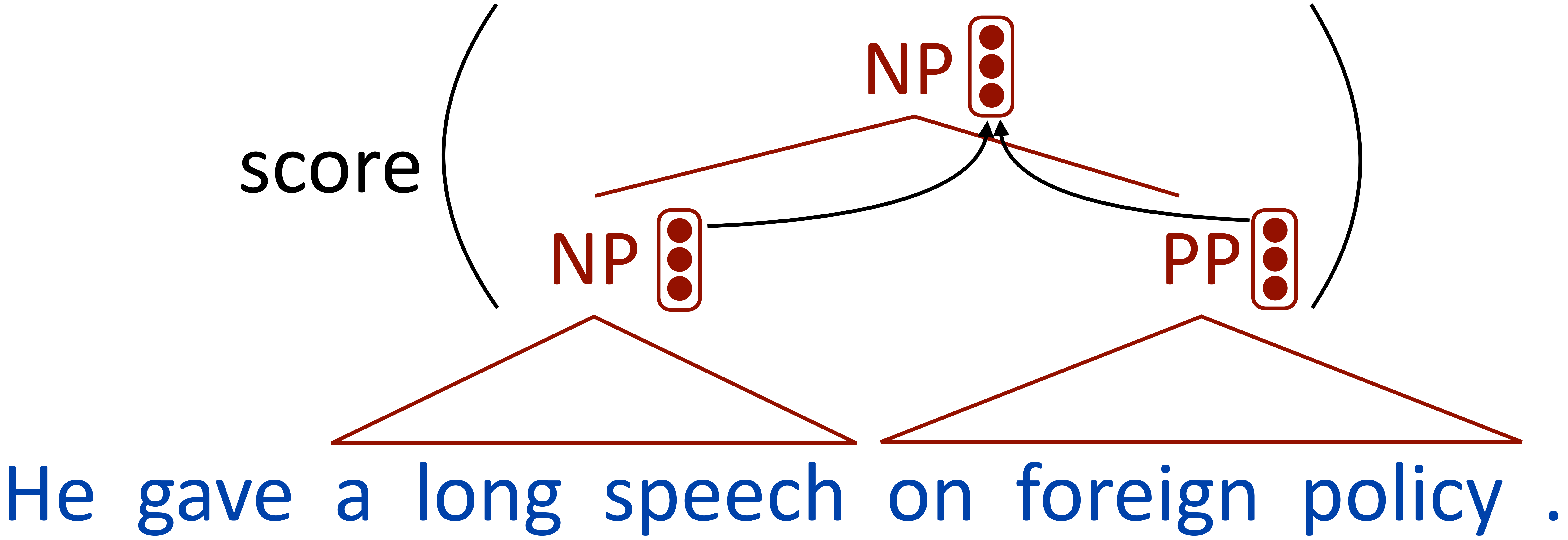


Continuous Structure



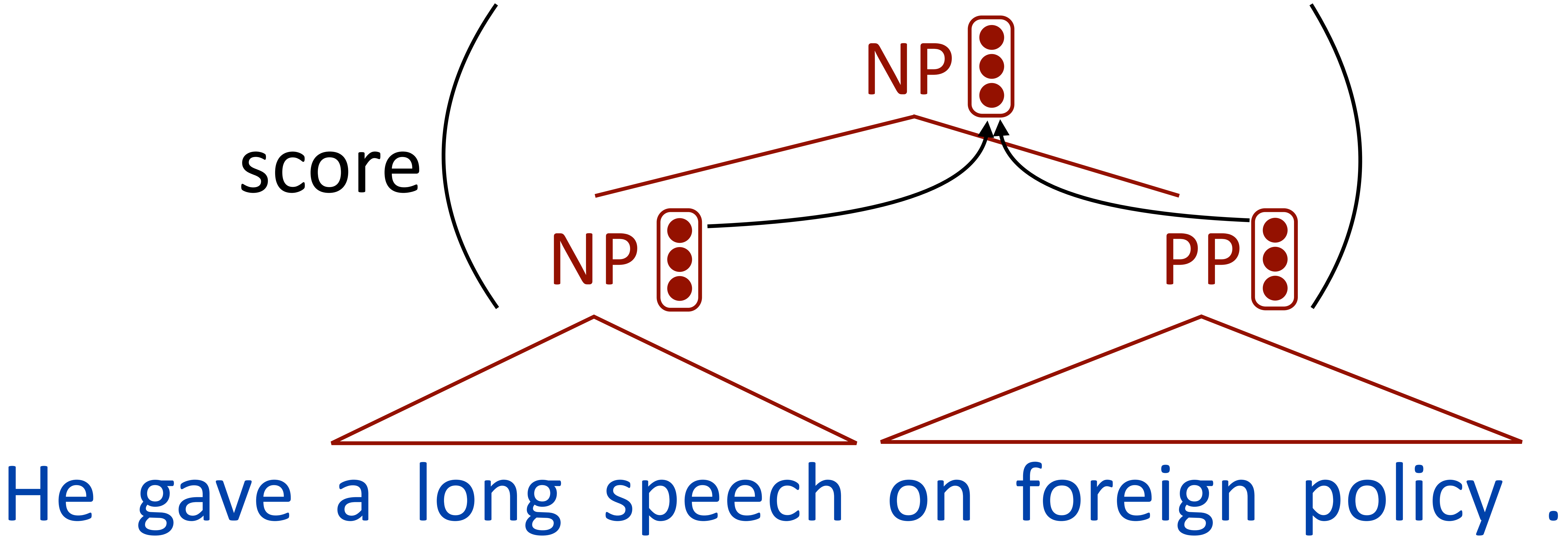
[Socher et al. (2013)]

Continuous Structure



[Socher et al. (2013)]

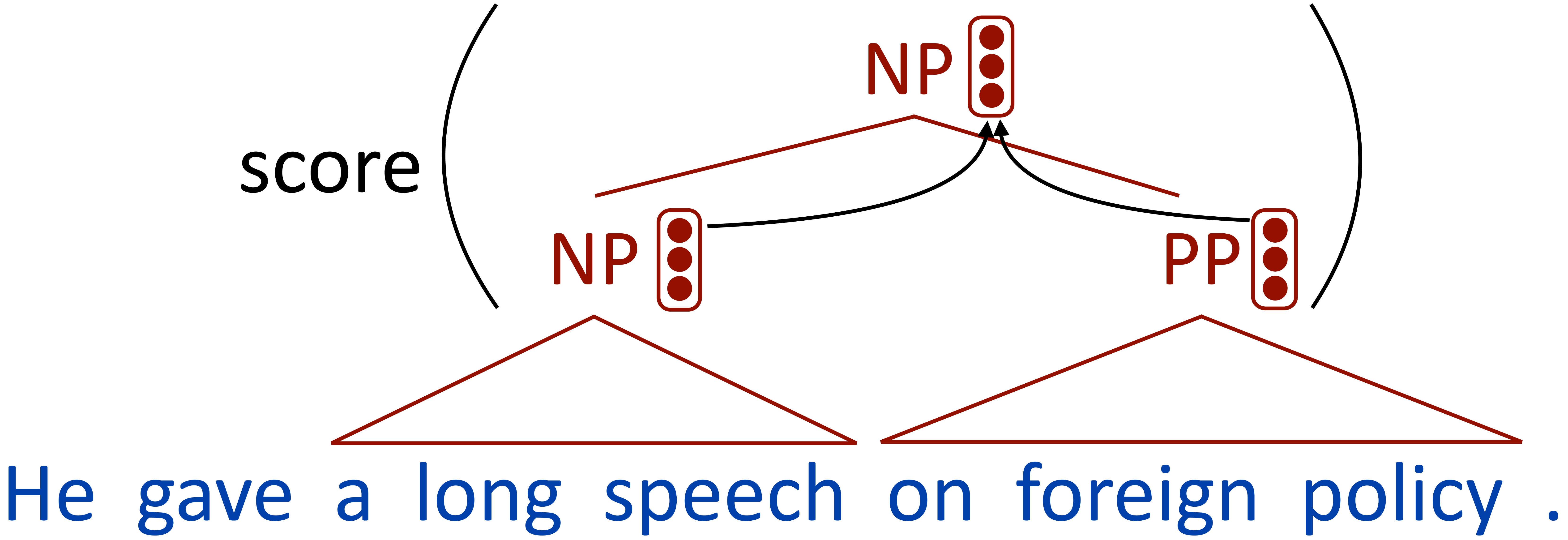
Continuous Structure



Powerful nonlinear featurization…

[Socher et al. (2013)]

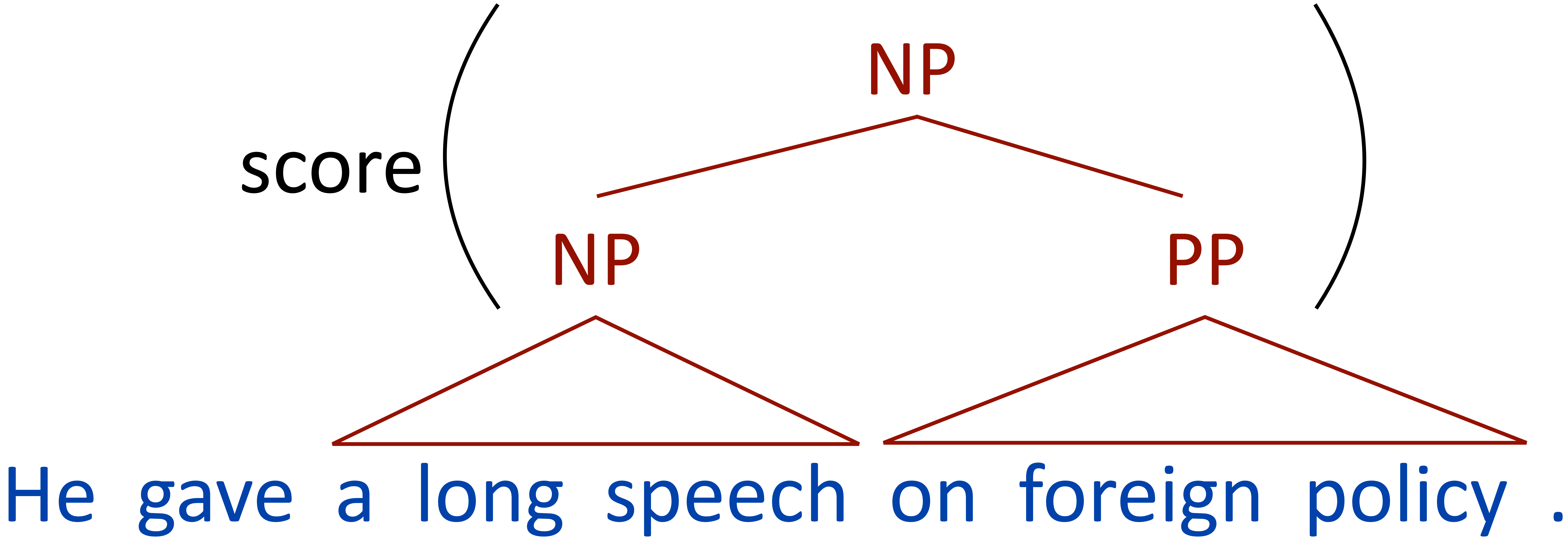
Continuous Structure



Powerful nonlinear featurization…but inference is now intractable.

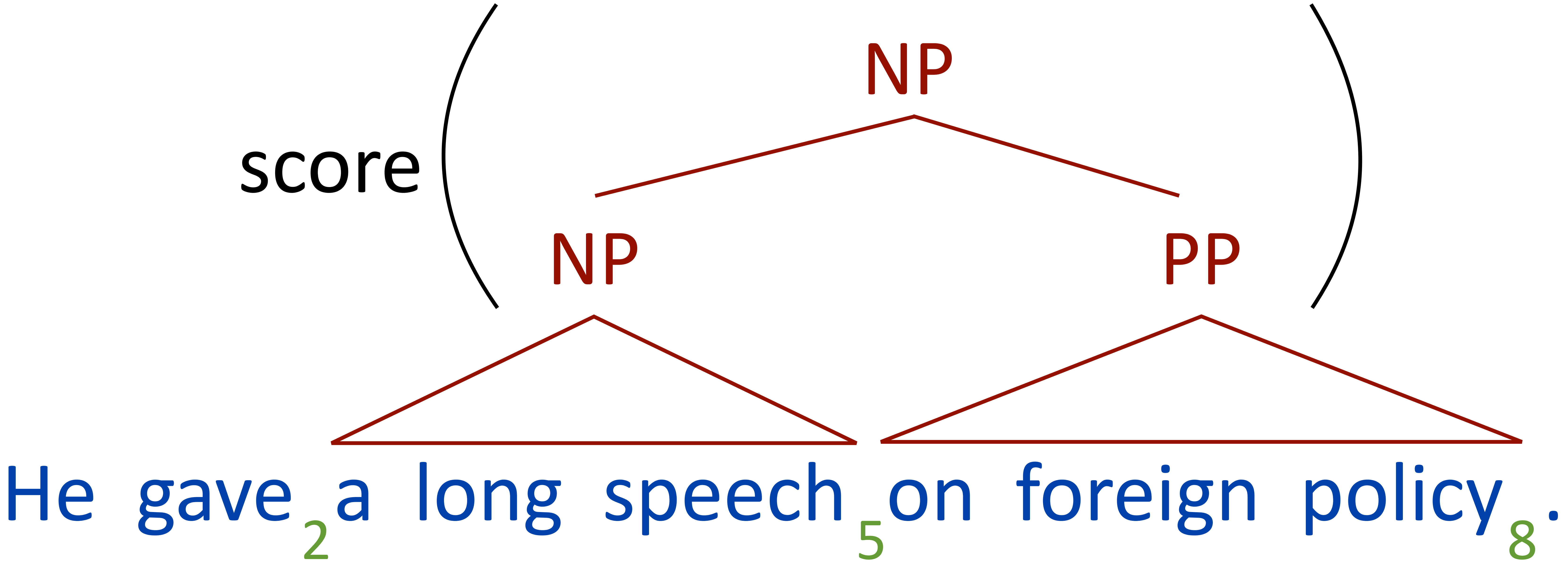
[Socher et al. (2013)]

CRF Parsing



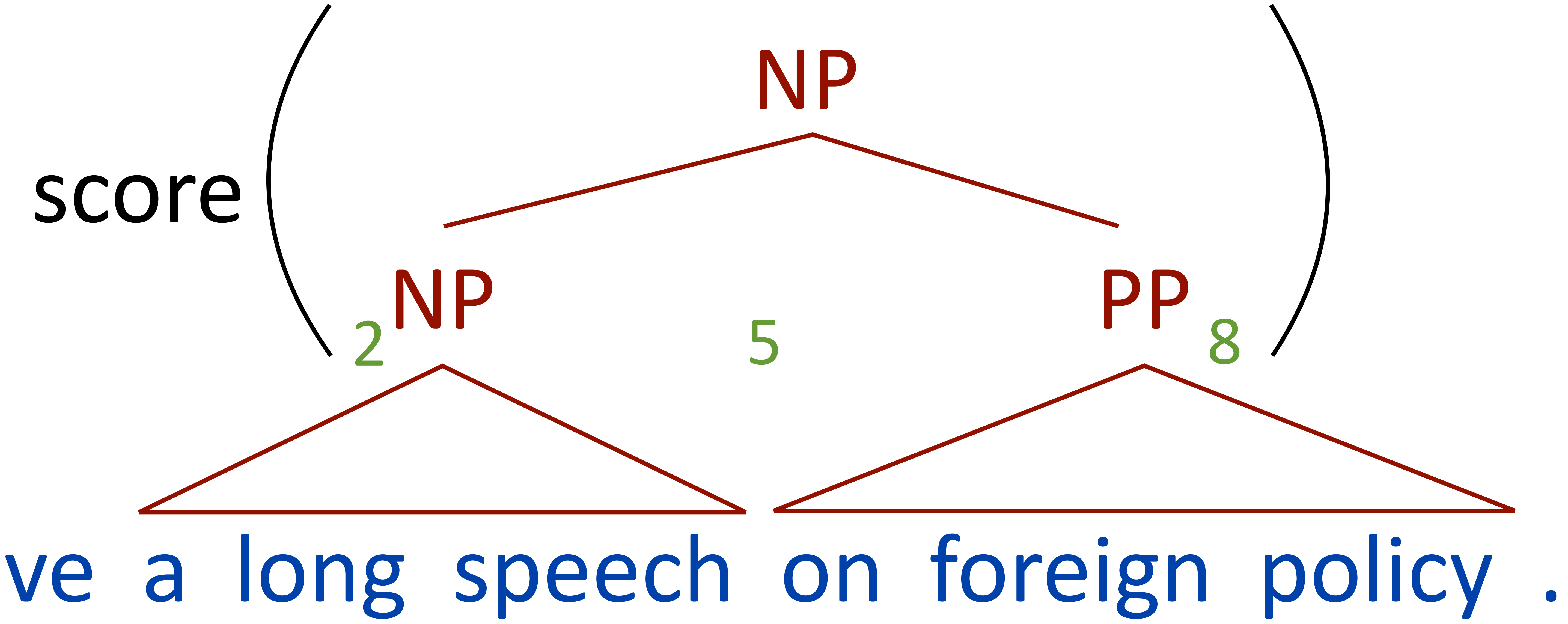
[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

CRF Parsing



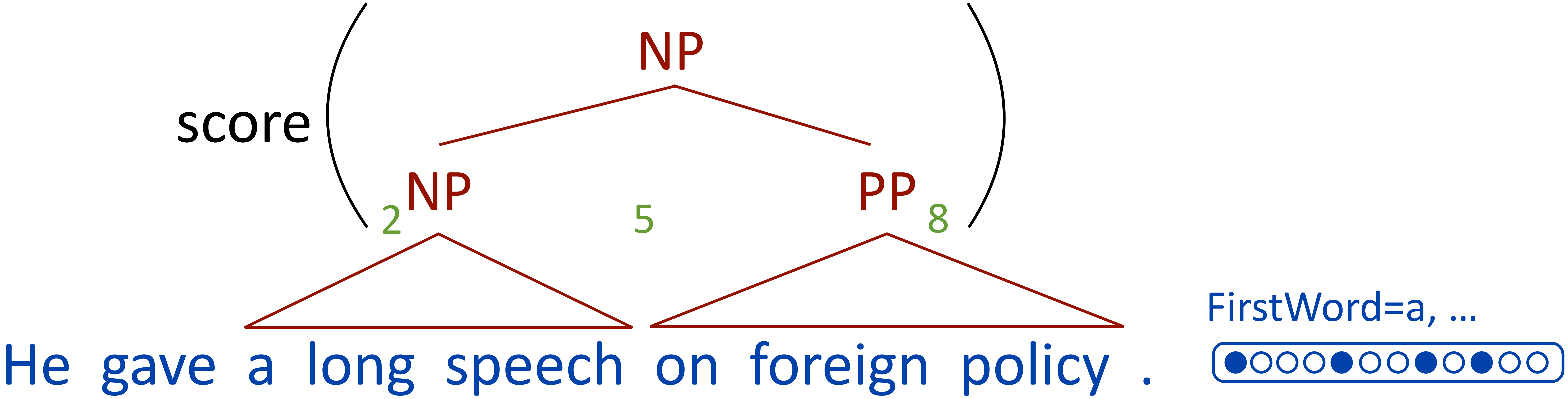
[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

CRF Parsing



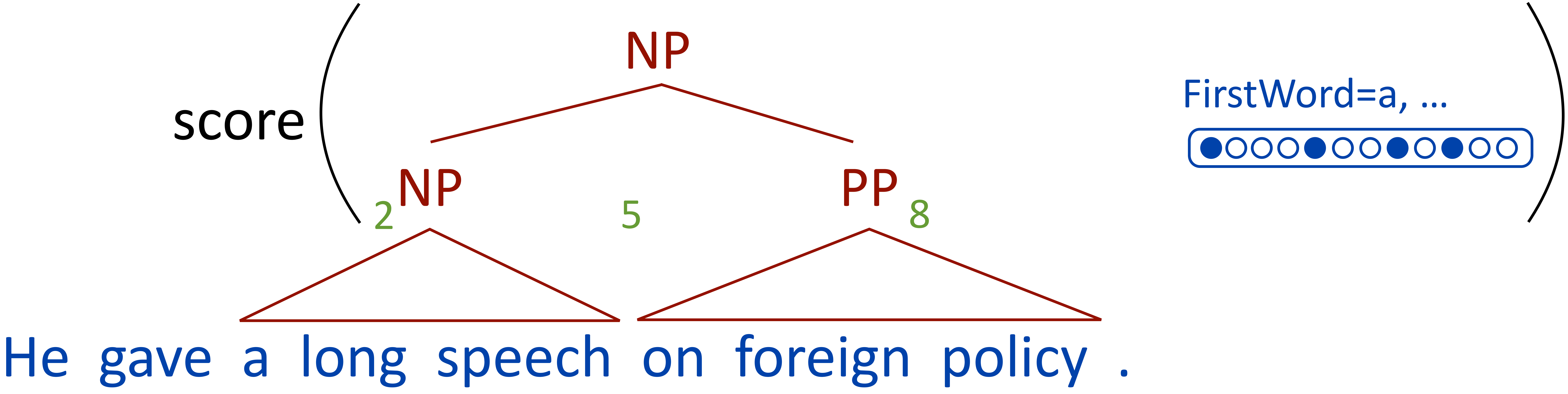
[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

CRF Parsing



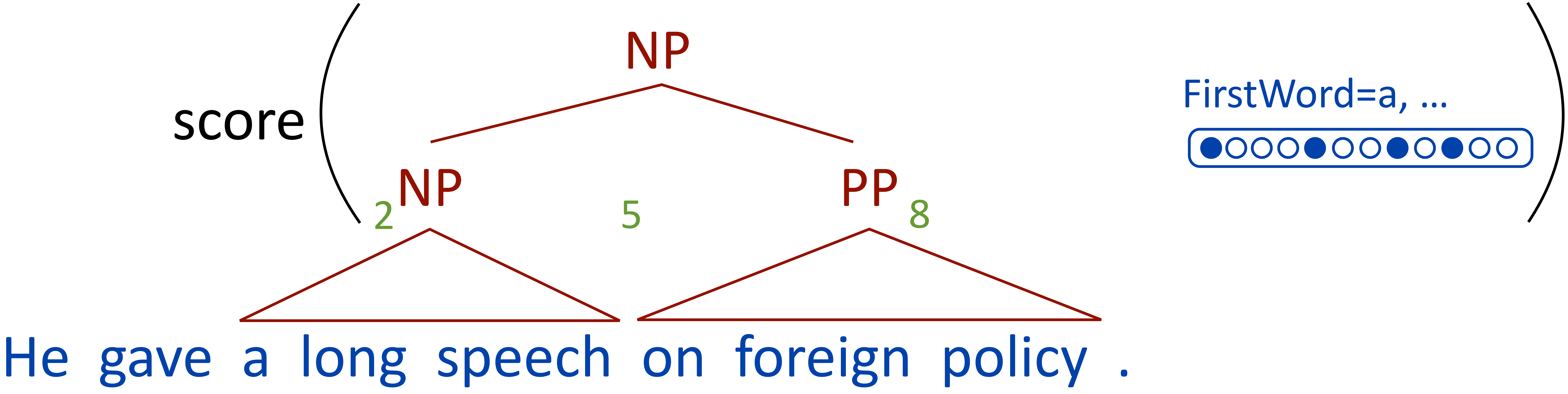
[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

CRF Parsing



[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

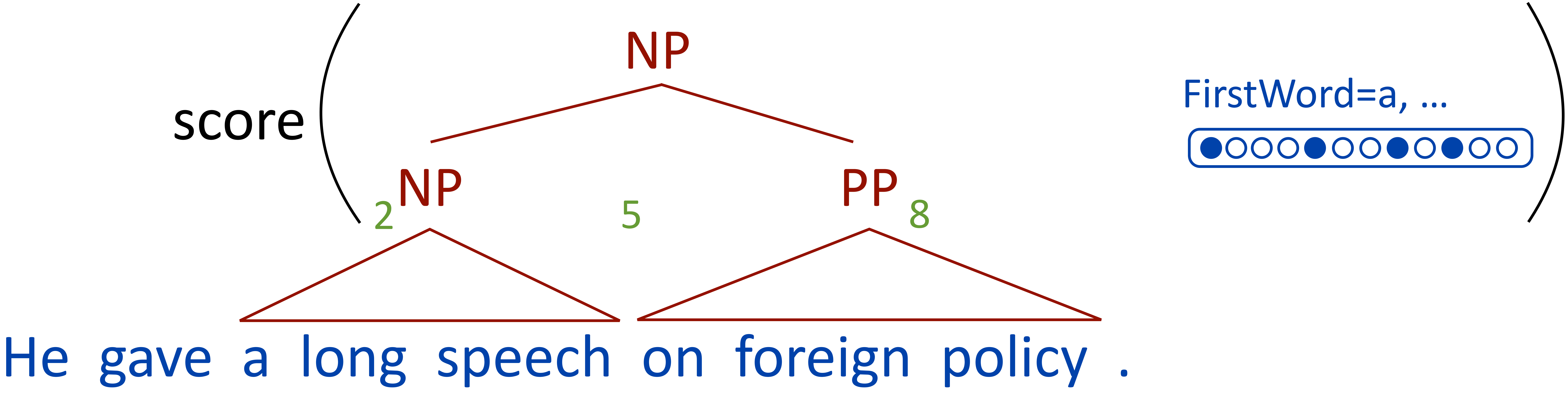
CRF Parsing



- Discrete structure with discrete features

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

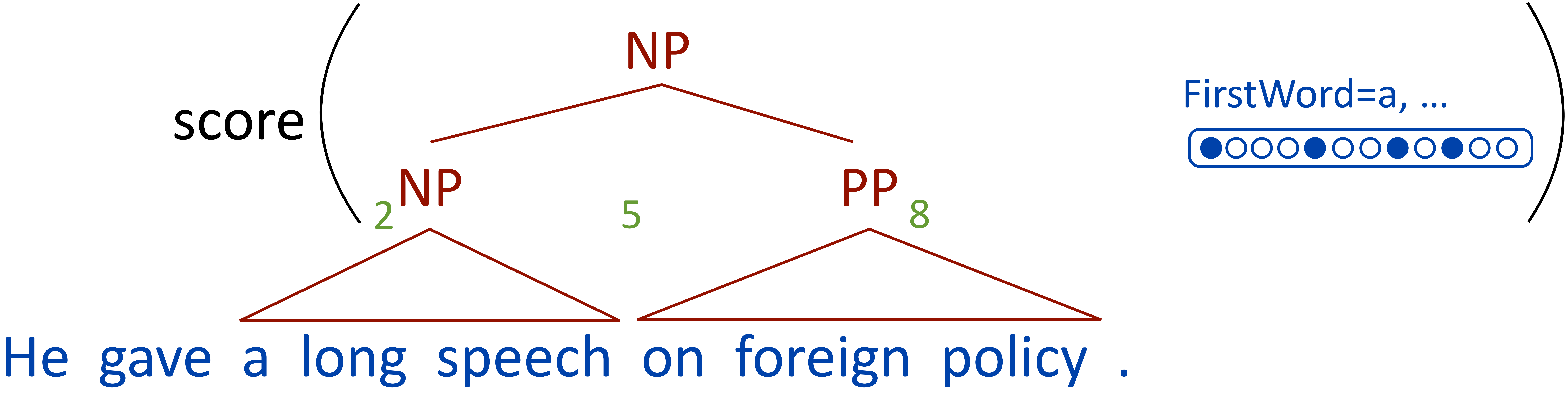
CRF Parsing



- Discrete structure with discrete features
- Efficient inference via basic CKY…

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

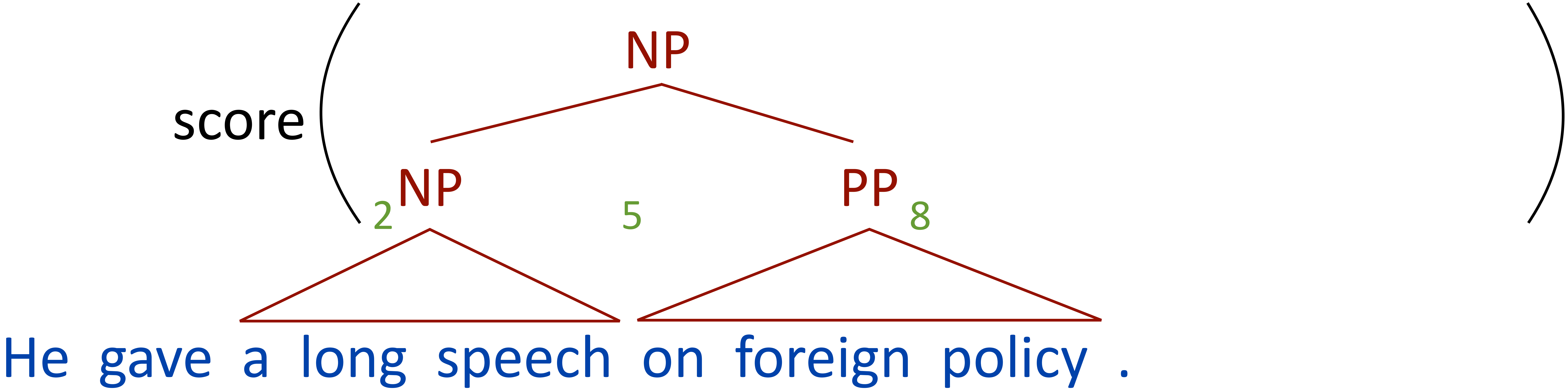
CRF Parsing



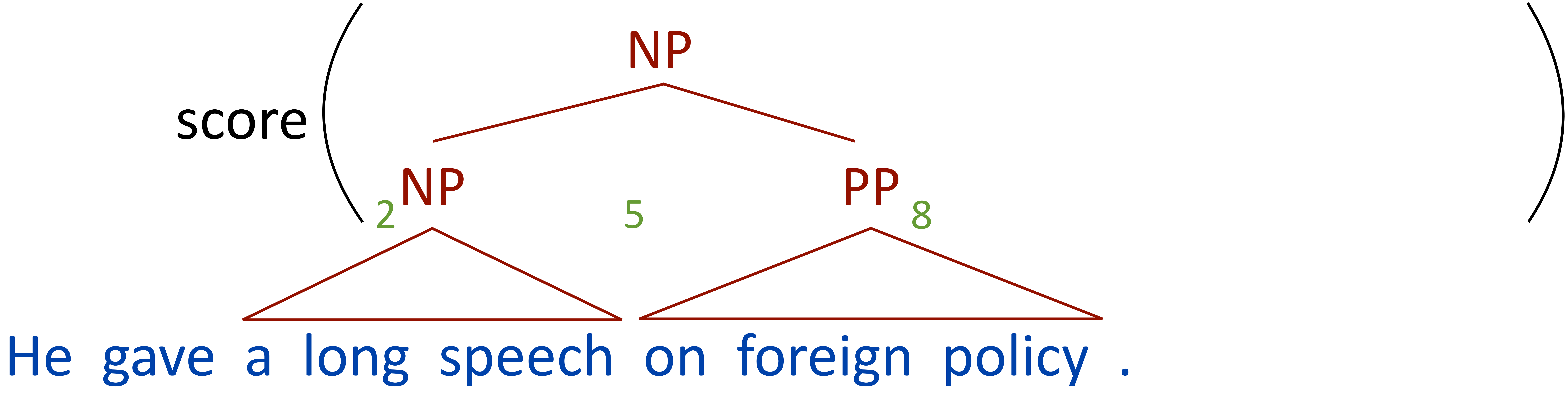
- Discrete structure with discrete features
- Efficient inference via basic CKY…but the model is typically linear

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]

Neural CRF Parsing

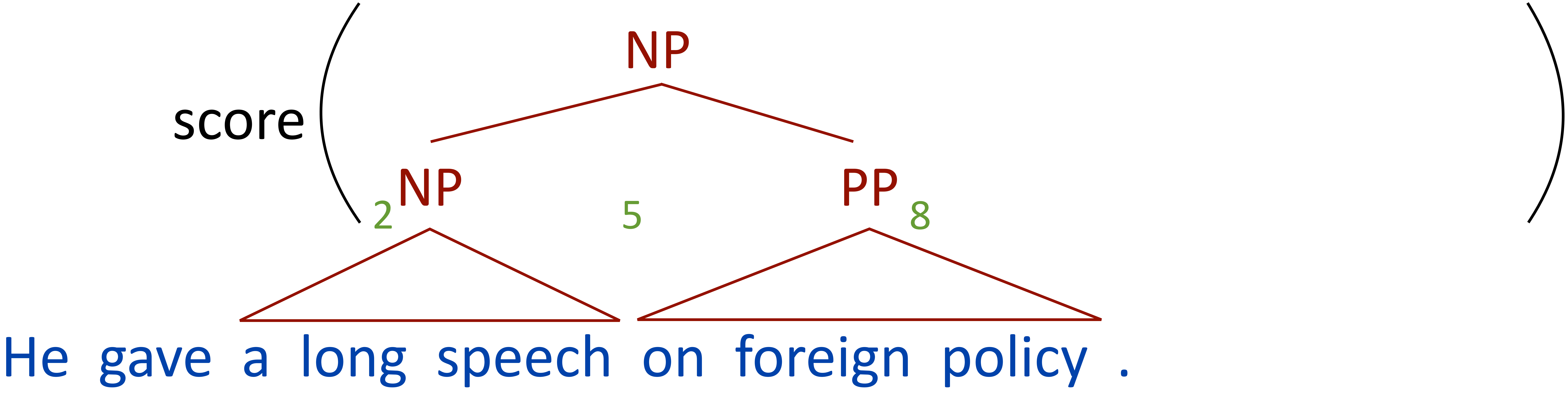


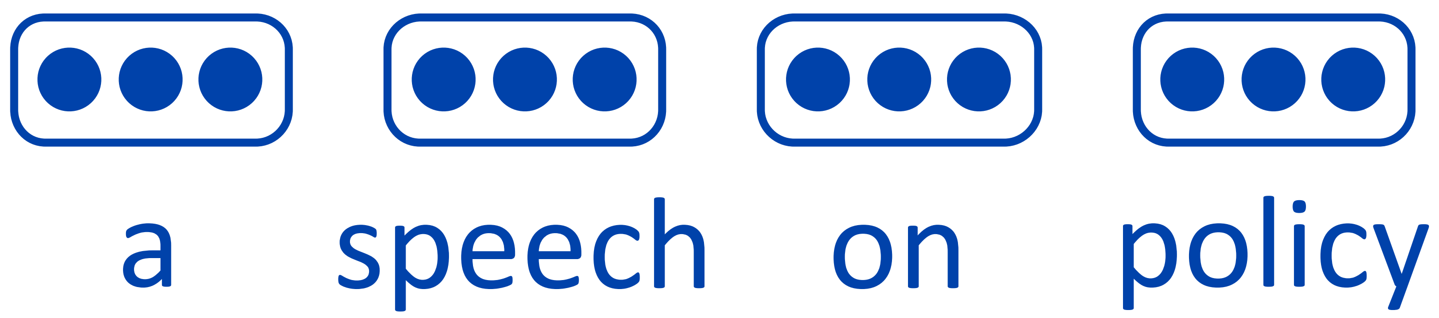
Neural CRF Parsing



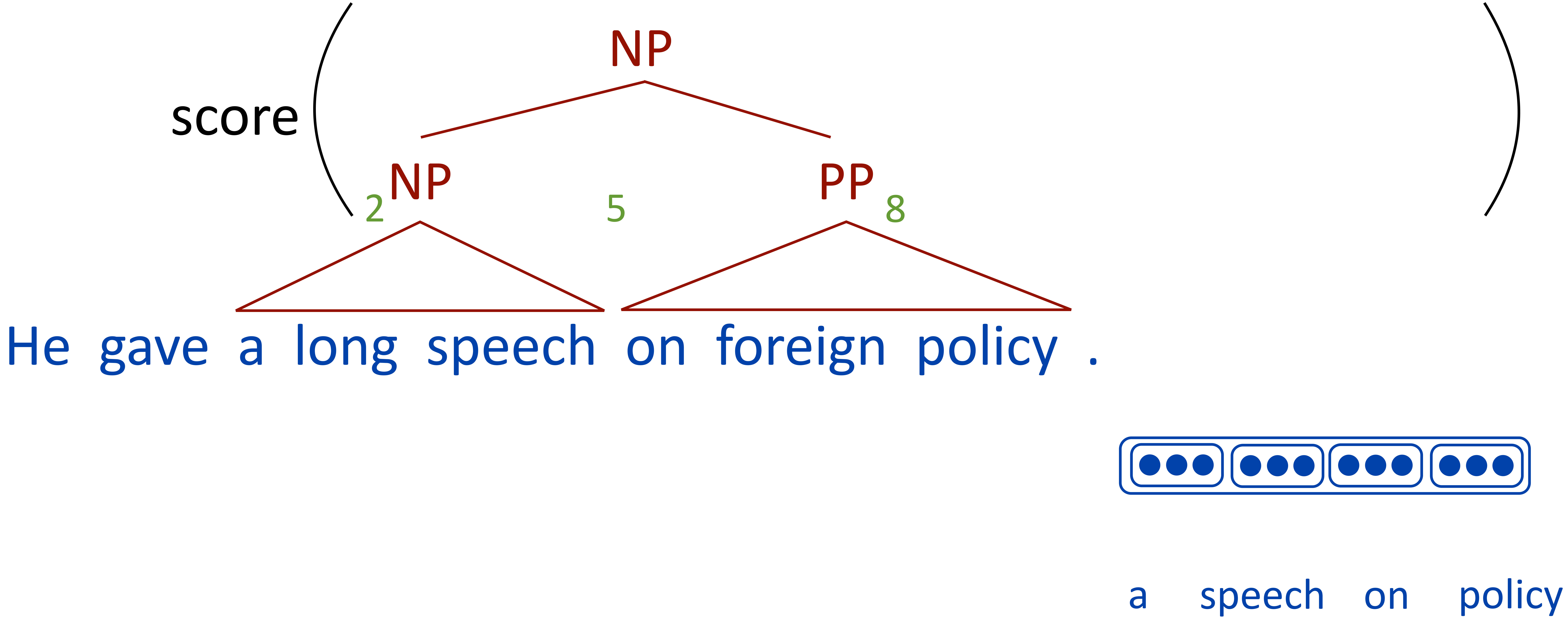
a speech on policy

Neural CRF Parsing

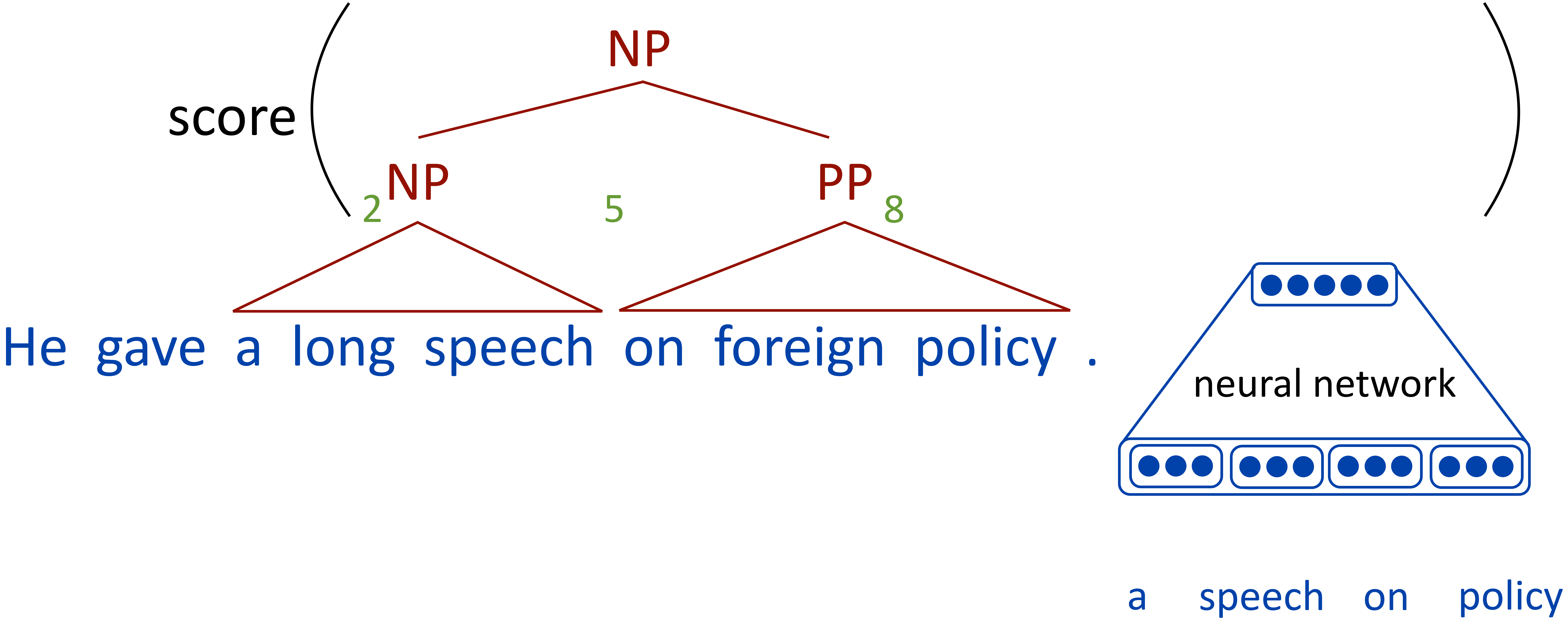




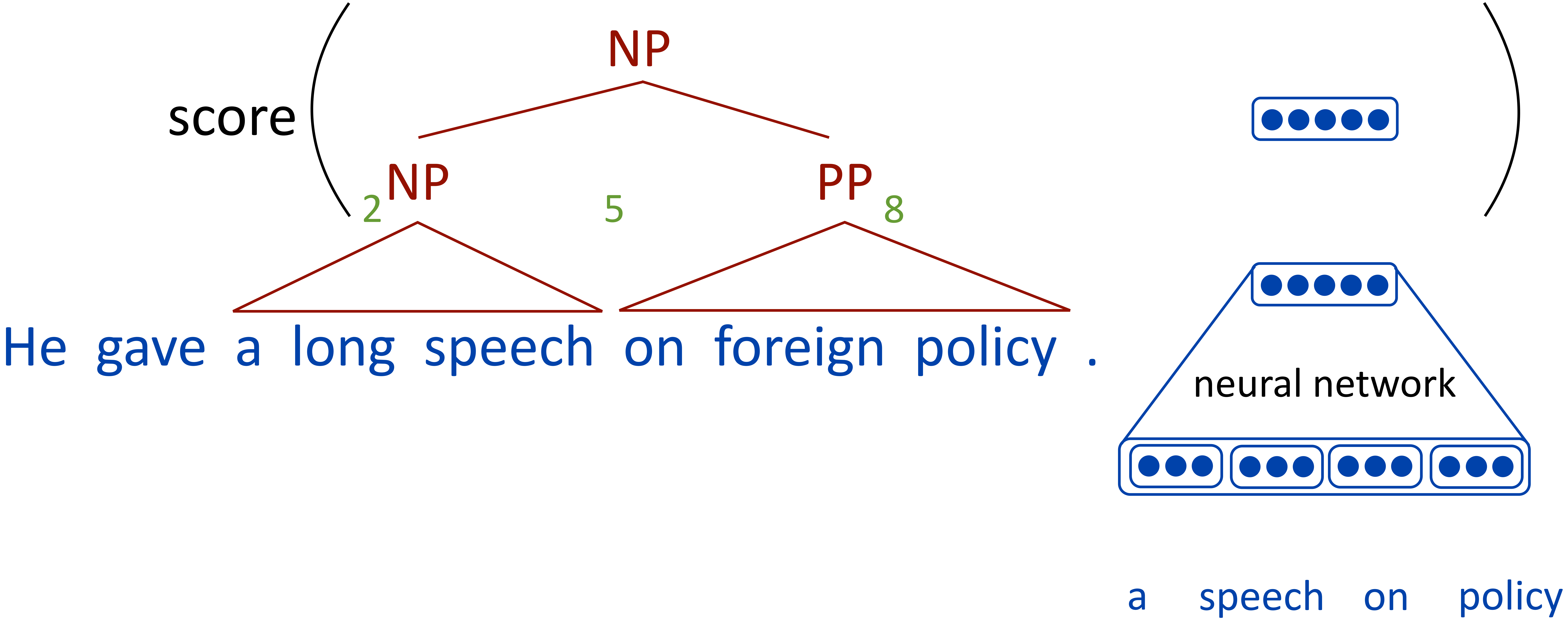
Neural CRF Parsing



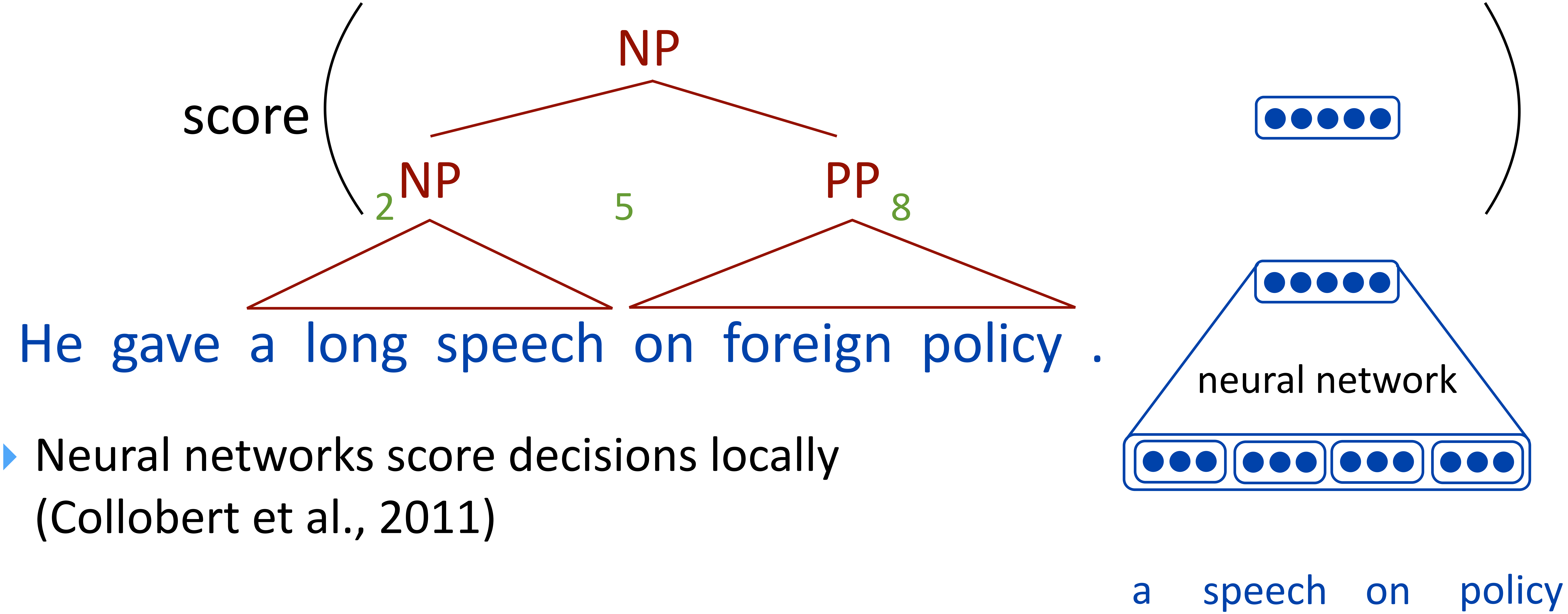
Neural CRF Parsing



Neural CRF Parsing

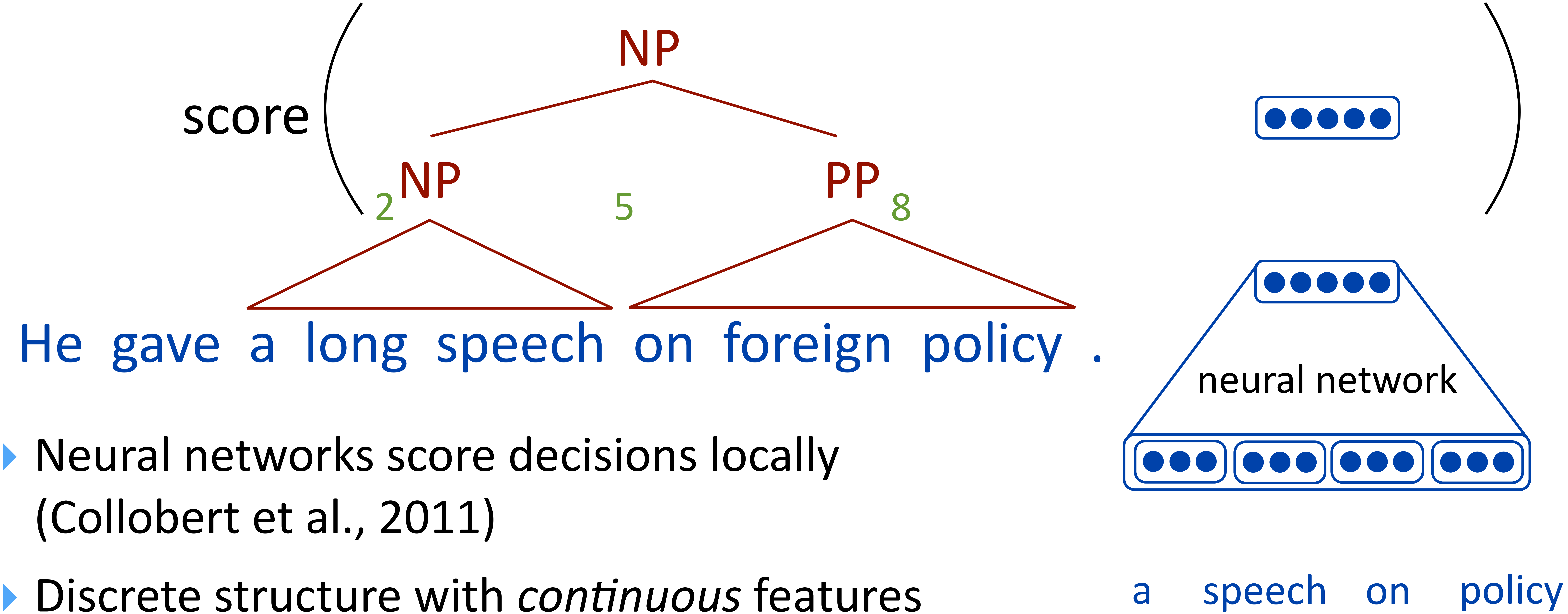


Neural CRF Parsing



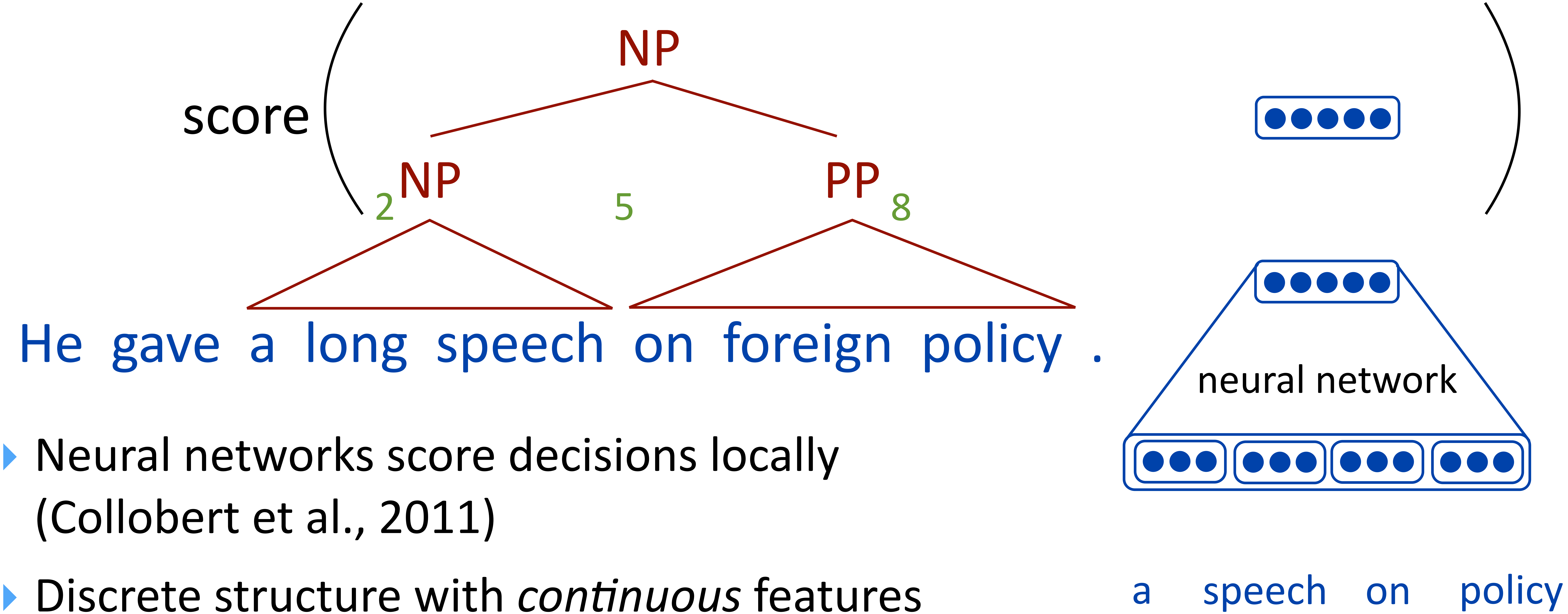
- Neural networks score decisions locally (Collobert et al., 2011)

Neural CRF Parsing



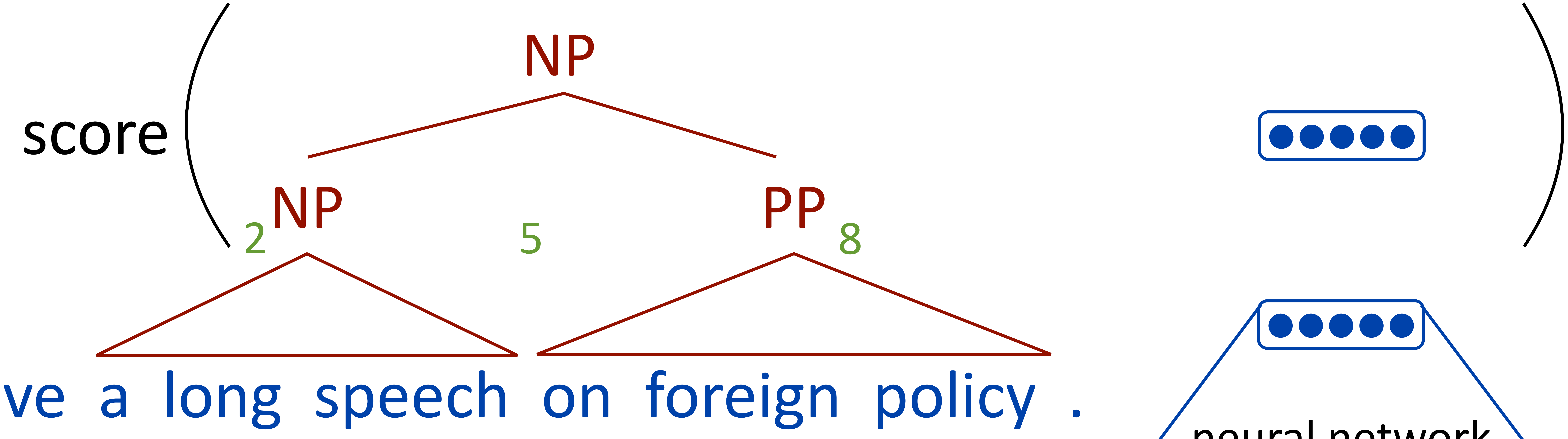
- Neural networks score decisions locally (Collobert et al., 2011)
- Discrete structure with *continuous* features

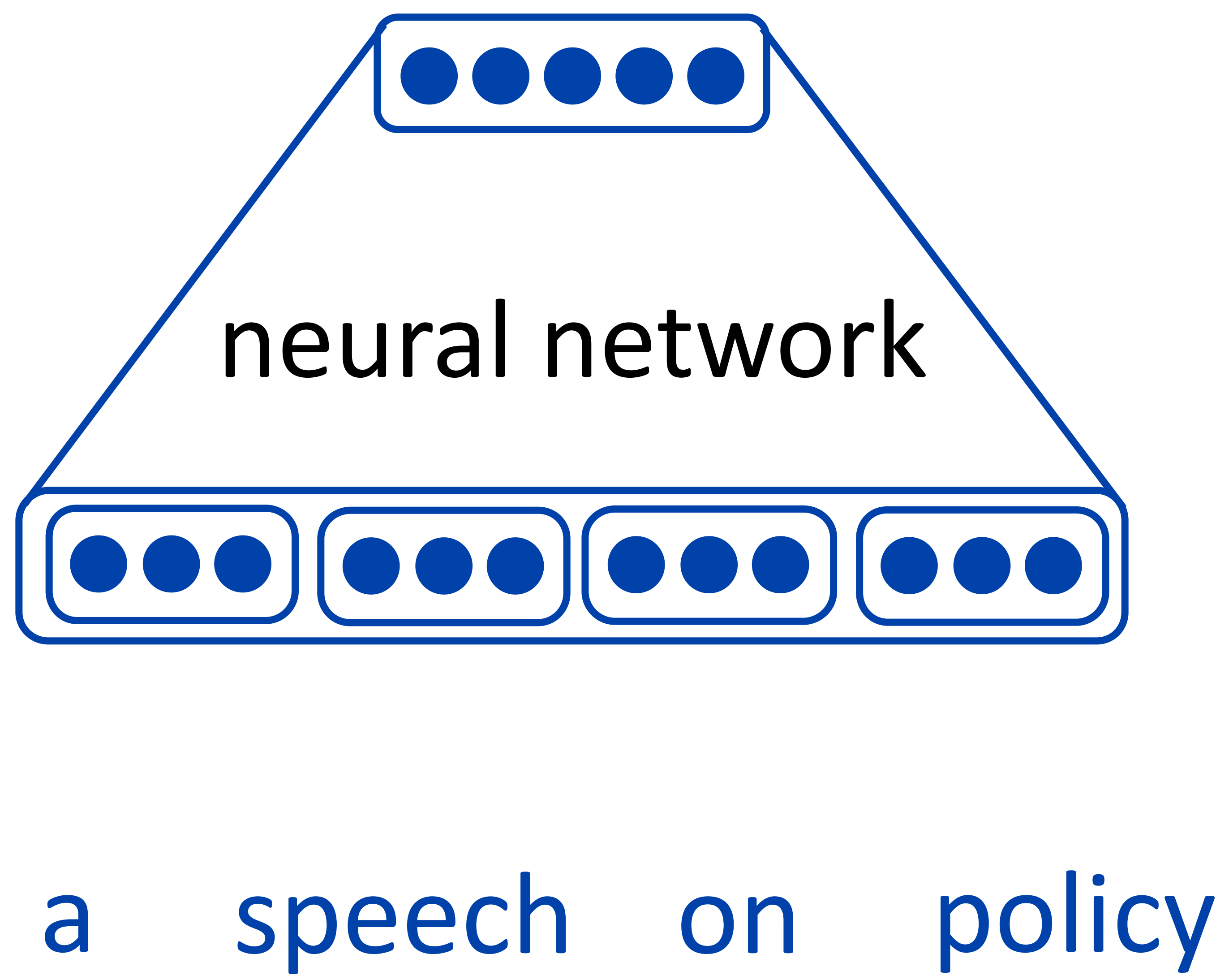
Neural CRF Parsing



- Neural networks score decisions locally (Collobert et al., 2011)
- Discrete structure with *continuous* features
- Inference is still efficient…

Neural CRF Parsing

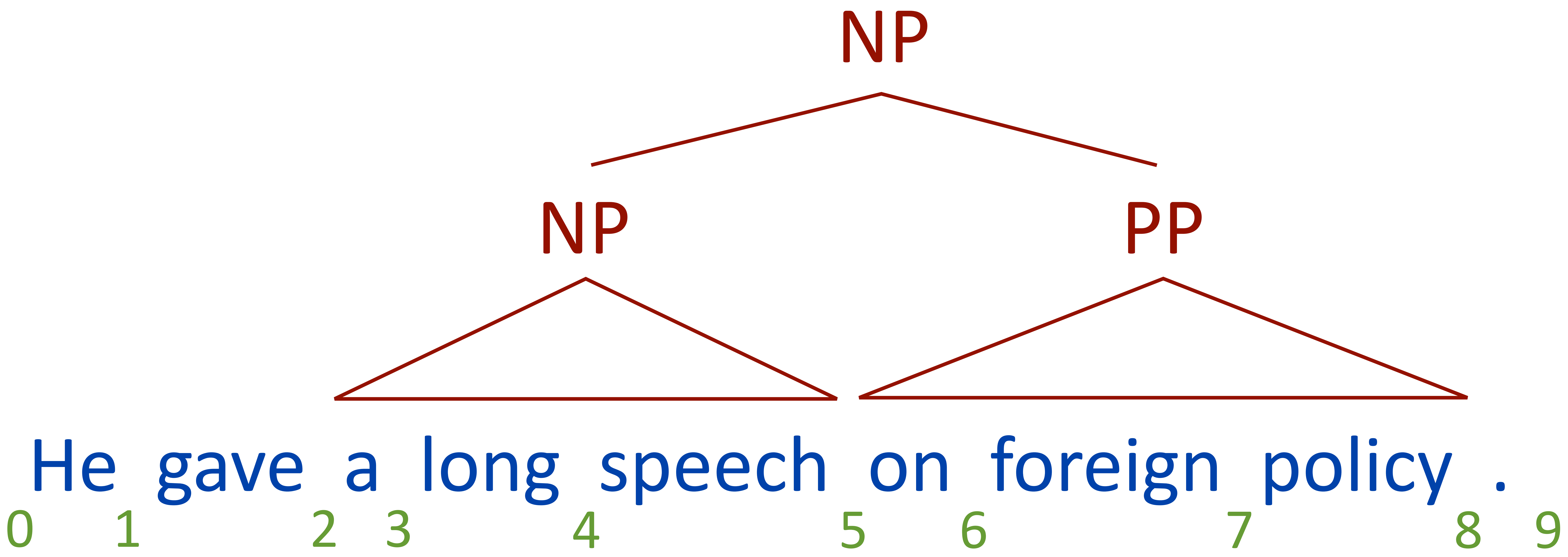




- Neural networks score decisions locally (Collobert et al., 2011)
- Discrete structure with *continuous* features
- Inference is still efficient…and we get nonlinear featurization!

Model

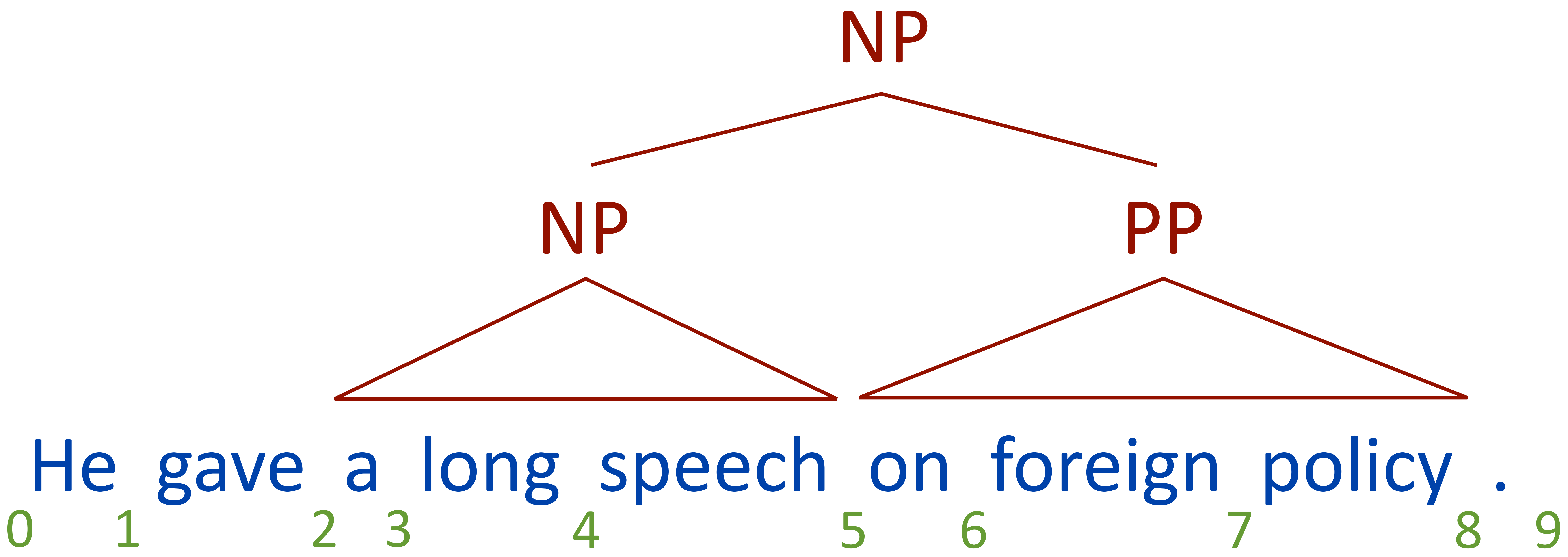
Basic CRF Model



[Hall, Durrett, Klein (2014)]

Basic CRF Model

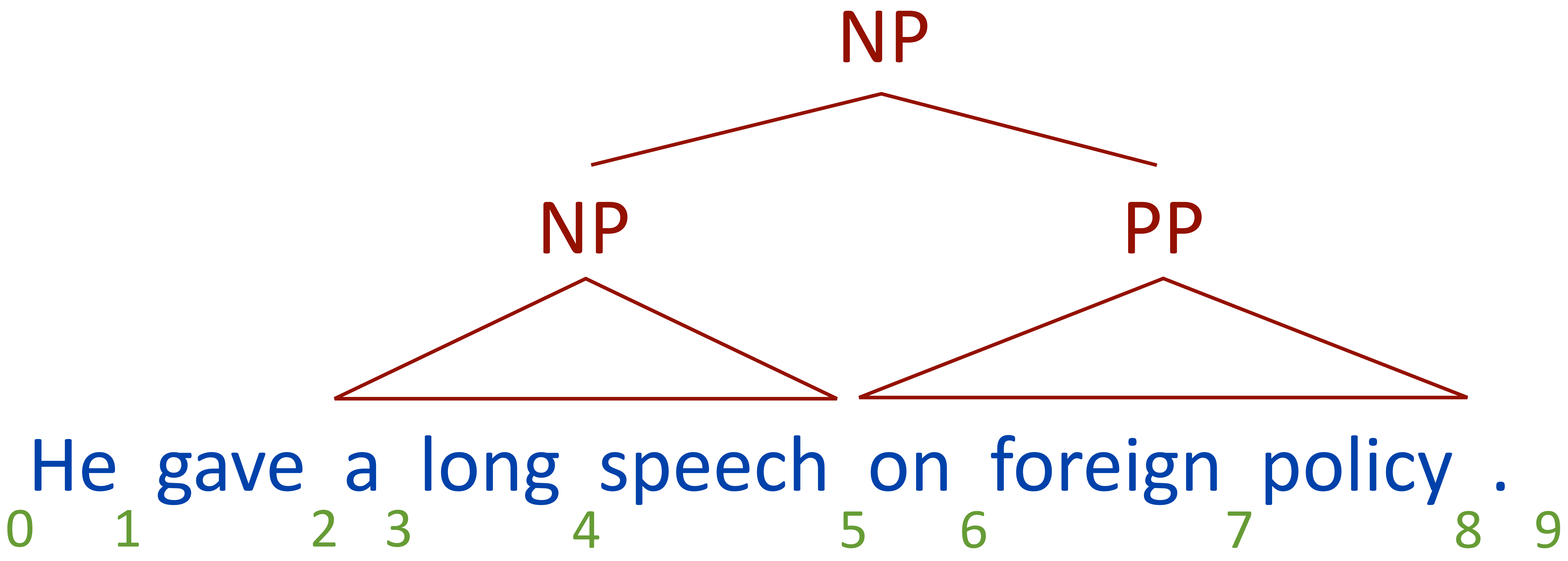
$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right)$$



[Hall, Durrett, Klein (2014)]

Basic CRF Model

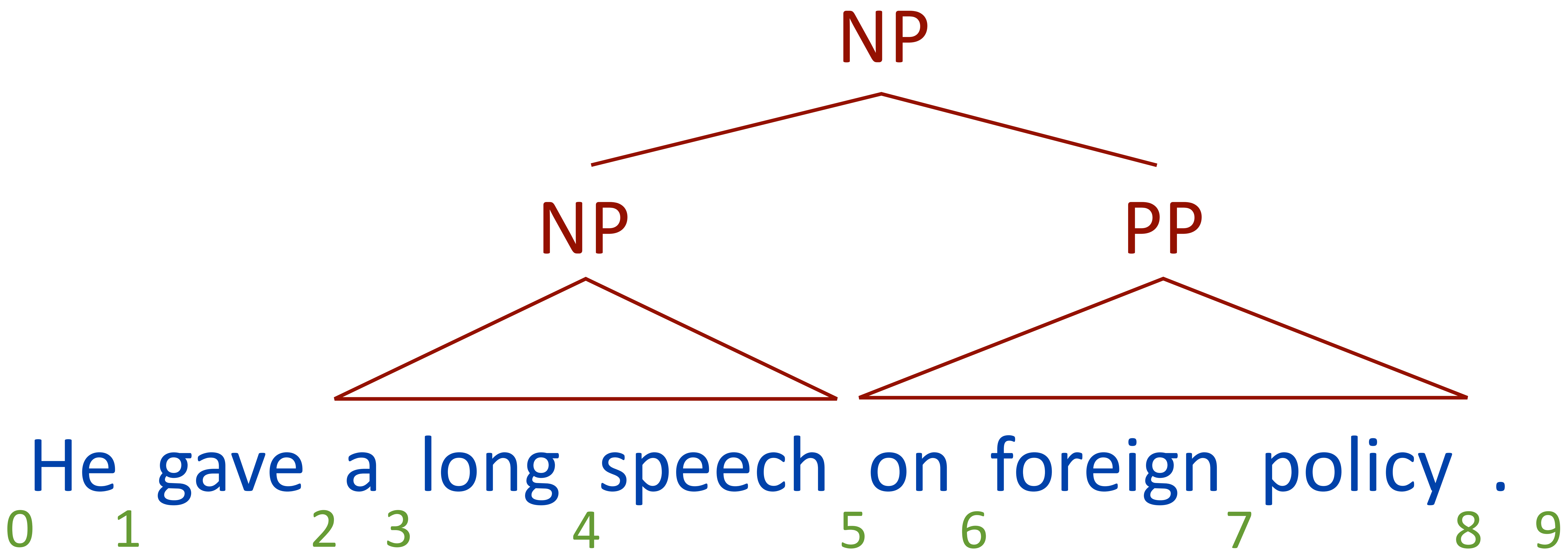
$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right) \qquad \text{score}\left(\begin{array}{c} \text{NP} \\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8 \end{array} \right)$$



[Hall, Durrett, Klein (2014)]

Basic CRF Model

$$P(T|x) \propto \prod_{r \in T} \exp\left(\mathrm{score}(r)\right) \qquad \mathrm{score}\left(\begin{array}{c}\mathrm{NP}\\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8}\end{array}\right) = w^\top f\left(\begin{array}{c}\mathrm{NP}\\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8}\end{array}\right)$$

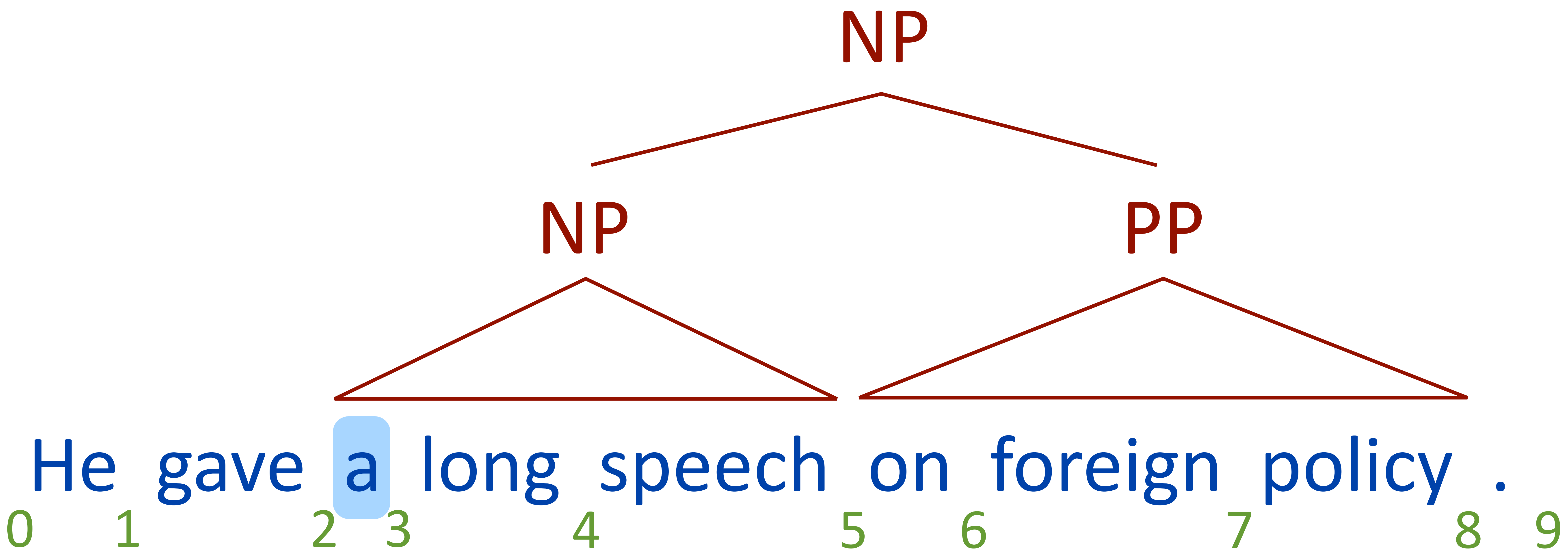


[Hall, Durrett, Klein (2014)]

Basic CRF Model

$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right)$$

$$\text{score}\left(\begin{array}{c} \mathrm{NP} \\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8} \end{array} \right) = w^\top f\left(\begin{array}{c} \mathrm{NP} \\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8} \end{array} \right)$$

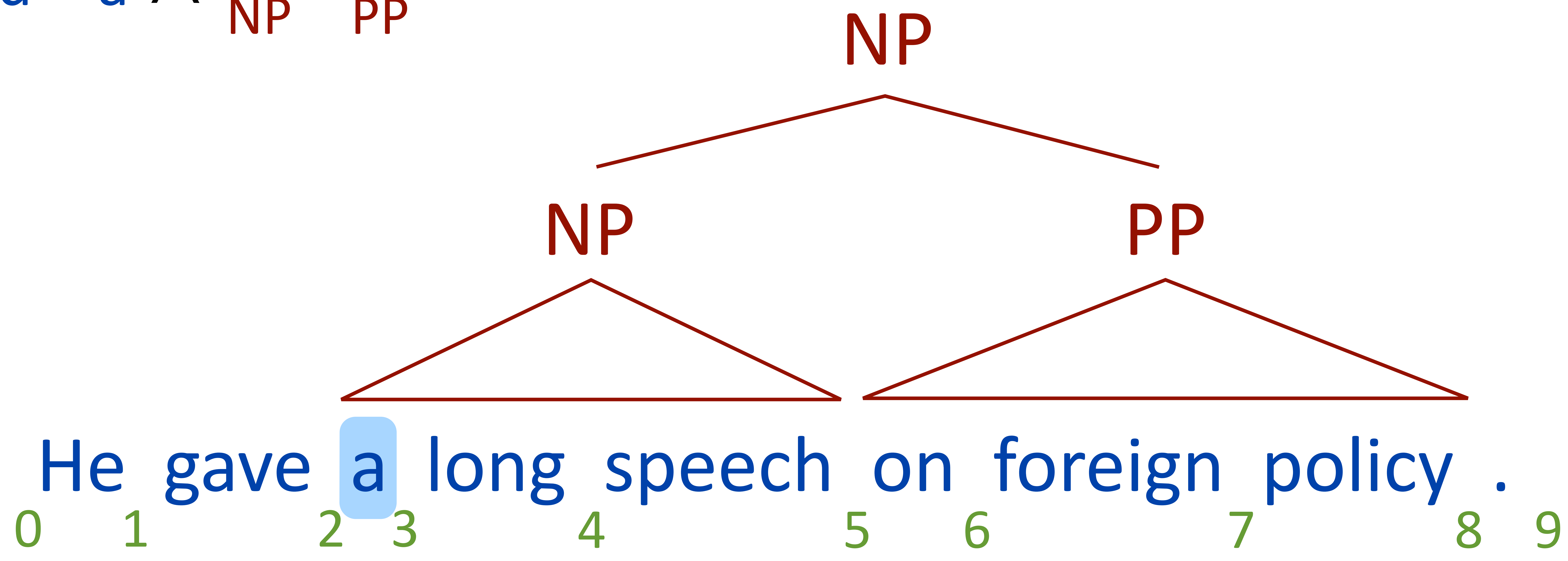


[Hall, Durrett, Klein (2014)]

Basic CRF Model

$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right) \qquad \text{score}\left(\begin{array}{c} \text{NP} \\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8 \end{array} \right) = w^\top f\left(\begin{array}{c} \text{NP} \\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8 \end{array} \right)$$

FirstWord = a $\wedge$ $\begin{array}{c} \text{NP} \\ \text{NP} \quad \text{PP} \end{array}$

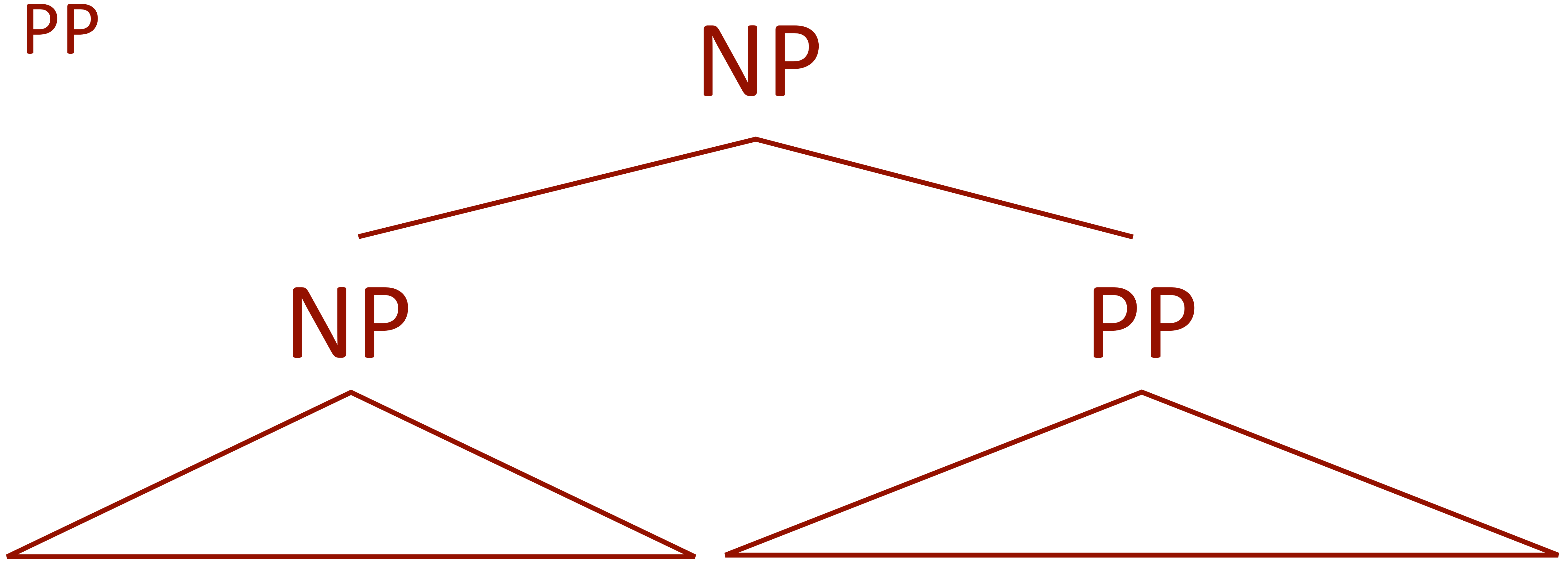


[Hall, Durrett, Klein (2014)]

Basic CRF Model

$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right) \qquad \text{score}\left(\text{NP} \to {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8\right) = w^\top f\left(\text{NP} \to {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8\right)$$

FirstWord = a $\wedge$ $\text{NP} \to \text{NP}\ \text{PP}$



He gave a long speech on foreign policy .

0 1 2 3 4 5 6 7 8 9

[Hall, Durrett, Klein (2014)]

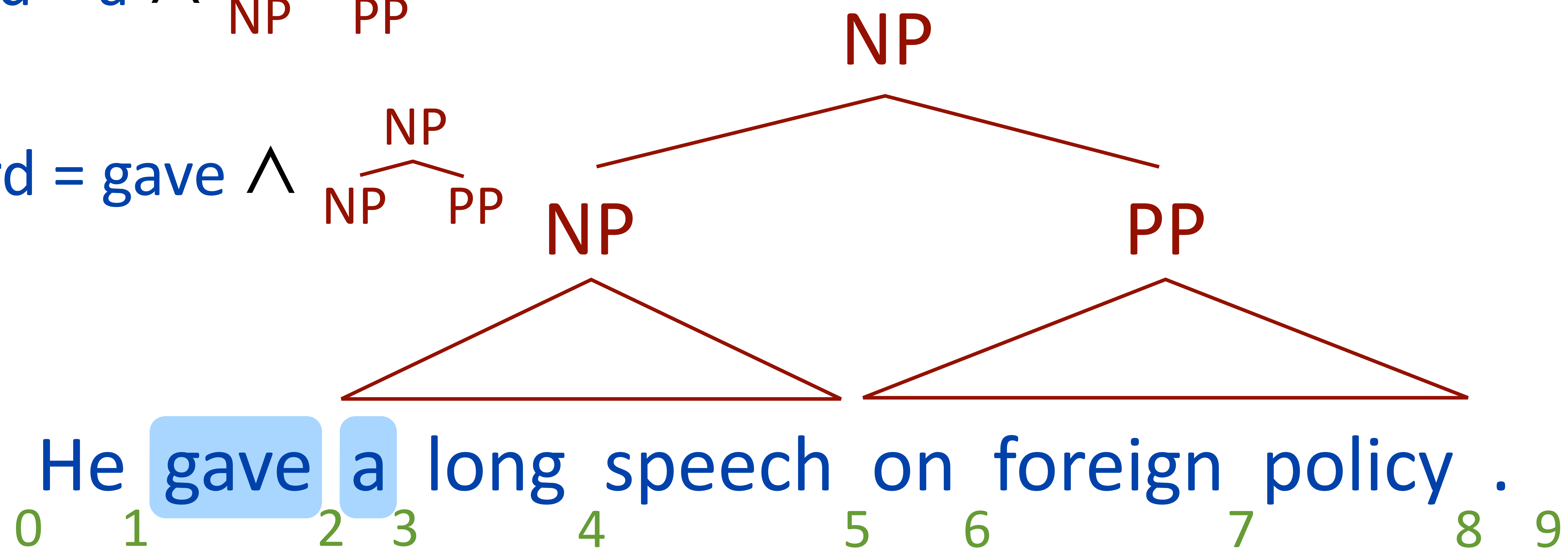
Basic CRF Model

$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right)$$

$$\text{score}\left(\mathrm{NP} \to {}_2\,\mathrm{NP}\,{}_5\,\mathrm{PP}\,{}_8\right) = w^\top f\left(\mathrm{NP} \to {}_2\,\mathrm{NP}\,{}_5\,\mathrm{PP}\,{}_8\right)$$

FirstWord = a $\wedge$ $\mathrm{NP} \to \mathrm{NP}\ \mathrm{PP}$

PrevWord = gave $\wedge$ $\mathrm{NP} \to \mathrm{NP}\ \mathrm{PP}$



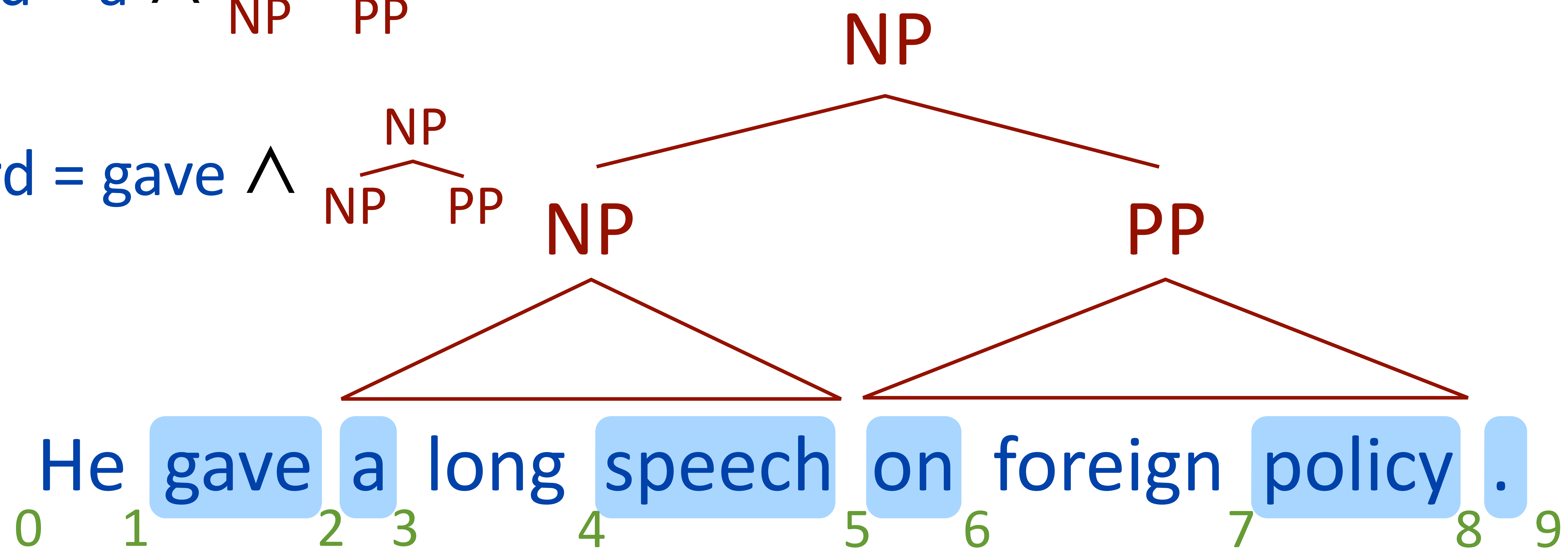
[Hall, Durrett, Klein (2014)]

Basic CRF Model

$$P(T|x) \propto \prod_{r \in T} \exp\left(\text{score}(r)\right) \qquad \text{score}\left(\begin{array}{c} \text{NP} \\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8 \end{array} \right) = w^\top f\left(\begin{array}{c} \text{NP} \\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8 \end{array} \right)$$

$$\text{FirstWord = a} \wedge \begin{array}{c} \text{NP} \\ \text{NP} \quad \text{PP} \end{array}$$

$$\text{PrevWord = gave} \wedge \begin{array}{c} \text{NP} \\ \text{NP} \quad \text{PP} \end{array}$$



[Hall, Durrett, Klein (2014)]

Basic CRF Model

$$\mathrm{score}\left(\begin{array}{c} \text{NP} \\ {}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8} \end{array}\right) = w^{\top} f\left(\begin{array}{c} \text{NP} \\ {}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8} \end{array}\right)$$

$$\text{FirstWord = a} \quad \wedge \quad \begin{array}{c} \text{NP} \\ \text{NP} \quad \text{PP} \end{array}$$

$$\text{PrevWord = gave} \quad \wedge \quad \begin{array}{c} \text{NP} \\ \text{NP} \quad \text{PP} \end{array}$$

Basic CRF Model

$$\mathrm{score}\left(\overset{\text{NP}}{{}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8}}\right) = w^\top f\left(\overset{\text{NP}}{{}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8}}\right)$$

Surface feature

FirstWord = a $\wedge$ $\overset{\text{NP}}{\text{NP} \quad \text{PP}}$

PrevWord = gave $\wedge$ $\overset{\text{NP}}{\text{NP} \quad \text{PP}}$

Basic CRF Model

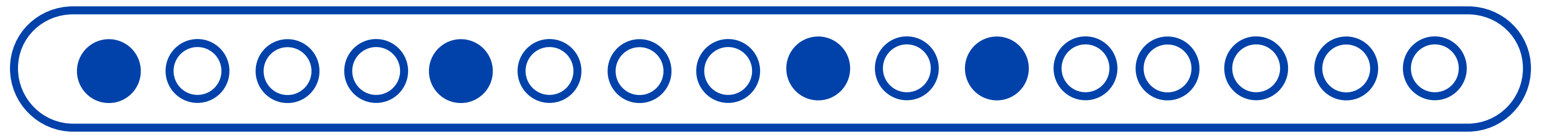
$$\mathrm{score}\left(\begin{array}{c} \mathrm{NP} \\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8} \end{array}\right) = w^{\top} f\left(\begin{array}{c} \mathrm{NP} \\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8} \end{array}\right)$$

Surface feature		Label feature
FirstWord = a	$\wedge$	$\begin{array}{c} \mathrm{NP} \\ \mathrm{NP} \quad \mathrm{PP} \end{array}$
PrevWord = gave	$\wedge$	$\begin{array}{c} \mathrm{NP} \\ \mathrm{NP} \quad \mathrm{PP} \end{array}$

Basic CRF Model

$$\mathrm{score}\left(\overset{\text{NP}}{\overbrace{{}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8}}}\right) = w^\top f\left(\overset{\text{NP}}{\overbrace{{}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8}}}\right)$$

Surface feature		Label feature
FirstWord = a	$\wedge$	$\overset{\text{NP}}{\overbrace{\text{NP} \quad \text{PP}}}$
PrevWord = gave	$\wedge$	$\overset{\text{NP}}{\overbrace{\text{NP} \quad \text{PP}}}$

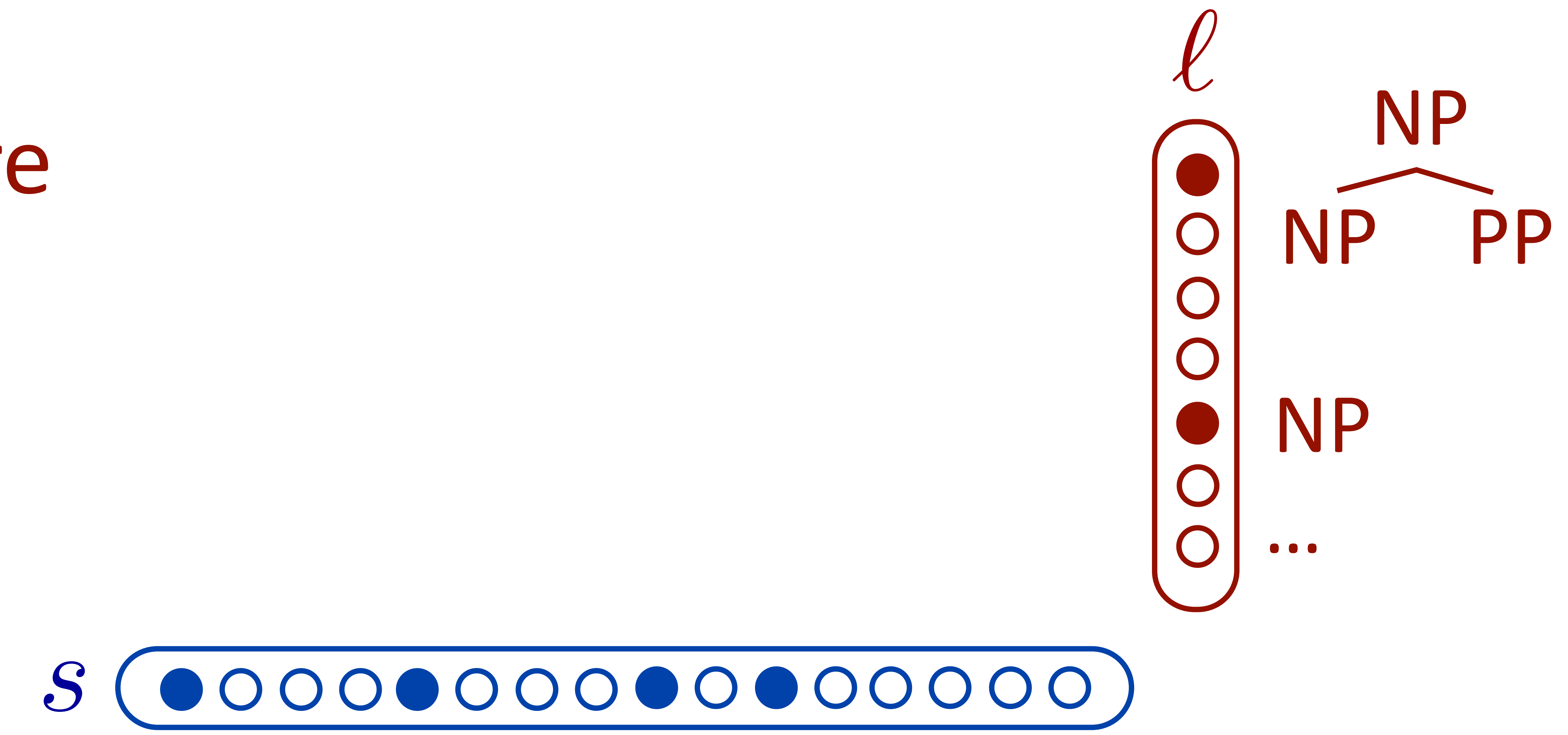
s 

First = a, Prev = gave, …

Basic CRF Model

$$\text{score}\left(\begin{array}{c} \text{NP} \\ \overbrace{{}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8} \end{array} \right) = w^\top f\left(\begin{array}{c} \text{NP} \\ \overbrace{{}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8} \end{array} \right)$$

Surface feature		Label feature
FirstWord = a	$\wedge$	NP → NP PP
PrevWord = gave	$\wedge$	NP → NP PP

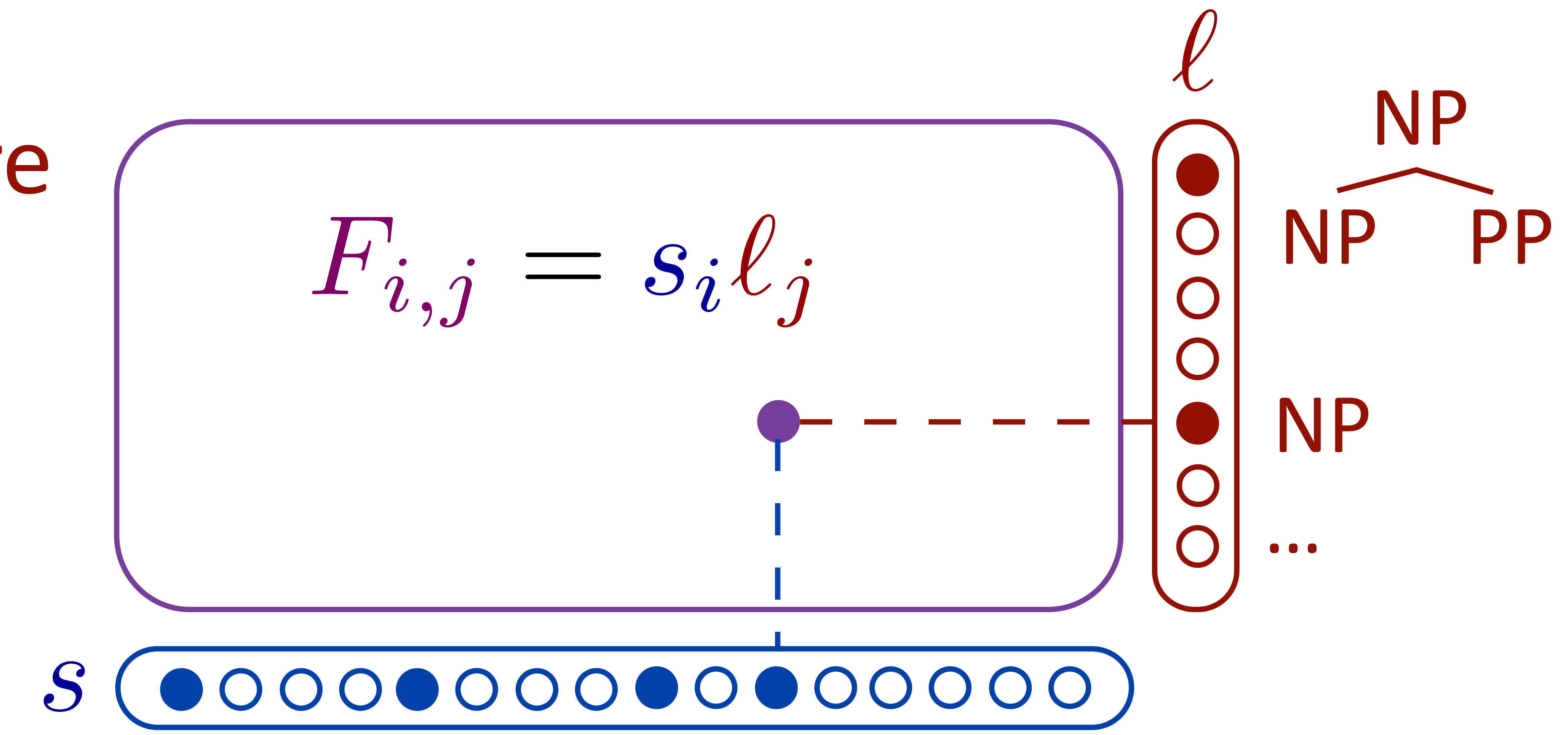


First = a, Prev = gave, ...

Basic CRF Model

$$\mathrm{score}\left(\overset{\text{NP}}{{}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8}\right) = w^\top f\left(\overset{\text{NP}}{{}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8}\right)$$

Surface feature		Label feature
FirstWord = a	$\wedge$	$\overset{\text{NP}}{\text{NP} \quad \text{PP}}$
PrevWord = gave	$\wedge$	$\overset{\text{NP}}{\text{NP} \quad \text{PP}}$

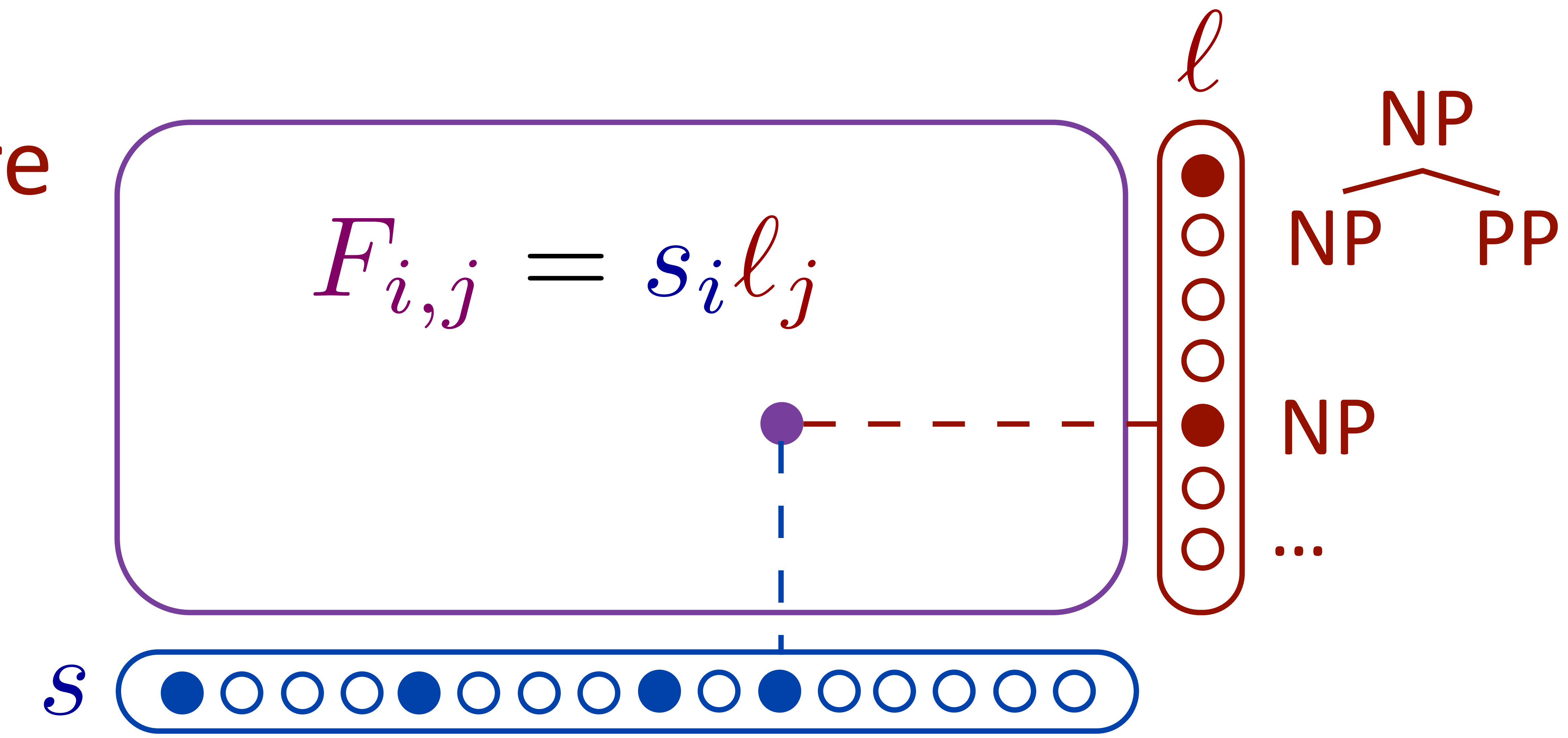


First = a, Prev = gave, ...

Basic CRF Model

$$\text{score}\left(\overset{\mathrm{NP}}{\overbrace{{}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8}}} \right) = w^{\top} f\left(\overset{\mathrm{NP}}{\overbrace{{}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8}}} \right) = W \odot$$

Surface feature		Label feature
FirstWord = a	$\wedge$	$\overset{\mathrm{NP}}{\overbrace{\mathrm{NP}\quad\mathrm{PP}}}$
PrevWord = gave	$\wedge$	$\overset{\mathrm{NP}}{\overbrace{\mathrm{NP}\quad\mathrm{PP}}}$

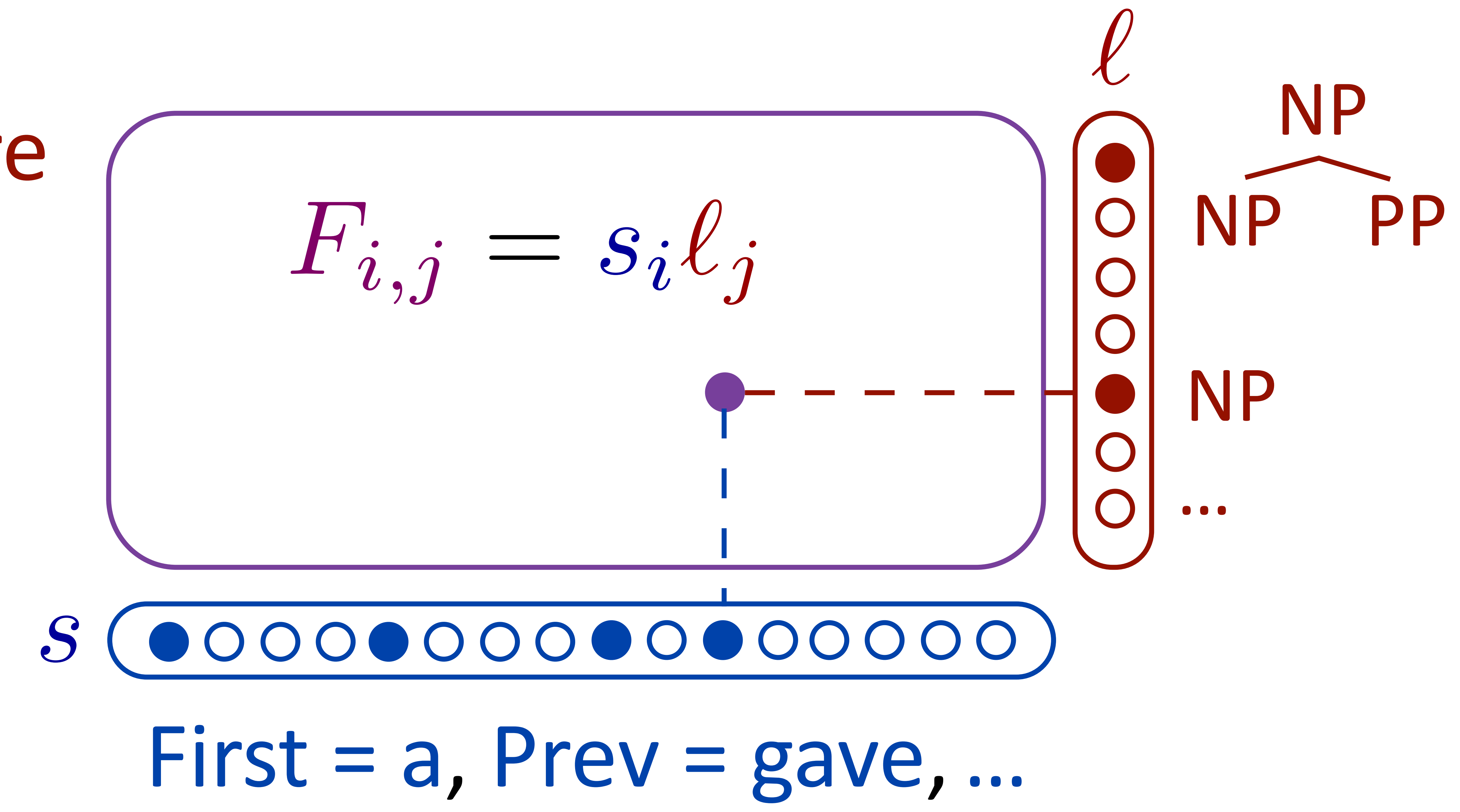


First = a, Prev = gave, …

Basic CRF Model

$$\text{score}\left({}_{2}\,\overset{\text{NP}}{\overbrace{\text{NP}\;{}_{5}\;\text{PP}}}\,{}_{8} \right) = w^\top f\left({}_{2}\,\overset{\text{NP}}{\overbrace{\text{NP}\;{}_{5}\;\text{PP}}}\,{}_{8} \right) = W \odot s\left({}_{2}\,\overset{\text{X}}{\overbrace{\text{X}\;{}_{5}\;\text{X}}}\,{}_{8} \right)$$

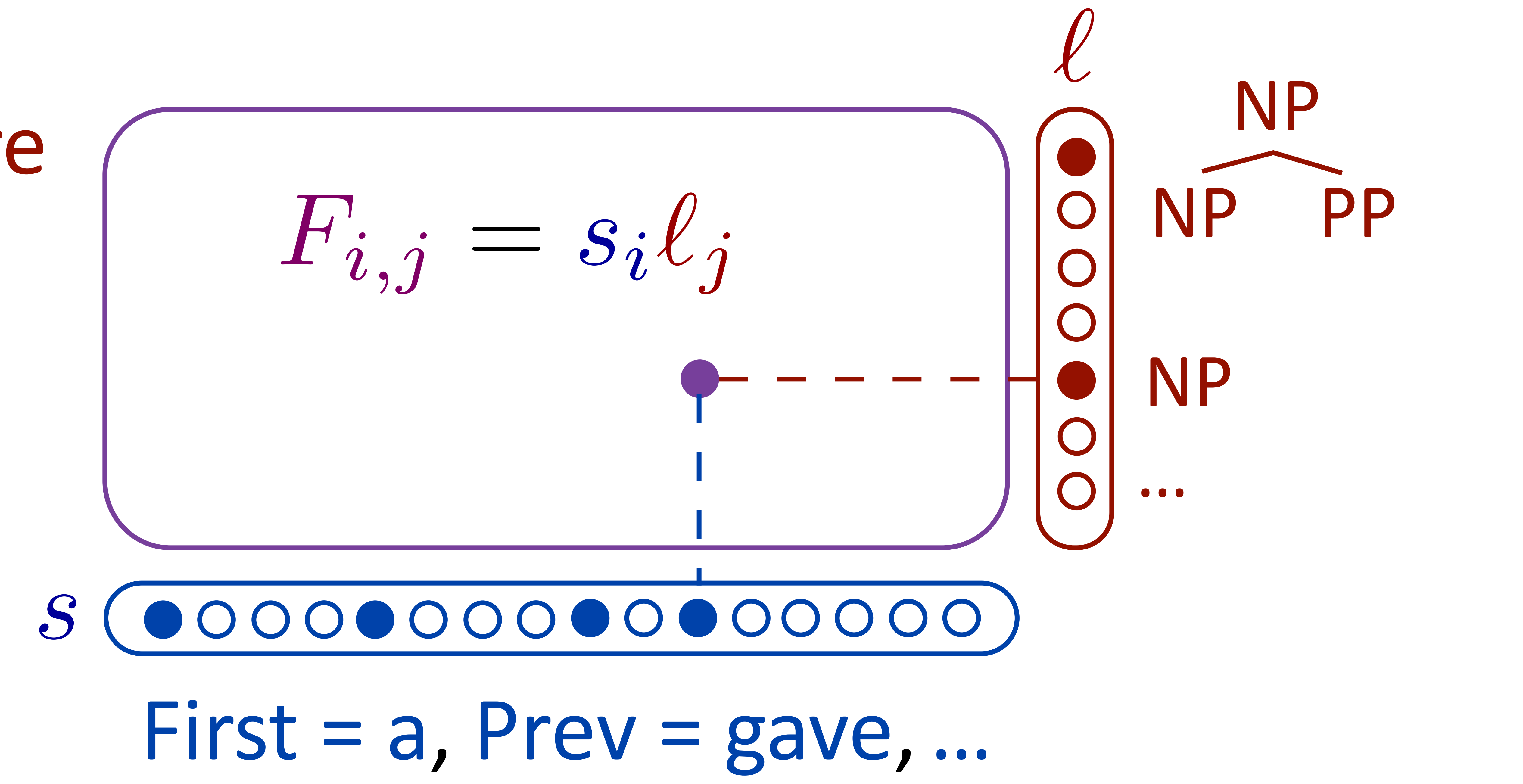
Surface feature		Label feature
FirstWord = a	$\wedge$	$\overset{\text{NP}}{\overbrace{\text{NP} \quad \text{PP}}}$
PrevWord = gave	$\wedge$	$\overset{\text{NP}}{\overbrace{\text{NP} \quad \text{PP}}}$



Basic CRF Model

$$\text{score}\left(\begin{array}{c}\text{NP}\\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8\end{array}\right) = w^\top f\left(\begin{array}{c}\text{NP}\\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8\end{array}\right) = W \odot s\left(\begin{array}{c}\text{X}\\ {}_2\,\text{X}\,{}_5\,\text{X}\,{}_8\end{array}\right)\ell^\top\left(\begin{array}{c}\text{NP}\\ \text{NP}\ \ \text{PP}\end{array}\right)$$

Surface feature		Label feature
FirstWord = a	$\wedge$	NP → NP PP
PrevWord = gave	$\wedge$	NP → NP PP



Neural CRF Model

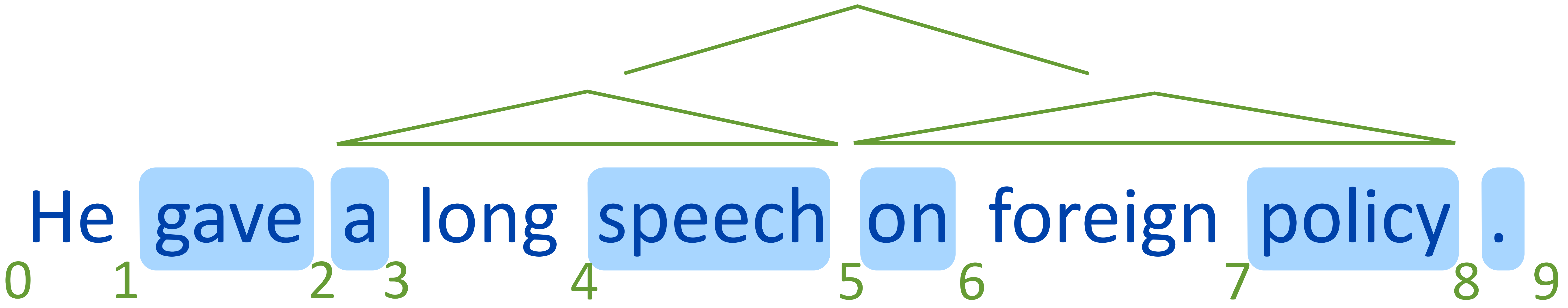
$$\mathrm{score}\left(\begin{array}{c} \mathrm{NP} \\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8} \end{array}\right) = W \odot s\left(\begin{array}{c} \mathrm{X} \\ {}_{2}\,\mathrm{X}\,{}_{5}\,\mathrm{X}\,{}_{8} \end{array}\right) \ell^{\top}\left(\begin{array}{c} \mathrm{NP} \\ \mathrm{NP} \quad \mathrm{PP} \end{array}\right)$$

Neural CRF Model

$$\text{score}\left(\overset{\text{NP}}{{}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8}} \right) = W \odot s\left(\overset{\text{X}}{{}_{2}\,\text{X}\,{}_{5}\,\text{X}\,{}_{8}} \right) \ell^{\top}\left(\overset{\text{NP}}{\text{NP}\quad\text{PP}} \right)$$

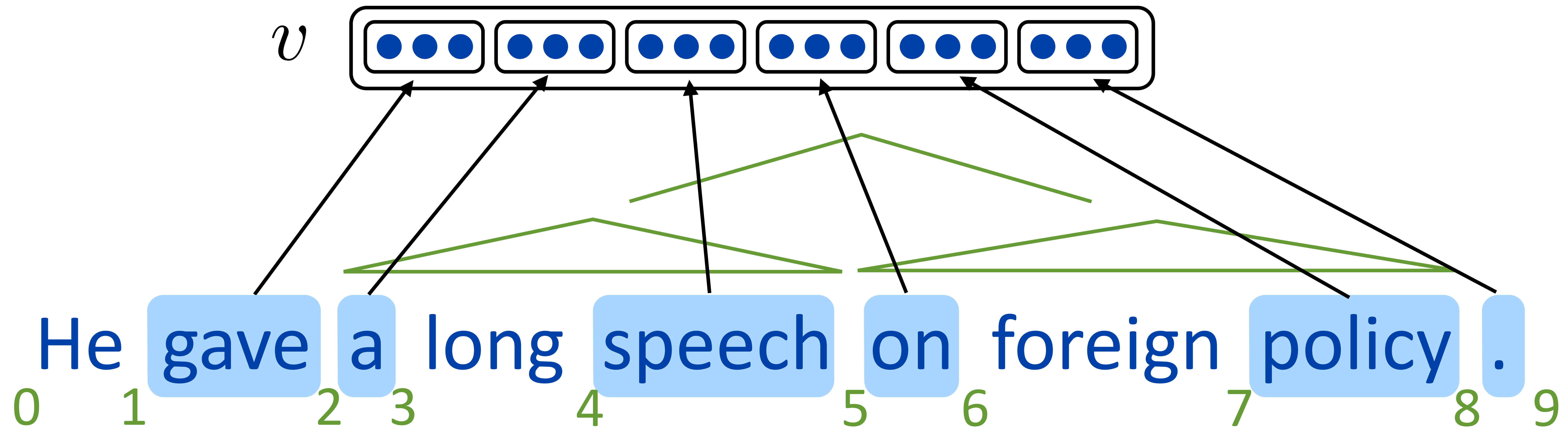
Neural CRF Model

$$\mathrm{score}\left(\overset{\mathrm{NP}}{{}_2\,\mathrm{NP}\,{}_5\,\mathrm{PP}\,{}_8}\right) = W \odot s\left(\overset{\mathrm{X}}{{}_2\,\mathrm{X}\,{}_5\,\mathrm{X}\,{}_8}\right)\ell^{\top}\left(\overset{\mathrm{NP}}{\mathrm{NP}\quad\mathrm{PP}}\right)$$



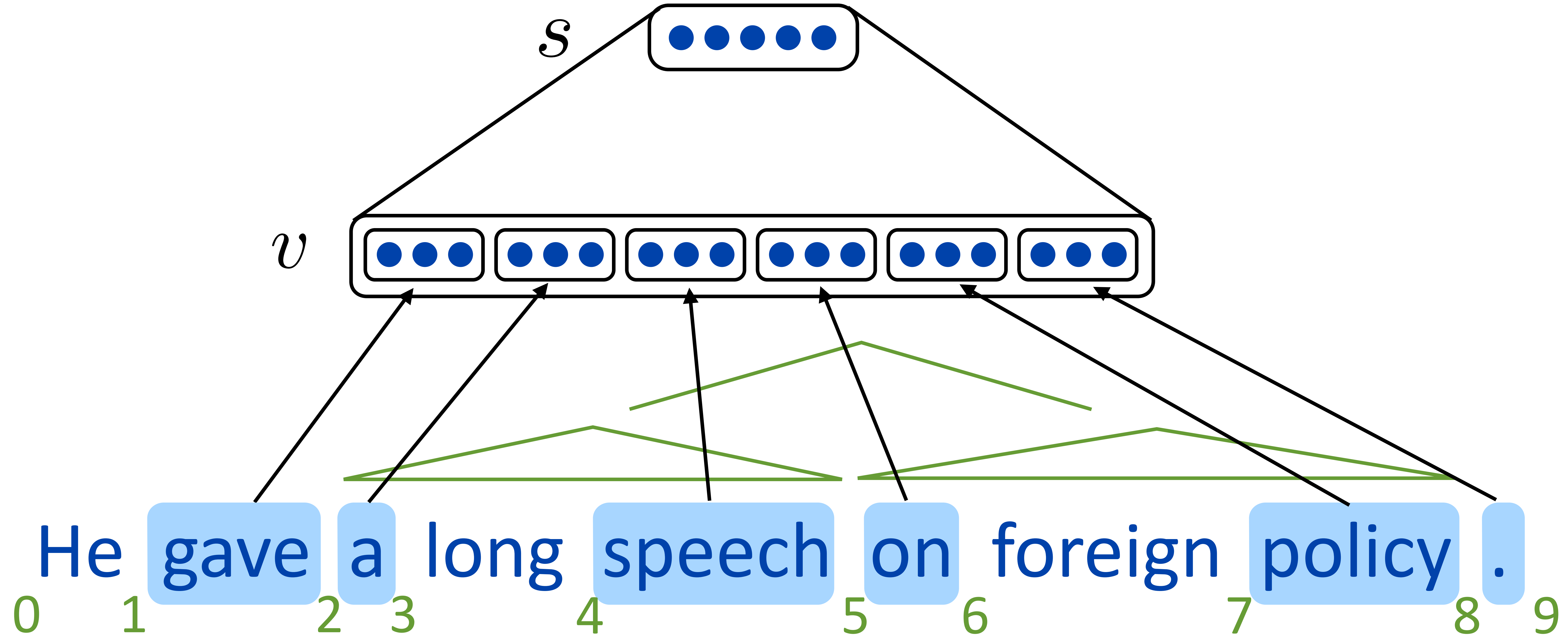
Neural CRF Model

$$\mathrm{score}\left({}_{2}\overset{\mathrm{NP}}{\mathrm{NP}\;{}_{5}\;\mathrm{PP}}{}_{8} \right) = W \odot s\left({}_{2}\overset{\mathrm{X}}{\mathrm{X}\;{}_{5}\;\mathrm{X}}{}_{8} \right) \ell^{\top}\left(\overset{\mathrm{NP}}{\mathrm{NP}\;\;\mathrm{PP}} \right)$$



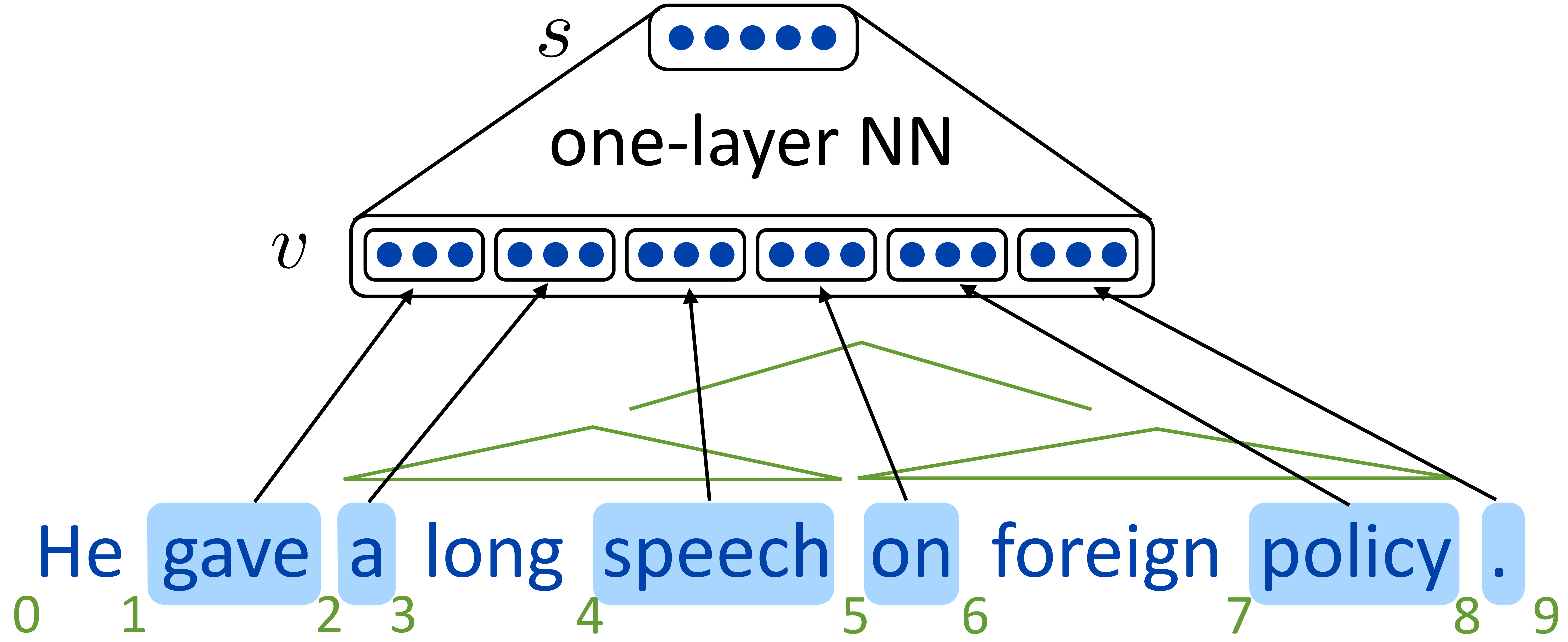
Neural CRF Model

$$\mathrm{score}\left(\begin{array}{c}\mathrm{NP}\\ {}_2\,\mathrm{NP}\,{}_5\,\mathrm{PP}\,{}_8\end{array}\right) = W \odot \boldsymbol{s}\left(\begin{array}{c}\mathrm{X}\\ {}_2\,\mathrm{X}\,{}_5\,\mathrm{X}\,{}_8\end{array}\right) \ell^\top\left(\begin{array}{c}\mathrm{NP}\\ \mathrm{NP}\;\;\mathrm{PP}\end{array}\right)$$



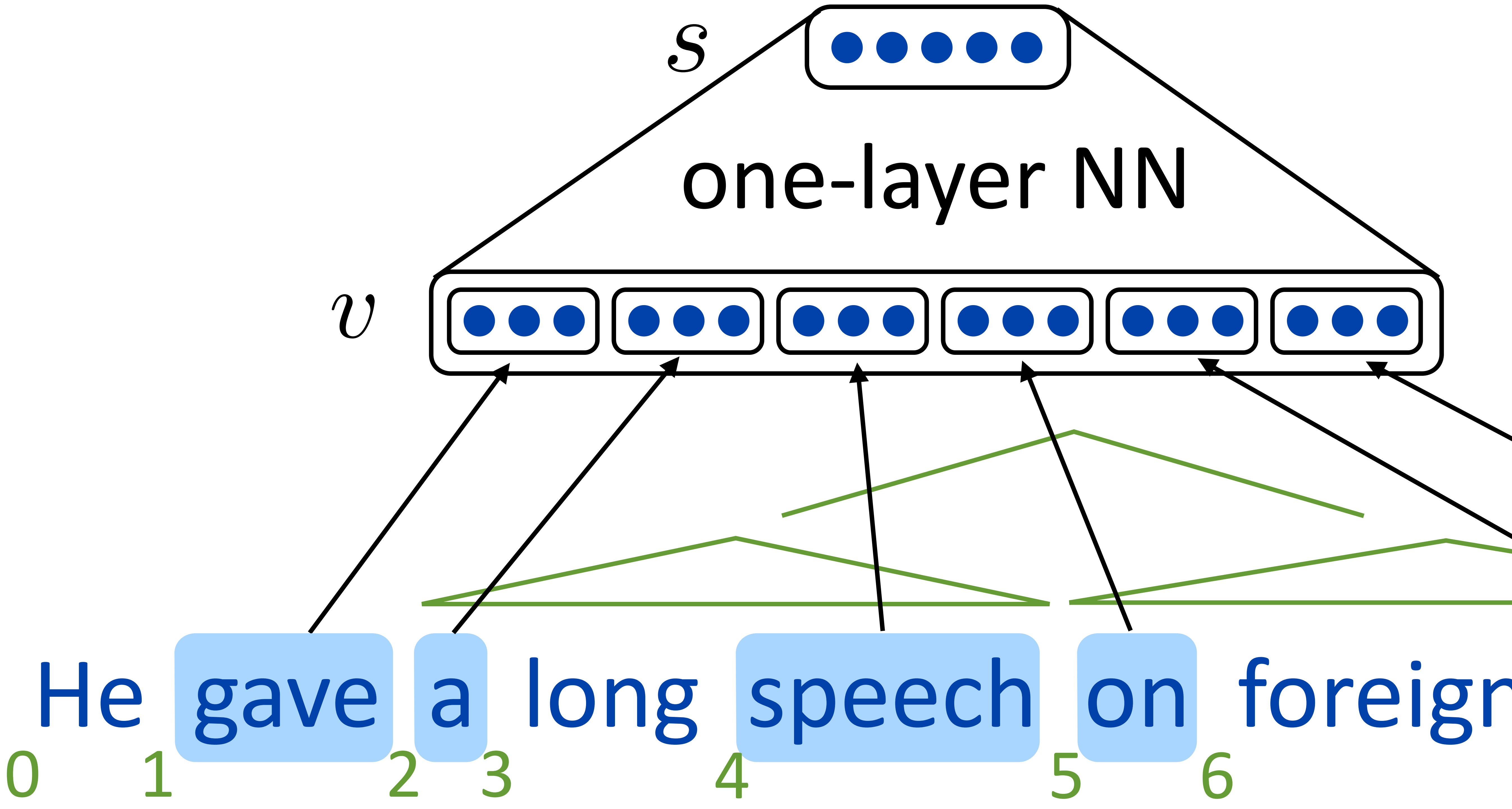
Neural CRF Model

$$\text{score}\left({}_2\mathrm{NP}_5\mathrm{PP}_8 \rightarrow \mathrm{NP} \right) = W \odot s\left({}_2\mathrm{X}_5\mathrm{X}_8 \rightarrow \mathrm{X} \right) \ell^\top \left(\mathrm{NP} \rightarrow \mathrm{NP}\ \mathrm{PP} \right)$$



Neural CRF Model

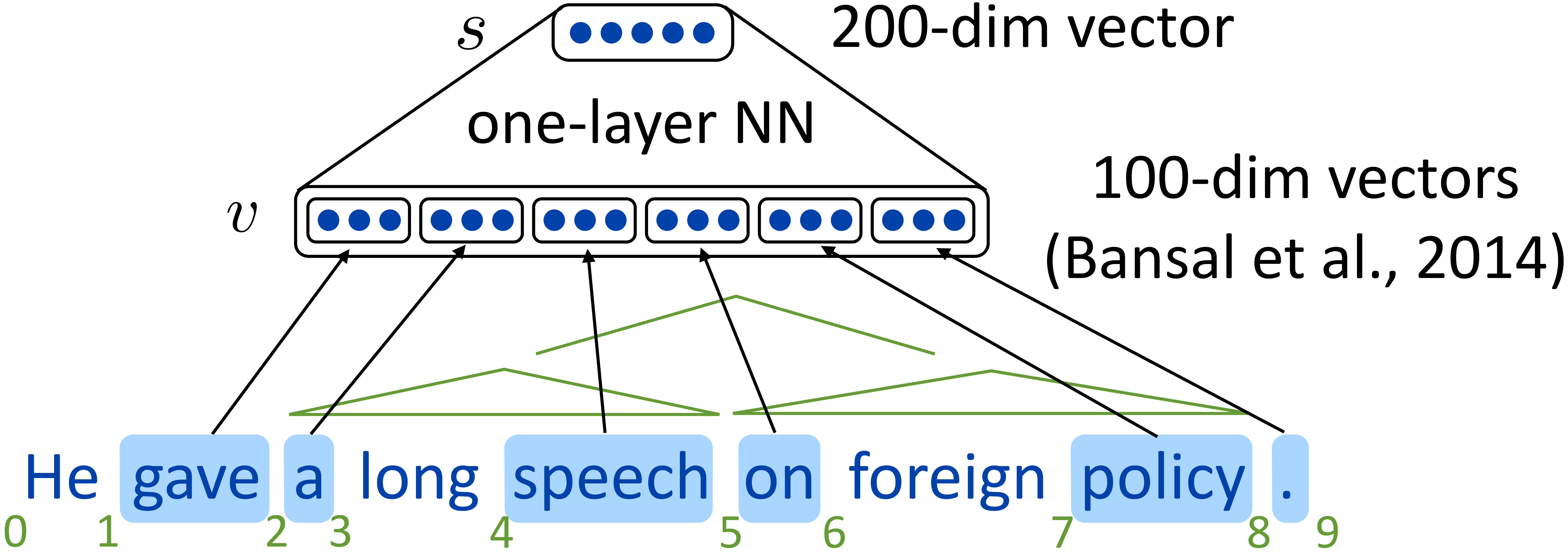
$$\text{score}\left(\begin{array}{c}\mathrm{NP}\\ {}_2\,\mathrm{NP}\,{}_5\,\mathrm{PP}\,{}_8\end{array}\right) = W \odot s\left(\begin{array}{c}\mathrm{X}\\ {}_2\,\mathrm{X}\,{}_5\,\mathrm{X}\,{}_8\end{array}\right)\ell^\top\left(\begin{array}{c}\mathrm{NP}\\ \mathrm{NP}\quad \mathrm{PP}\end{array}\right)$$



100-dim vectors
(Bansal et al., 2014)

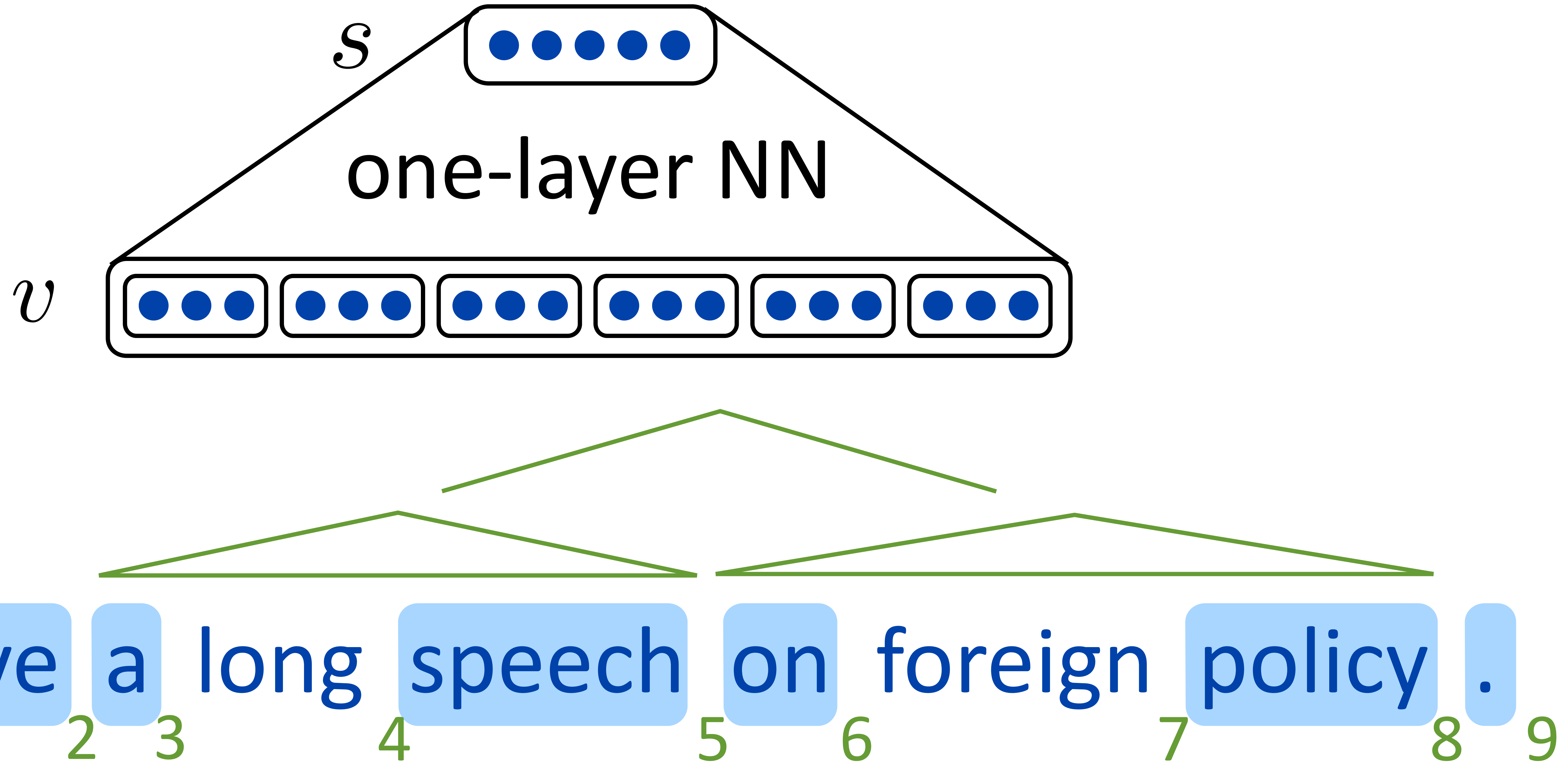
Neural CRF Model

$$\text{score}\left(\begin{array}{c}\text{NP}\\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8\end{array}\right) = W \odot s\left(\begin{array}{c}\text{X}\\ {}_2\,\text{X}\,{}_5\,\text{X}\,{}_8\end{array}\right) \ell^\top\left(\begin{array}{c}\text{NP}\\ \text{NP}\quad\text{PP}\end{array}\right)$$



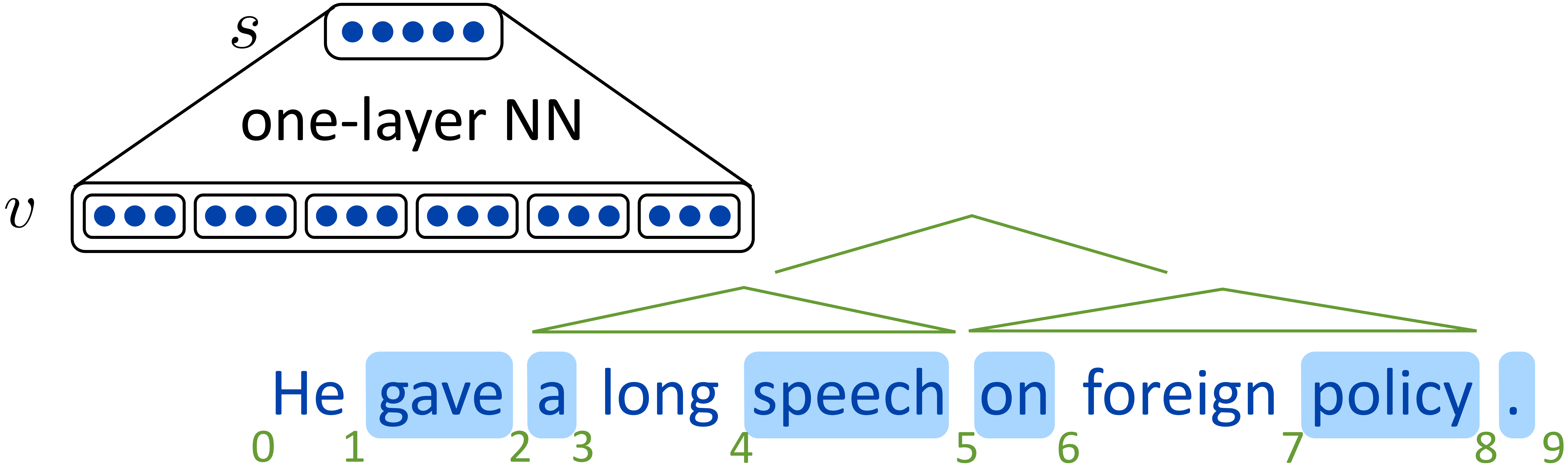
Neural CRF Model

$$\mathrm{score}\left(\begin{array}{c}\mathrm{NP}\\ {}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8}\end{array}\right) = W \odot s\left(\begin{array}{c}\mathrm{X}\\ {}_{2}\,\mathrm{X}\,{}_{5}\,\mathrm{X}\,{}_{8}\end{array}\right)\ell^{\top}\left(\begin{array}{c}\mathrm{NP}\\ \mathrm{NP}\quad\mathrm{PP}\end{array}\right)$$



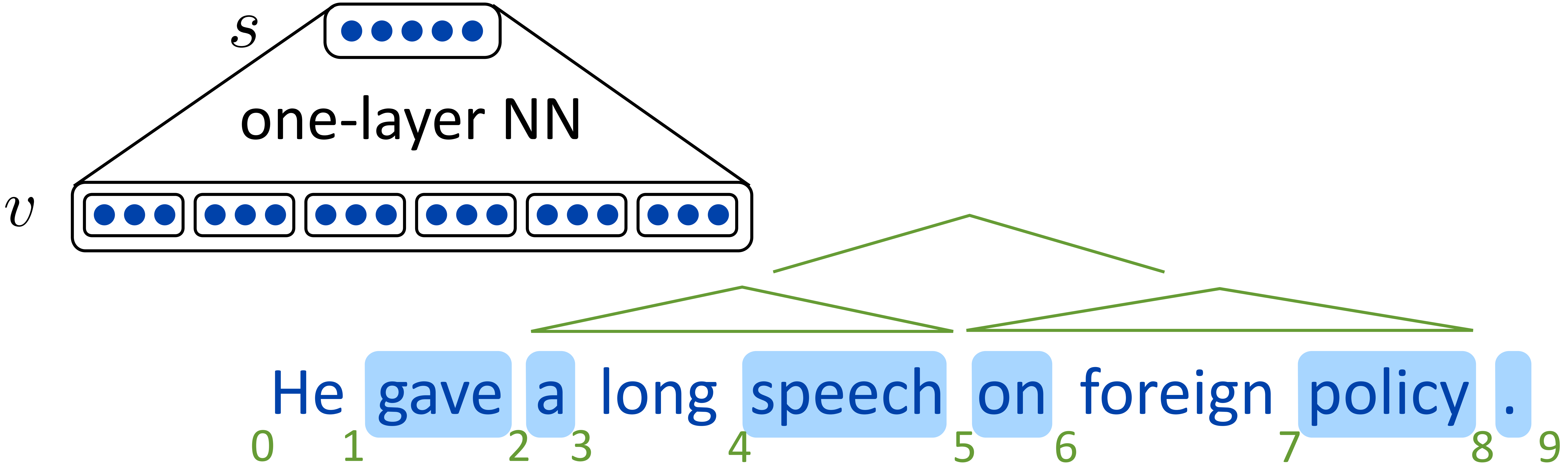
Neural CRF Model

$$\mathrm{score}\left({}_{2}\,\mathrm{NP}\,{}_{5}\,\mathrm{PP}\,{}_{8} \rightarrow \mathrm{NP} \right) = W \odot s\left({}_{2}\,X\,{}_{5}\,X\,{}_{8} \rightarrow X \right) \ell^{\top}\left(\mathrm{NP}\ \mathrm{PP} \rightarrow \mathrm{NP} \right)$$



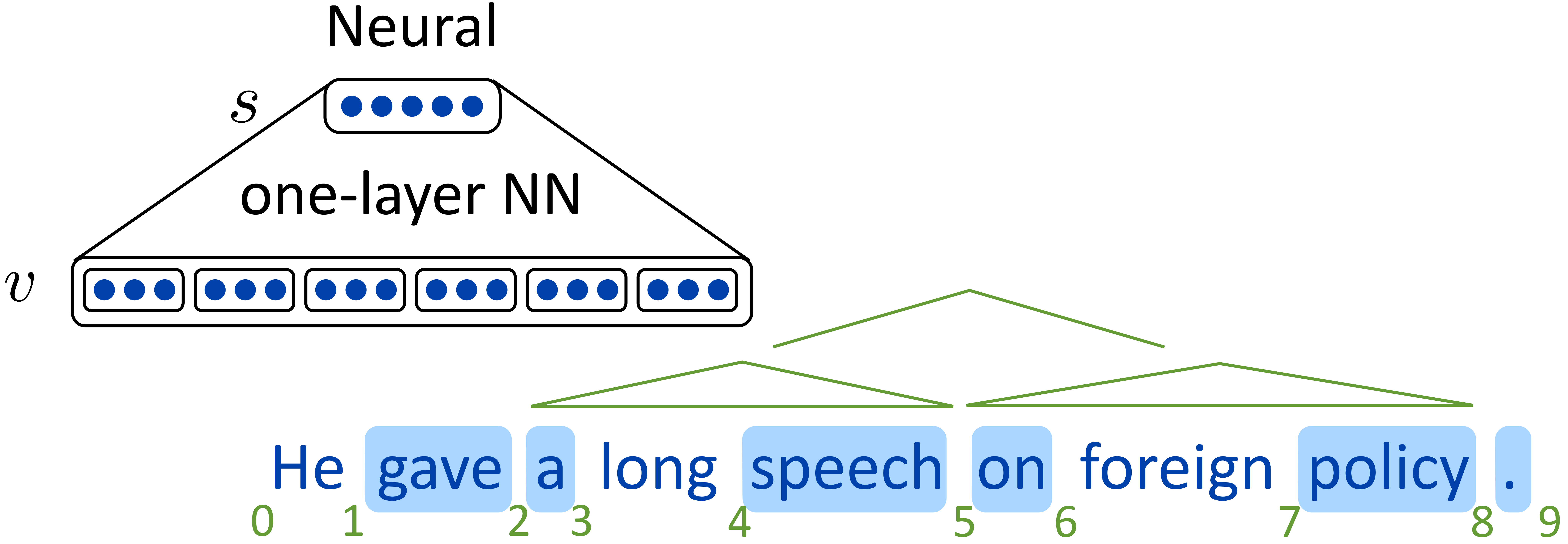
Neural CRF Model

$$\text{score}\left({}_2\text{NP}_5\,\text{PP}_8 \to \text{NP} \right) = W \odot s\left({}_2\text{X}_5\,\text{X}_8 \to \text{X} \right) \ell^\top \left(\text{NP}\ \text{PP} \to \text{NP} \right)$$



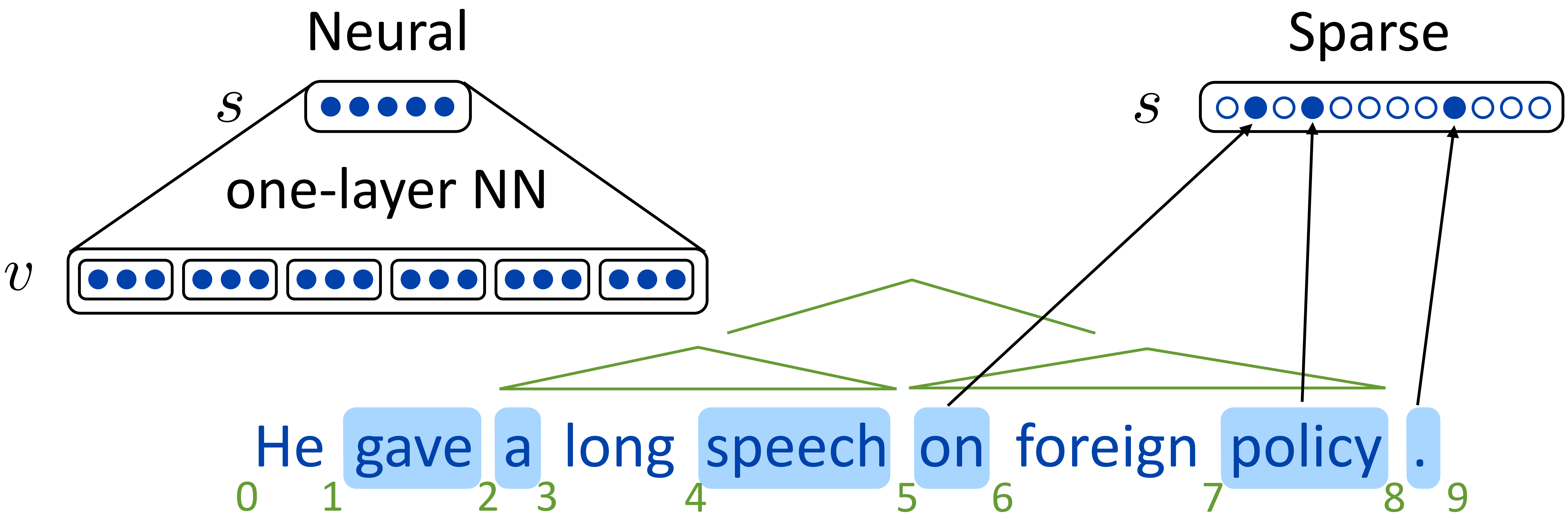
Neural CRF Model

$$\mathrm{score}\left(\begin{array}{c} \text{NP} \\ {}_2\,\text{NP}\,{}_5\,\text{PP}\,{}_8 \end{array} \right) = W \odot s\left(\begin{array}{c} \text{X} \\ {}_2\,\text{X}\,{}_5\,\text{X}\,{}_8 \end{array} \right) \ell^\top \left(\begin{array}{c} \text{NP} \\ \text{NP} \;\; \text{PP} \end{array} \right)$$



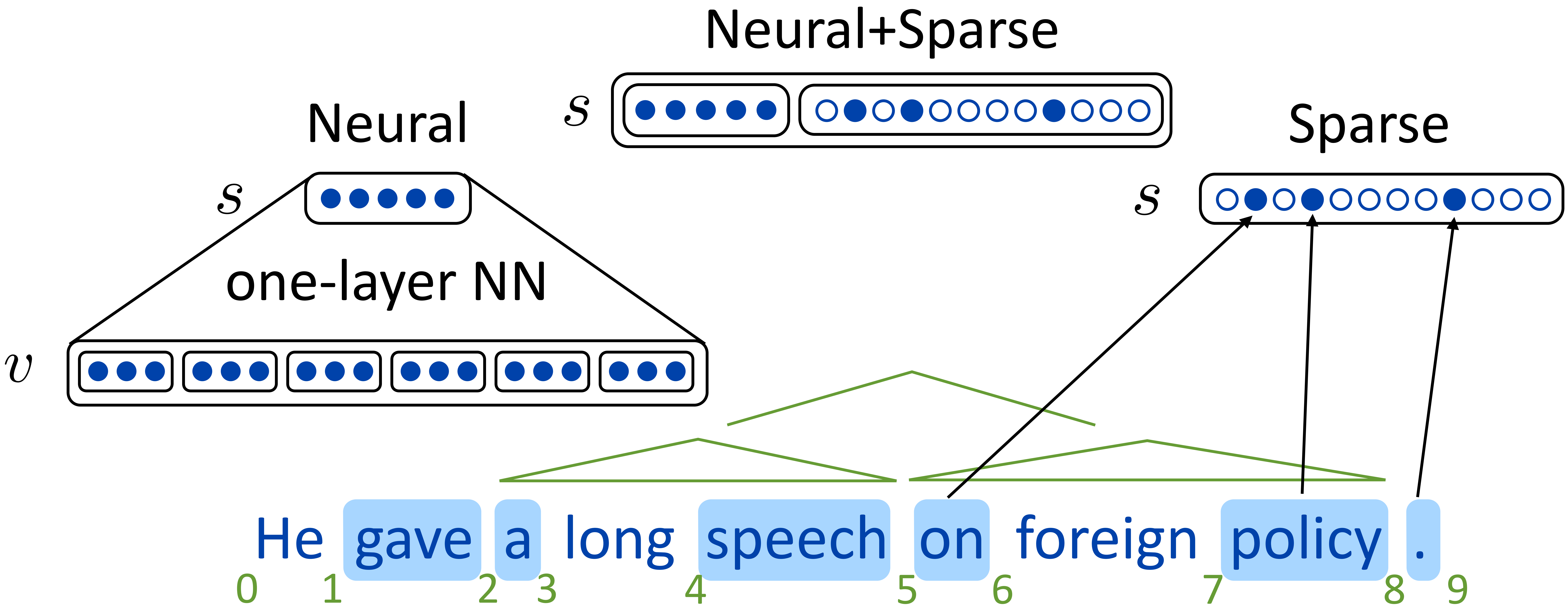
Neural CRF Model

$$\mathrm{score}\left(\begin{array}{c}\text{NP}\\ {}_{2}\,\text{NP}\,{}_{5}\,\text{PP}\,{}_{8}\end{array}\right) = W \odot s\left(\begin{array}{c}\text{X}\\ {}_{2}\,\text{X}\,{}_{5}\,\text{X}\,{}_{8}\end{array}\right) \ell^{\top}\left(\begin{array}{c}\text{NP}\\ \text{NP}\;\;\text{PP}\end{array}\right)$$



Neural CRF Model

$$\text{score}\left(\mathrm{NP} \rightarrow {}_2\,\mathrm{NP}\,{}_5\,\mathrm{PP}\,{}_8\right) = W \odot s\left(\mathrm{X} \rightarrow {}_2\,\mathrm{X}\,{}_5\,\mathrm{X}\,{}_8\right)\,\ell^\top\left(\mathrm{NP} \rightarrow \mathrm{NP}\;\mathrm{PP}\right)$$



Inference

Inference

Just CKY!

Inference

Just CKY!

… with coarse pruning and caching of neural net operations

(Goodman, 1997) (Chen and Manning, 2014)

Inference

Just CKY!

… with coarse pruning and caching of neural net operations

(Goodman, 1997) (Chen and Manning, 2014)

Roughly 2x slower than with sparse features alone

Learning

Just Maximum Likelihood!

Learning

Just Maximum Likelihood!

… with backpropagation through each local neural network

Learning

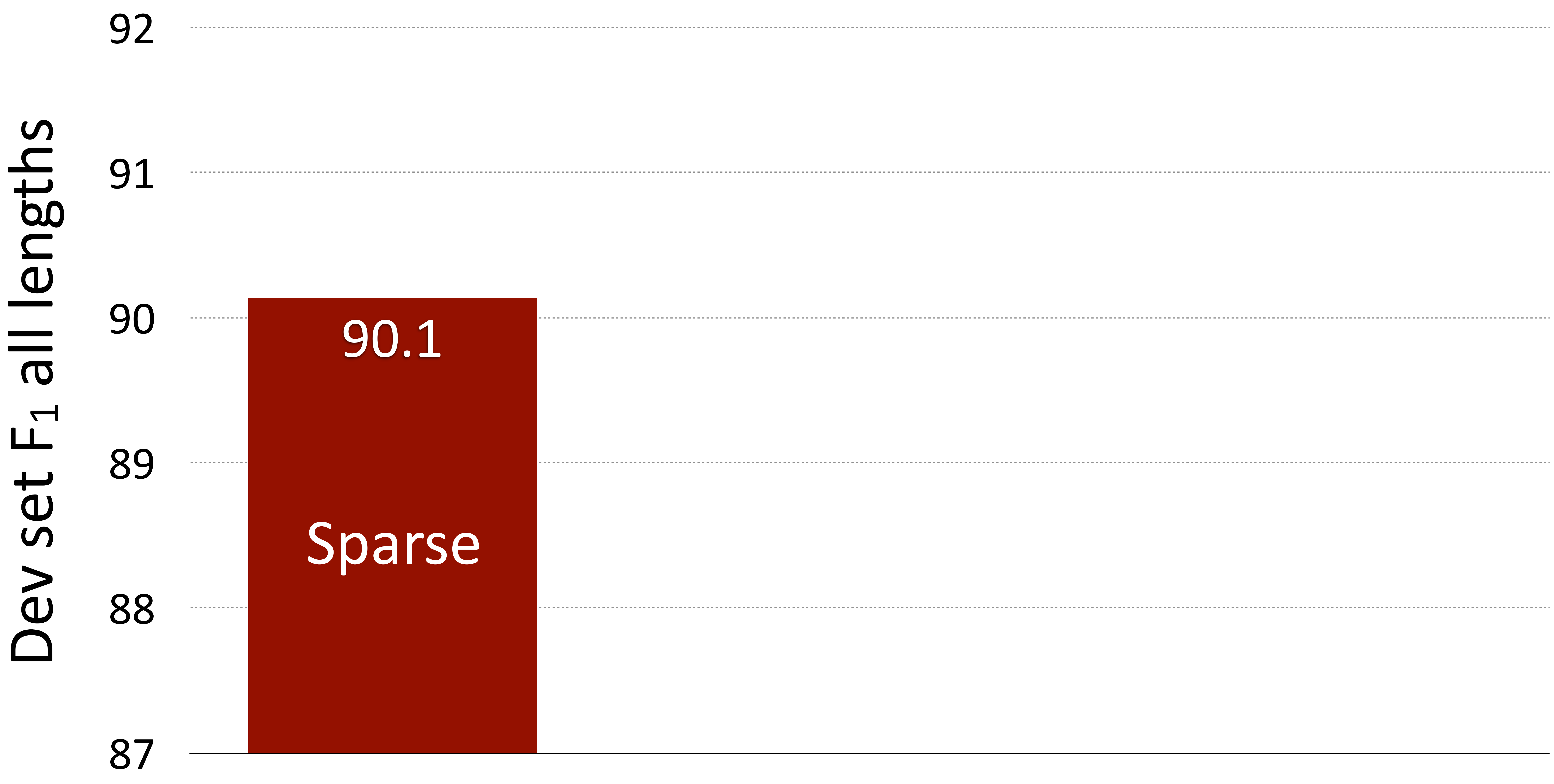
Just Maximum Likelihood!

… with backpropagation through each local neural network

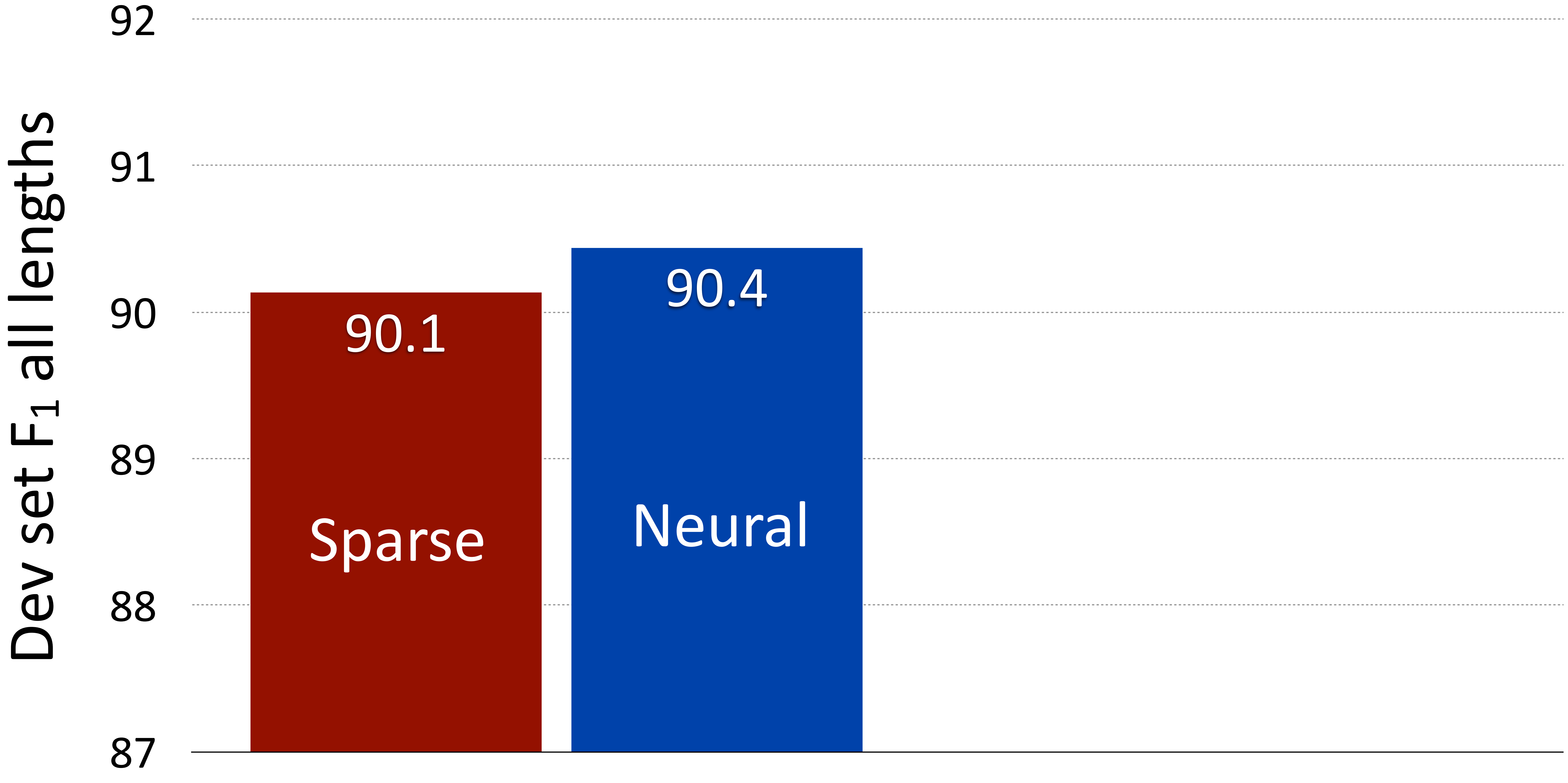
Optimization: Adadelta (Zeiler, 2012) worked slightly better than Adagrad (Duchi et al., 2011)

Results

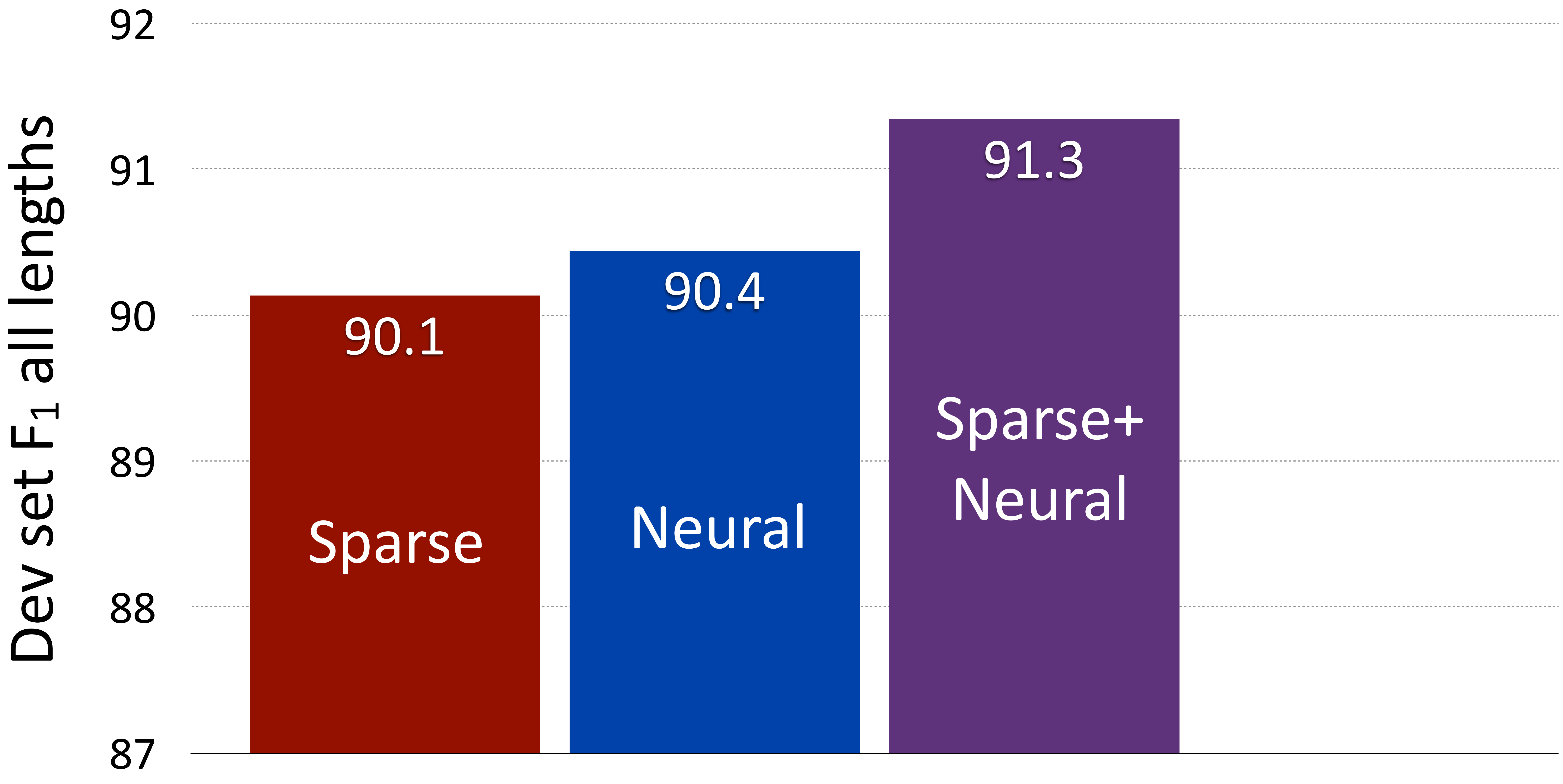
Results: English Treebank (Dev)



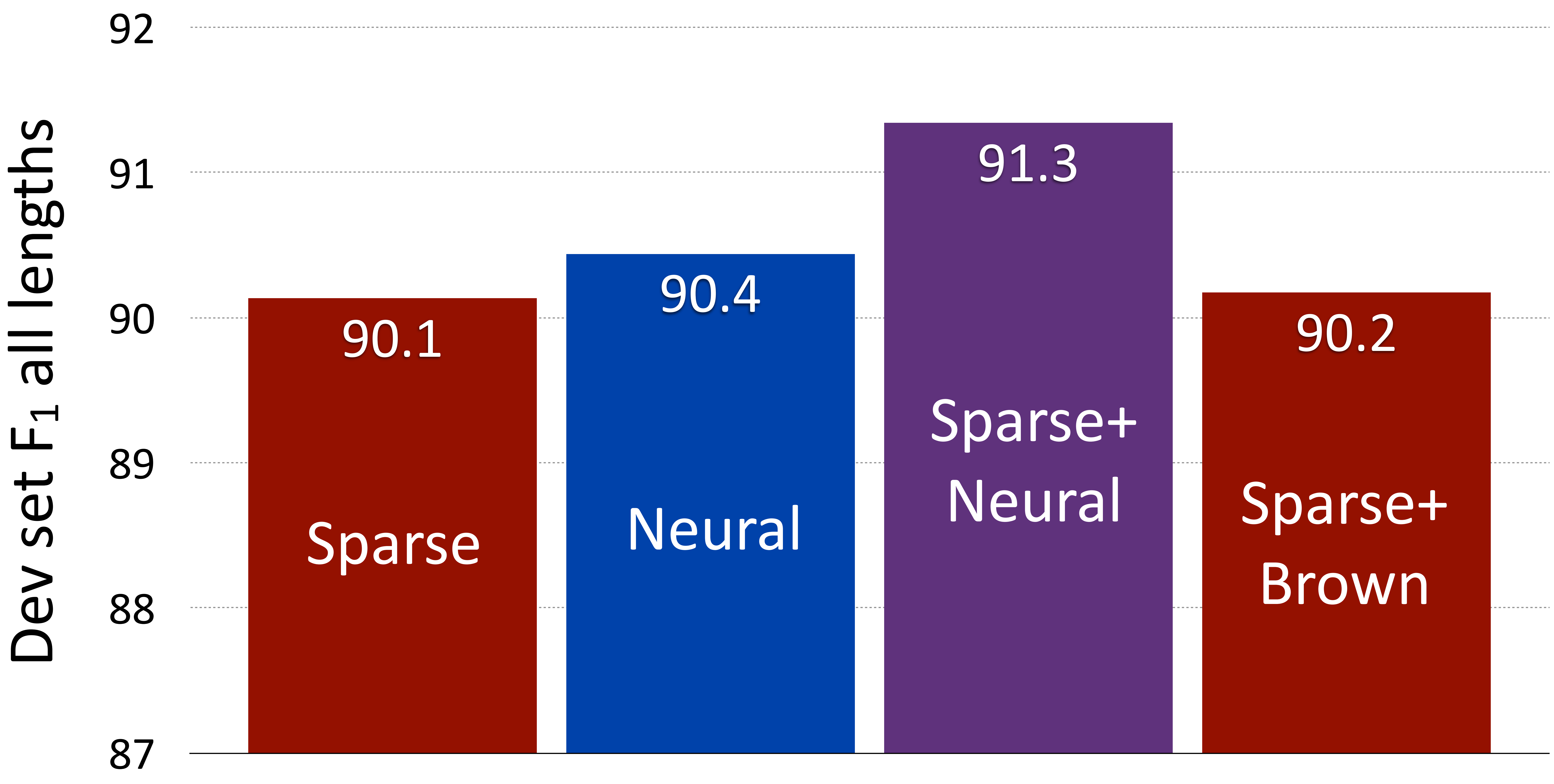
Results: English Treebank (Dev)



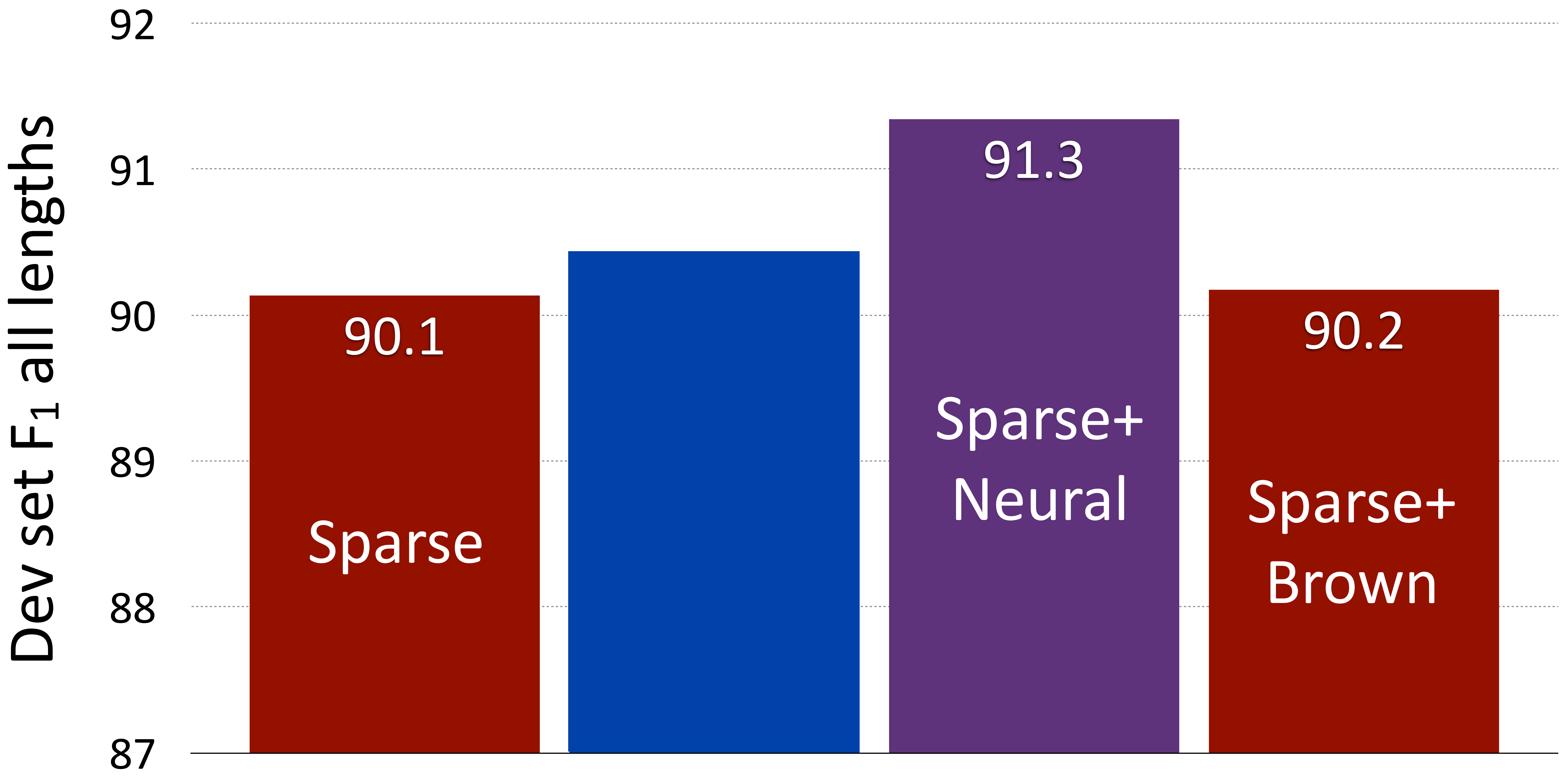
Results: English Treebank (Dev)



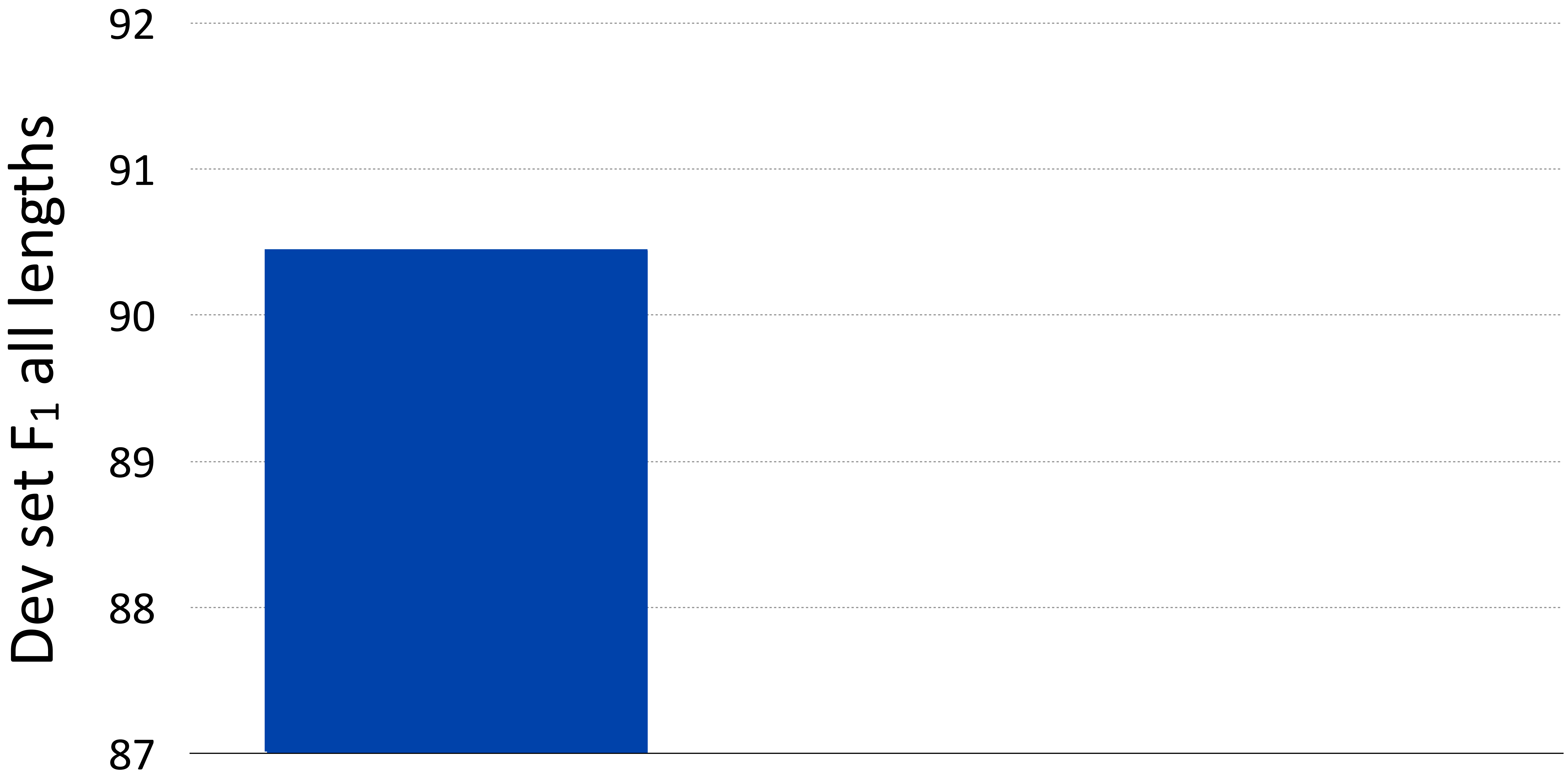
Results: English Treebank (Dev)



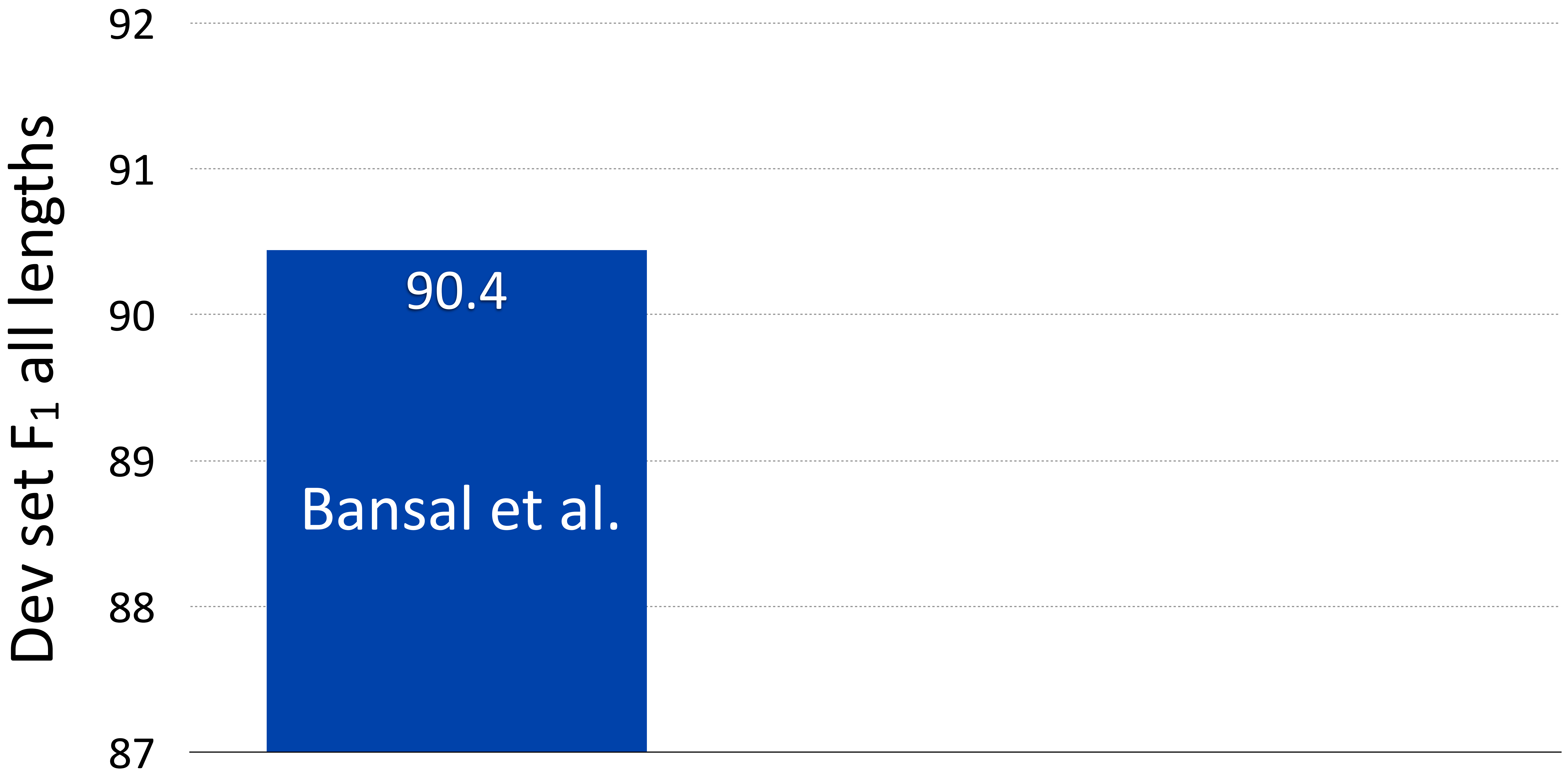
Results: English Treebank (Dev)



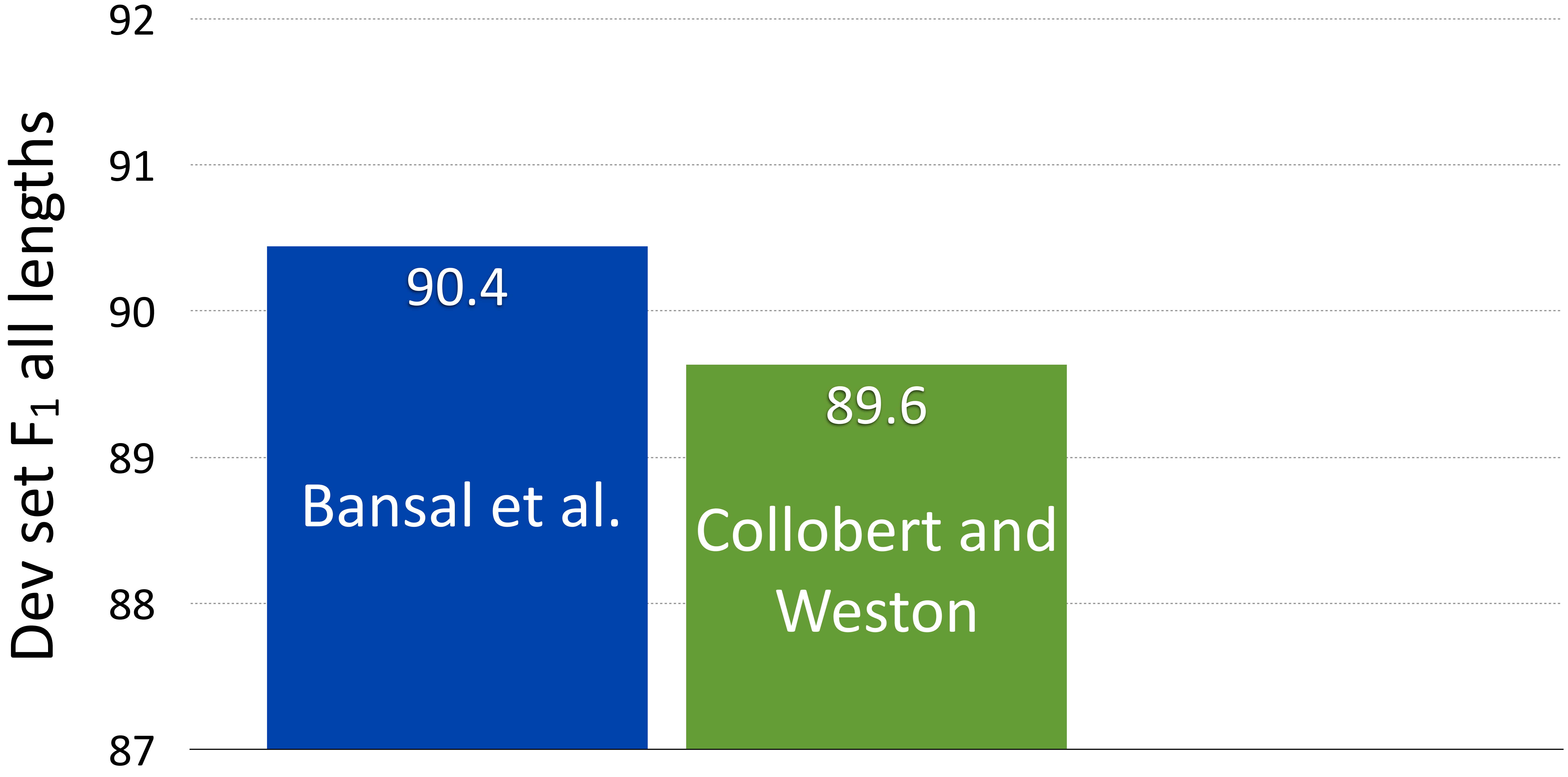
Word Vectors



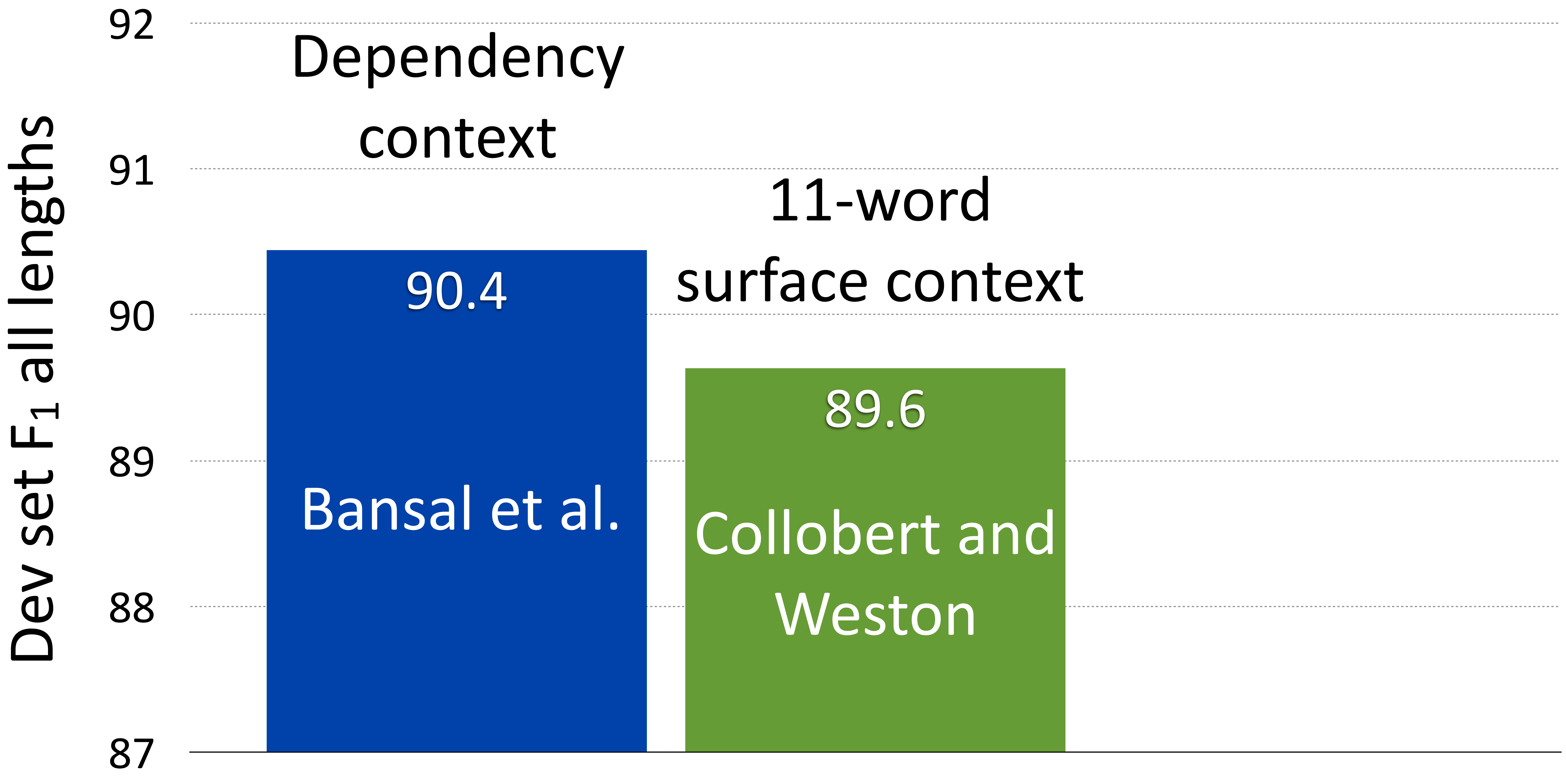
Word Vectors



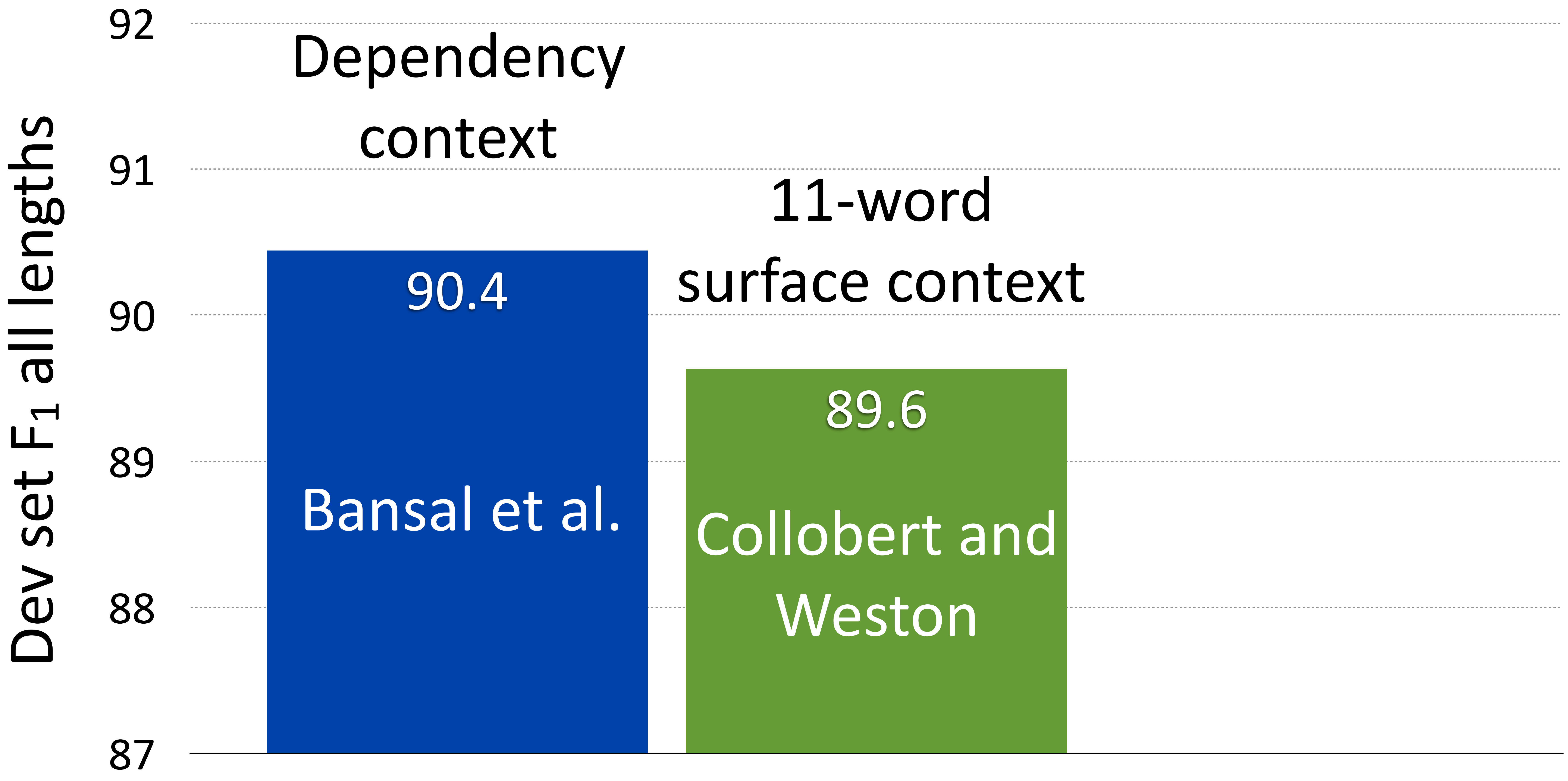
Word Vectors



Word Vectors

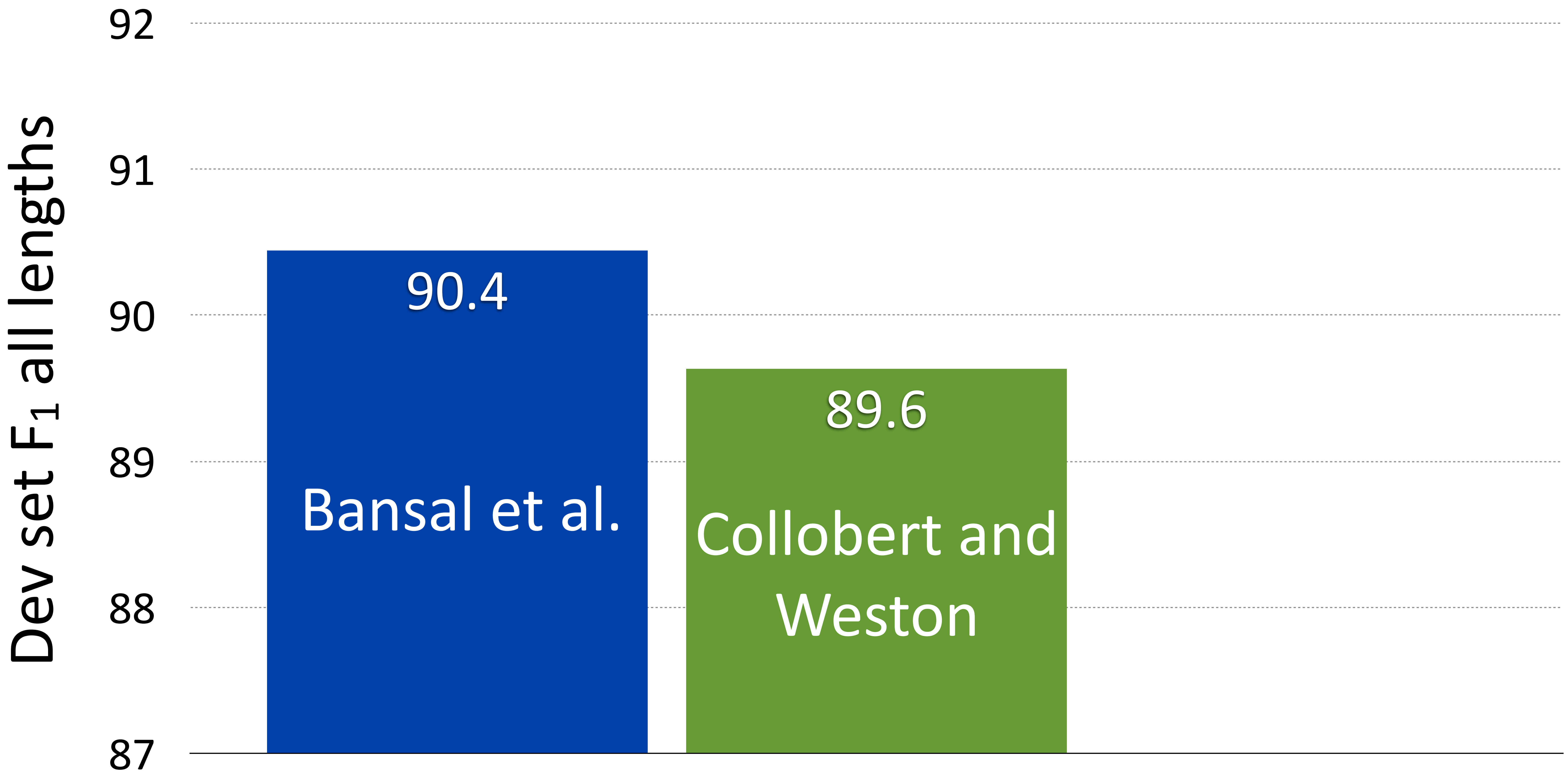


Word Vectors



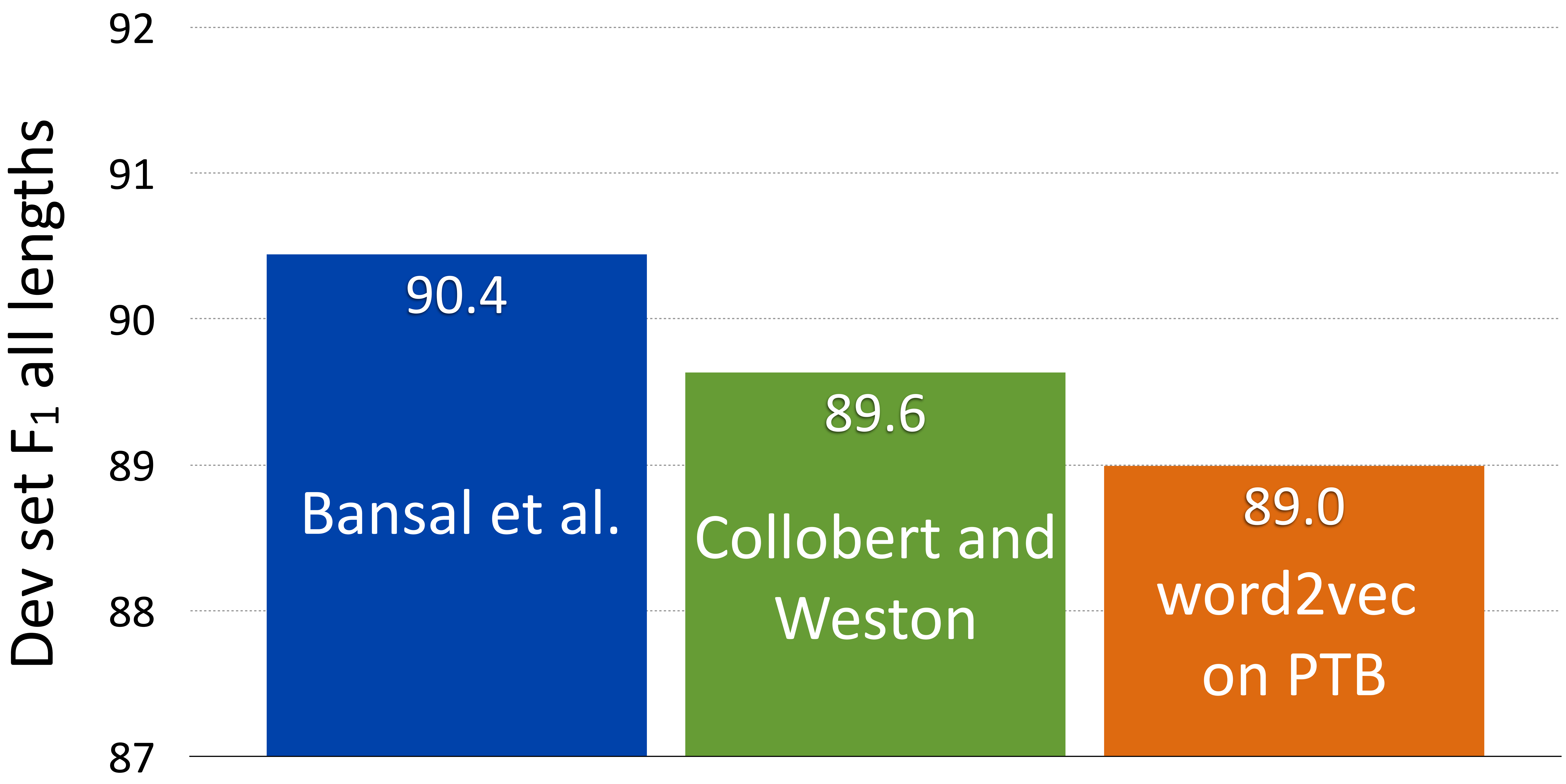
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Word Vectors



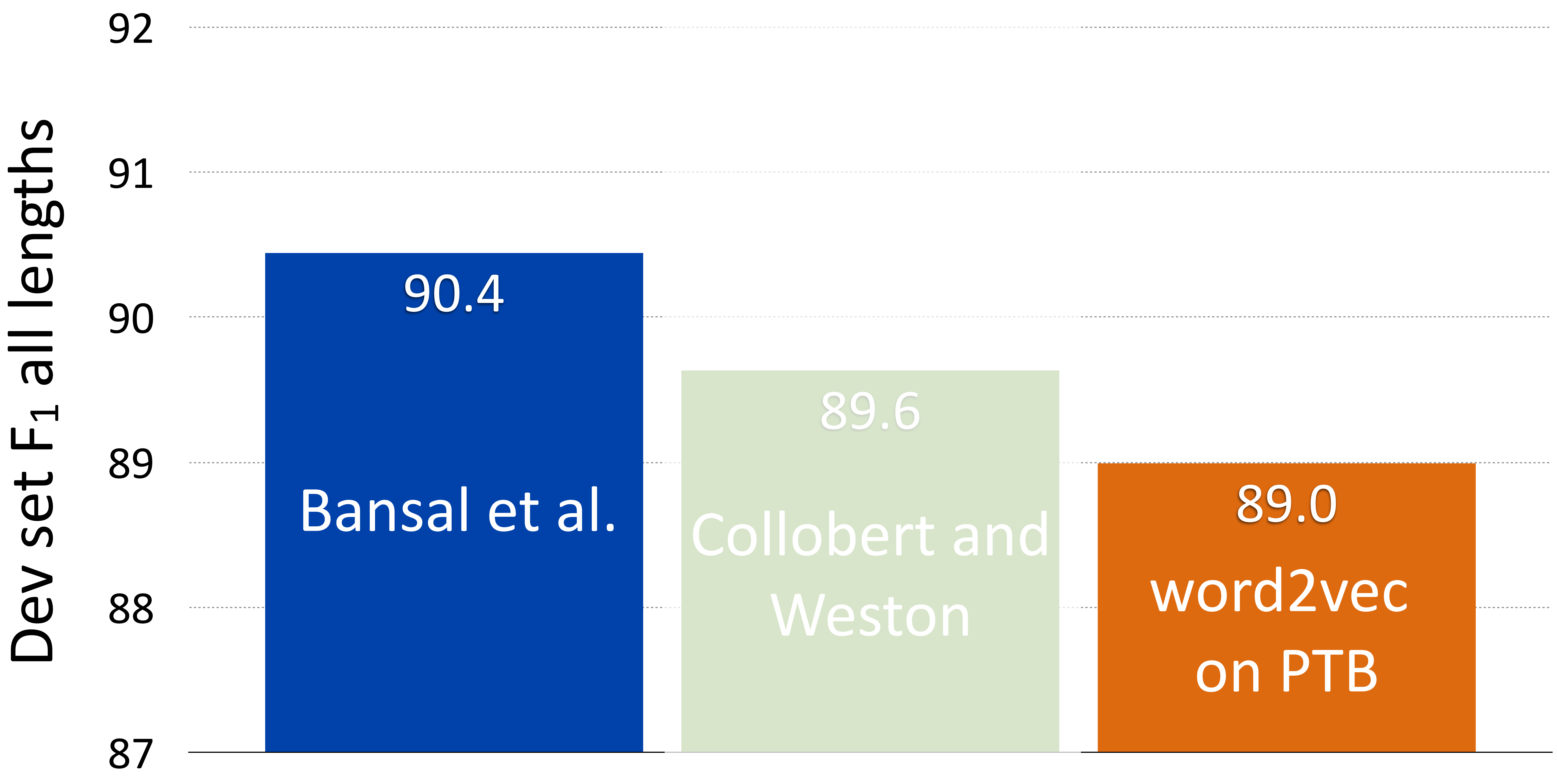
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Word Vectors



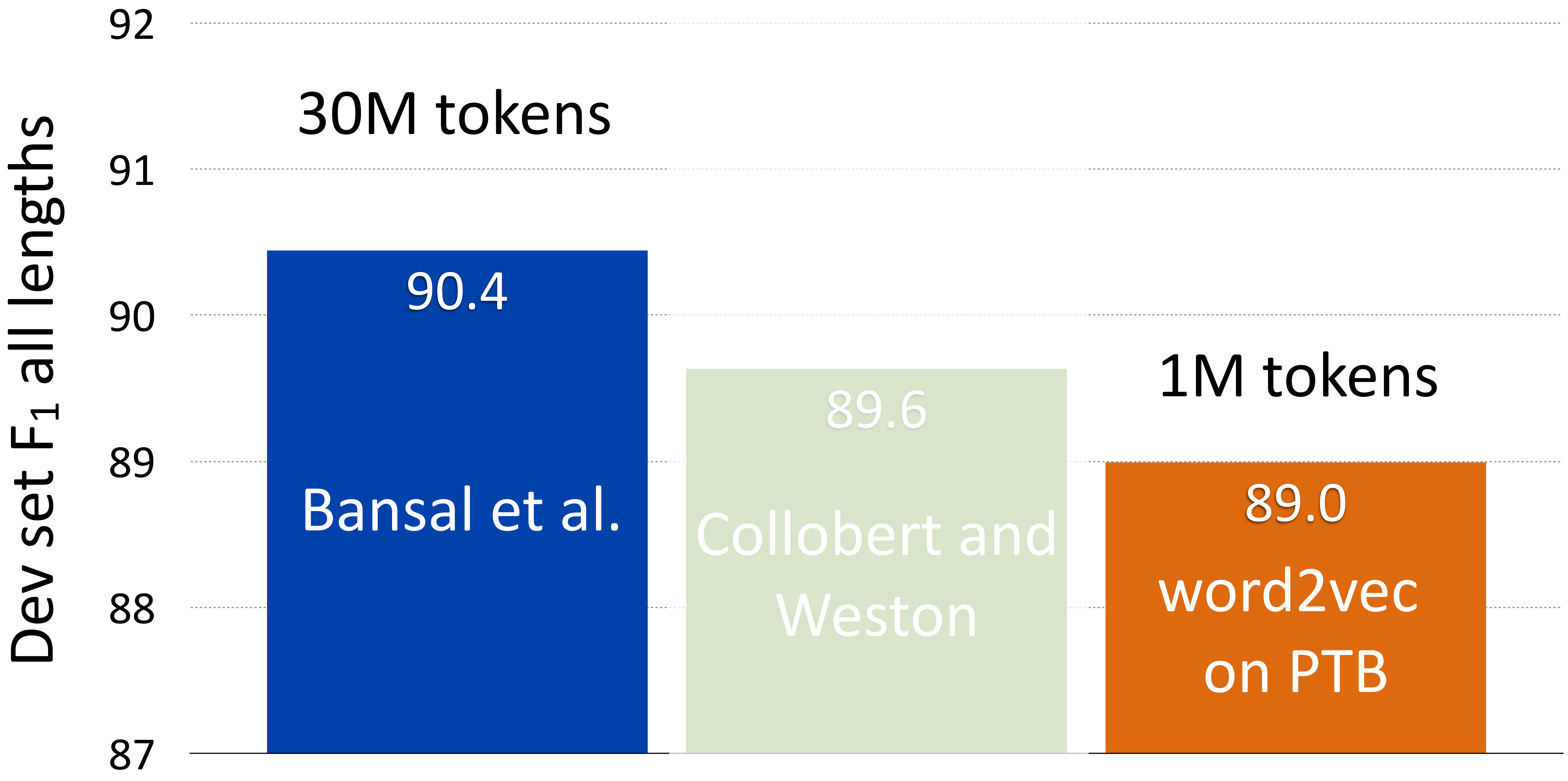
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Word Vectors



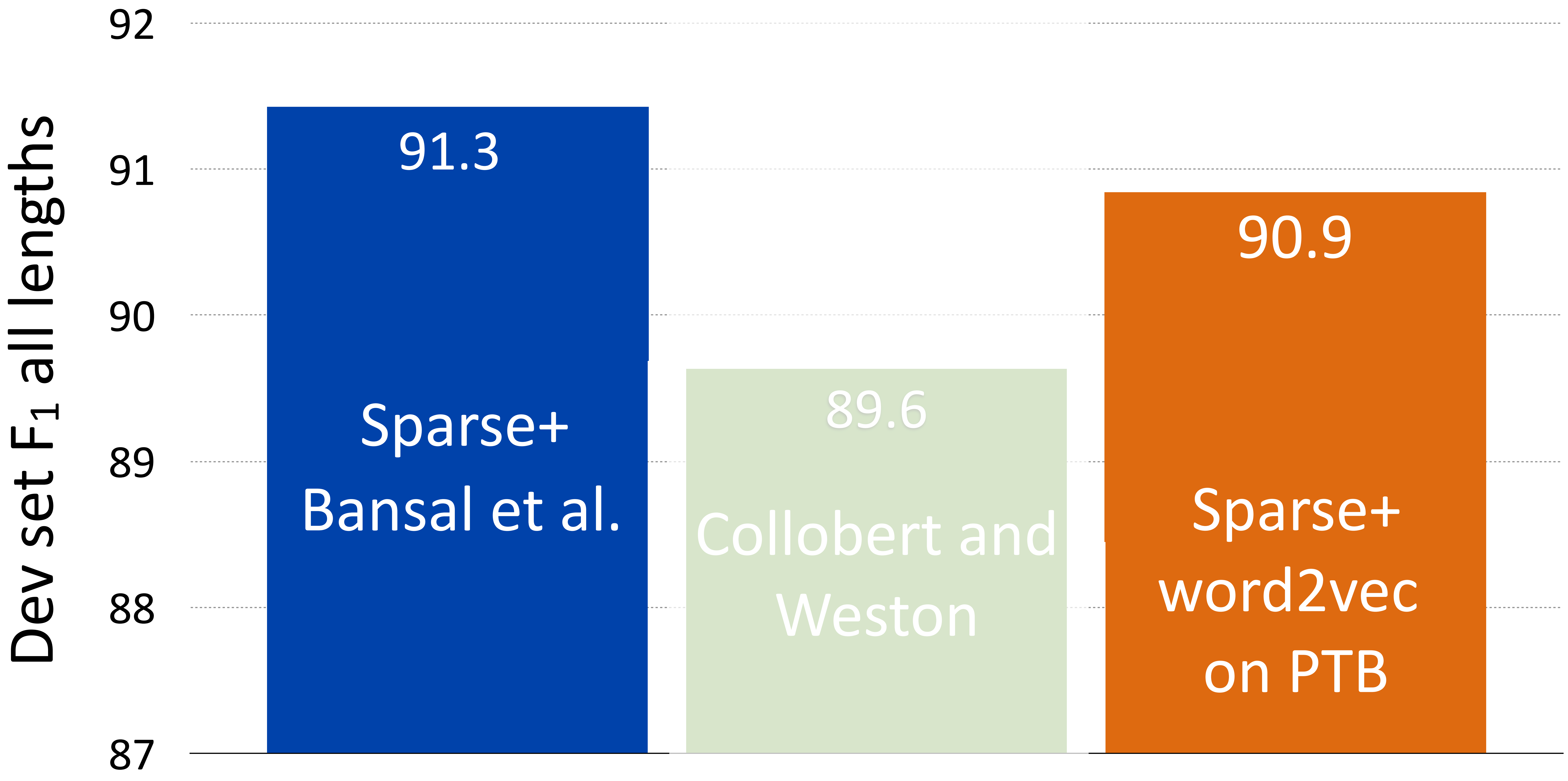
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Word Vectors



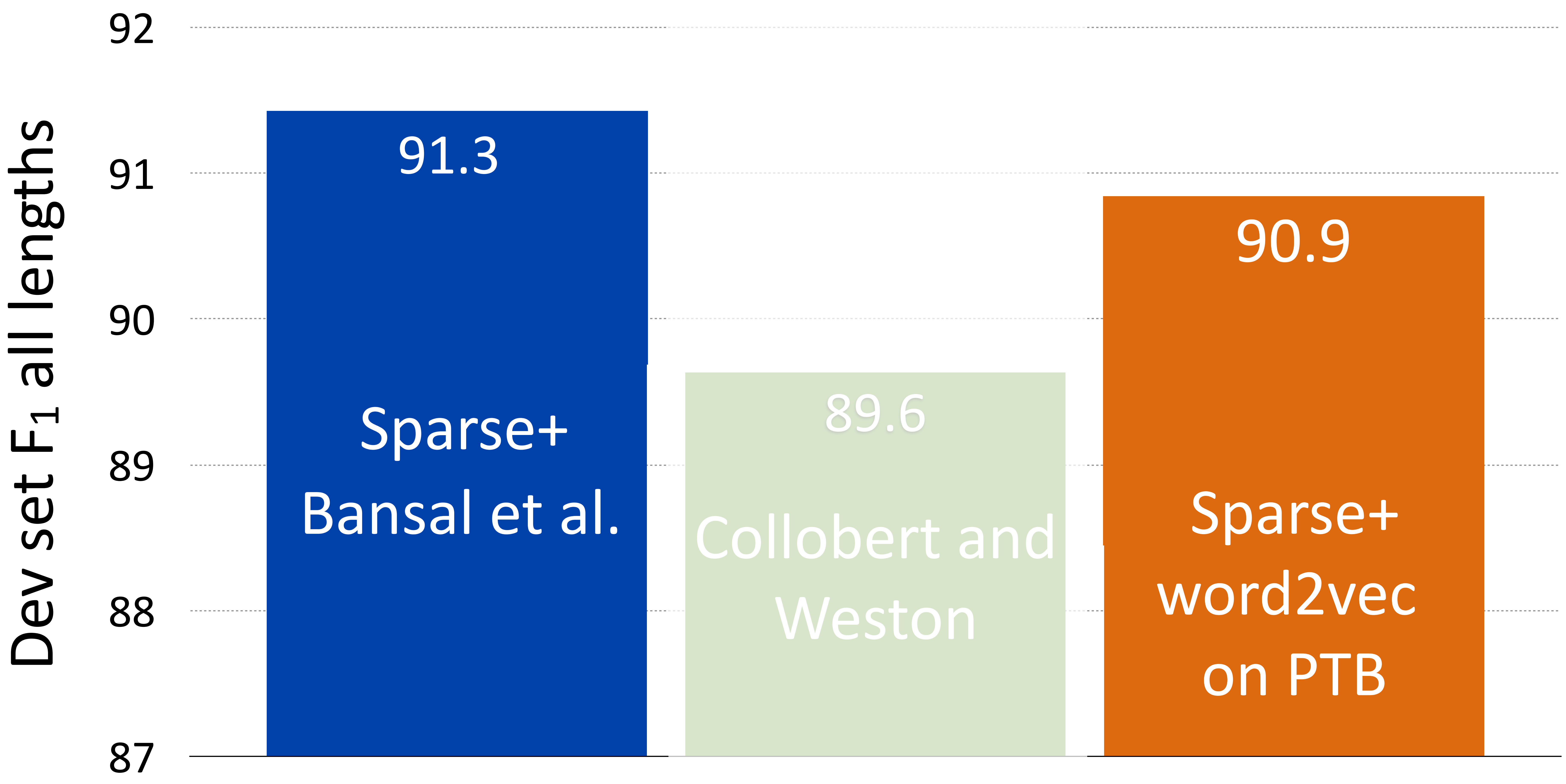
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Word Vectors



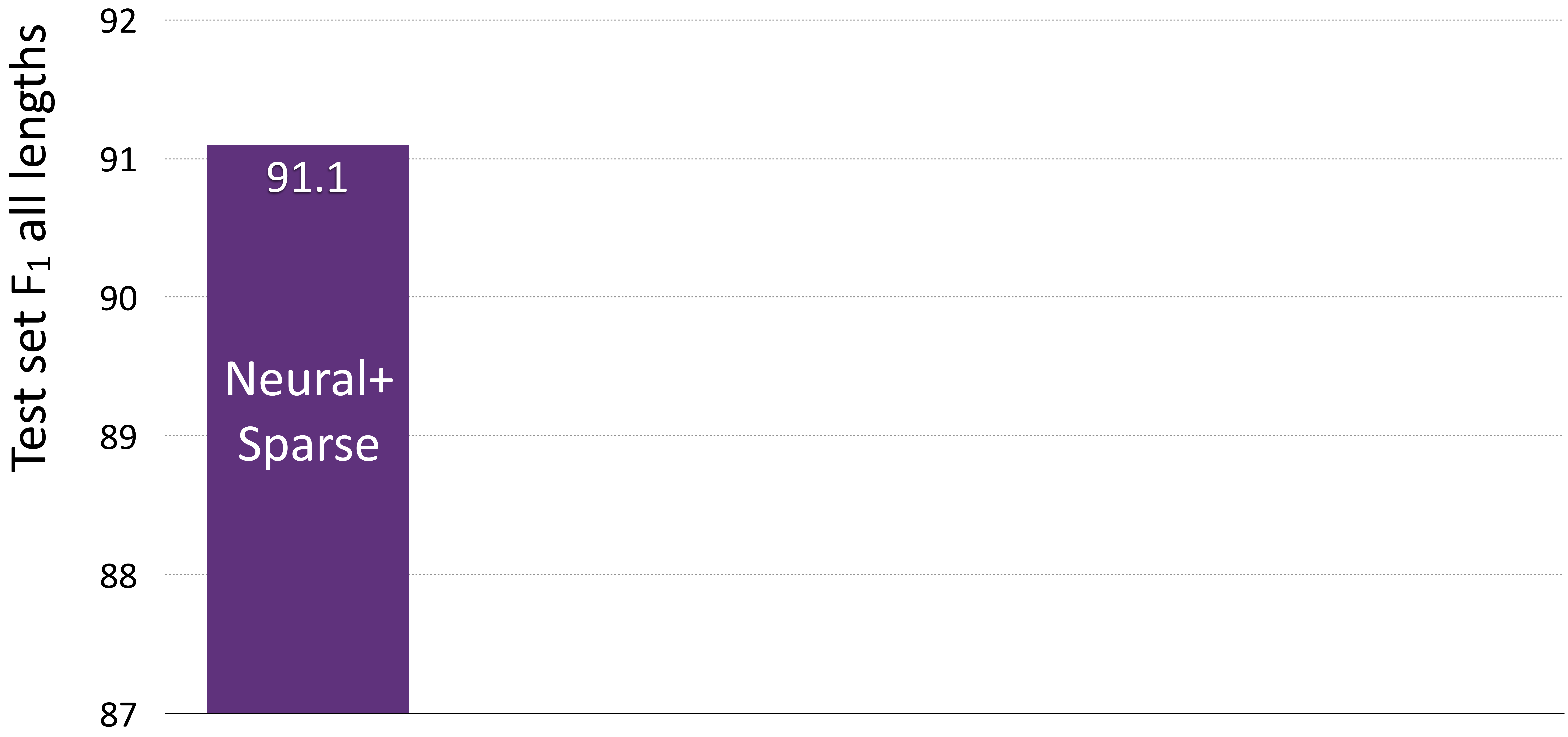
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Word Vectors

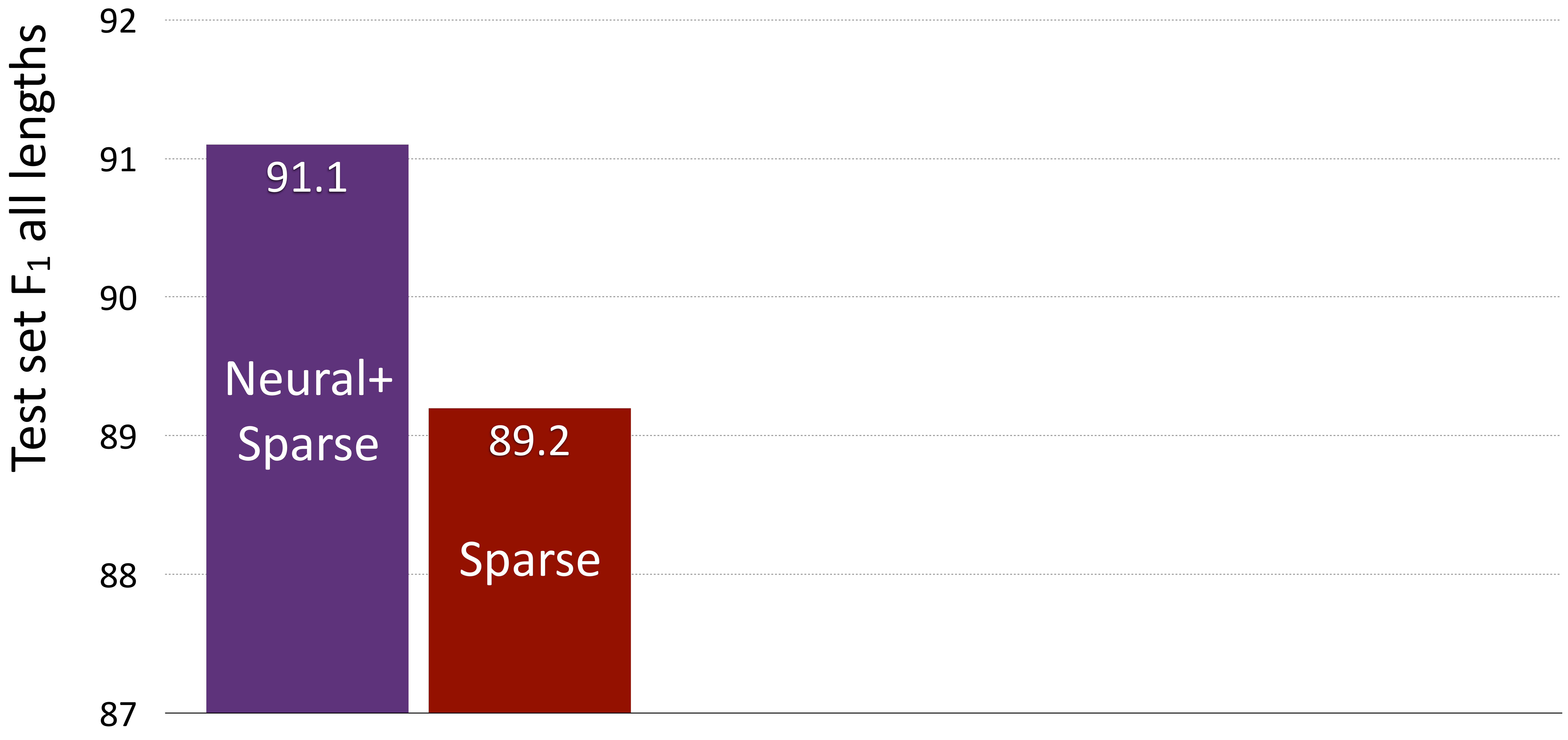


- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
- Don't need huge unlabeled corpora for these methods to be effective

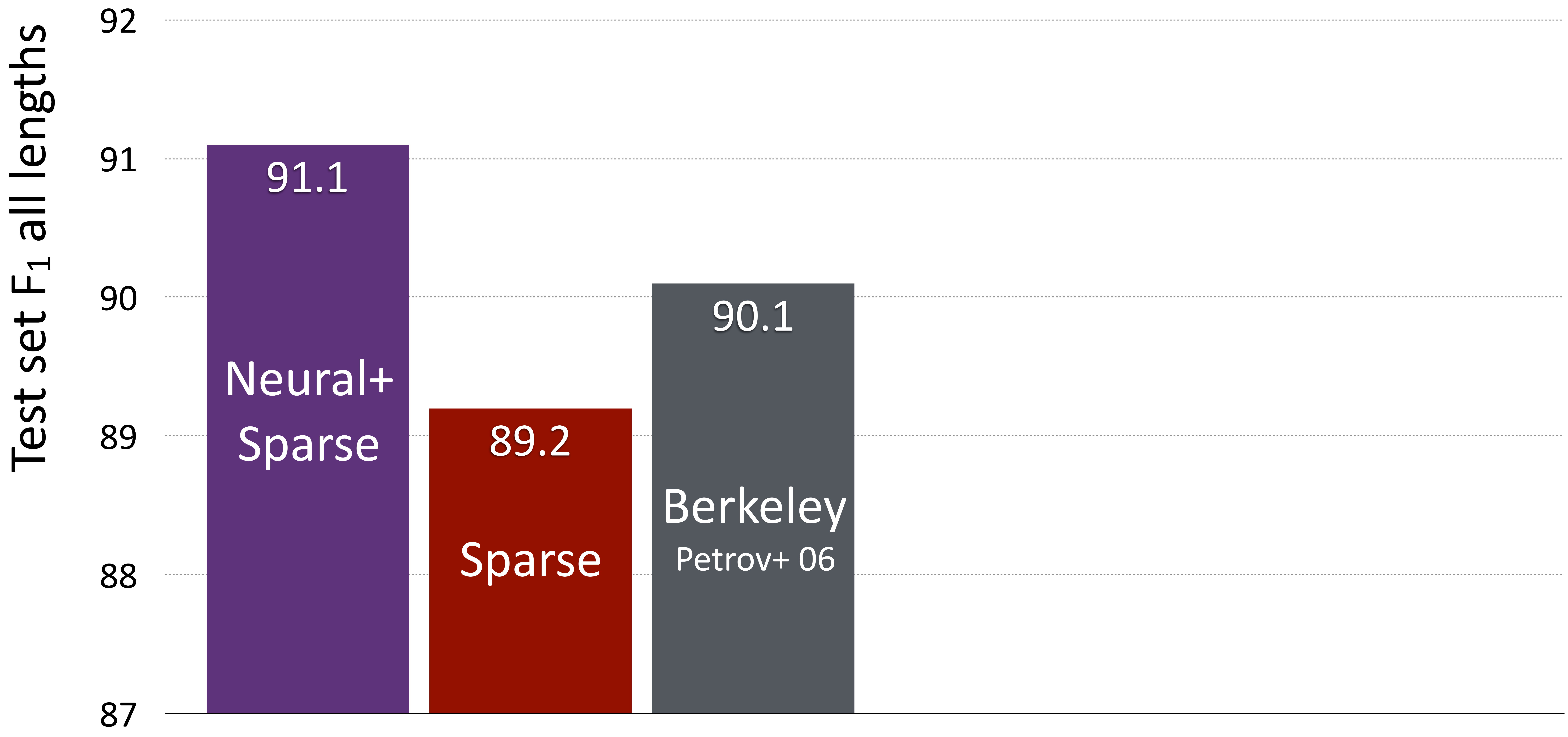
Results: English Treebank (Test)



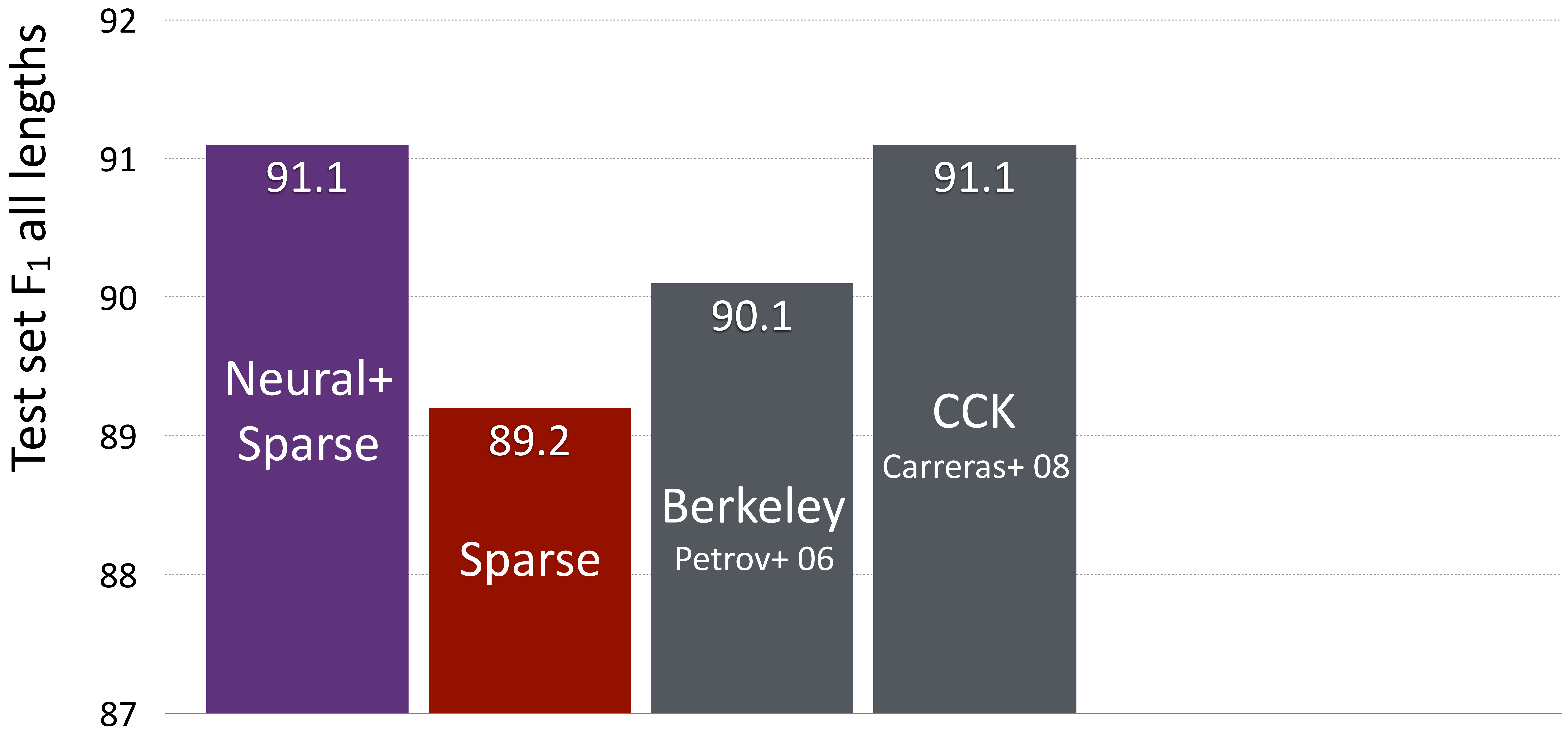
Results: English Treebank (Test)



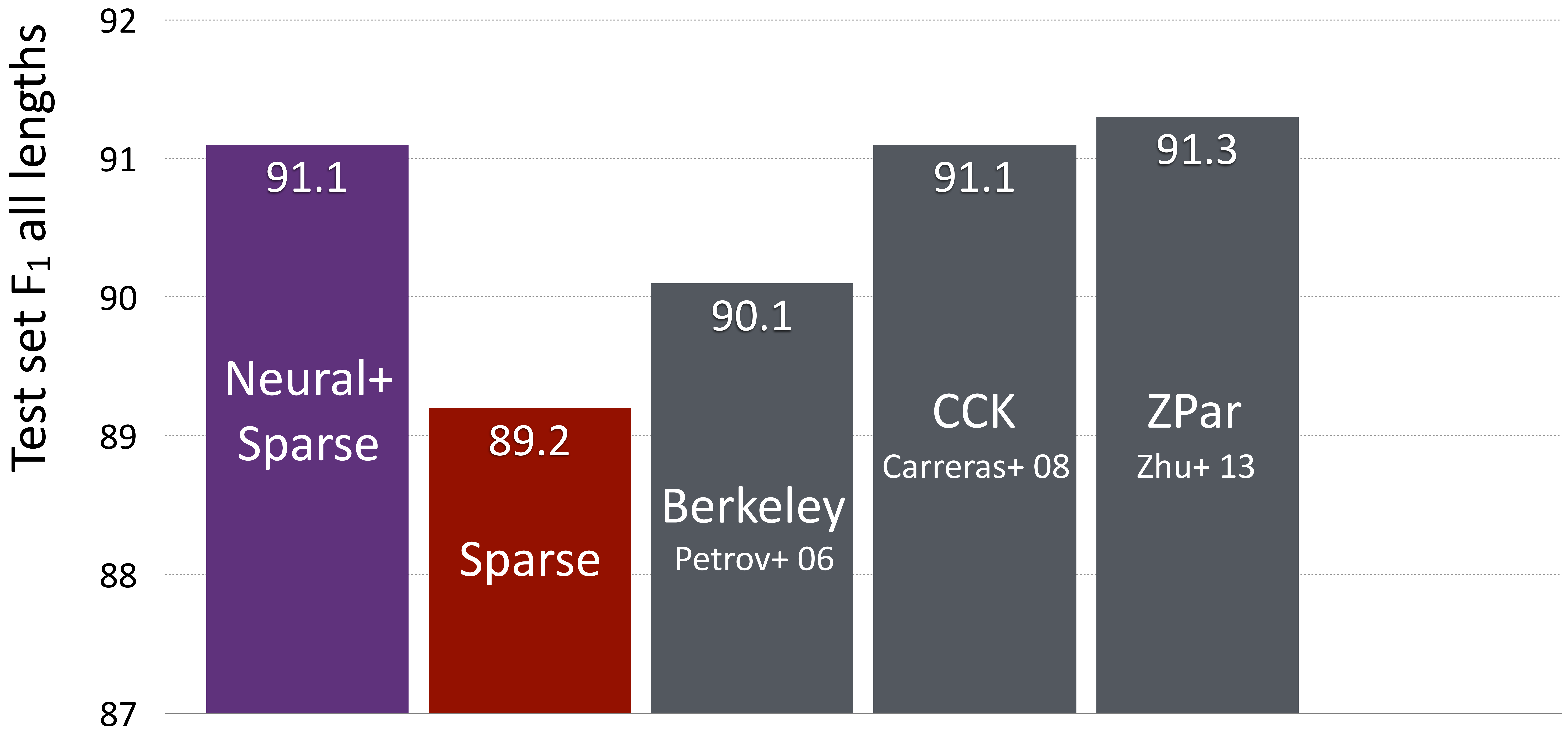
Results: English Treebank (Test)



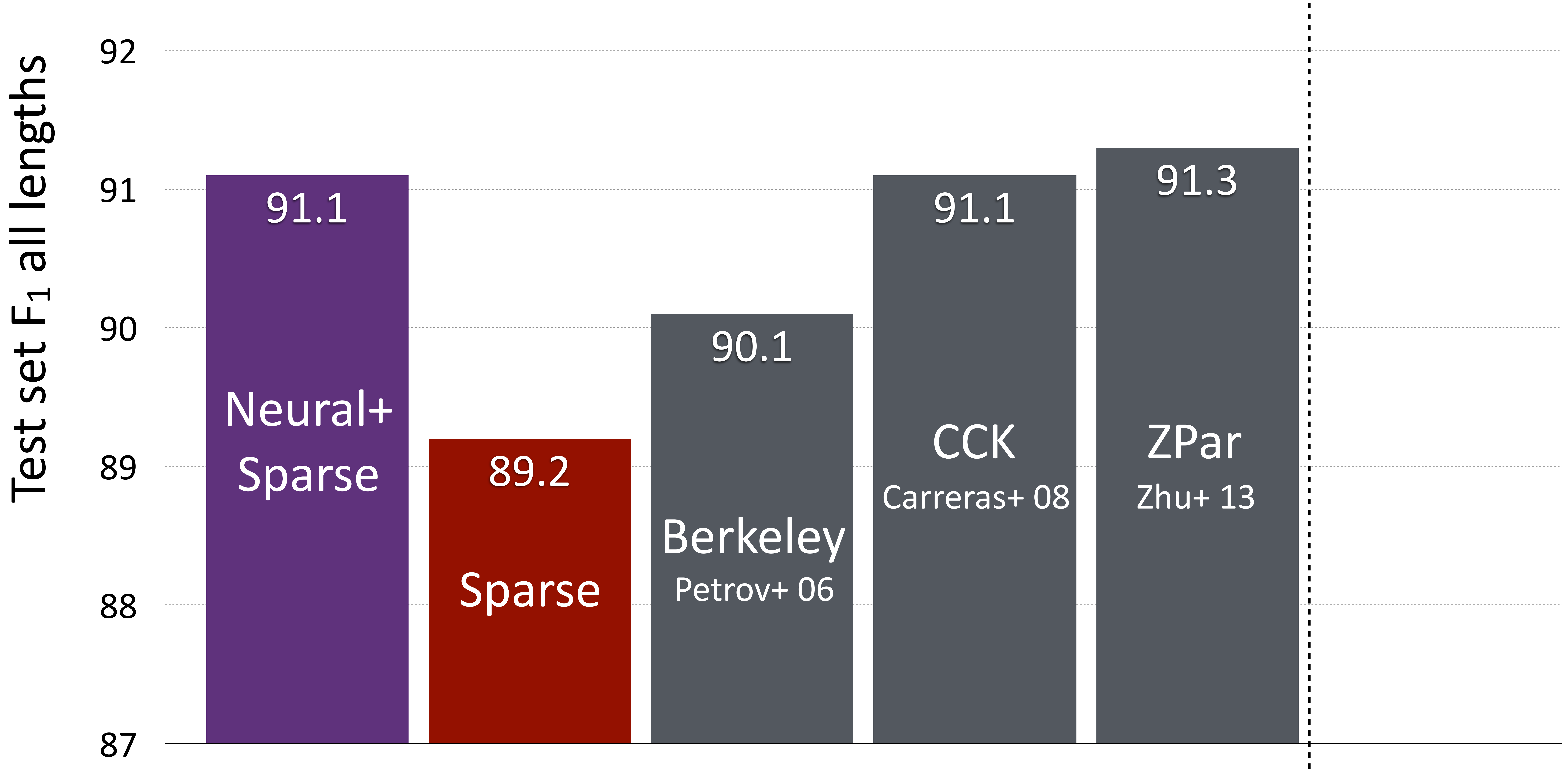
Results: English Treebank (Test)



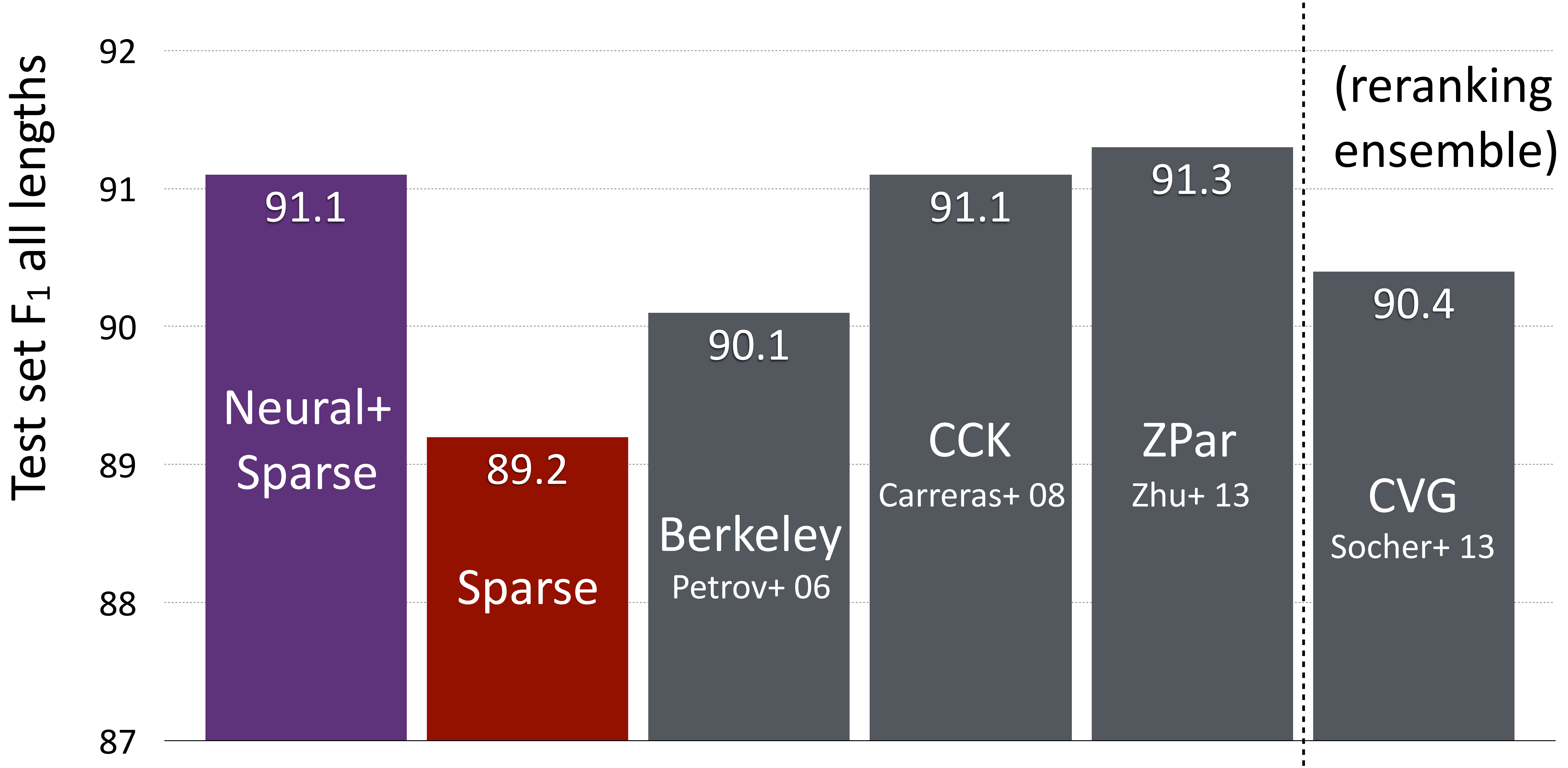
Results: English Treebank (Test)



Results: English Treebank (Test)



Results: English Treebank (Test)



Related Work

Related Work

- Transition-based neural parsers: Henderson (2003), Chen and Manning (2014)

Related Work

- Transition-based neural parsers: Henderson (2003), Chen and Manning (2014)
- Local decisions only: Belinkov et al. (2014)

Related Work

- Transition-based neural parsers: Henderson (2003), Chen and Manning (2014)
- Local decisions only: Belinkov et al. (2014)
- Sequence-to-sequence LSTM: Vinyals et al. (2014)

Results: Other Languages

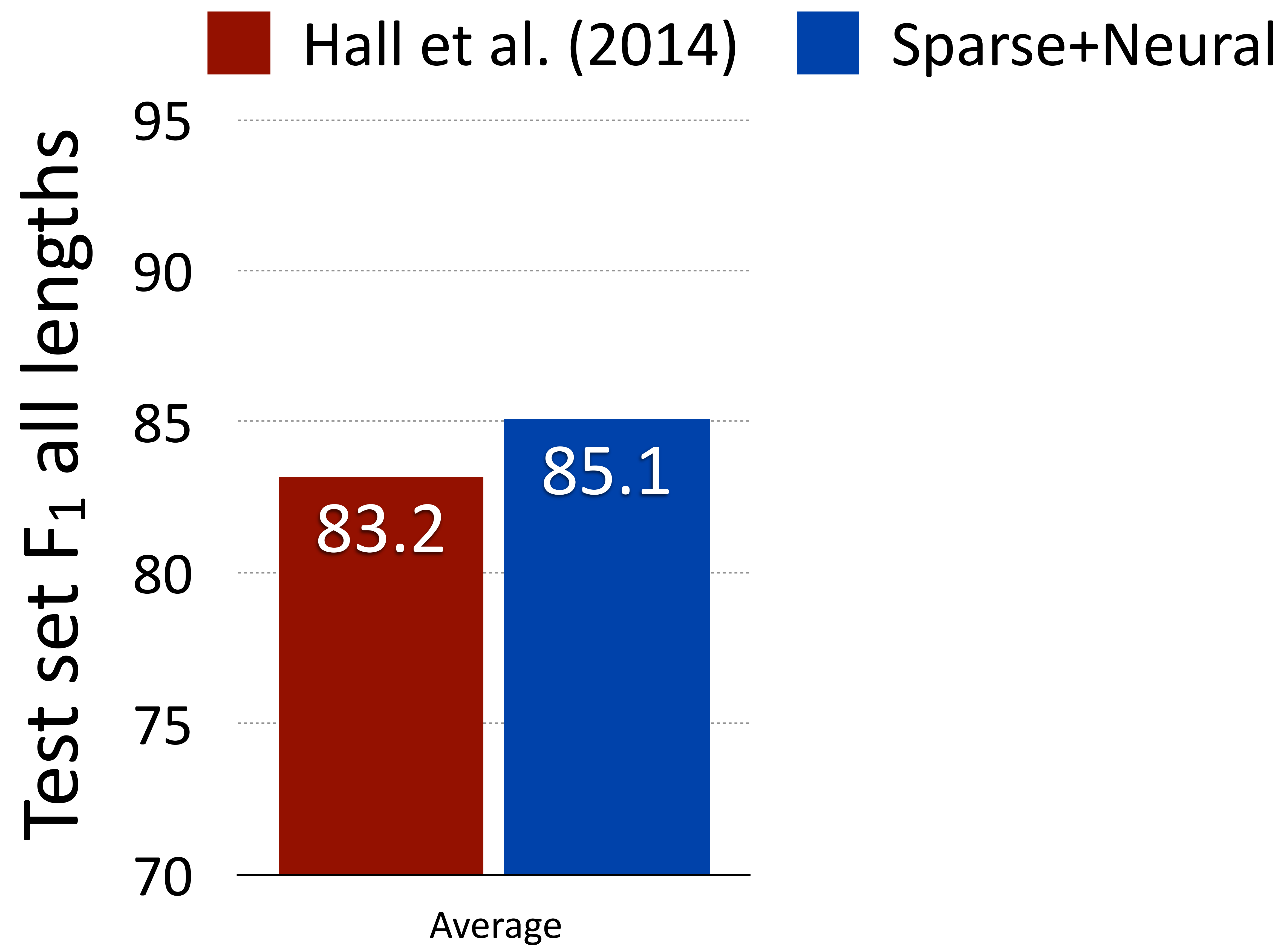
Results: Other Languages

- Nine morphologically-rich languages from the SPMRL shared task

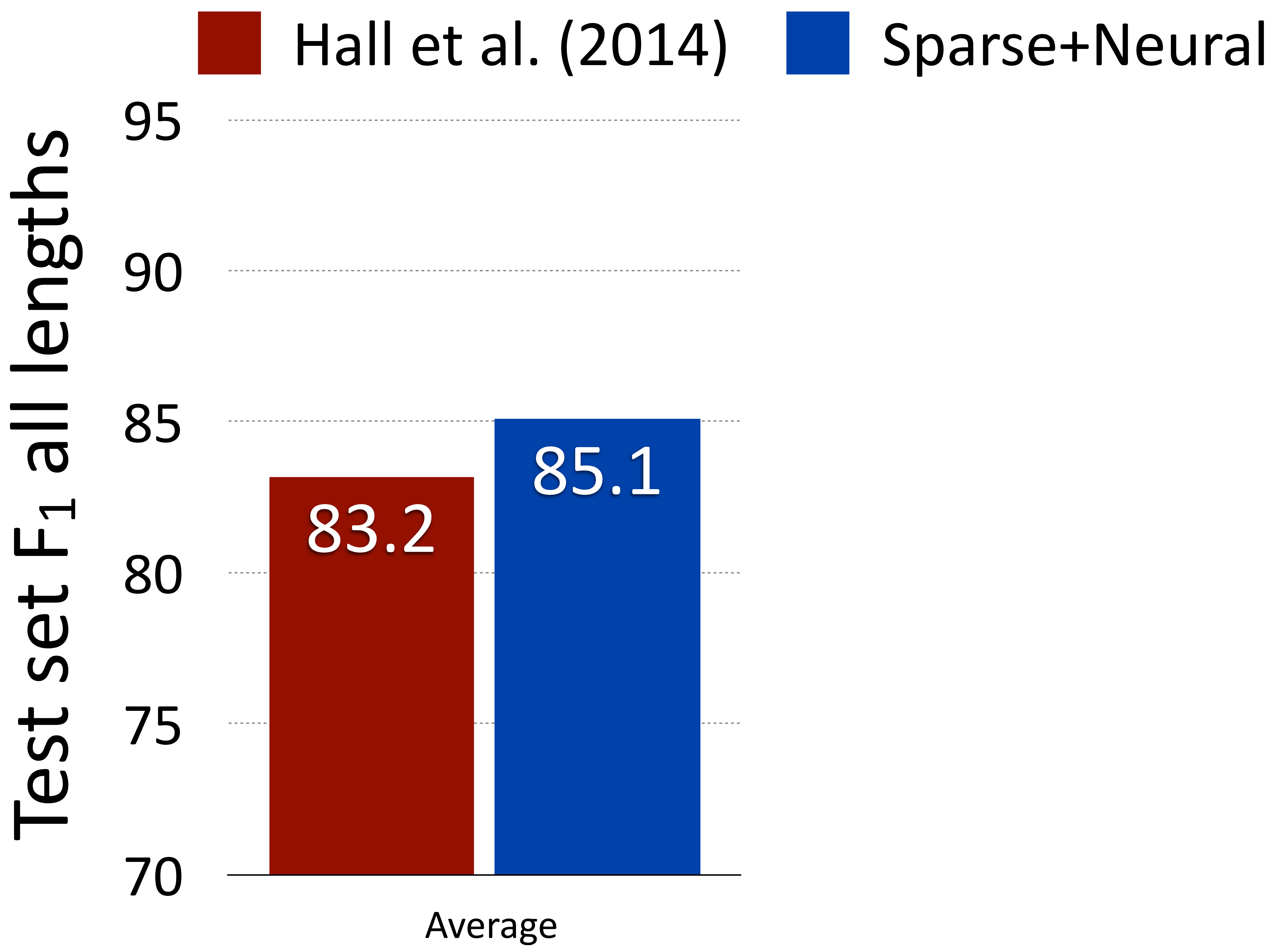
Results: Other Languages

- Nine morphologically-rich languages from the SPMRL shared task
- Word vectors trained on SPMRL monolingual data with word2vec (approximately 100M tokens per language)

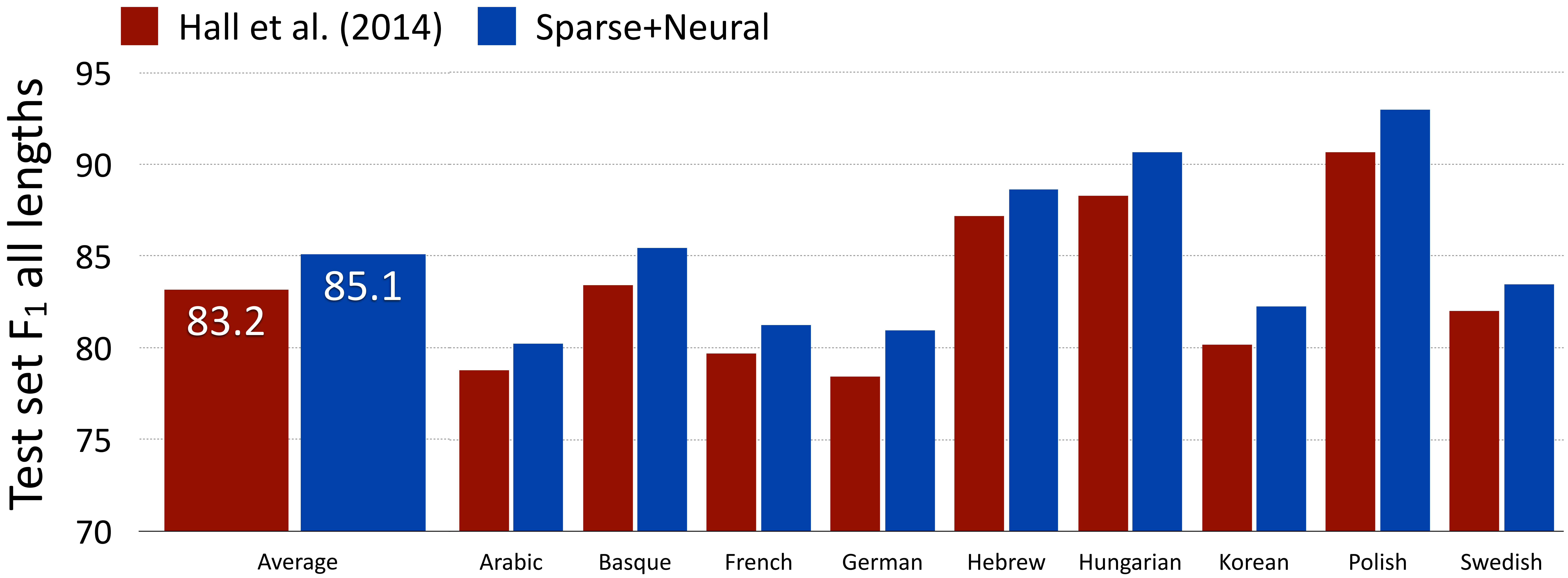
Results: Other Languages



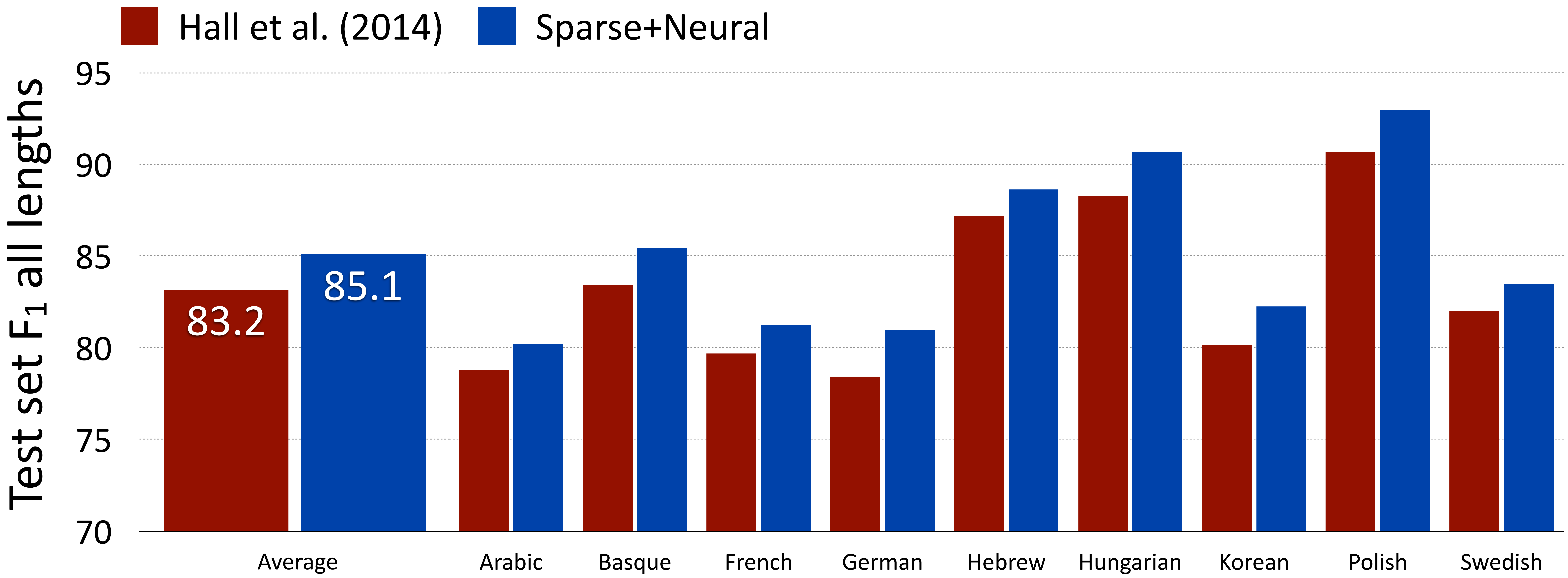
Results: Other Languages



Results: Other Languages



Results: Other Languages



- Works well even on smaller treebanks

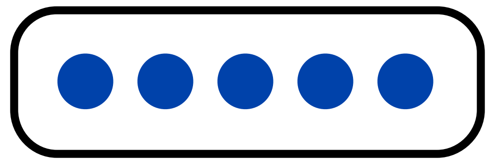
Conclusion

Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models

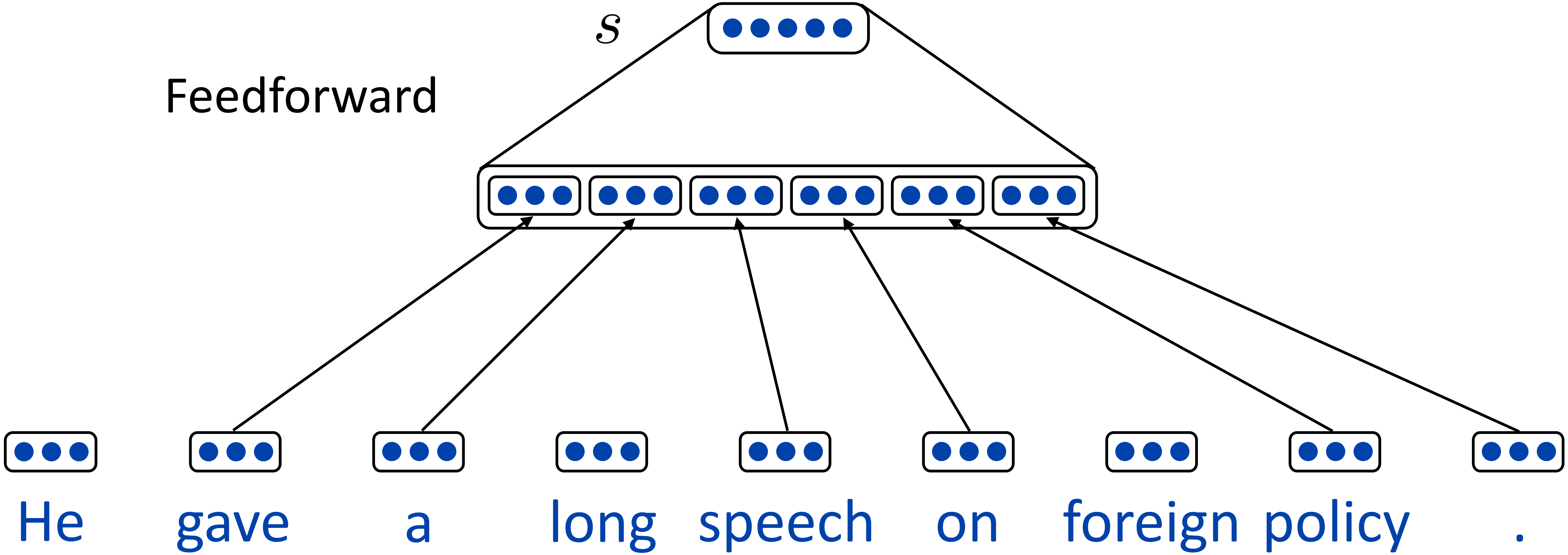
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models

s 

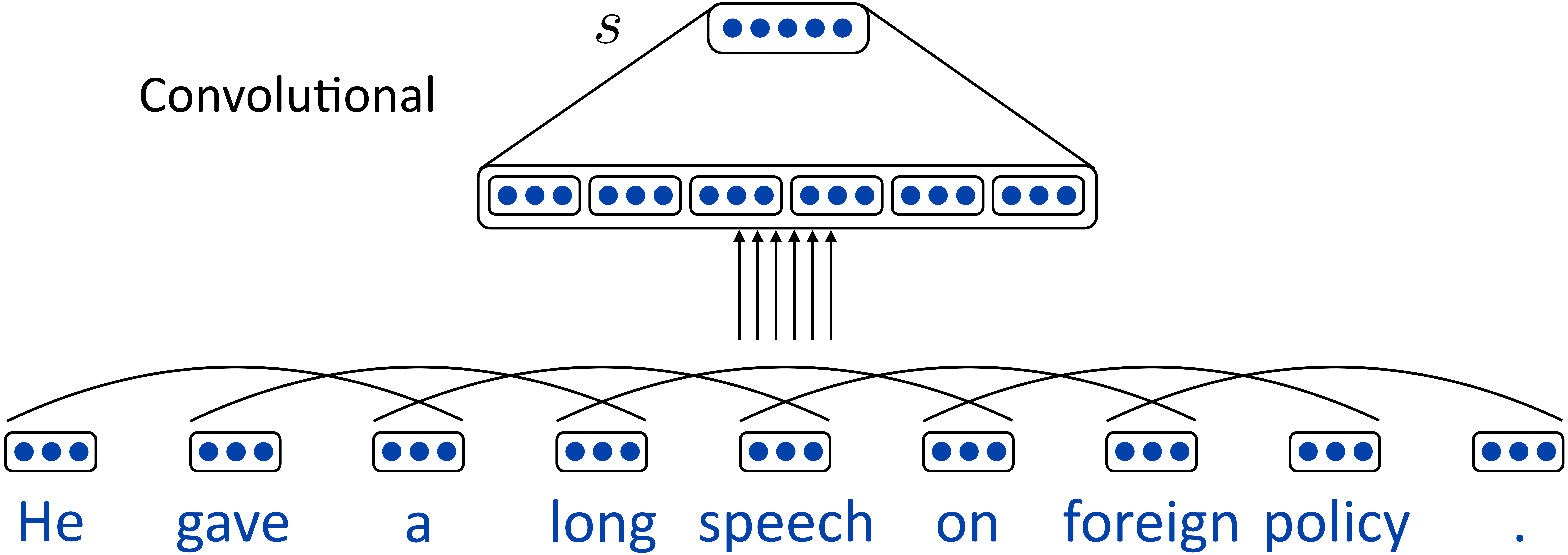
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models



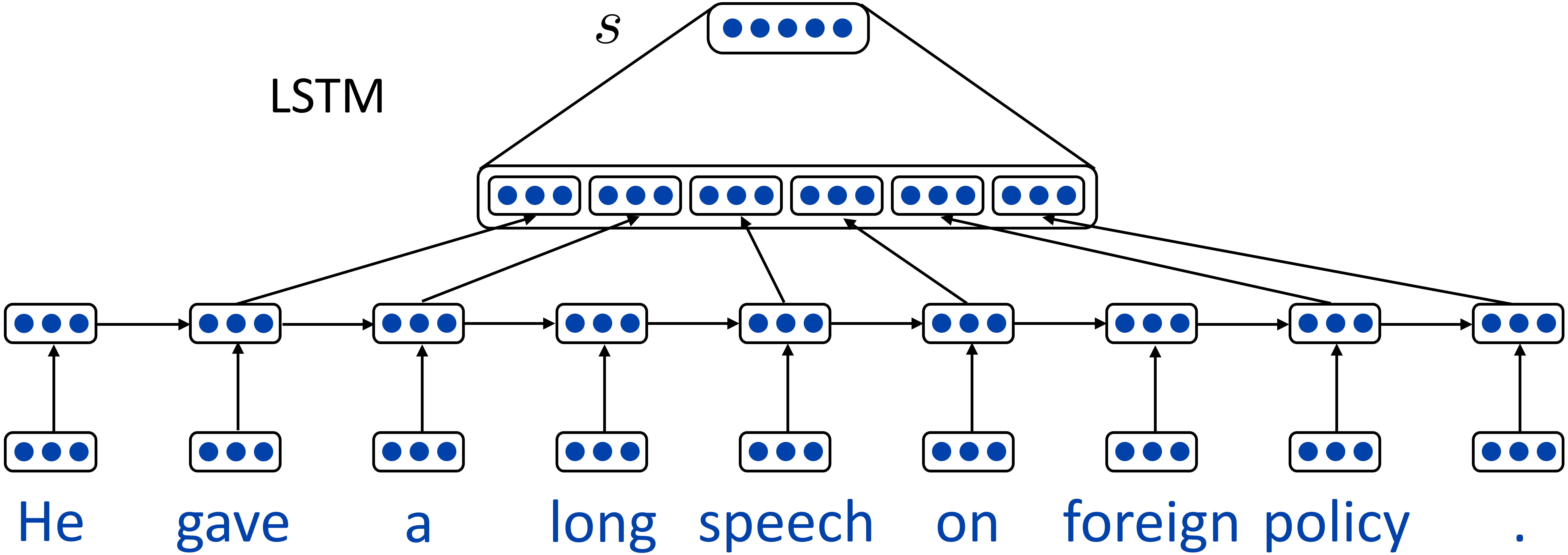
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models



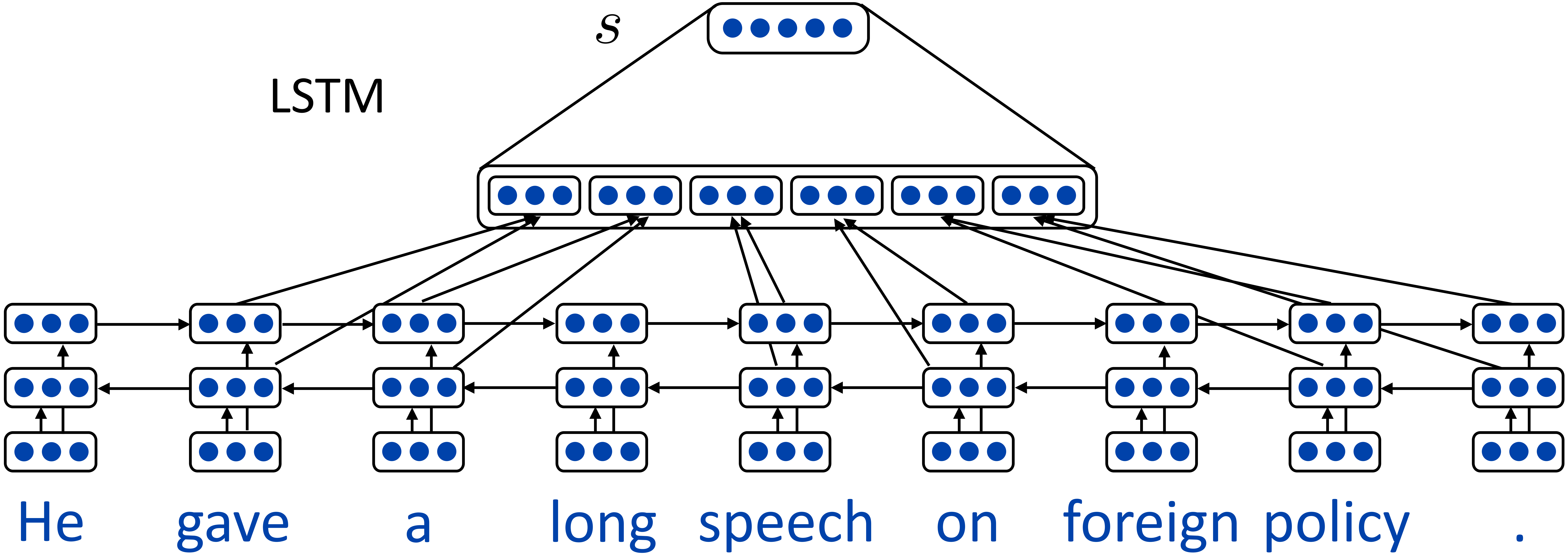
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models



Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models



Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models

Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models
- Inference and learning are unchanged from the purely discrete model

Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models
- Inference and learning are unchanged from the purely discrete model
- High performance on constituency parsing for a range of languages

Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models
- Inference and learning are unchanged from the purely discrete model
- High performance on constituency parsing for a range of languages

`nlp.cs.berkeley.edu/projects/neuralcrf.shtml`

Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models
- Inference and learning are unchanged from the purely discrete model
- High performance on constituency parsing for a range of languages

`nlp.cs.berkeley.edu/projects/neuralcrf.shtml`

Thank you!