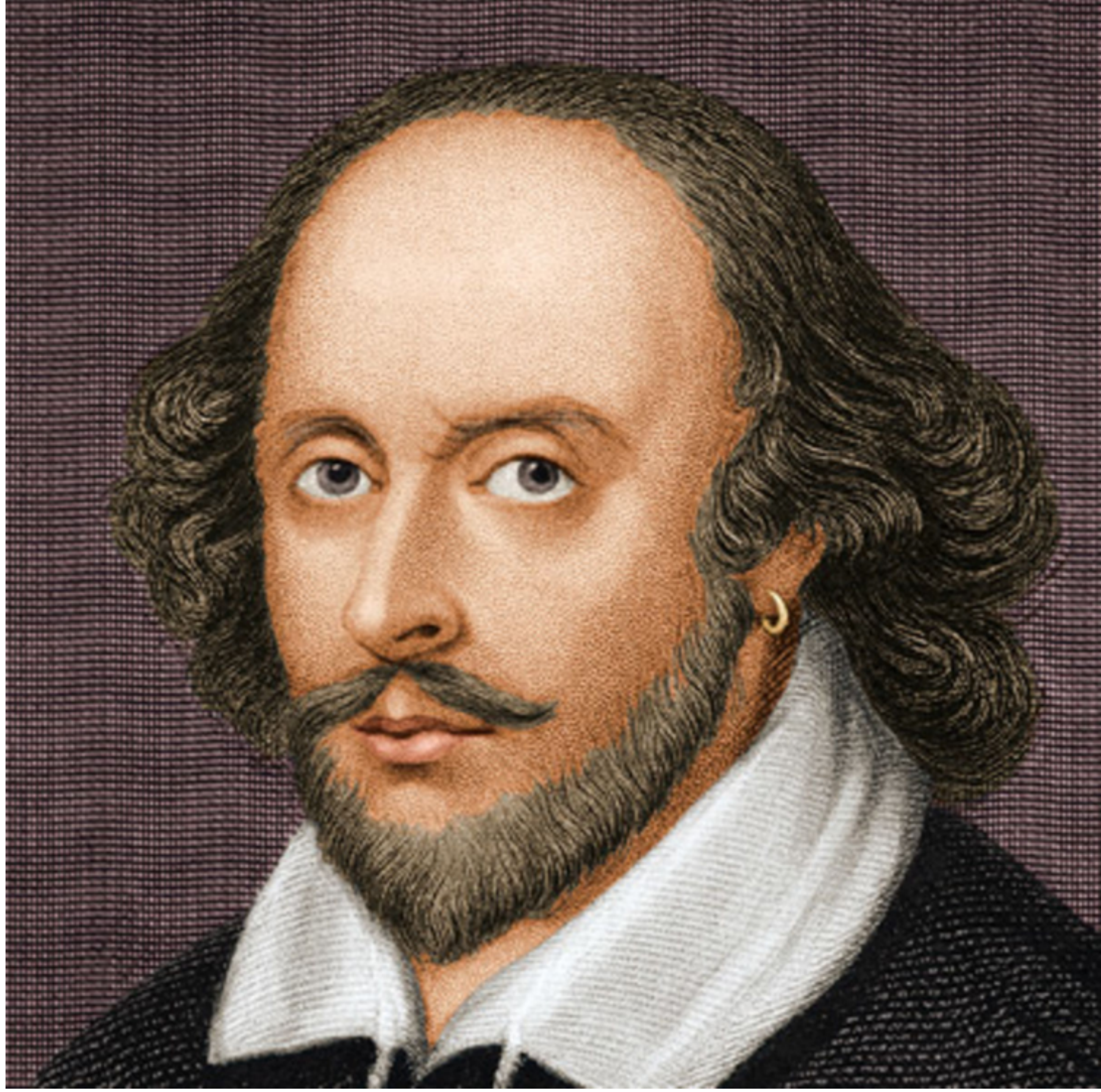
Part-of-Speech Tagging for Historical English

Yi Yang and Jacob Eisenstein

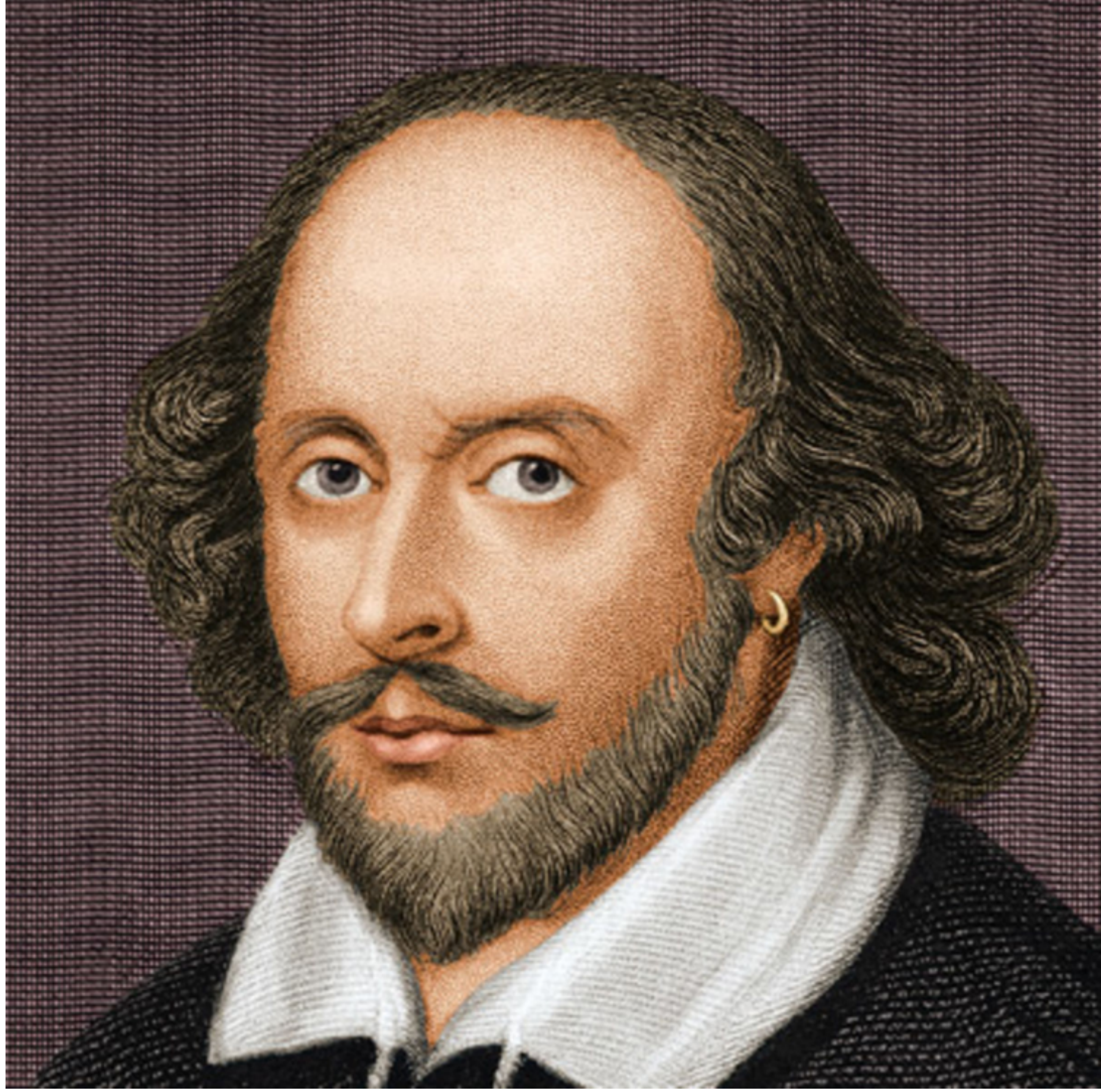
Georgia Tech

- Digital humanities research
  - How does the portrayal of men and women differ in Shakespeare's plays?
  - What's the language use patterns in North American slave narratives?



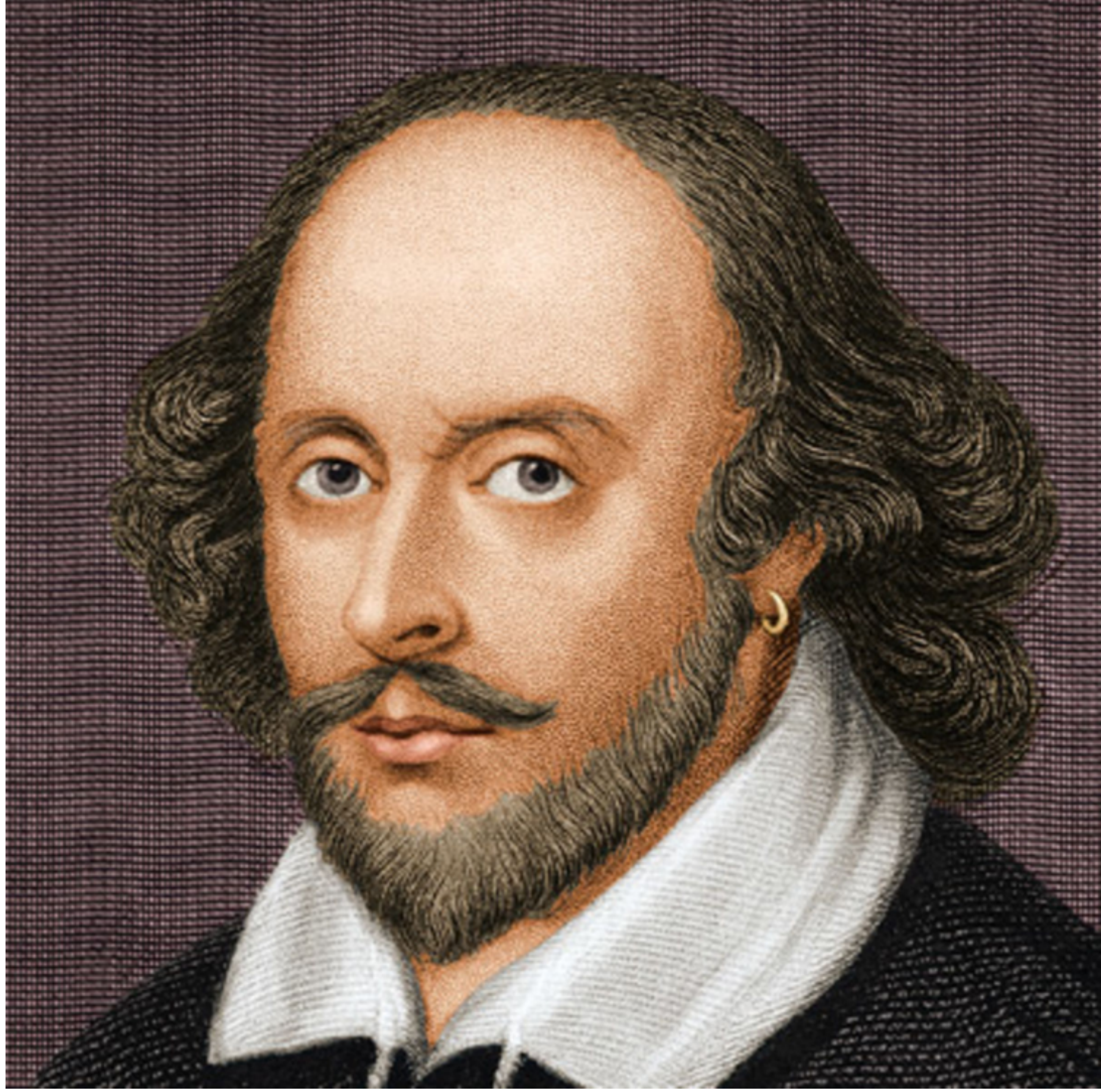
[Muralidharan and Hearst, 2011&2012]

- Digital humanities research
  - How does the portrayal of men and women differ in Shakespeare's plays?
  - What's the language use patterns in North American slave narratives?
- NLP can help!



[Muralidharan and Hearst, 2011&2012]

- Digital humanities research
  - How does the portrayal of men and women differ in Shakespeare's plays?
  - What's the language use patterns in North American slave narratives?
- NLP can help!
- Only if NLP works for historical texts …



[Muralidharan and Hearst, 2011&2012]

Early Modern English

Hee said nobody had said anything agt mee .

[Henry Oxinden, 1660]

Early Modern English

He against me

Hee said nobody had said anything agt mee .

- Spelling variation

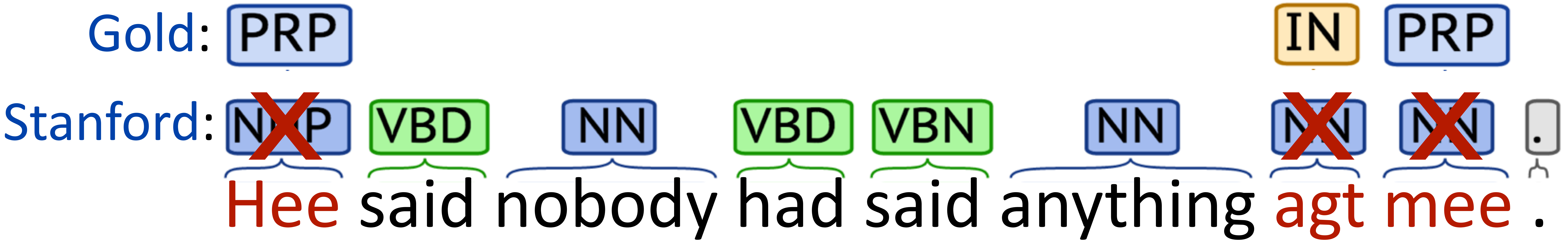
[Henry Oxinden, 1660]

Stanford POS Tagger

Stanford:	NNP	VBD	NN	VBD	VBN	NN	NN	NN	.
	Hee	said	nobody	had	said	anything	agt	mee	.

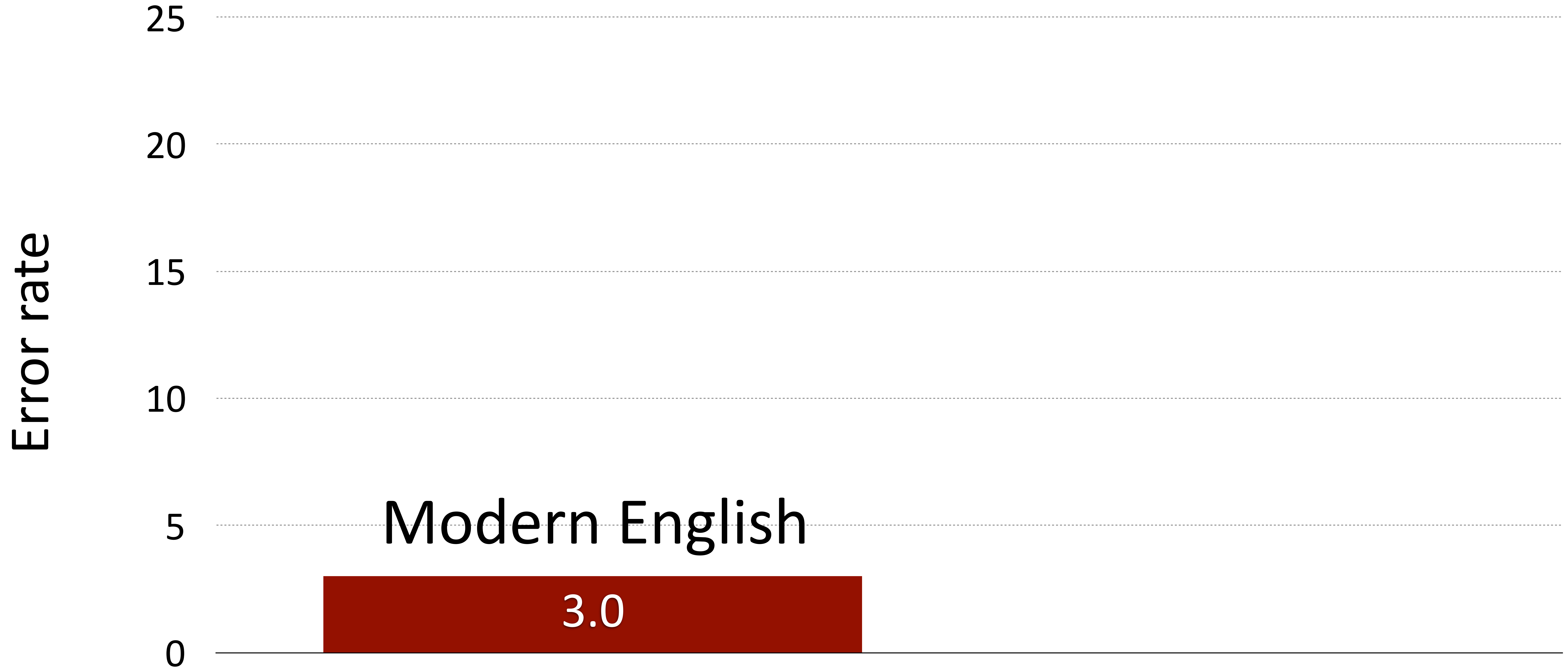
- Spelling variation

Stanford POS Tagger



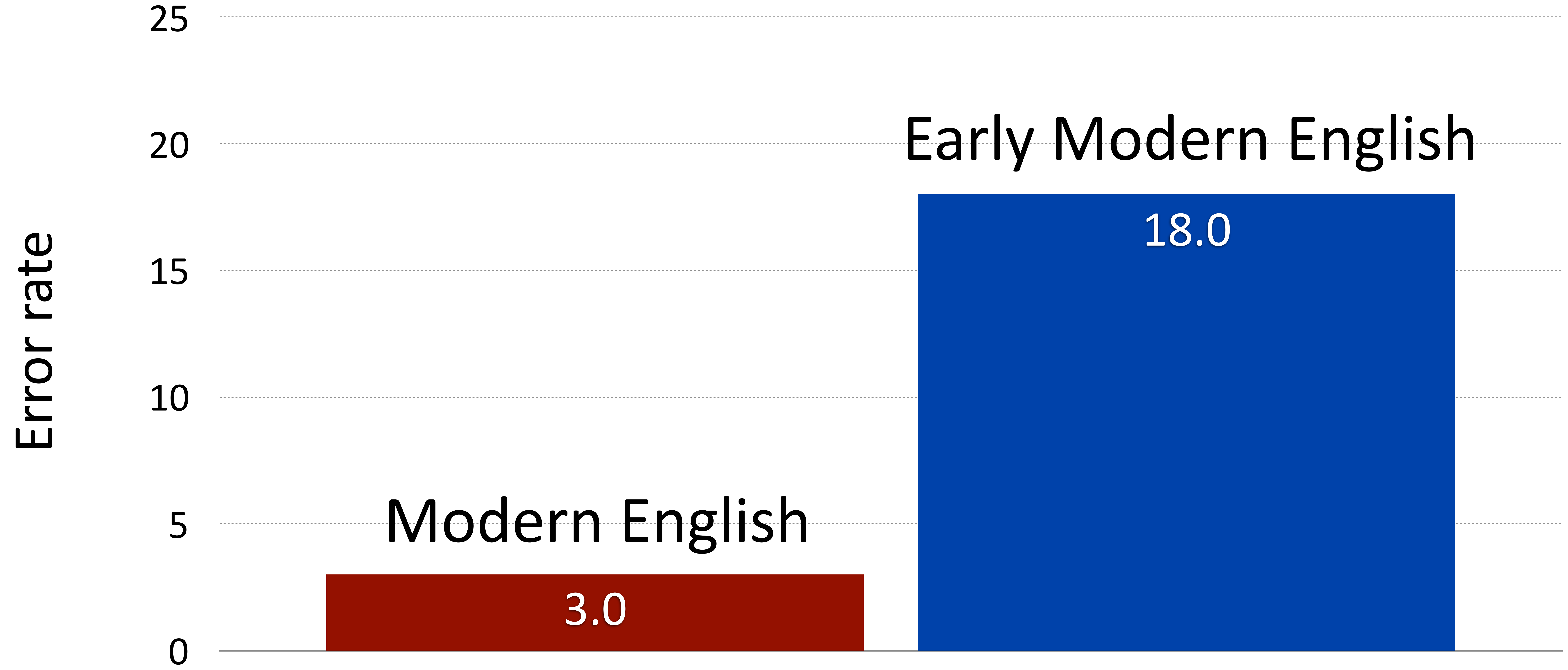
- Spelling variation

Transfer Loss for POS Tagging



[Rayson et al., 2007]

Transfer Loss for POS Tagging



[Rayson et al., 2007]

Approaches

- Spelling normalization
  - Map from historical spellings to contemporary forms.

Rayson et al. (2007)
Scheible et al. (2011)
Bollmann (2011)

Approaches

- Spelling normalization
  - Map from historical spellings to contemporary forms.

Rayson et al. (2007)
Scheible et al. (2011)
Bollmann (2011)

- Domain adaptation (this work)
  - Build robust NLP systems with representation learning.

Yang & Eisenstein (2014)
Yang & Eisenstein (2015)

Spelling Normalization

Original: Hee said nobody had said anything agt mee .

Normalized: Hee said nobody had said anything aged me .

[VARD; Baron and Rayson, 2008]

Spelling Normalization

Original: Hee said nobody had said anything agt mee .

Normalized: Hee said nobody had said anything aged me .

- Correct normalization

[VARD; Baron and Rayson, 2008]

Spelling Normalization

against

Original: Hee said nobody had said anything agt mee .

Normalized: Hee said nobody had said anything aged me .

✗ ✓

- Correct normalization
- Incorrect normalization

[VARD; Baron and Rayson, 2008]

Spelling Normalization

He against

Original: Hee said nobody had said anything agt mee .

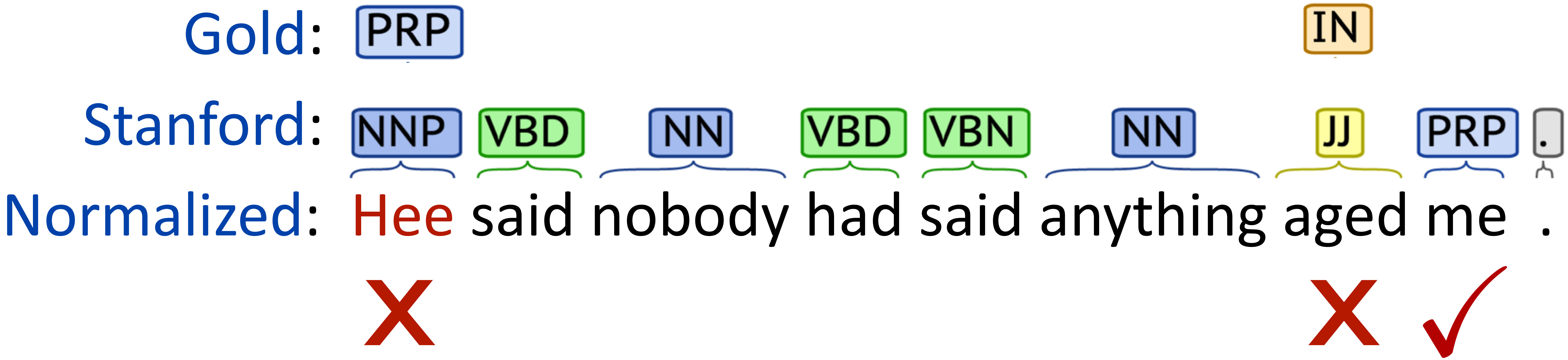
Normalized: Hee said nobody had said anything aged me .

✗ ✗ ✓

- Correct normalization
- Incorrect normalization
- False negative

[VARD; Baron and Rayson, 2008]

Spelling Normalization



[VARD; Baron and Rayson, 2008]

Spelling Normalization

Gold: PRP IN

Stanford: ~~NNP~~ VBD NN VBD VBN NN ~~JJ~~ PRP .

Normalized: Hee said nobody had said anything aged me .

✗ ✗ ✓

[VARD; Baron and Rayson, 2008]

Representation Learning

Hee said nobody had said anything agt mee .

Representation Learning

PRP VBD NN VBD VBN NN IN PRP .

Hee said nobody had said anything agt mee .

Representation Learning

Hee said nobody had said anything agt mee .

Representation Learning

Hee said nobody had said anything agt mee .

OOV	Context	IV	Context
Hee	said was came told …	He I We …	said was came told …

Model

Feature Embeddings

Hee said nobody had said anything agt mee .

[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

Hee said nobody had said anything agt mee .

[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

Hee said nobody had said anything agt mee .

features

CurrWord = hee 1

NextWord = said 2

Prefix1 = h 3

Suffix1 = e 4

…

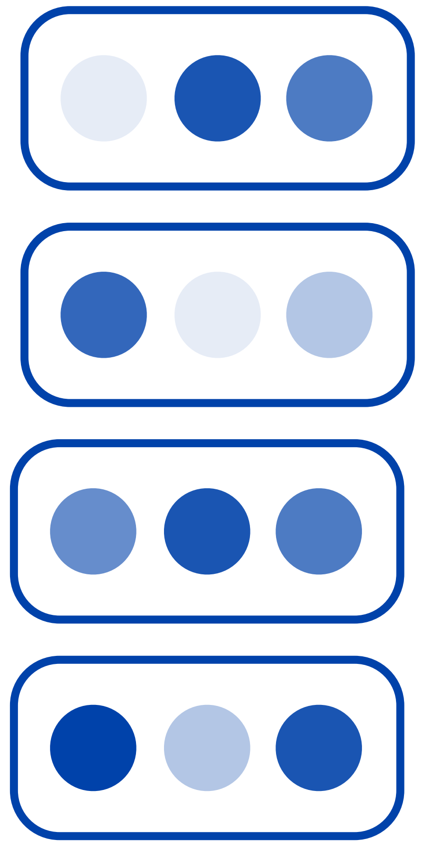
[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

Hee said nobody had said anything agt mee .

features

- CurrWord = hee — 1
- NextWord = said — 2
- Prefix1 = h — 3
- Suffix1 = e — 4
- …



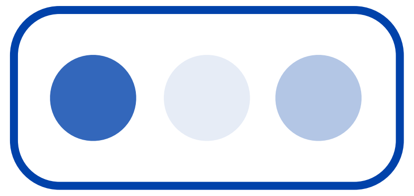
[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

Hee said nobody had said anything agt mee .

features

- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …



[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

Hee said nobody had said anything agt mee .

features

CurrWord = hee 1

NextWord = said 2

Prefix1 = h 3

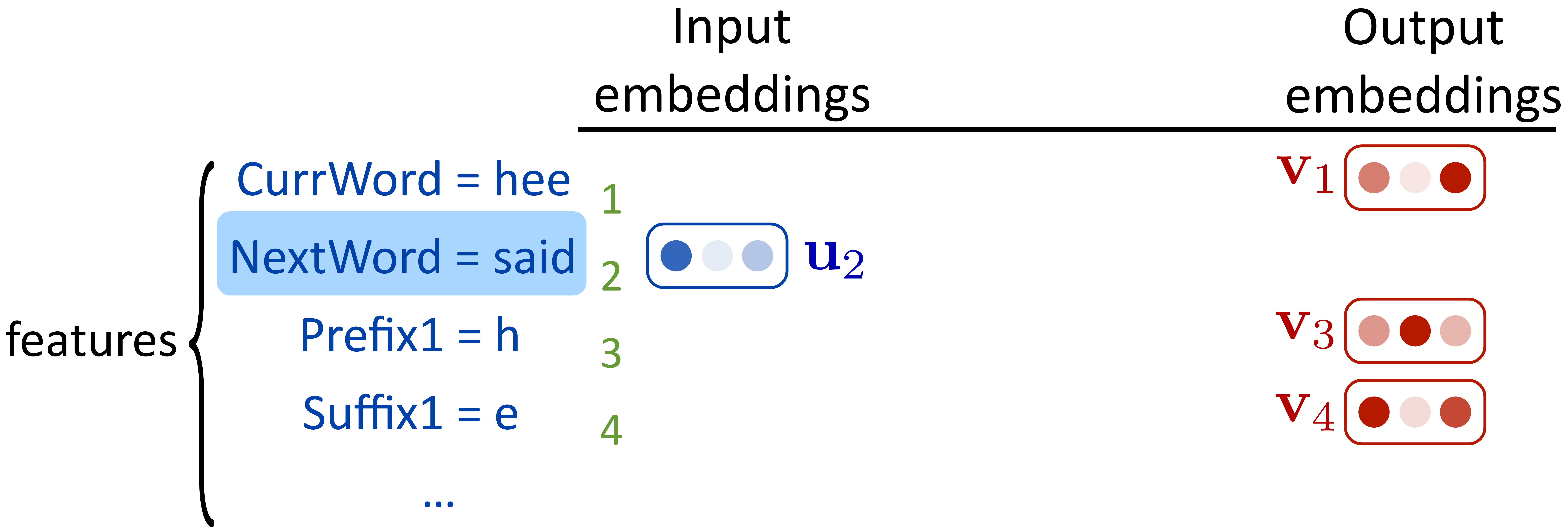
Suffix1 = e 4

...

[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

$$\mathbf{p}(\mathrm{f}_t|\mathrm{f}_2) \propto \exp\left(\mathbf{u}_2{}^\top \mathbf{v}_t\right)$$

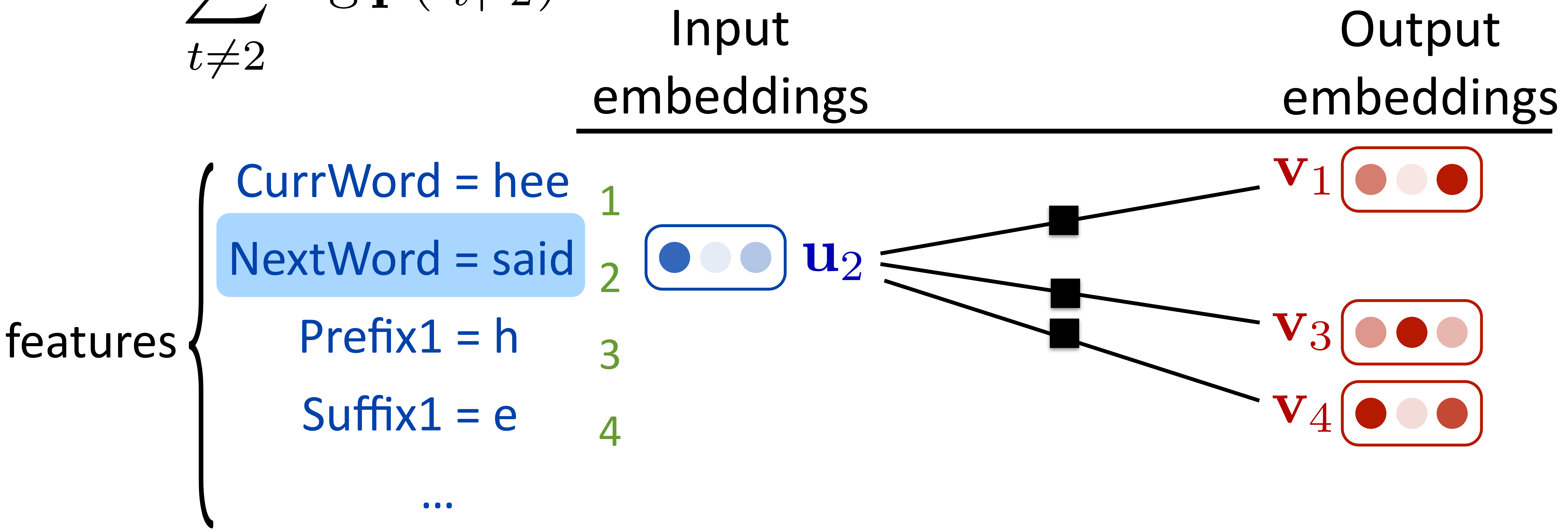


[FEMA; Yang and Eisenstein, 2015]

Feature Embeddings

$$\mathbf{p}(\mathrm{f}_t|\mathrm{f}_2) \propto \exp\left(\mathbf{u}_2^\top \mathbf{v}_t\right)$$

$$\ell = \sum_{t \neq 2}^{T} \log \mathbf{p}(\mathrm{f}_t|\mathrm{f}_2)$$



[FEMA; Yang and Eisenstein, 2015]

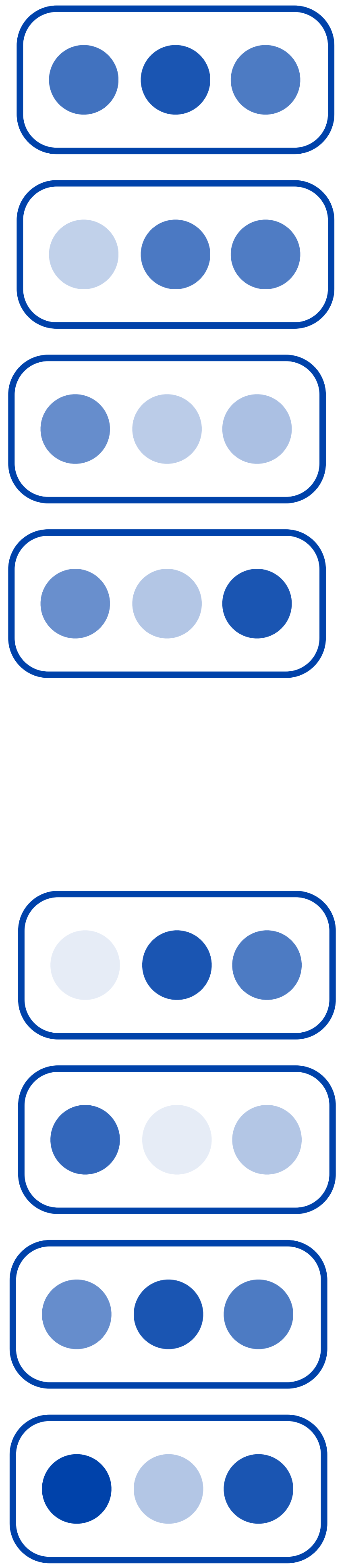
Word Embeddings

words

- hee 1
- said 2
- nobody 3
- had 4
- …

features

- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …



- Word embeddings
- Feature embeddings

[word2vec; Mikolov et al., 2013]

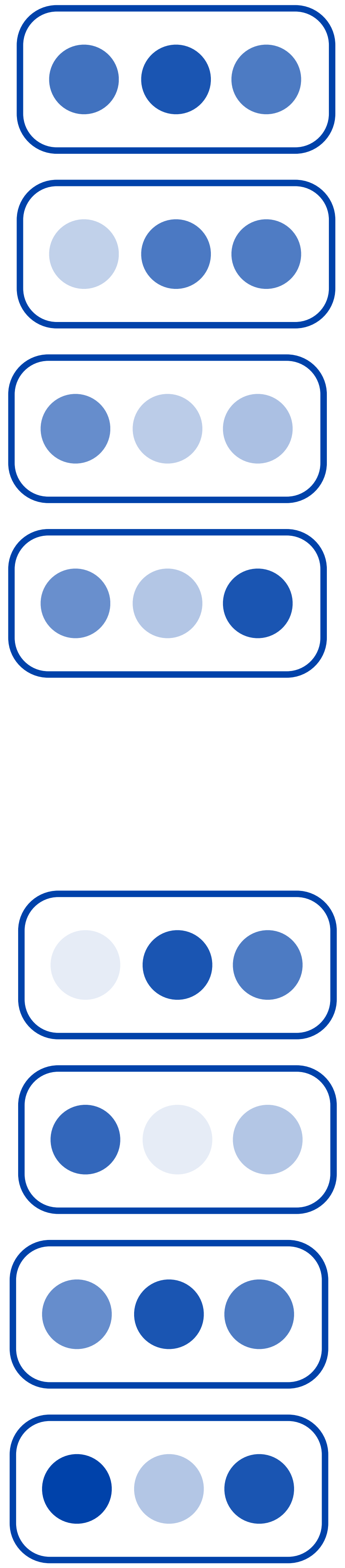
Word Embeddings

words
- hee
- said
- nobody
- had
- …

features
- CurrWord = hee
- NextWord = said
- Prefix1 = h
- Suffix1 = e
- …

1
2
3
4

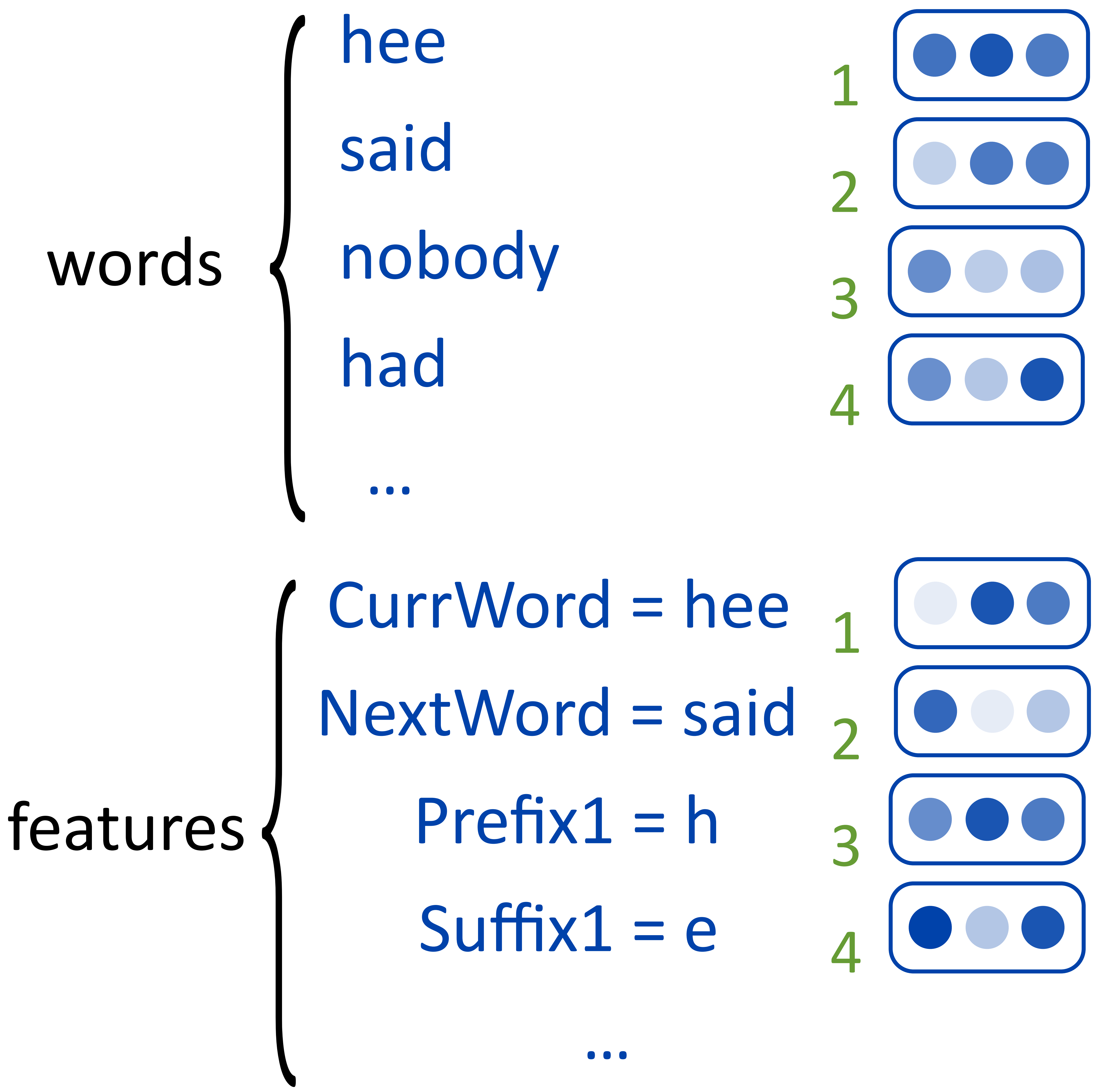
1
2
3
4



- Word embeddings
  - Generic representations
- Feature embeddings

[word2vec; Mikolov et al., 2013]

Word Embeddings



- Word embeddings
  - Generic representations
- Feature embeddings
  - Task-specific representations

[word2vec; Mikolov et al., 2013]

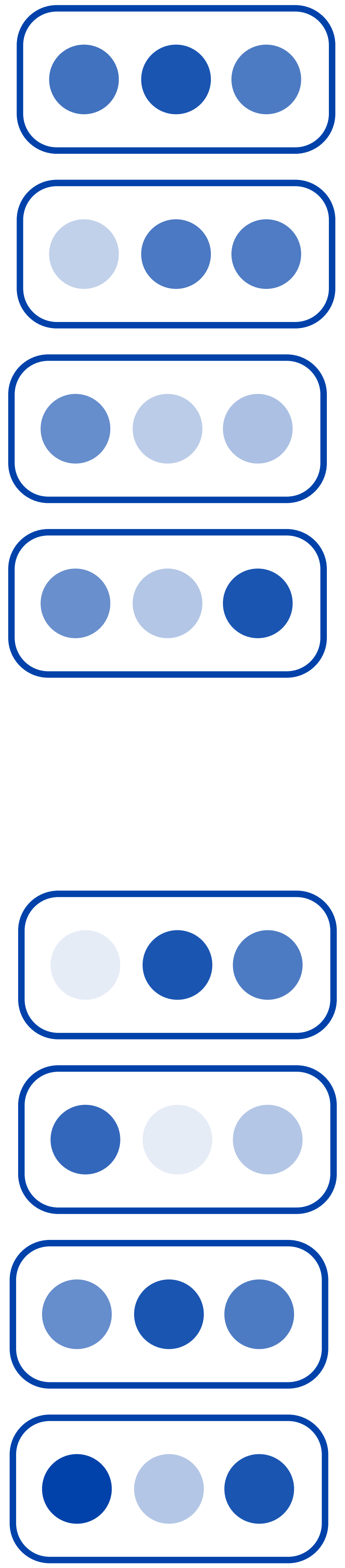
Word Embeddings

words
- hee
- said
- nobody
- had
- …

features
- CurrWord = hee
- NextWord = said
- Prefix1 = h
- Suffix1 = e
- …

1
2
3
4

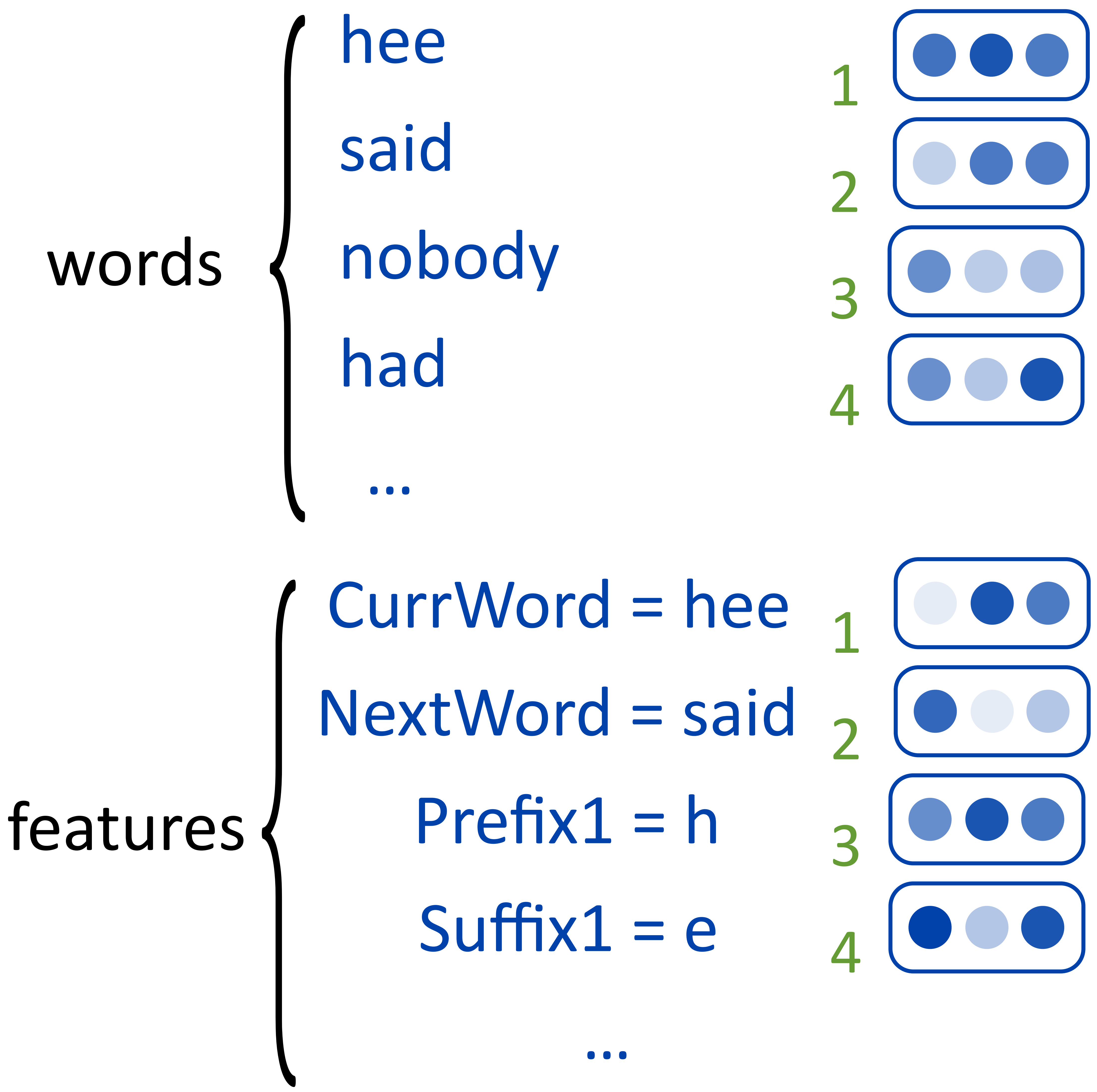
1
2
3
4



- Word embeddings
  - Generic representations
  - Word co-occurrences
- Feature embeddings
  - Task-specific representations

[word2vec; Mikolov et al., 2013]

Word Embeddings



- Word embeddings
  - Generic representations
  - Word co-occurrences
- Feature embeddings
  - Task-specific representations
  - Feature co-occurrences

[word2vec; Mikolov et al., 2013]

Learning from Multiple Domains

- Previous work on unsupervised domain adaptation involves in two domains.

[FEMA; Yang and Eisenstein, 2015]

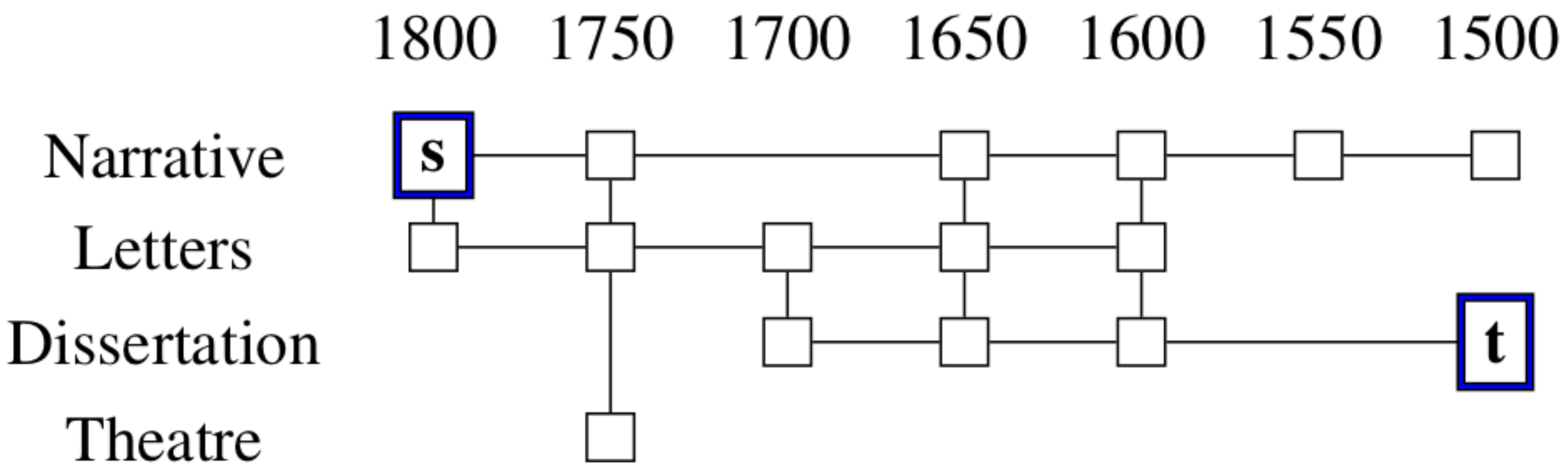
Learning from Multiple Domains

- Previous work on unsupervised domain adaptation involves in two domains.
- Unsupervised multi-domain adaptation

[FEMA; Yang and Eisenstein, 2015]

Learning from Multiple Domains

- Previous work on unsupervised domain adaptation involves in two domains.
- Unsupervised multi-domain adaptation



[FEMA; Yang and Eisenstein, 2015]

Multiple Feature Embeddings

Hee said nobody had said anything agt mee .

[FEMA; Yang and Eisenstein, 2015]

Multiple Feature Embeddings

Domain Attributes: Genre Epoch

Hee said nobody had said anything agt mee .

[FEMA; Yang and Eisenstein, 2015]

Multiple Feature Embeddings

Domain Attributes: Genre Epoch

letters 1600+

Hee said nobody had said anything agt mee .

[FEMA; Yang and Eisenstein, 2015]

Multiple Feature Embeddings

Domain Attributes: Genre Epoch

letters 1600+

Hee said nobody had said anything agt mee .

features
- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …

[FEMA; Yang and Eisenstein, 2015]

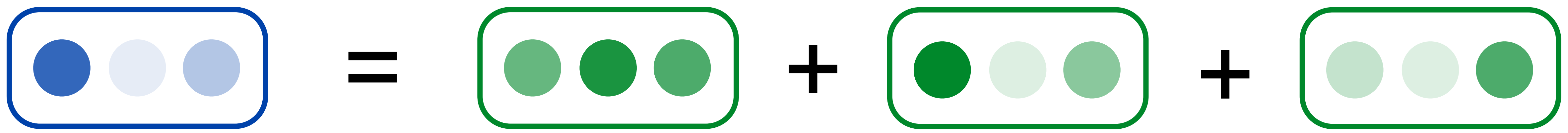
Multiple Feature Embeddings

Domain Attributes: Genre | Epoch

letters | 1600+

Hee said nobody had said anything agt mee .

features
- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- ...



(shared) (letters) (1600+)

[FEMA; Yang and Eisenstein, 2015]

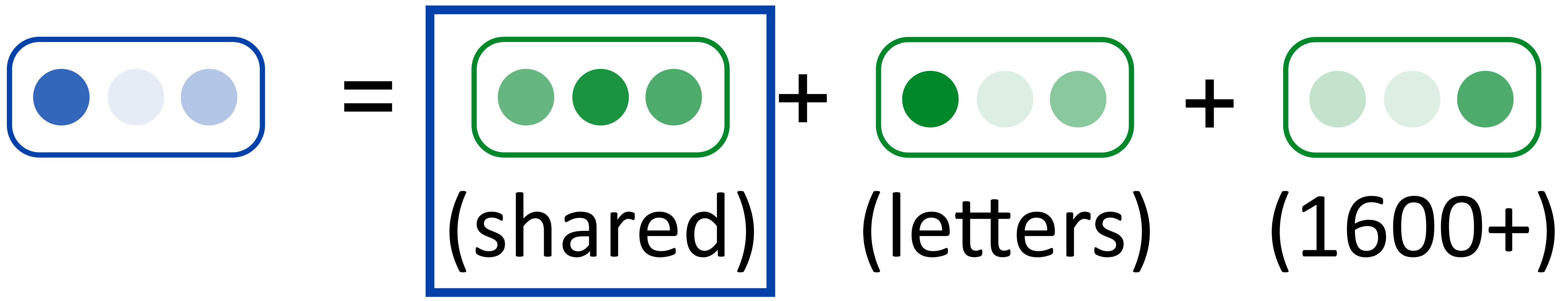
Multiple Feature Embeddings

Domain Attributes: Genre Epoch

letters 1600+

Hee said nobody had said anything agt mee .

features
- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …

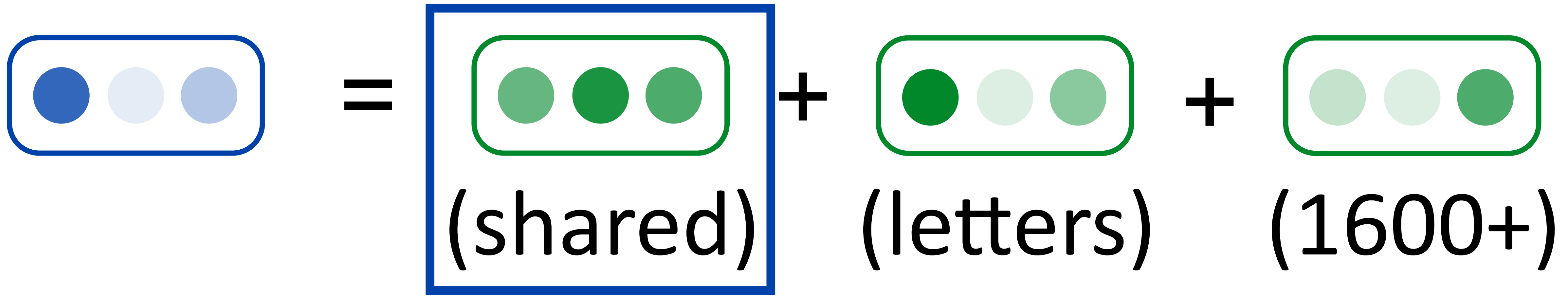


[FEMA; Yang and Eisenstein, 2015]

Multiple Feature Embeddings

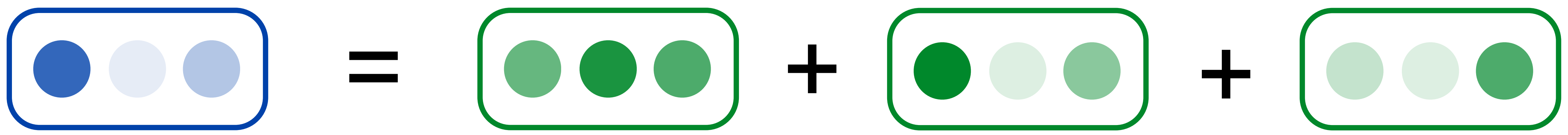
Hee said nobody had said anything agt mee .

features
- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …



[FEMA; Yang and Eisenstein, 2015]

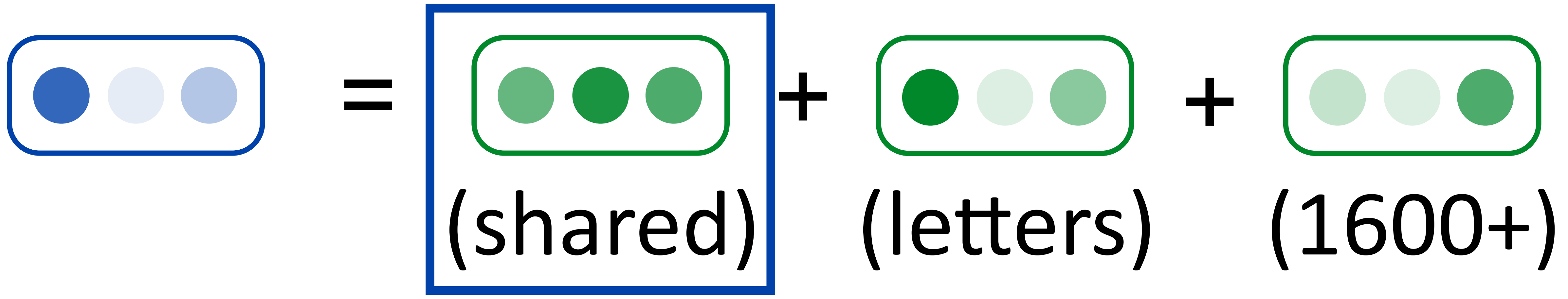
Multiple Feature Embeddings



$$\mathbf{u}_2 = \mathbf{h}_2^{(\text{shared})} + \mathbf{h}_2^{(\text{letters})} + \mathbf{h}_2^{(1600+)}$$

Hee said nobody had said anything agt mee .

features:
- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …



[FEMA; Yang and Eisenstein, 2015]

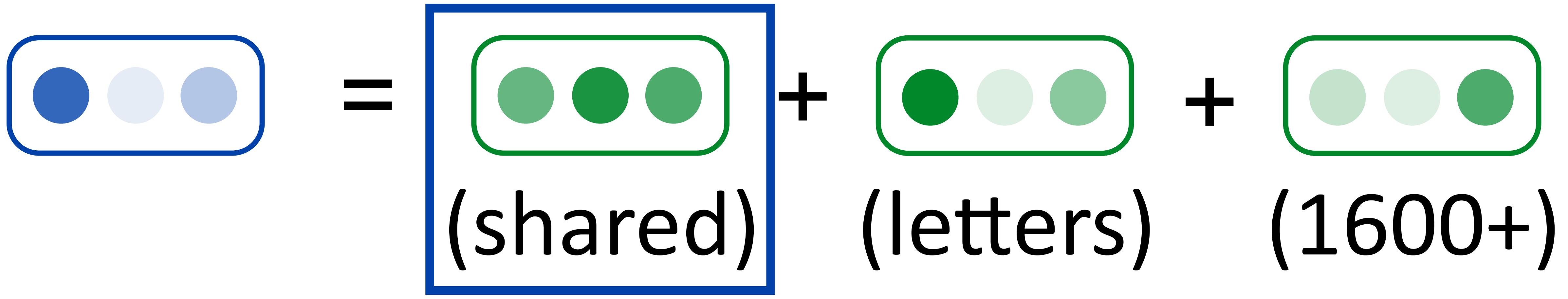
Multiple Feature Embeddings

$$\mathbf{p}(\mathrm{f}_t|\mathrm{f}_2) \propto \exp\left(\mathbf{u_2}^\top \mathbf{v}_t\right)$$

$$\mathbf{u}_2 = \mathbf{h}_2^{(\text{shared})} + \mathbf{h}_2^{(\text{letters})} + \mathbf{h}_2^{(1600+)}$$

Hee said nobody had said anything agt mee .

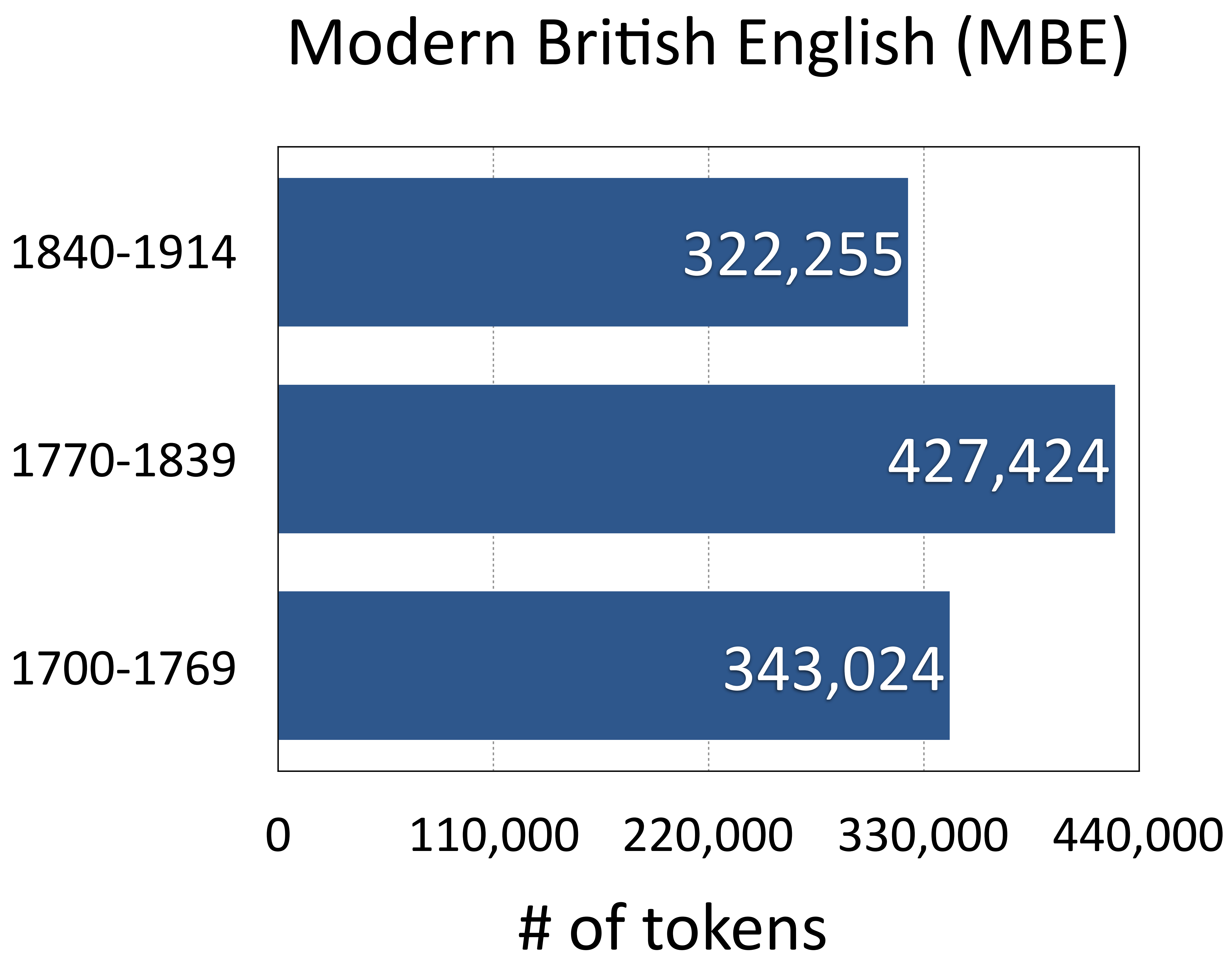
features
- CurrWord = hee 1
- NextWord = said 2
- Prefix1 = h 3
- Suffix1 = e 4
- …

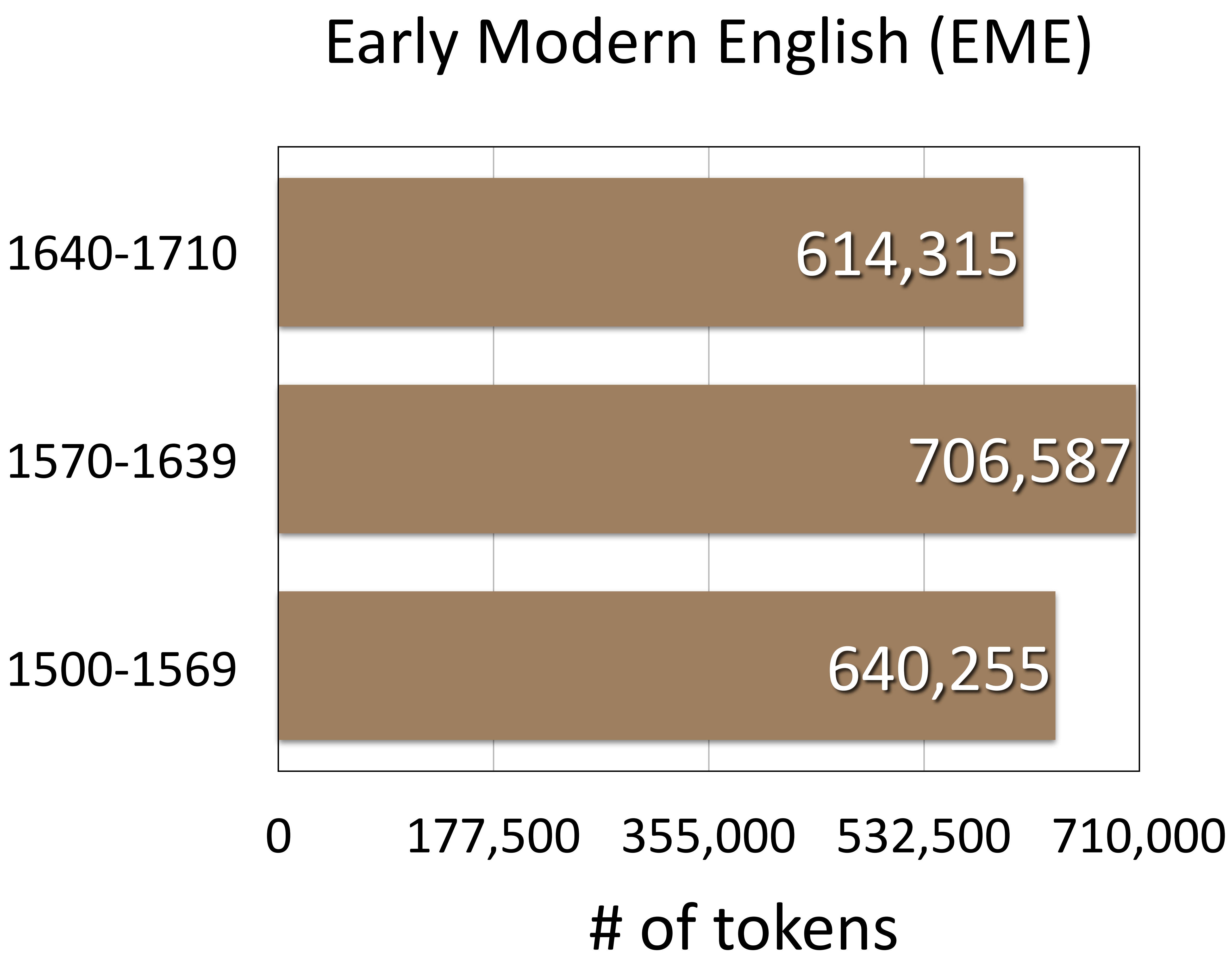


[FEMA; Yang and Eisenstein, 2015]

Experiments

Penn Corpora of Historical English





[Krochand Taylor, 2000; Kroch et al., 2004]

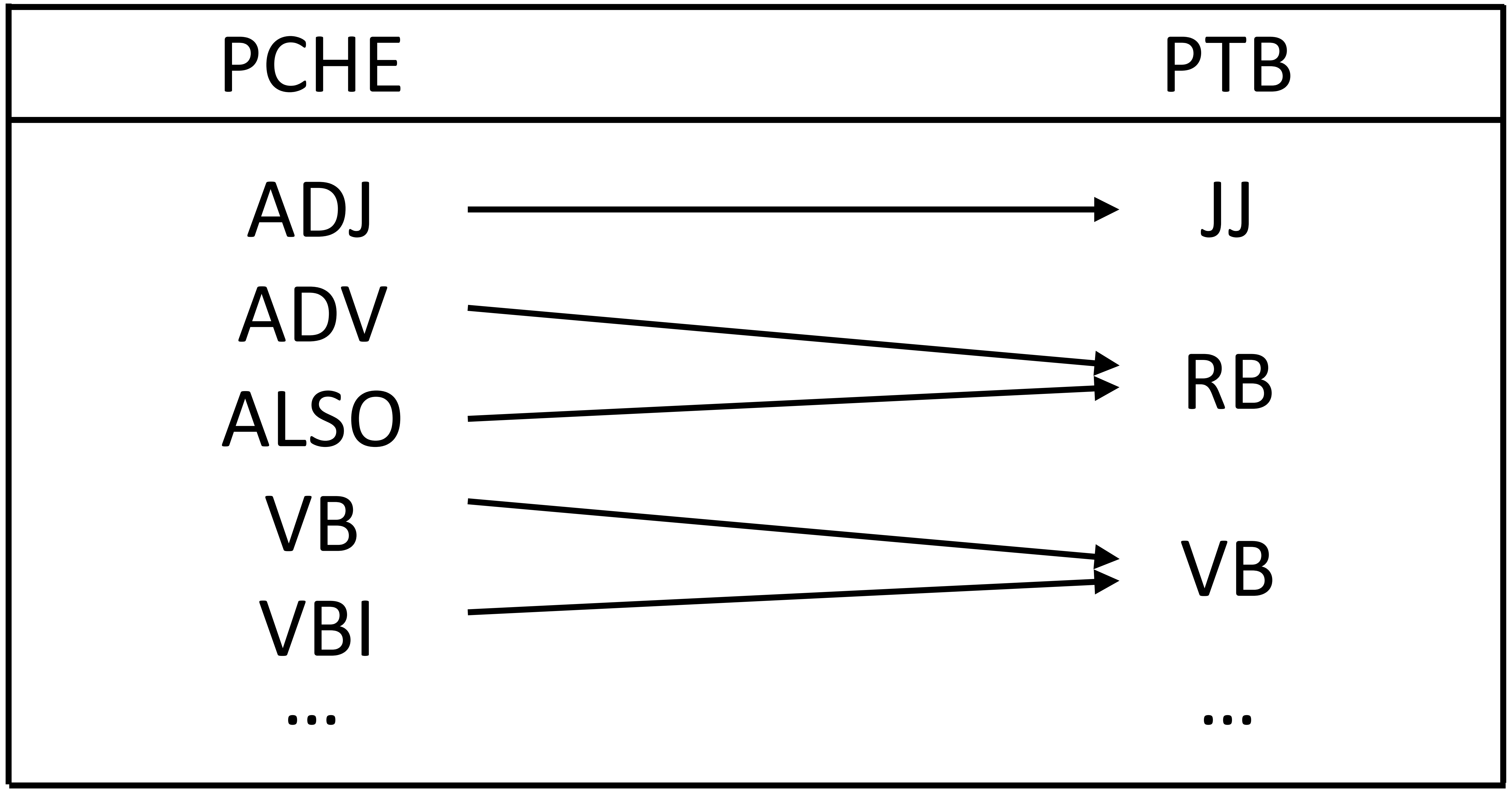
Tagset Mappings

- Penn Corpora of Historical English (PCHE) tagset: 83 tags
- Penn Treebank (PTB) tagset: 45 tags

[Moon and Baldridge, 2007]

Tagset Mappings

- Penn Corpora of Historical English (PCHE) tagset: 83 tags
- Penn Treebank (PTB) tagset: 45 tags



[Moon and Baldridge, 2007]

Systems

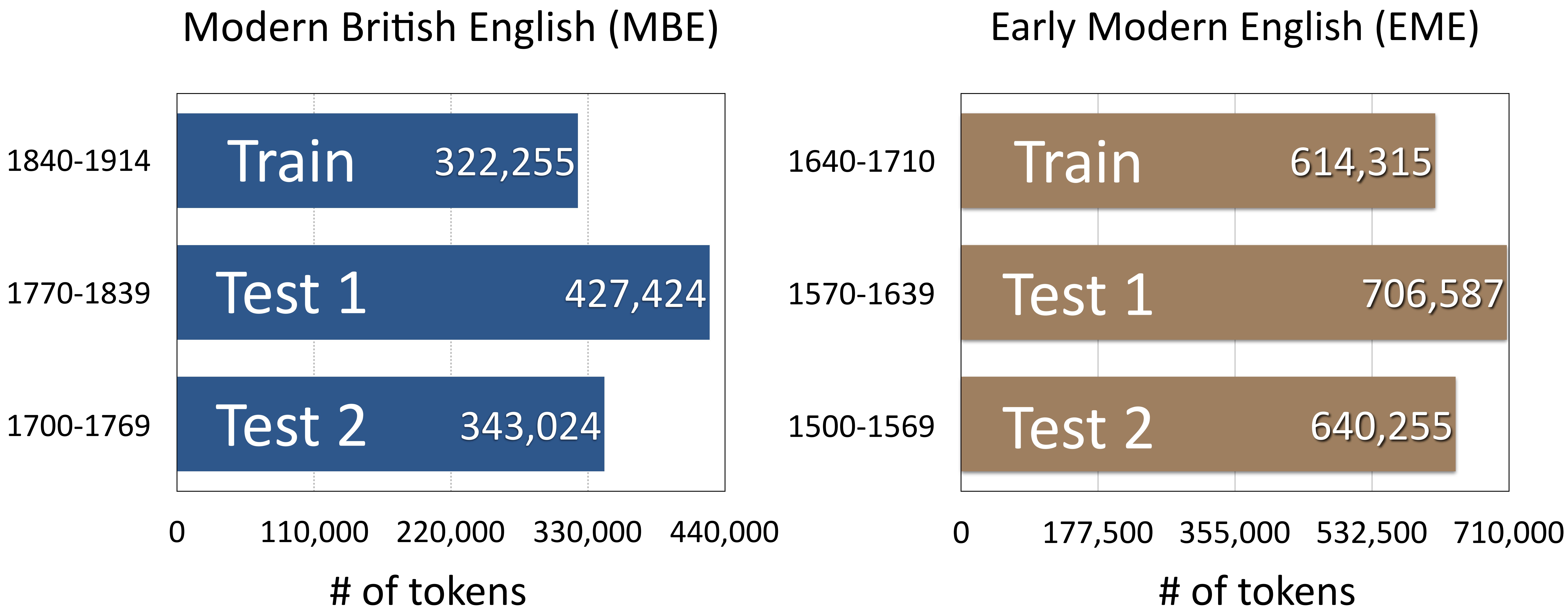
- Support vector machine (SVM) tagger
  - Sixteen basic feature templates by Ratnaparkhi (1996)

Systems

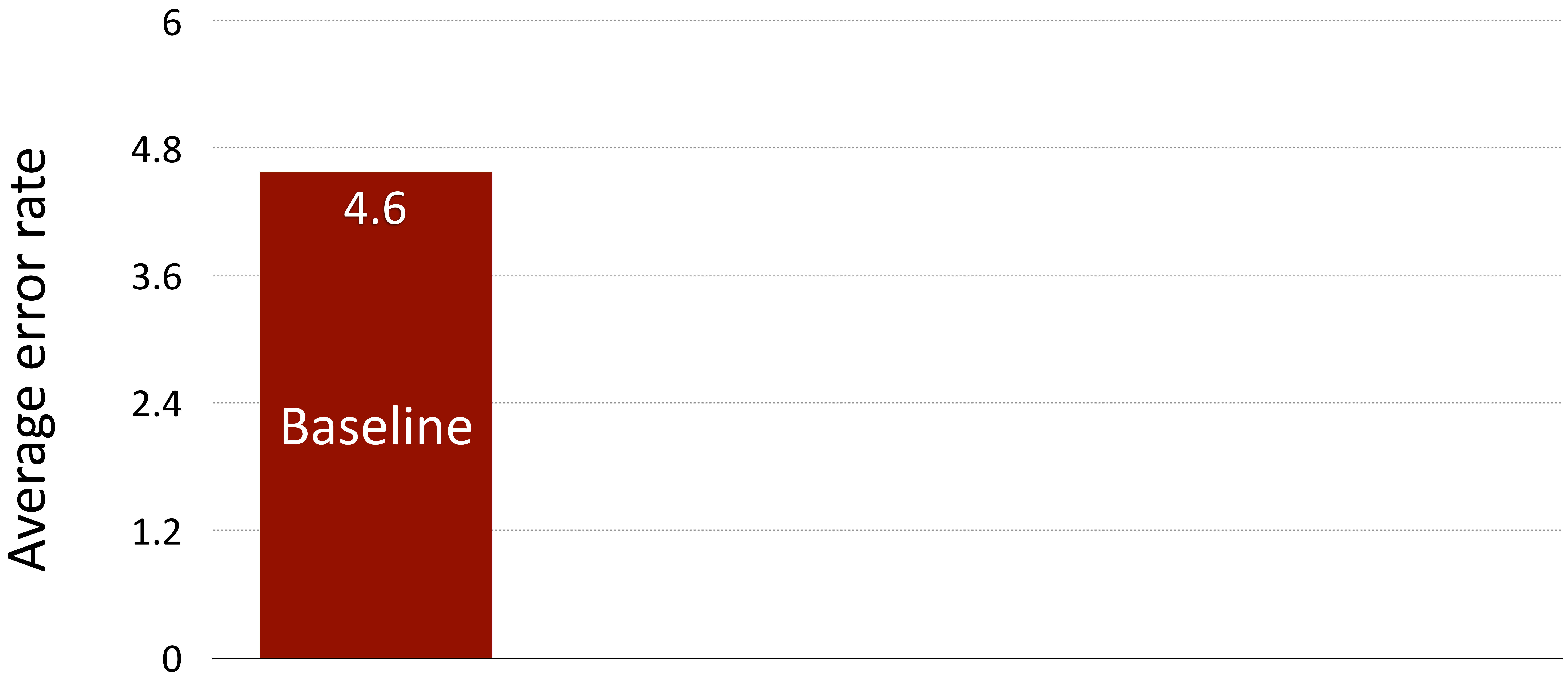
- Support vector machine (SVM) tagger
  - Sixteen basic feature templates by Ratnaparkhi (1996)
- Representation learning methods
  - Structural correspondence learning (SCL)
  - Brown clustering
  - word2vec embeddings
  - Multiple feature embeddings (FEMA)

[Blitzer et al., 2006; Brown et al., 1992; Mikolov et al.,2013]

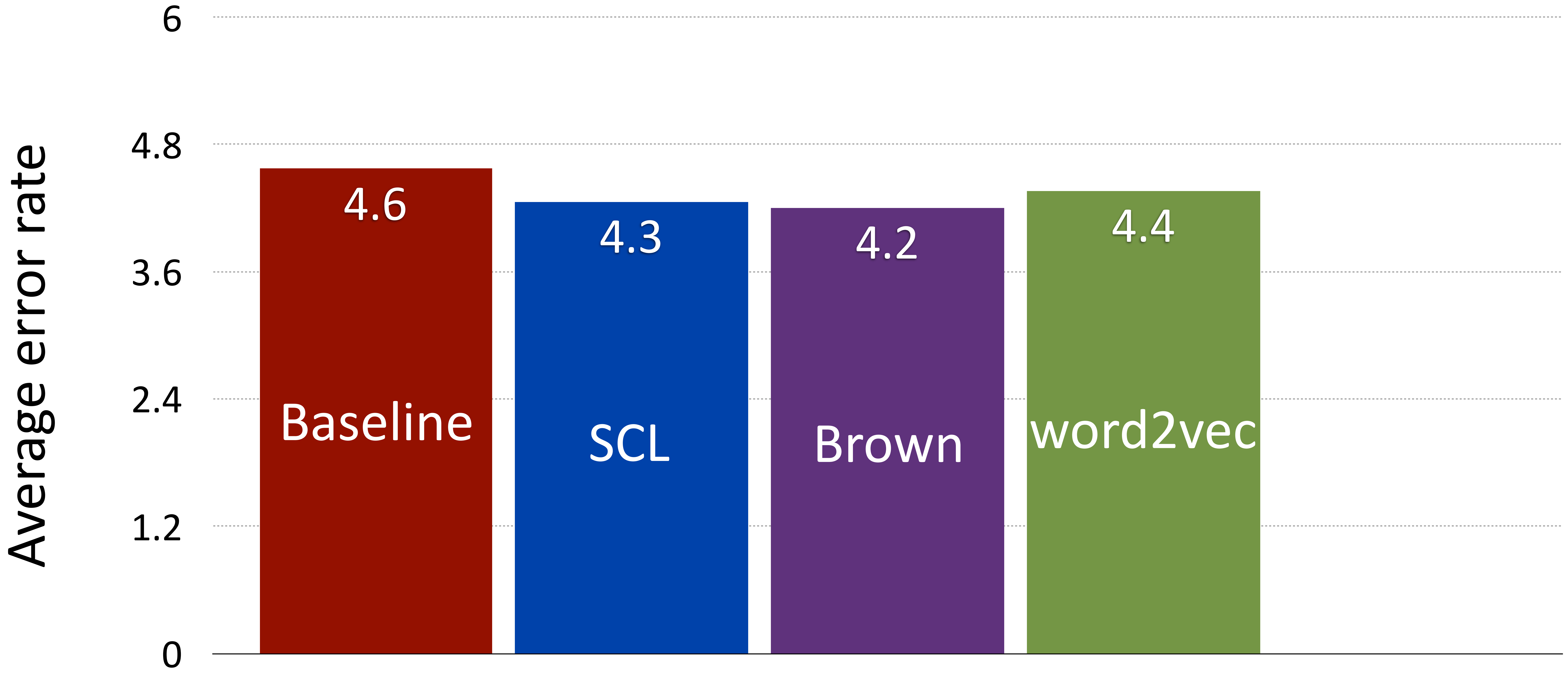
Temporal Adaptation



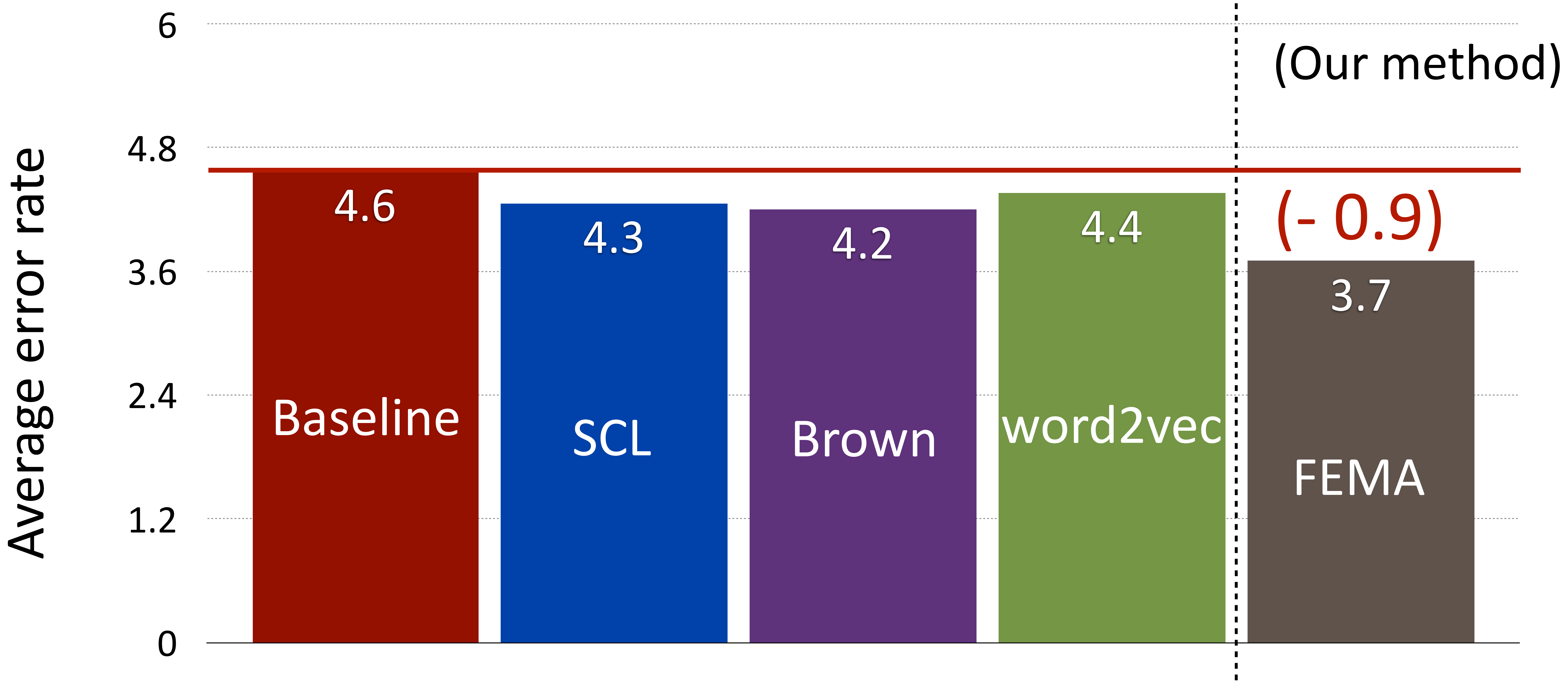
Results: Modern British English



Results: Modern British English

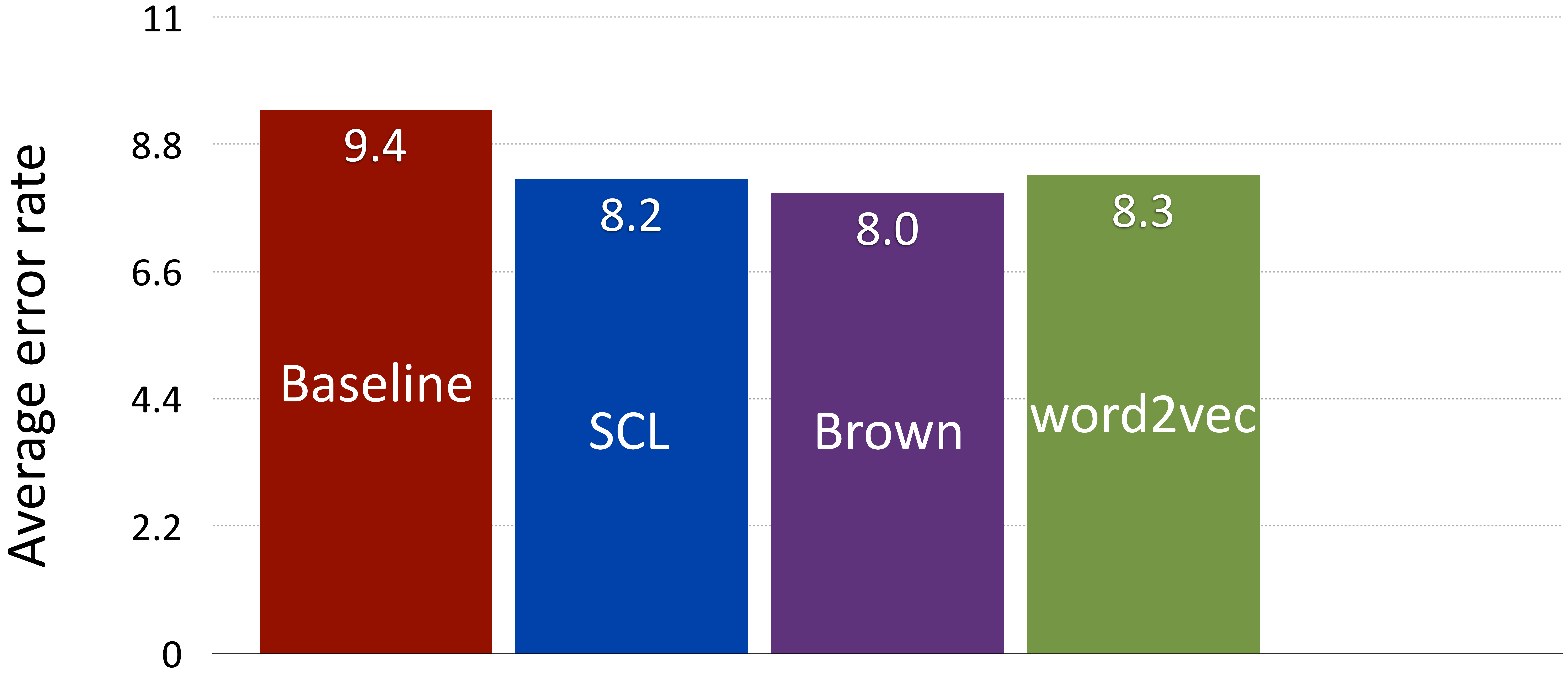


Results: Modern British English

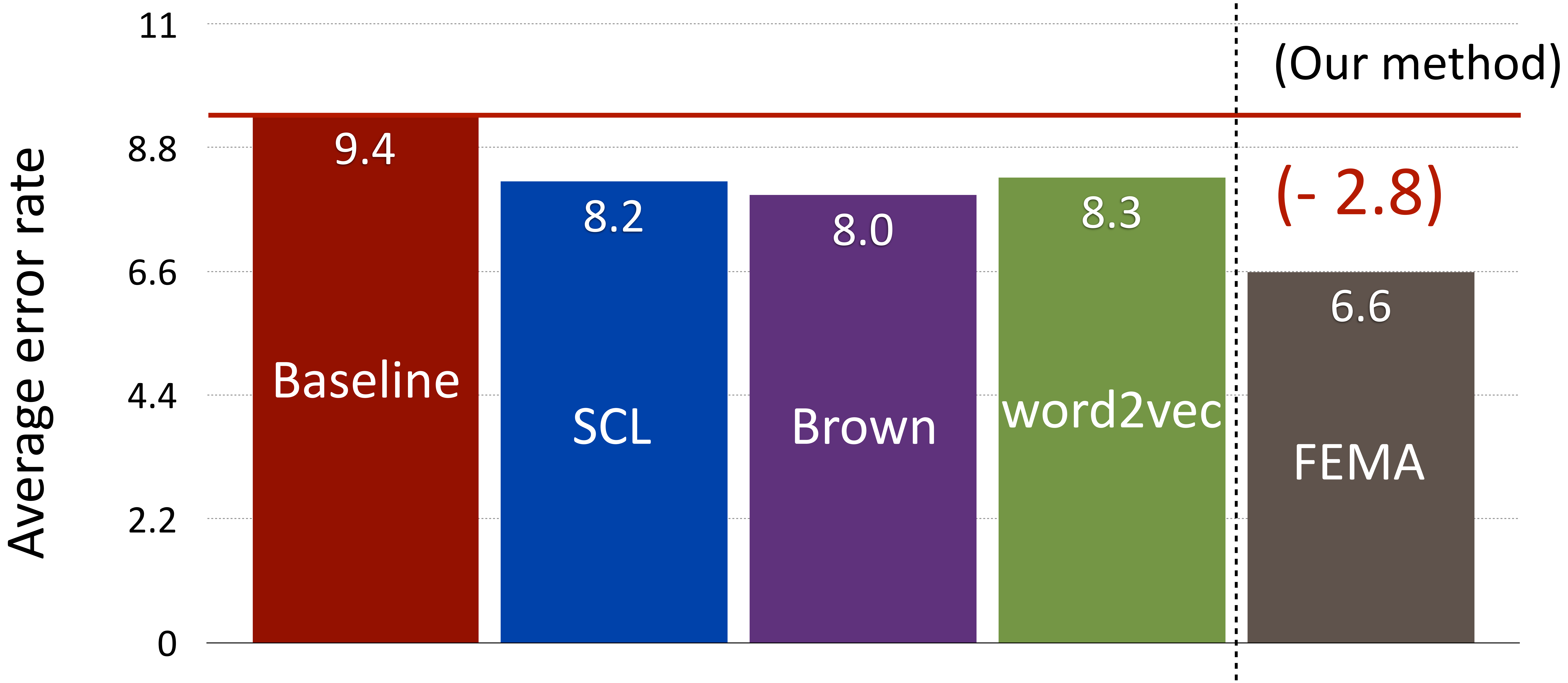


Results: Early Modern English

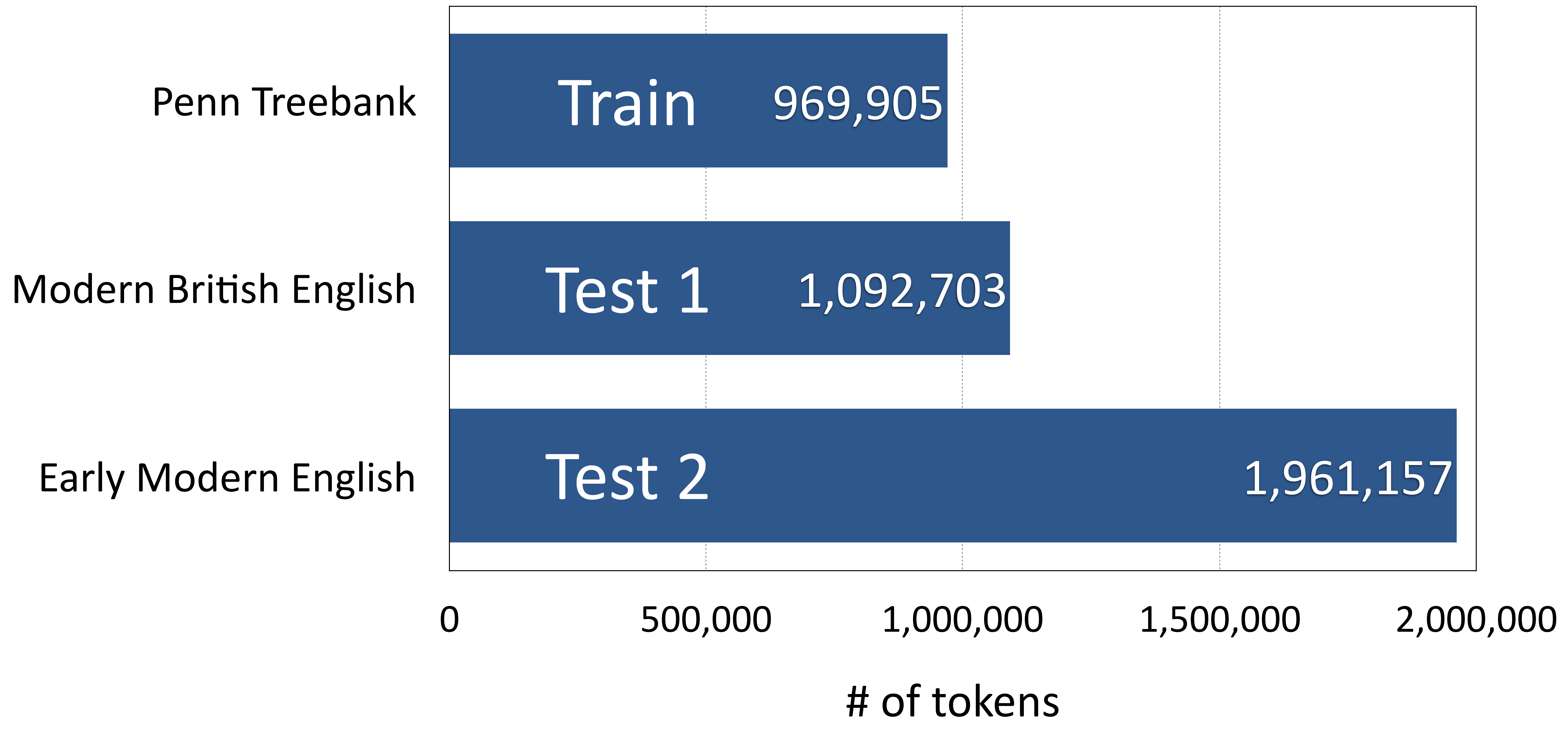
Results: Early Modern English



Results: Early Modern English



Adaptation from PTB



Adaptation from PTB

Standard evaluation scenario for English POS tagging.

WALL STREET JOURNAL

Adaptation from PTB

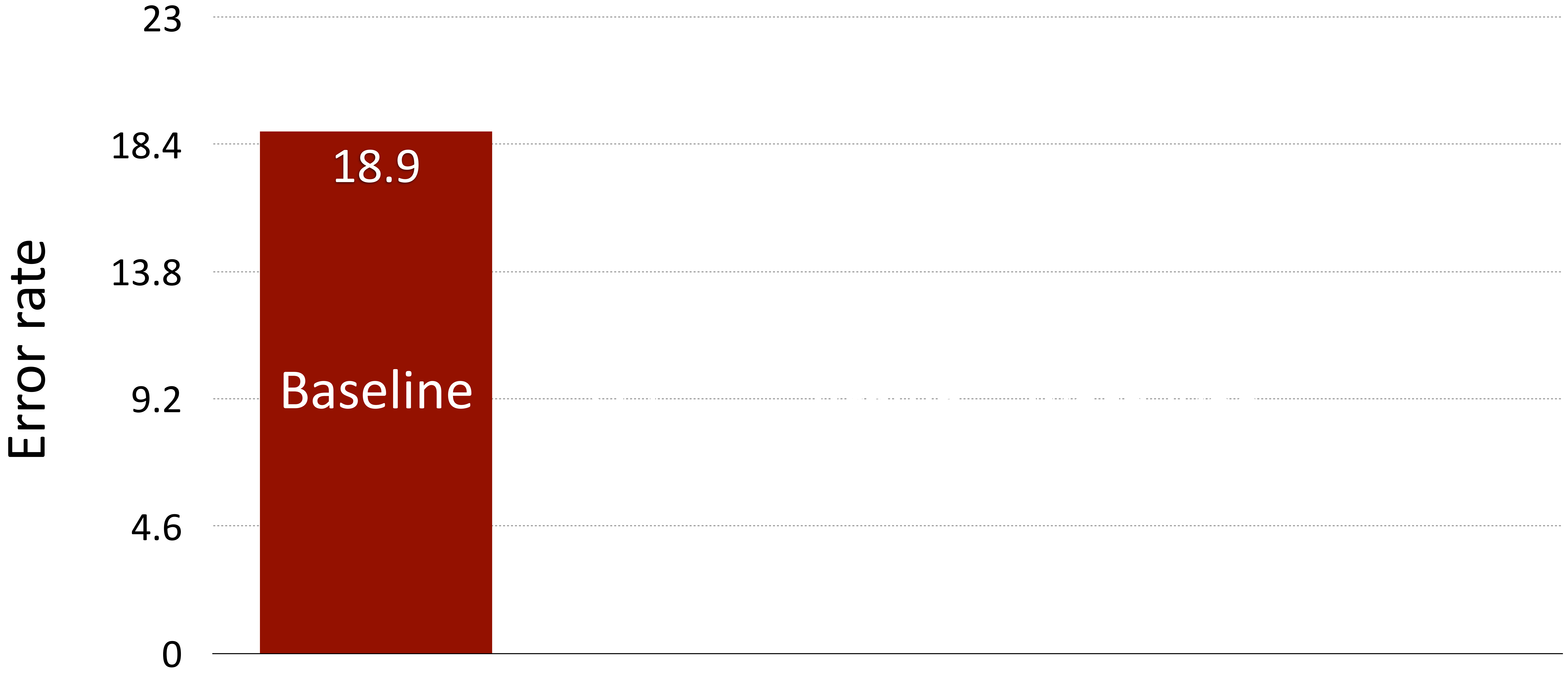
Standard evaluation scenario for English POS tagging.

WALL STREET JOURNAL

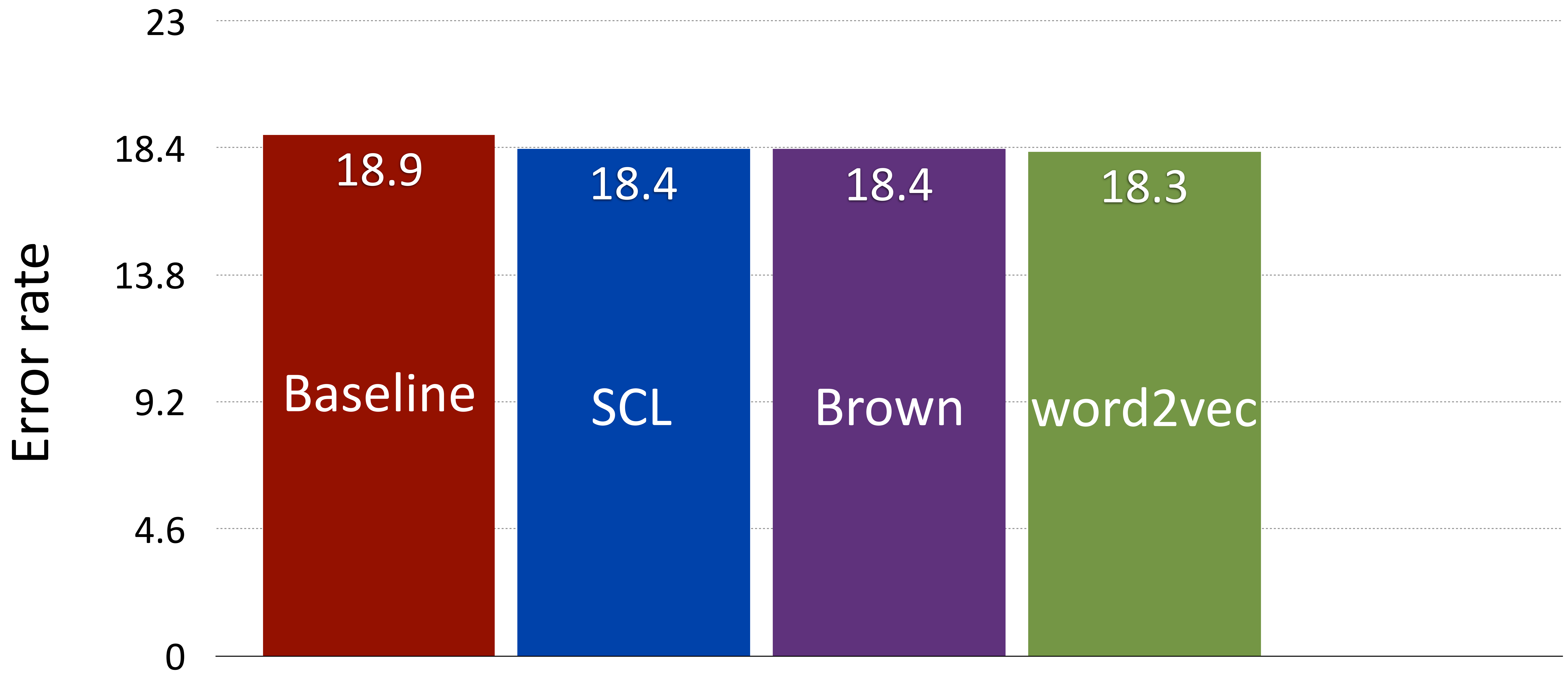
Insufficient data annotation for historical texts.

- Low resource languages
- Specific genres, styles, or epochs

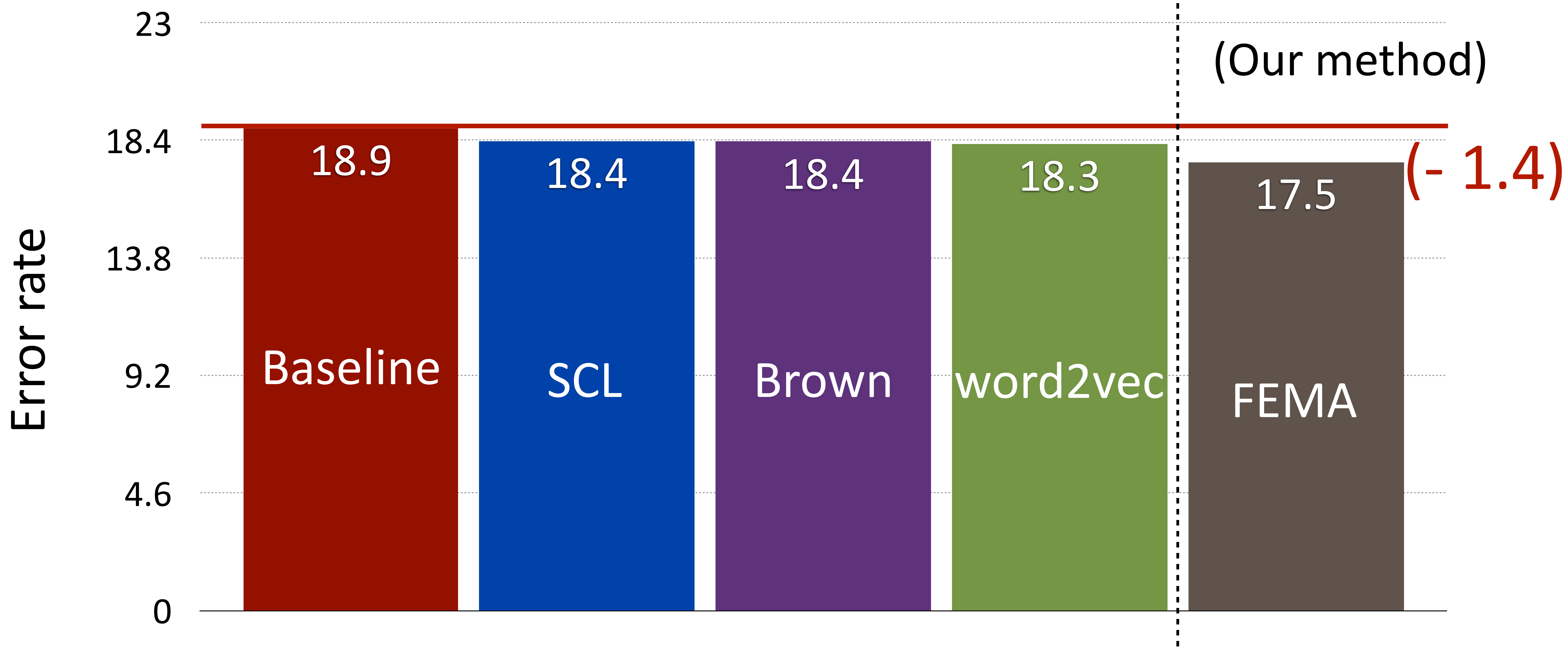
Results: Modern British English



Results: Modern British English

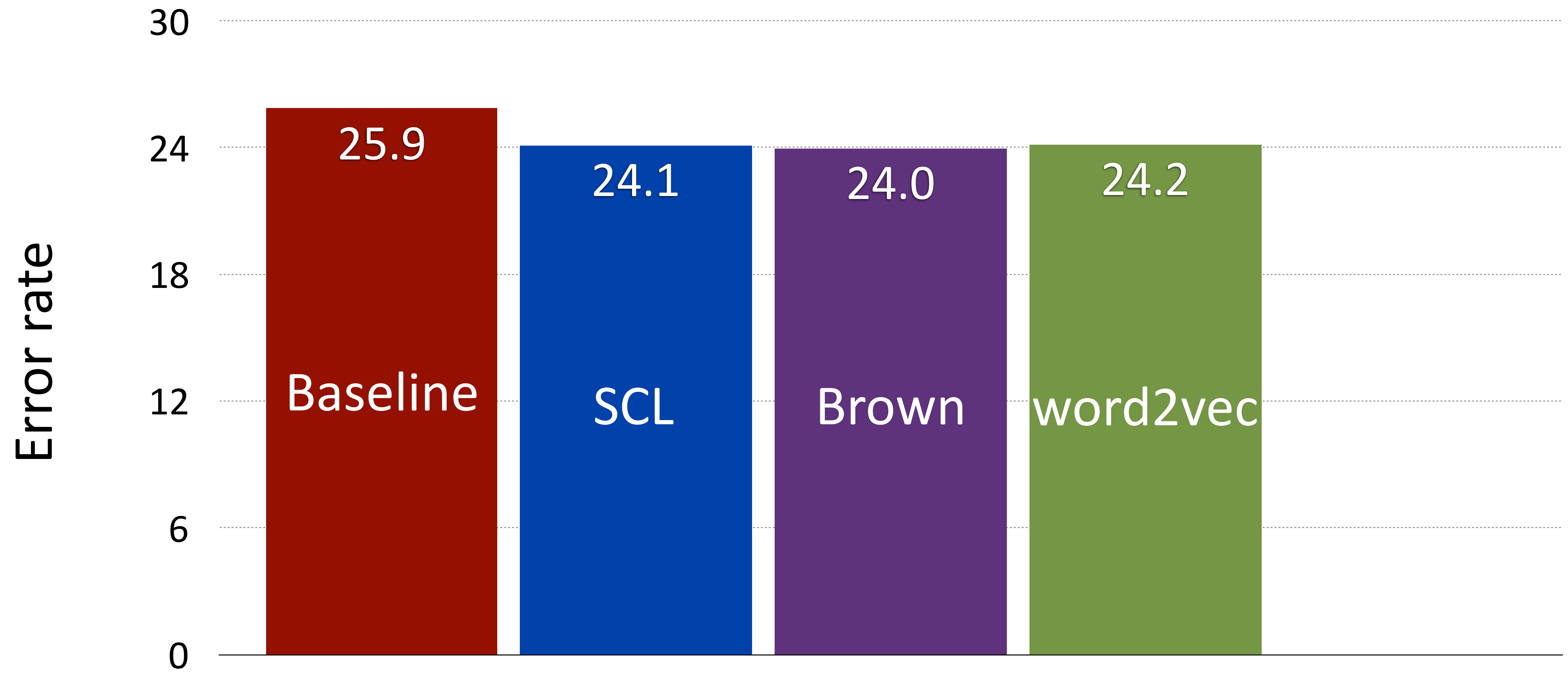


Results: Modern British English

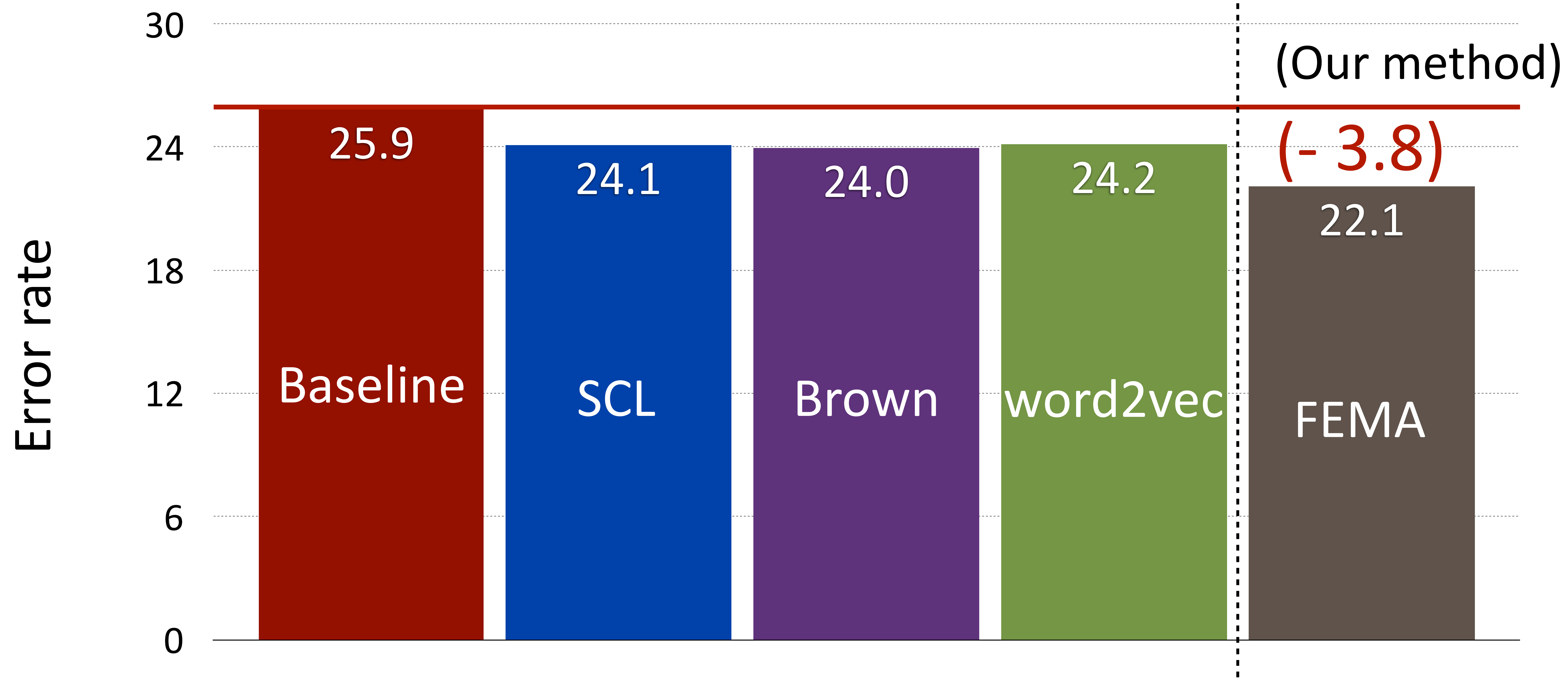


Results: Early Modern English

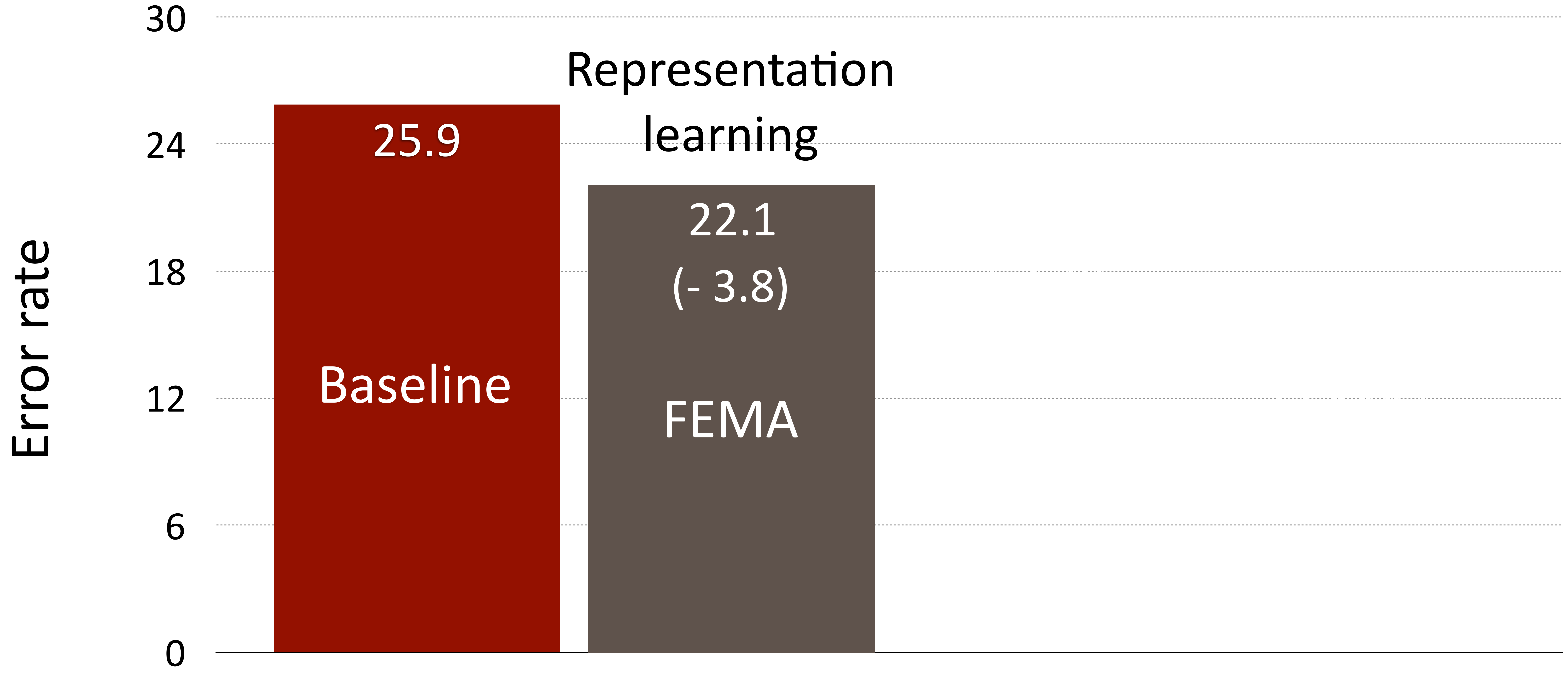
Results: Early Modern English



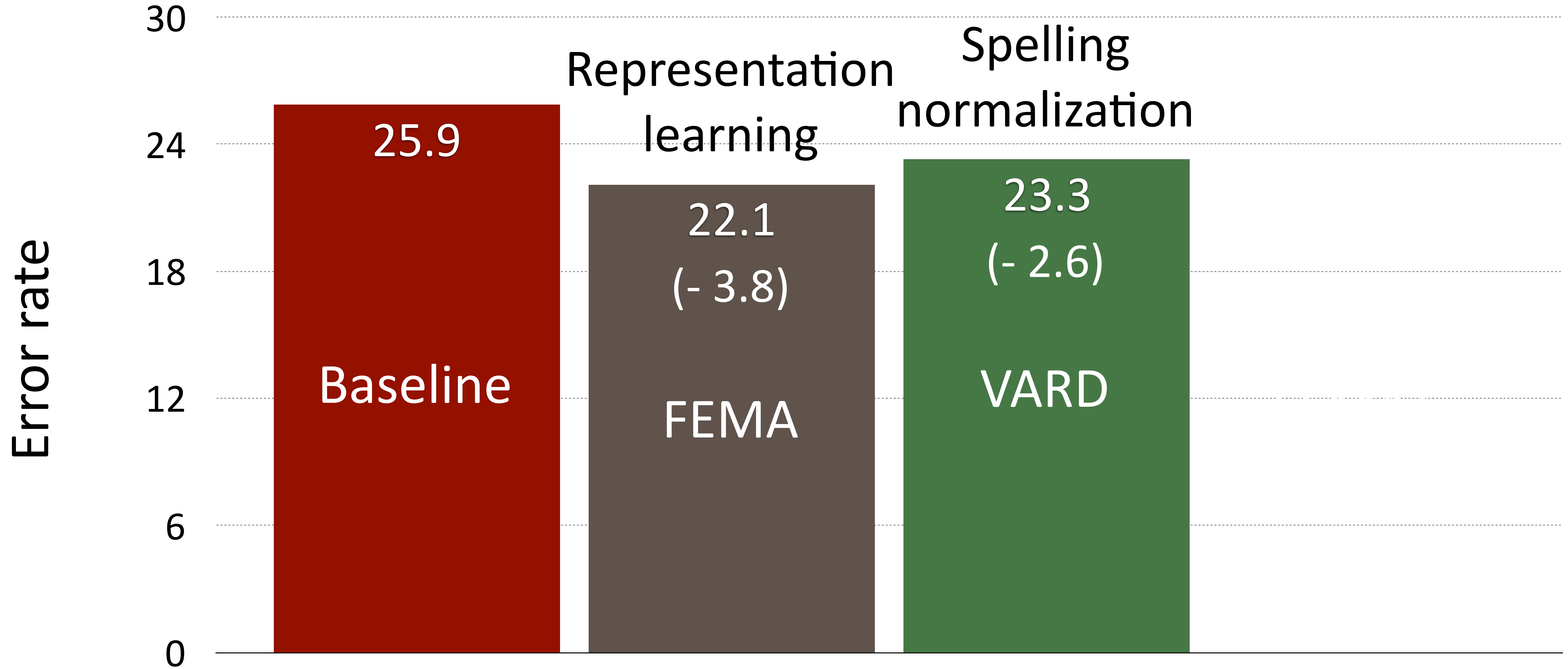
Results: Early Modern English



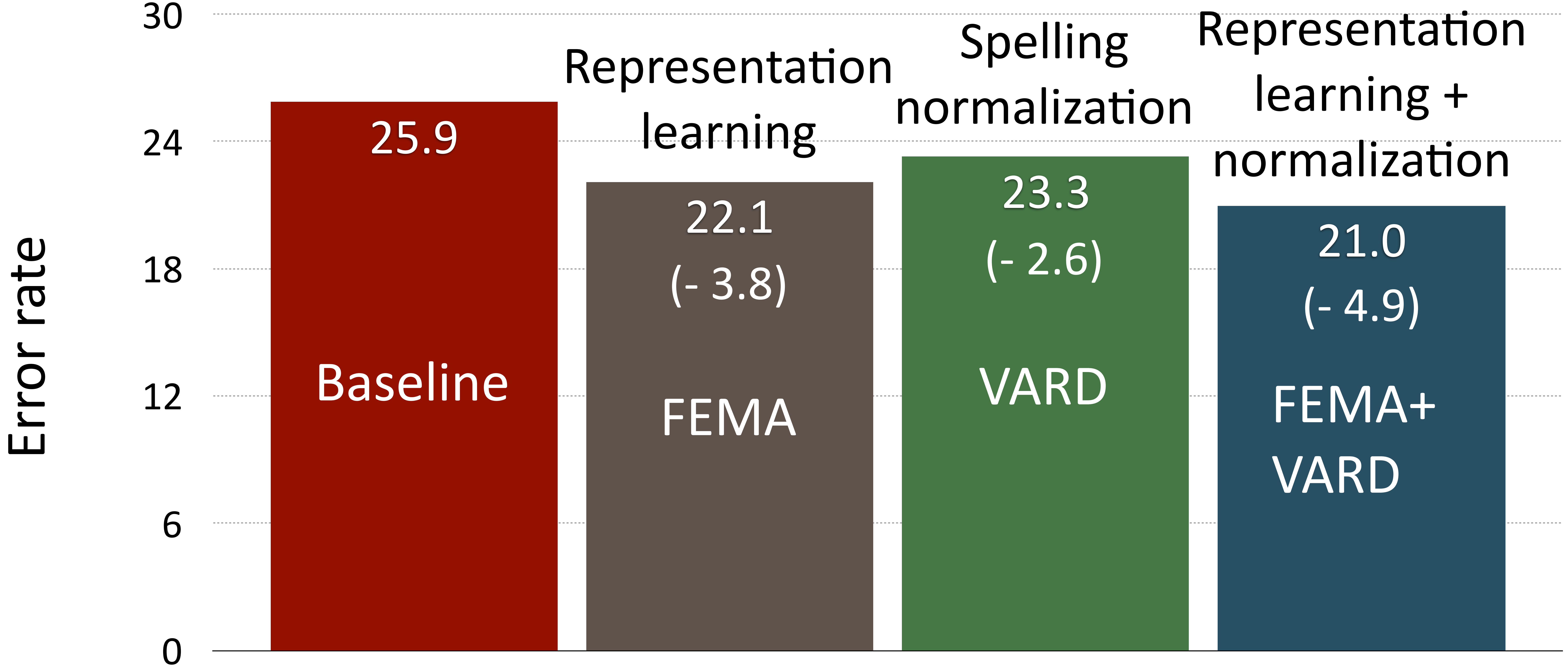
Normalization vs. Representation Learning



Normalization vs. Representation Learning



Normalization vs. Representation Learning



Error Analysis

- Annotation inconsistencies and tagset mismatches

token	annotations in PCHE	annotations in PTB
, (comma)	, (comma; 83.4%) . (period; 16.6%)	, (comma)

Error Analysis

- Annotation inconsistencies and tagset mismatches

token	annotations in PCHE	annotations in PTB
, (comma)	, (comma; 83.4%) . (period; 16.6%)	, (comma)
. (period)	, (comma; 12.3%) . (period; 87.7%)	. (period)

Error Analysis

- Annotation inconsistencies and tagset mismatches

token	annotations in PCHE	annotations in PTB
, (comma)	, (comma; 83.4%) . (period; 16.6%)	, (comma)
. (period)	, (comma; 12.3%) . (period; 87.7%)	. (period)
to	TO (54.6%) IN (44.3%)	TO

Error Analysis

- Annotation inconsistencies and tagset mismatches

token	annotations in PCHE	annotations in PTB
, (comma)	, (comma; 83.4%) . (period; 16.6%)	, (comma)
. (period)	, (comma; 12.3%) . (period; 87.7%)	. (period)
to	TO (54.6%) IN (44.3%)	TO
all/any/every	JJ (quantifier)	DT

Conclusions

Conclusions

- Feature embeddings outperform word embeddings by exploiting task-specific information in feature templates.

Conclusions

- Feature embeddings outperform word embeddings by exploiting task-specific information in feature templates.
- Representation learning and spelling normalization are complementary for improving tagging performance.

Conclusions

- Feature embeddings outperform word embeddings by exploiting task-specific information in feature templates.
- Representation learning and spelling normalization are complementary for improving tagging performance.
- Tagset mismatches make it hard to evaluate modern POS taggers for historical English.