Easy Victories and Uphill Battles in Coreference Resolution

Greg Durrett and Dan Klein

UC Berkeley
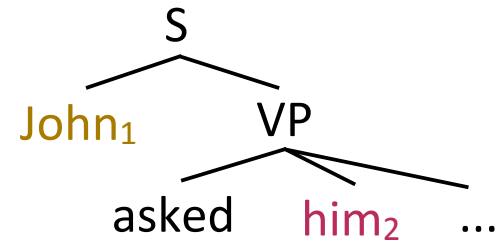
Aspects of Coreference

Aspects of Coreference
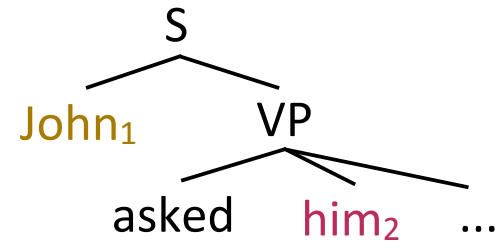
Binding theory (Chomsky, 1981)

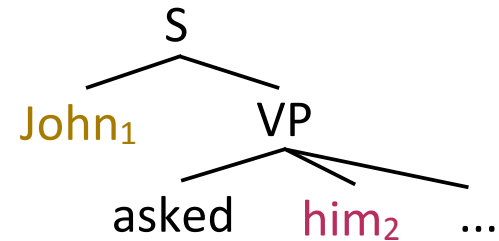# Aspects of Coreference

Binding theory (Chomsky, 1981)

S
├── John$_1$
└── VP
    ├── asked
    └── him$_2$ ...

Pronoun agreement (Hobbs, 1977)

MALE   FEMALE   FEMALE
John$_1$ talked to Jane$_2$. She$_2$ asked...

Aspects of Coreference

Binding theory (Chomsky, 1981)

S
John$_1$ VP
          asked   him$_2$ ...

Pronoun agreement (Hobbs, 1977)

MALE   FEMALE FEMALE
John$_1$ talked to Jane$_2$. She$_2$ asked...
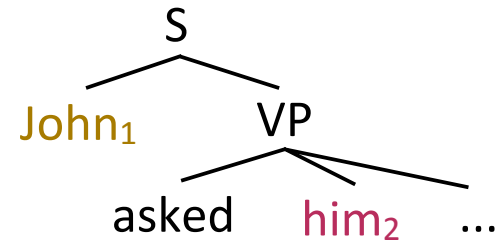
Centering (Grosz et al., 1995)

SUBJ.   OBJ. SUBJ.
John$_1$ talked to Bill$_2$. He$_1$ asked...

Aspects of Coreference

Binding theory (Chomsky, 1981)

$$S$$

John$_1$ — VP

asked   him$_2$ ...

Pronoun agreement (Hobbs, 1977)

MALE   FEMALE FEMALE
John$_1$ talked to Jane$_2$. She$_2$ asked...

Centering (Grosz et al., 1995)

SUBJ.   OBJ. SUBJ.
John$_1$ talked to Bill$_2$. He$_1$ asked...

Semantic compatibility

The president$_1$ ... The leader$_1$

Aspects of Coreference

Binding theory (Chomsky, 1981)

S
John$_1$ VP
asked   him$_2$ ...

Pronoun agreement (Hobbs, 1977)

MALE   FEMALE FEMALE
John$_1$ talked to Jane$_2$. She$_2$ asked...

Centering (Grosz et al., 1995)
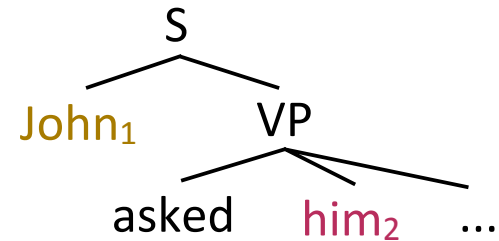
SUBJ.   OBJ. SUBJ.
John$_1$ talked to Bill$_2$. He$_1$ asked...

Semantic compatibility

The president$_1$ ... The leader$_1$

Definiteness

The president

A president

Definiteness: Classical

Soon et al. (2001), Ng et al. (2002), etc.

Definiteness: Classical

The president

Soon et al. (2001), Ng et al. (2002), etc.

Definiteness: Classical

The president

```
if (startsWith("the"))
    DEFINITE
else if (startsWith("a"))
    INDEFINITE
else
    NO_ART
```

Soon et al. (2001), Ng et al. (2002), etc.

Definiteness: Classical

The president

```
if (startsWith("the"))
    DEFINITE
else if (startsWith("a"))
    INDEFINITE
else
    NO_ART
```

DEFINITE

INDEFINITE   NO_ART

Soon et al. (2001), Ng et al. (2002), etc.

Definiteness: Data-Driven

The president

Definiteness: Data-Driven

The president

↓

The

Definiteness: Data-Driven

The president

↓

The   *these*

that

this   *U.S.*   *all*   *his*

a   *no*

an   *some*   *more*   *those*

John

its

their

Barack   *Israeli*   *new*

...

Approaches

Classical approach

Approaches

Classical approach

▶ Learning with heuristic features ⎰ Soon et al. (2001), inter alia

Approaches

Classical approach

▶ Learning with heuristic features ⎰ Soon et al. (2001), inter alia

▶ Rule-based ⎰ Haghighi and Klein (2009), Lee et al. (2011)

Approaches

Classical approach

▸ Learning with heuristic features { Soon et al. (2001), inter alia

▸ Rule-based { Haghighi and Klein (2009), Lee et al. (2011)

Mixed approach

▸ Add data-driven features on a few axes { Bengtson and Roth (2008), Rahman and Ng (2011), Björkelund and Nugues (2011)

Approaches

Classical approach

▸ Learning with heuristic features $\Big\{$ Soon et al. (2001), inter alia

▸ Rule-based $\left\{\begin{array}{l} \text{Haghighi and Klein (2009),} \\ \text{Lee et al. (2011)} \end{array}\right.$

Mixed approach

▸ Add data-driven features on a few axes $\left\{\begin{array}{l} \text{Bengtson and Roth (2008),} \\ \text{Rahman and Ng (2011),} \\ \text{Björkelund and Nugues (2011)} \end{array}\right.$

Data-driven approach (this work)

▸ >400,000 comprehensive, uniform features

Mention-Ranking Architecture

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture

[Voters]$_1$ agree when [they]$_1$ are given [a chance]$_2$ to decide if [they]$_1$ ...

Denis and Baldridge (2008), Durrett et al. (2013)
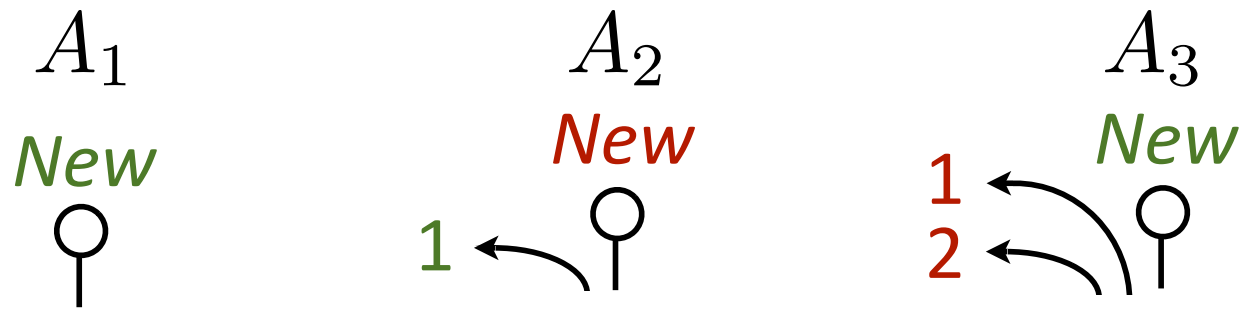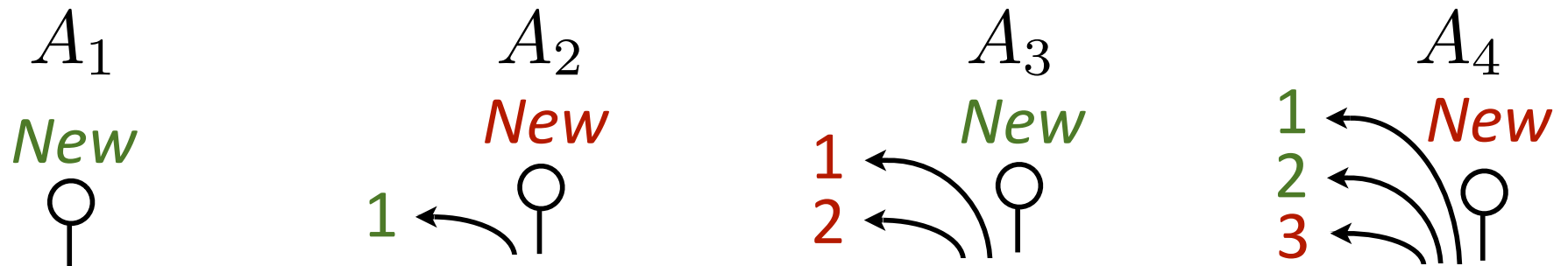
Mention-Ranking Architecture

A_1

New

○

[Voters]₁ agree when [they]₁ are given [a chance]₂ to decide if [they]₁ ...

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture
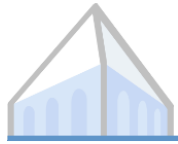
A_1

New

A_2

New

1 ←

[Voters]₁ agree when [they]₁ are given [a chance]₂ to decide if [they]₁ ...

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture

A_1

New

A_2
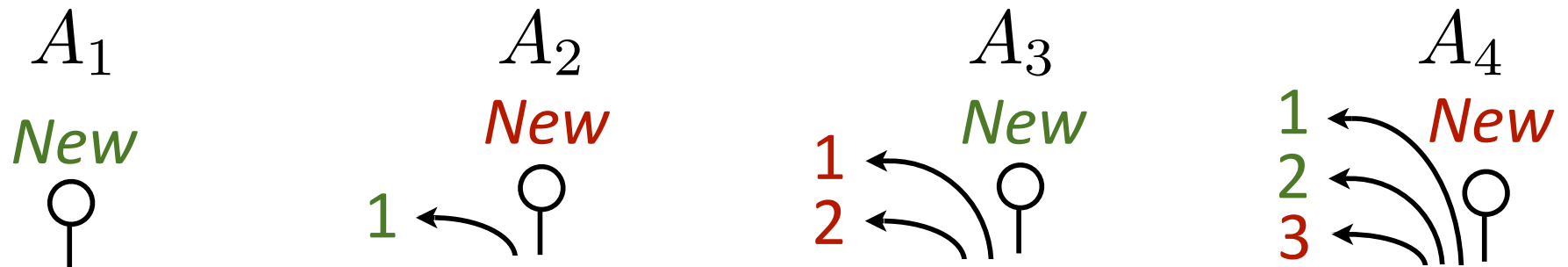
New

1

A_3

New

1
2

[Voters]₁ agree when [they]₁ are given [a chance]₂ to decide if [they]₁ …

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture

A_1 A_2 A_3 A_4

New   *New*   *New*   *New*

[Voters]₁ agree when [they]₁ are given [a chance]₂ to decide if [they]₁ …

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture

$$Pr(A_i = a|x) \propto \exp(w^\top f(i, a, x))$$

[Voters]$_1$ agree when [they]$_1$ are given [a chance]$_2$ to decide if [they]$_1$ ...

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture

$$Pr(A_i = a|x) \propto \exp(w^\top f(i, a, x))$$

[*Voters-they*]
[Nom-Pronoun]
...

A_1

A_2

A_3

A_4

New
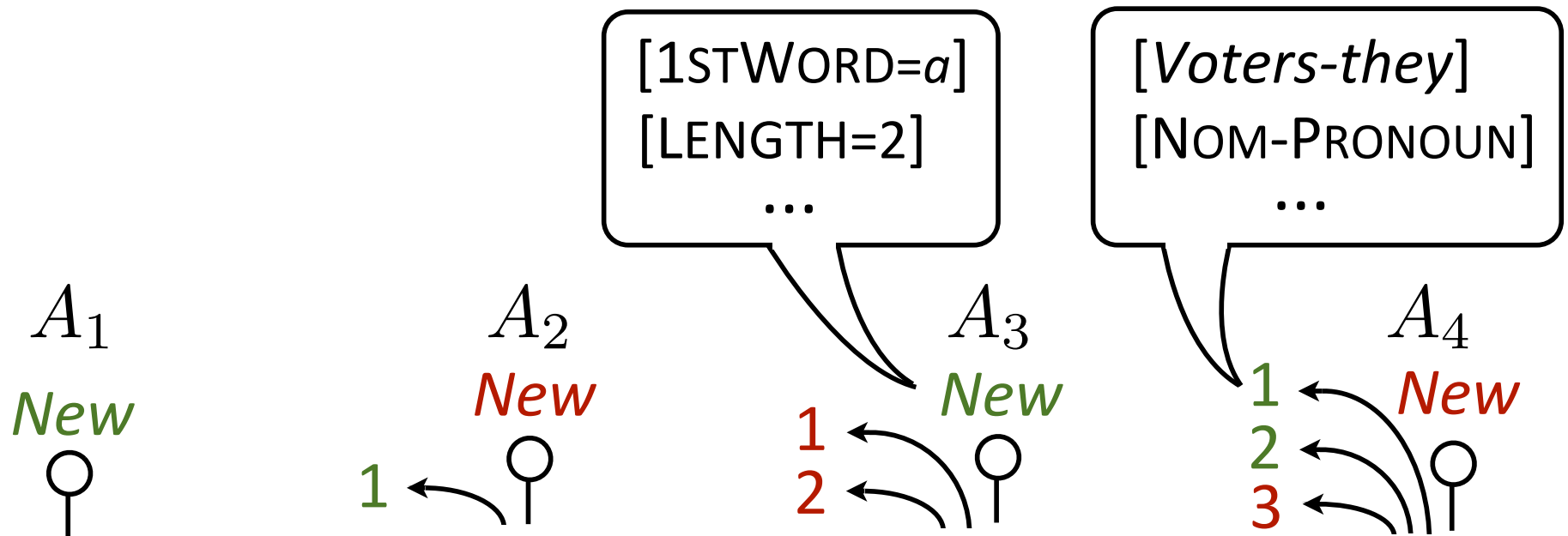
New

New

New

1

1
2

1
2
3

[Voters]₁ agree when [they]₁ are given [a chance]₂ to decide if [they]₁ ...

Denis and Baldridge (2008), Durrett et al. (2013)

Mention-Ranking Architecture

$$Pr(A_i = a|x) \propto \exp(w^\top f(i, a, x))$$

[1STWORD=*a*]
[LENGTH=2]
...

[*Voters-they*]
[NOM-PRONOUN]
...

A_1

New

A_2

New

1

A_3

New

1
2

A_4

1
2
3

New

[Voters]₁ agree when [they]₁ are given [a chance]₂ to decide if [they]₁ ...

Denis and Baldridge (2008), Durrett et al. (2013)

BASIC Features

BASIC Features

New

[*Voters*]₁ *agree when* [*they*]₁ *...*

BASIC Features

Type

New

[*Voters*]$_1$ *agree when* [*they*]$_1$ *...*
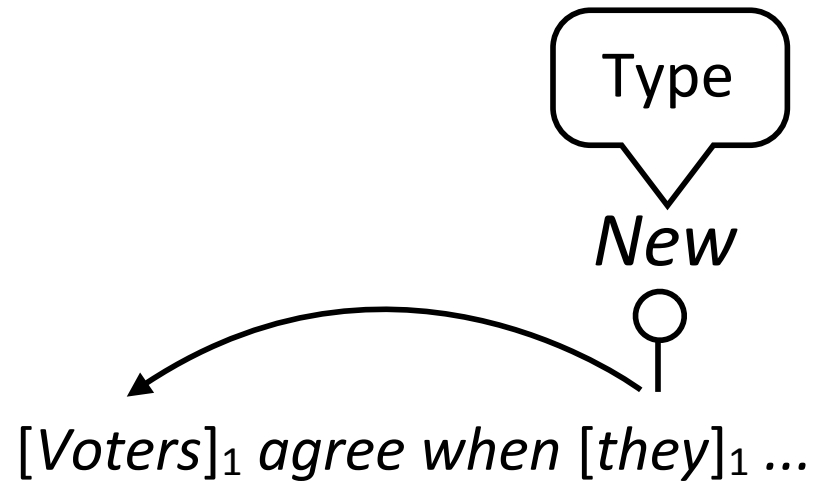
BASIC Features

Types
Sentence distance (+ types)
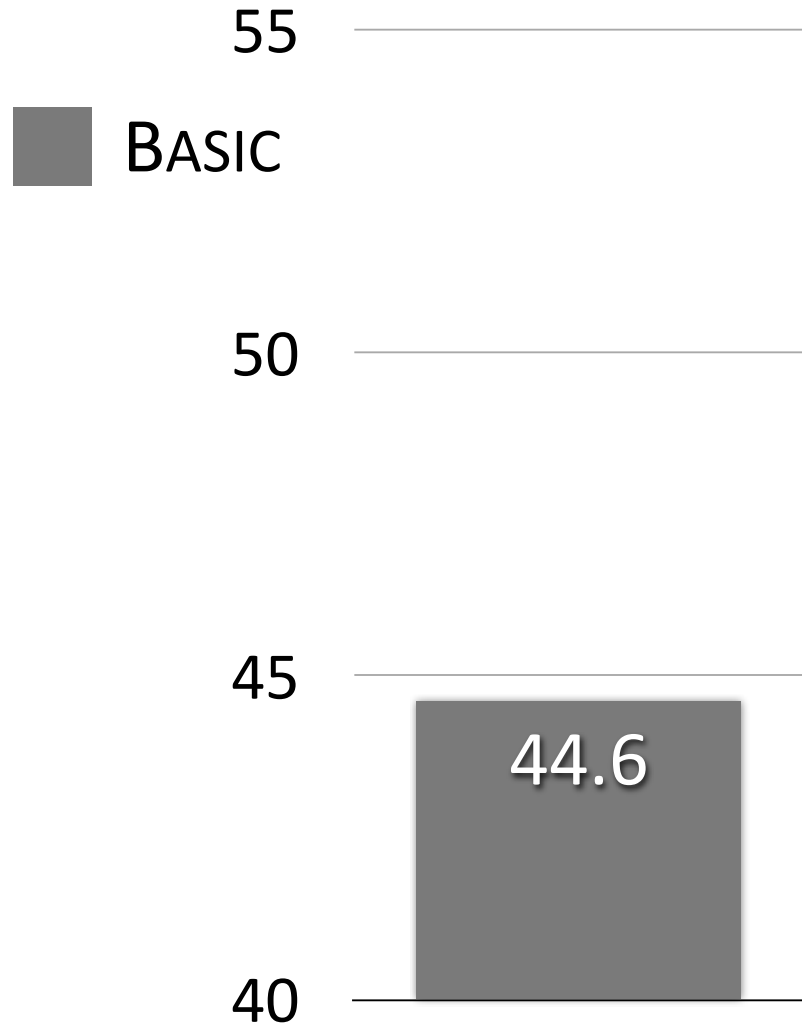Mention distance (+ types)
Head match

Type

New

[*Voters*]$_1$ *agree when* [*they*]$_1$ ...

BASIC Features

(Dev set CoNLL-F1, predicted mentions)

Definiteness

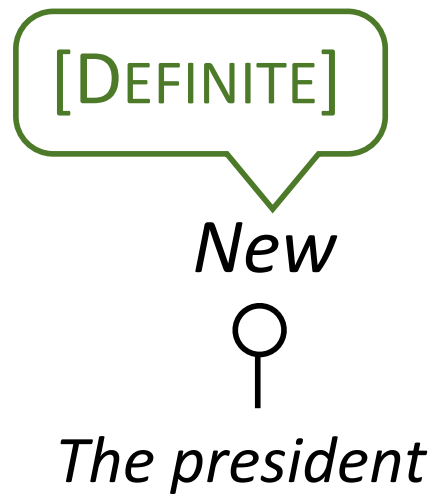Soon et al. (2001), Ng et al. (2002), etc.

Definiteness

New

○

The president

Soon et al. (2001), Ng et al. (2002), etc.

Definiteness

[DEFINITE]

New

The president

Soon et al. (2001), Ng et al. (2002), etc.

Definiteness

[1stWord=*The*]

New

○

The president

Definiteness

55

50

45

40

■ BASIC

44.6

(Dev set CoNLL-F1, predicted mentions)

Definiteness

(Dev set CoNLL-F1, predicted mentions)

Definiteness

BASIC

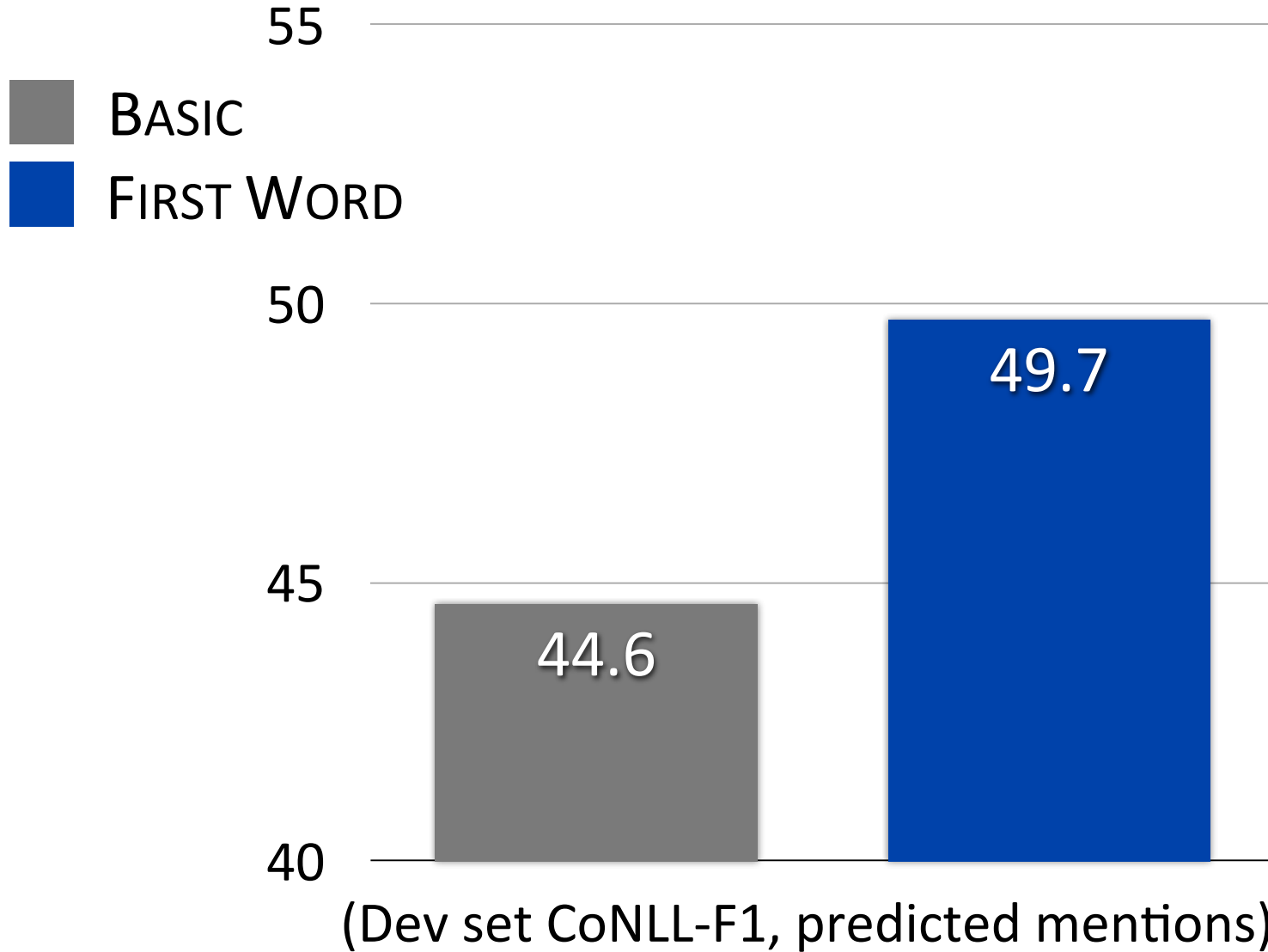FIRST WORD

49.7

44.6

(Dev set CoNLL-F1, predicted mentions)

Definiteness

DEFINITENESS | 47.0

Definiteness

DEFINITENESS **47.0**

FIRST WORD: "the, a, an" **47.0**
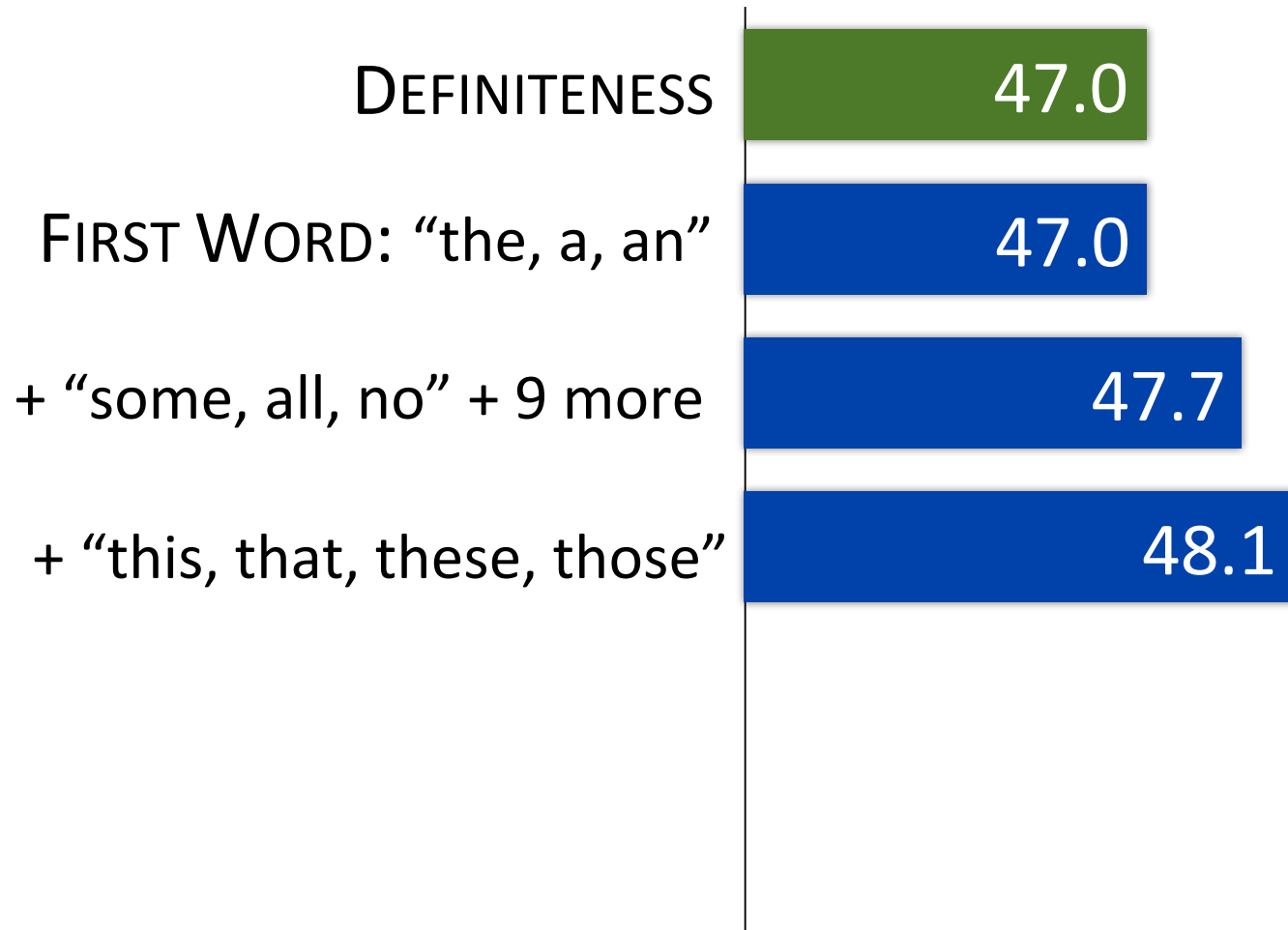
Definiteness

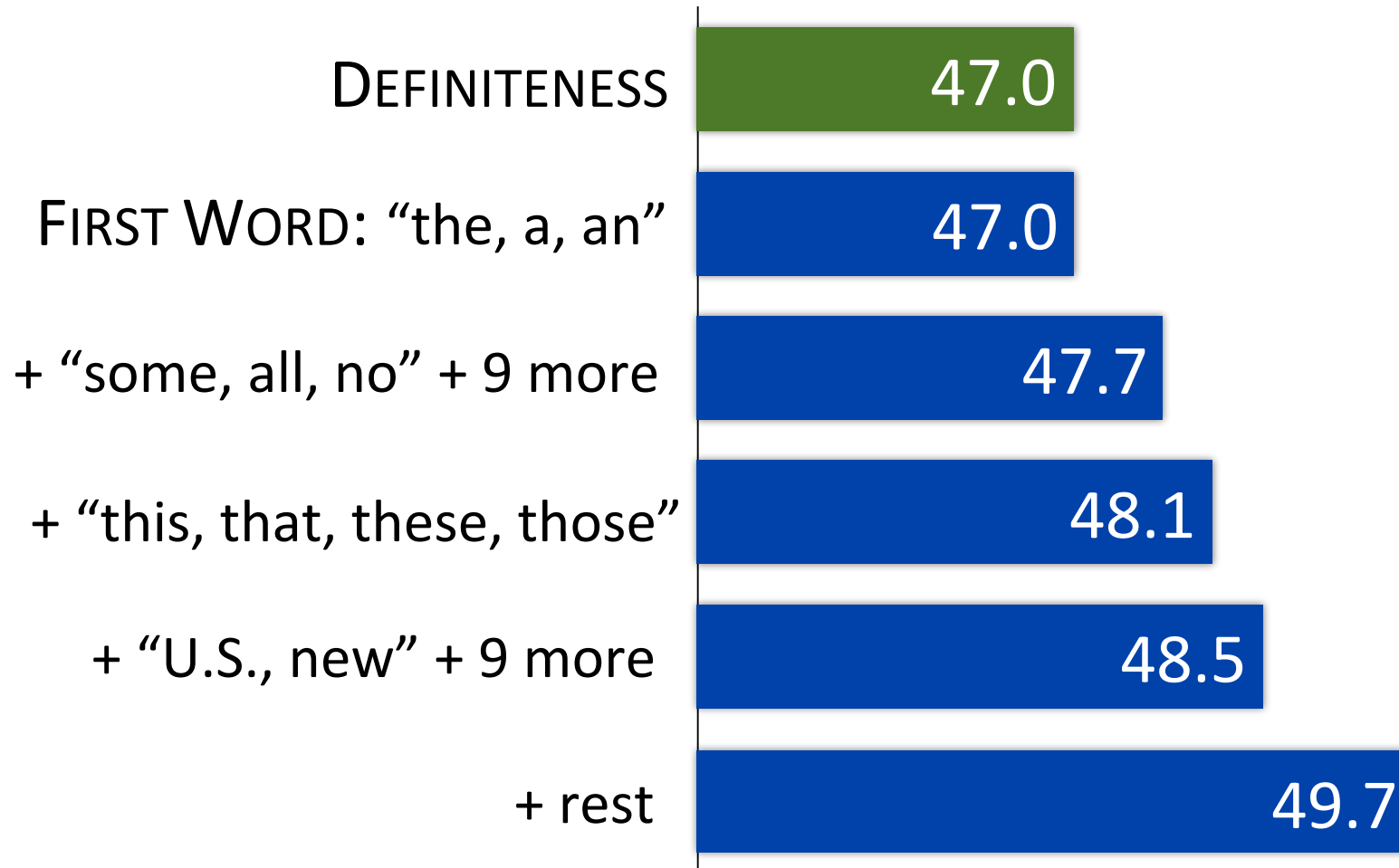|	
DEFINITENESS	47.0
FIRST WORD: "the, a, an"	47.0
+ "some, all, no" + 9 more	47.7

Definiteness

DEFINITENESS	47.0
FIRST WORD: "the, a, an"	47.0
+ "some, all, no" + 9 more	47.7
+ "this, that, these, those"	48.1
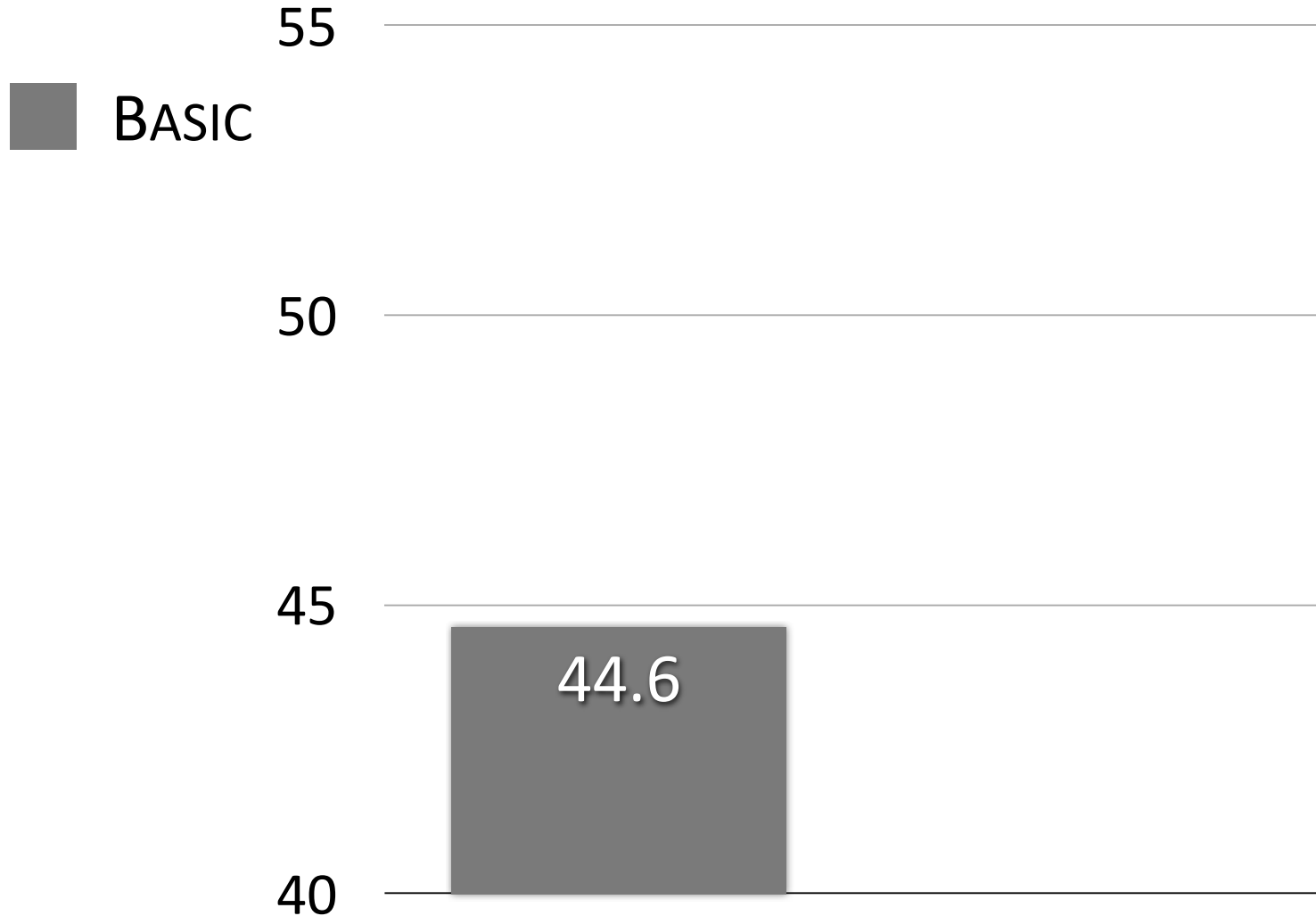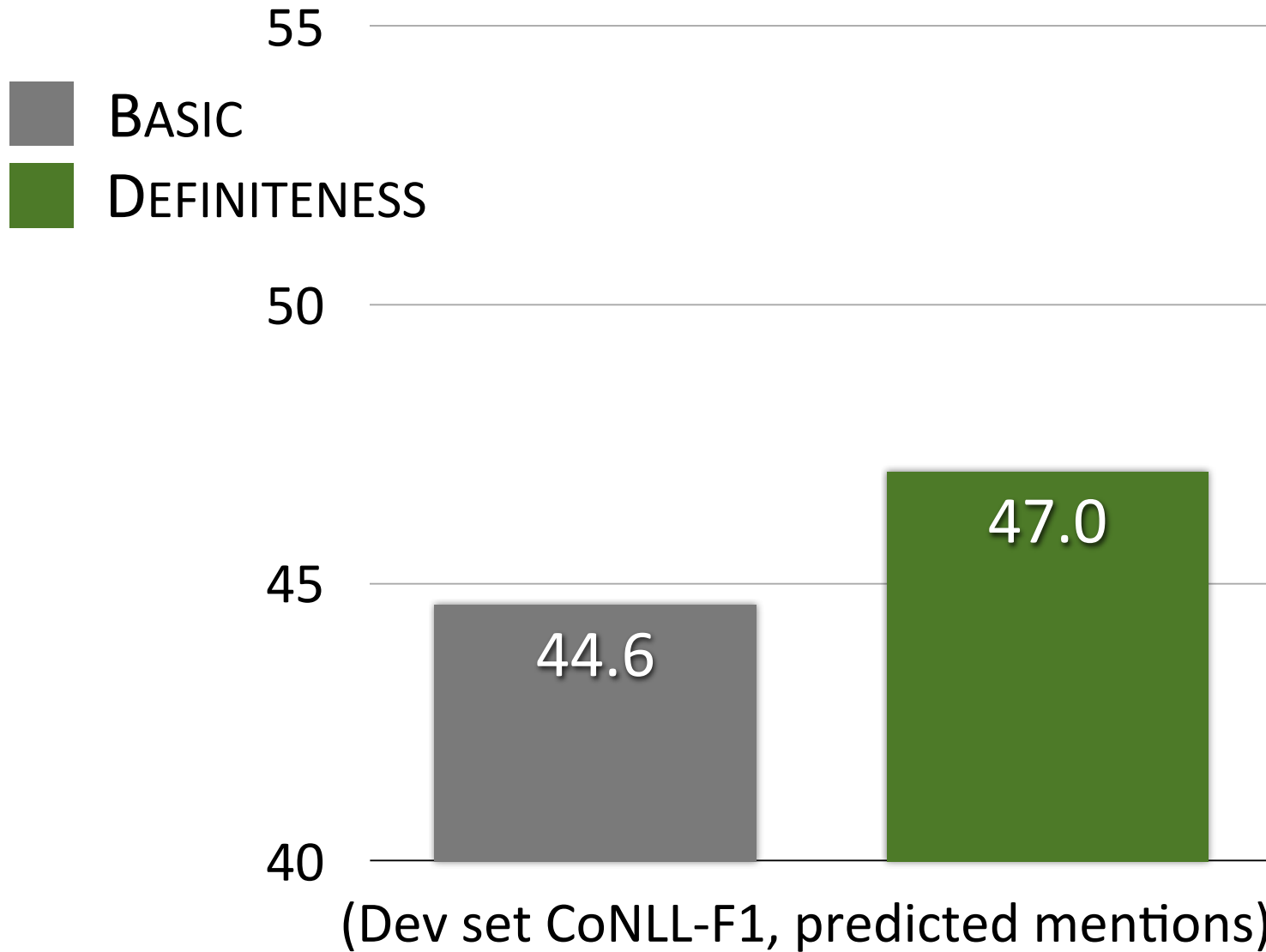
Definiteness

Definiteness

DEFINITENESS	47.0
FIRST WORD: "the, a, an"	47.0
+ "some, all, no" + 9 more	47.7
+ "this, that, these, those"	48.1
+ "U.S., new" + 9 more	48.5
+ rest	49.7
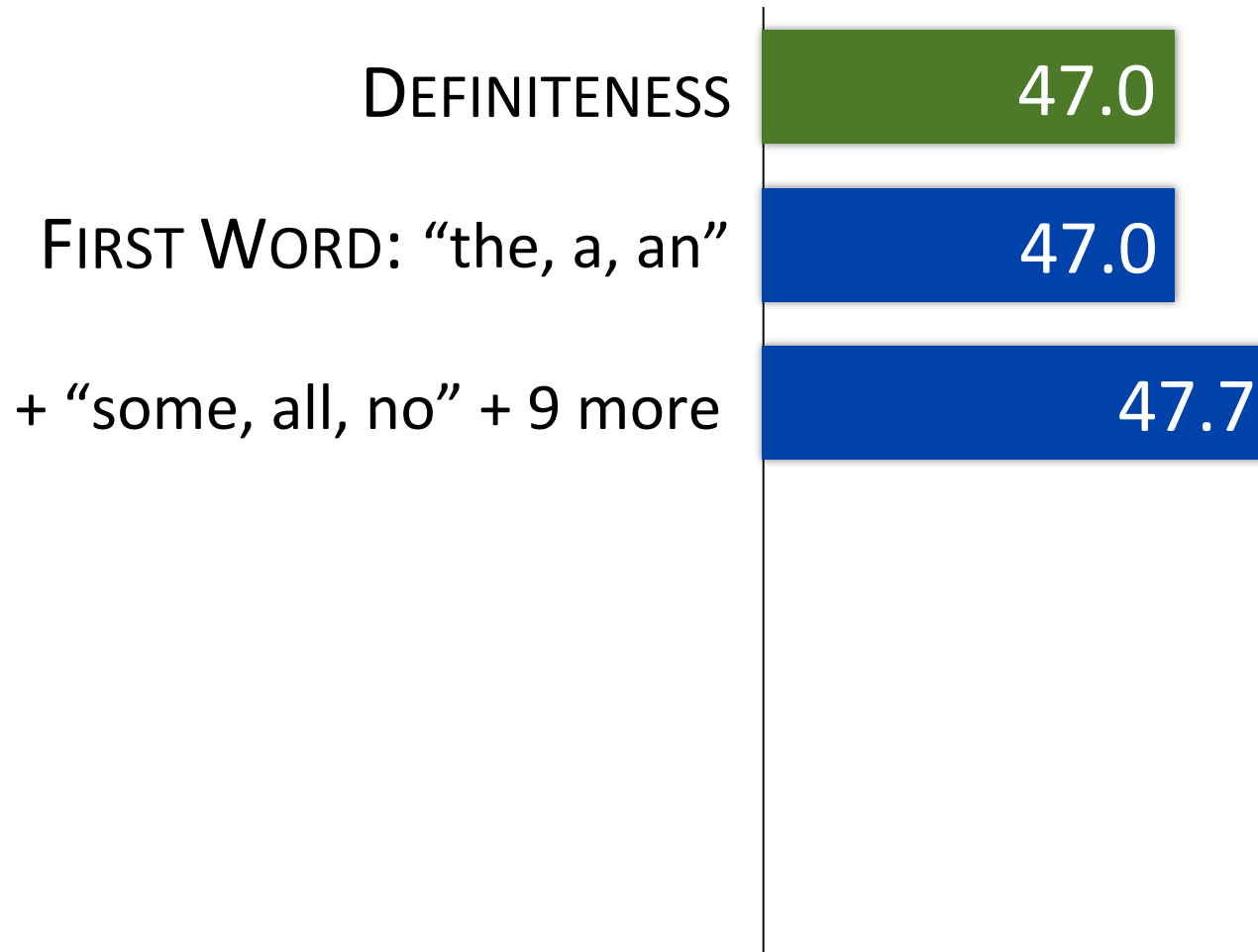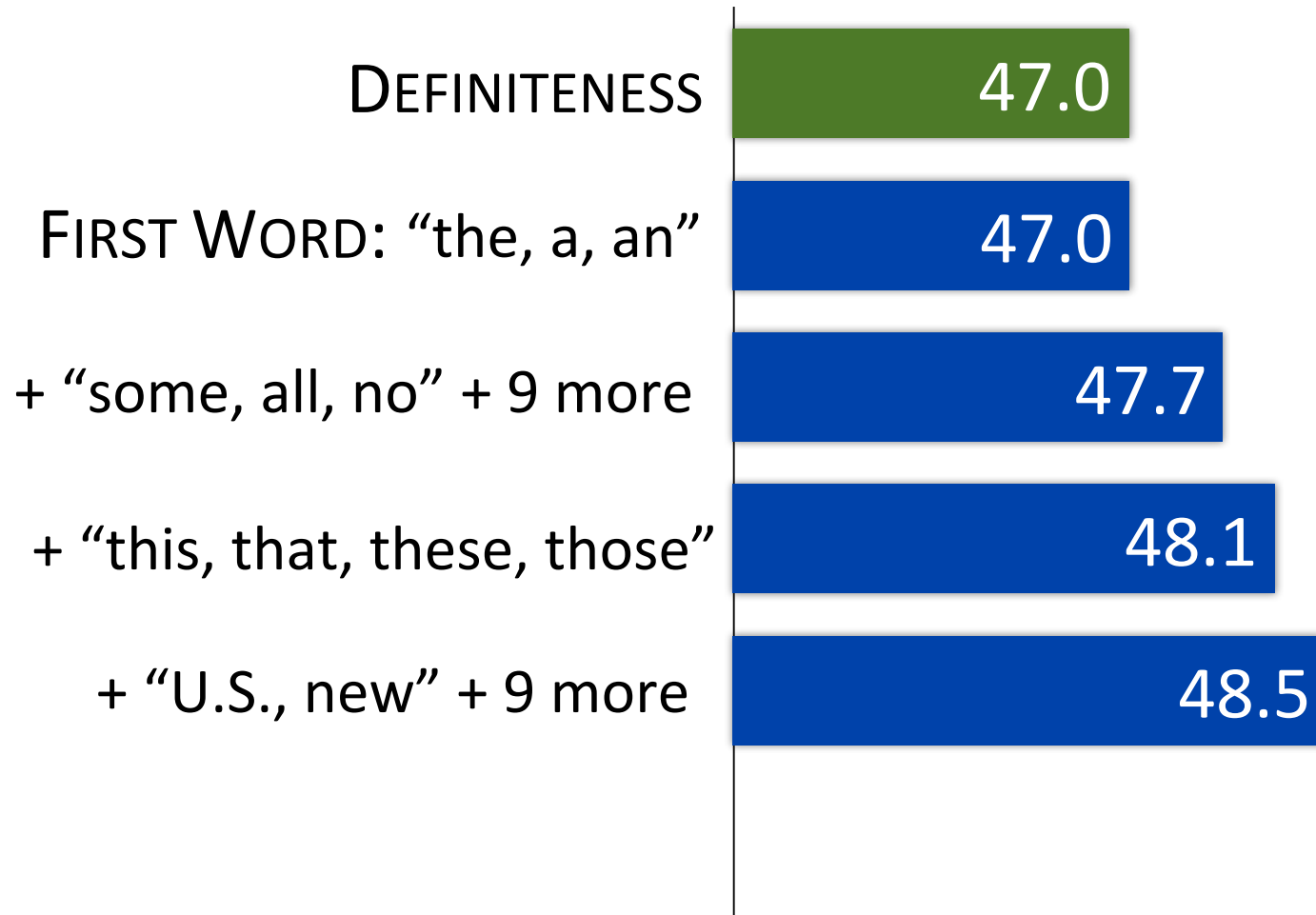
Centering

Grosz al. (1995)

Centering

[*Barack Obama*]$_1$ *met with* [*David Cameron*]$_2$ *.* [*He*]$_1$ *said ...*

Grosz al. (1995)

Centering

[*Barack Obama*]$_1$ *met with* [*David Cameron*]$_2$ *.* [*He*]$_1$ *said ...*

Grosz al. (1995)

Centering

SUBJECT–SUBJECT

OBJECT–SUBJECT

[*Barack Obama*]$_1$ *met with* [*David Cameron*]$_2$ *.* [*He*]$_1$ *said ...*

Grosz al. (1995)

Centering

SUBJECT–SUBJECT

[*Barack Obama*]$_1$ *met with* [*David Cameron*]$_2$. [*He*]$_1$ *said …*

Grosz al. (1995)

Centering

[*Barack Obama*]$_1$ *met with* [*David Cameron*]$_2$ *.* [*He*]$_1$ *said ...*

Haghighi and Klein (2010)

Centering

[*Barack Obama*]₁ *met with* [*David Cameron*]₂ *. [He]₁ said ...*

SUBJECT   OBJECT   SUBJECT

Haghighi and Klein (2010)

Centering

[OBJECT–SUBJECT]

[*Barack Obama*]₁ *met with* [*David Cameron*]₂ . [*He*]₁ *said …*

SUBJECT   OBJECT   SUBJECT

Haghighi and Klein (2010)

Centering

[*Barack Obama*]$_1$ *met with* [*David Cameron*]$_2$. [*He*]$_1$ *said* ...

Centering

[*Barack Obama*]₁ *met with* [*David Cameron*]₂ *.* [*He*]₁ *said …*
with [X] *.*

Centering

[*Barack Obama*]₁ *met with* [*David Cameron*]₂ . [*He*]₁ *said ...*

with [X] .   . [X] *said*

Centering

[*with* X − . Y]
[*with* X − Y *said*]
...

[*Barack Obama*]₁ *met with* [*David Cameron*]₂ . [*He*]₁ *said* ...

with [X] .   . [X] *said*

Centering

So Far

50.9
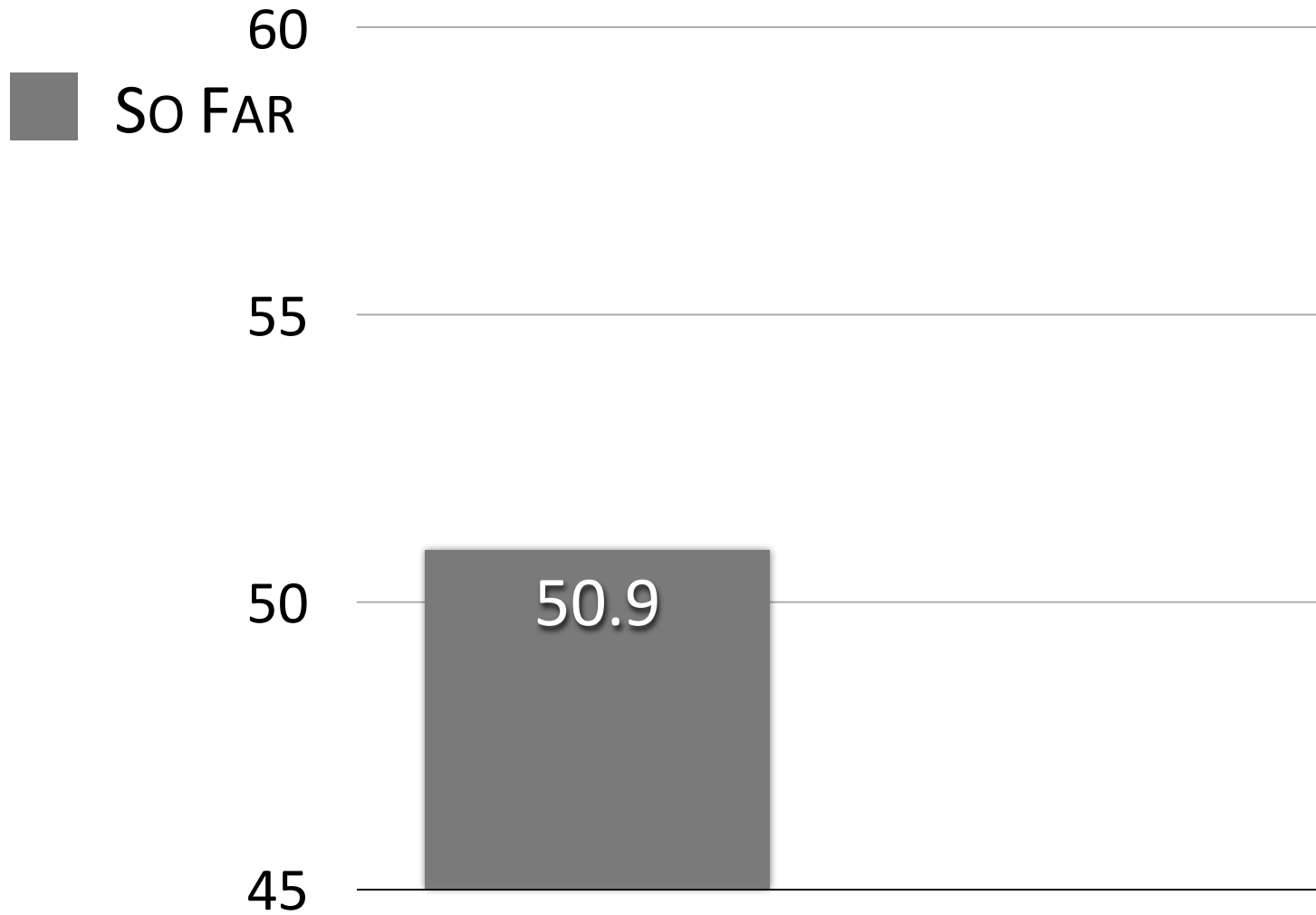
(Dev set CoNLL-F1, predicted mentions)

Centering

(Dev set CoNLL-F1, predicted mentions)

Legend:
- So Far: 50.9
- Position: 51.9

Centering
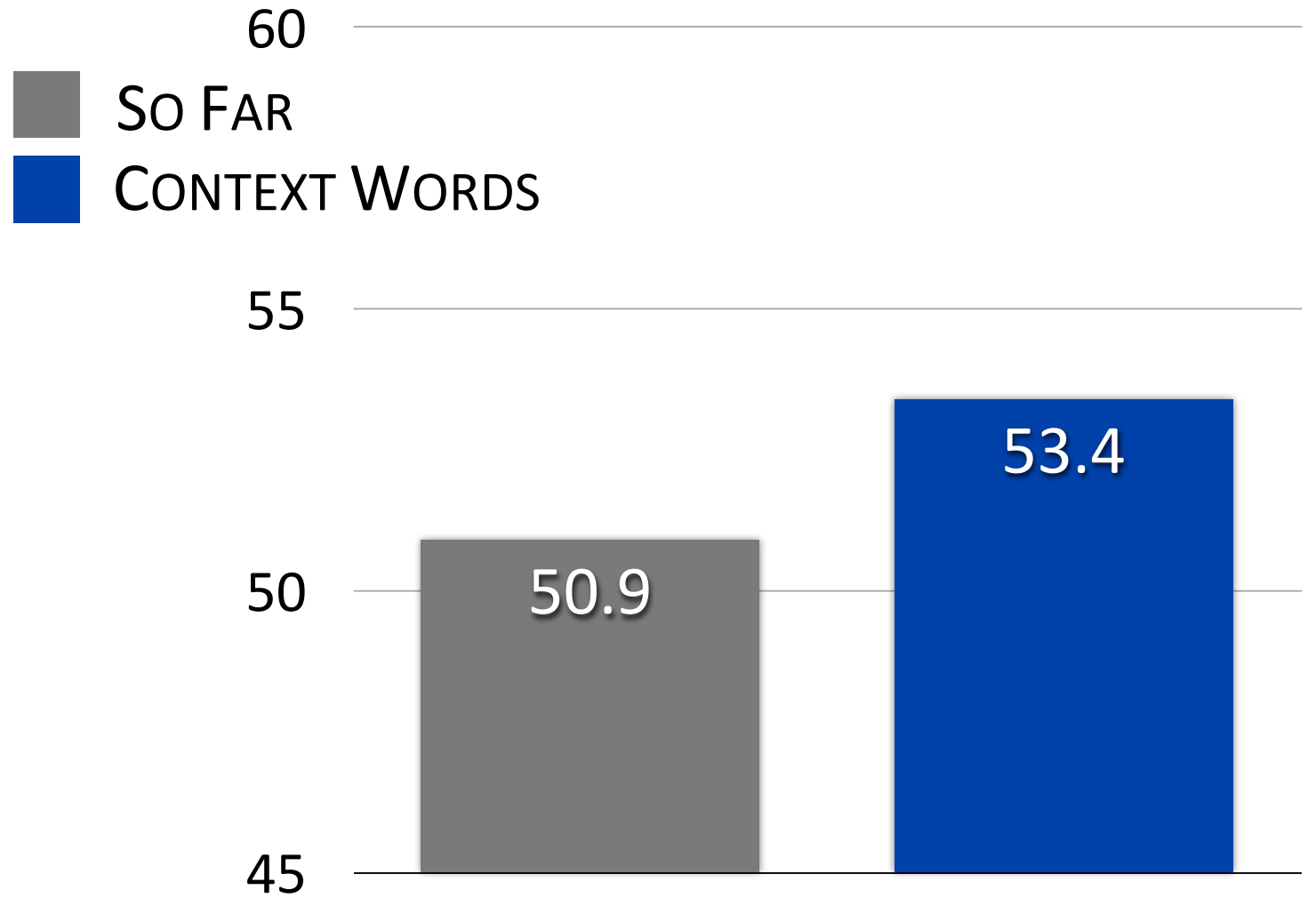
(Dev set CoNLL-F1, predicted mentions)
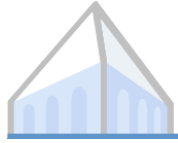
Pronoun Agreement

Bergsma and Lin (2006)
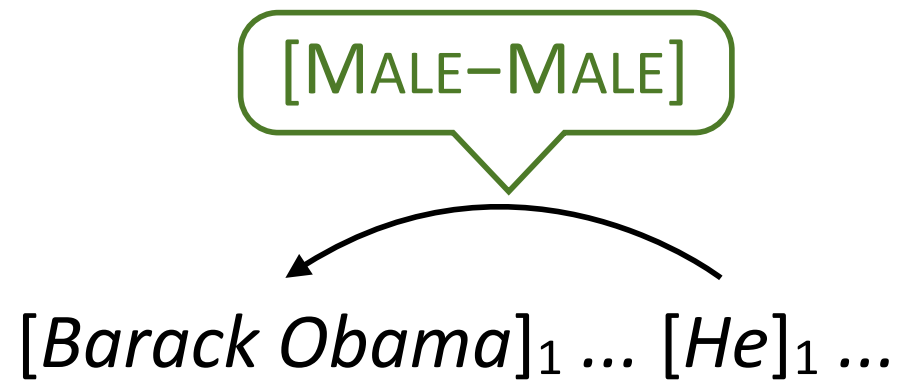
Pronoun Agreement

[*Barack Obama*]$_1$ ... [*He*]$_1$ ...

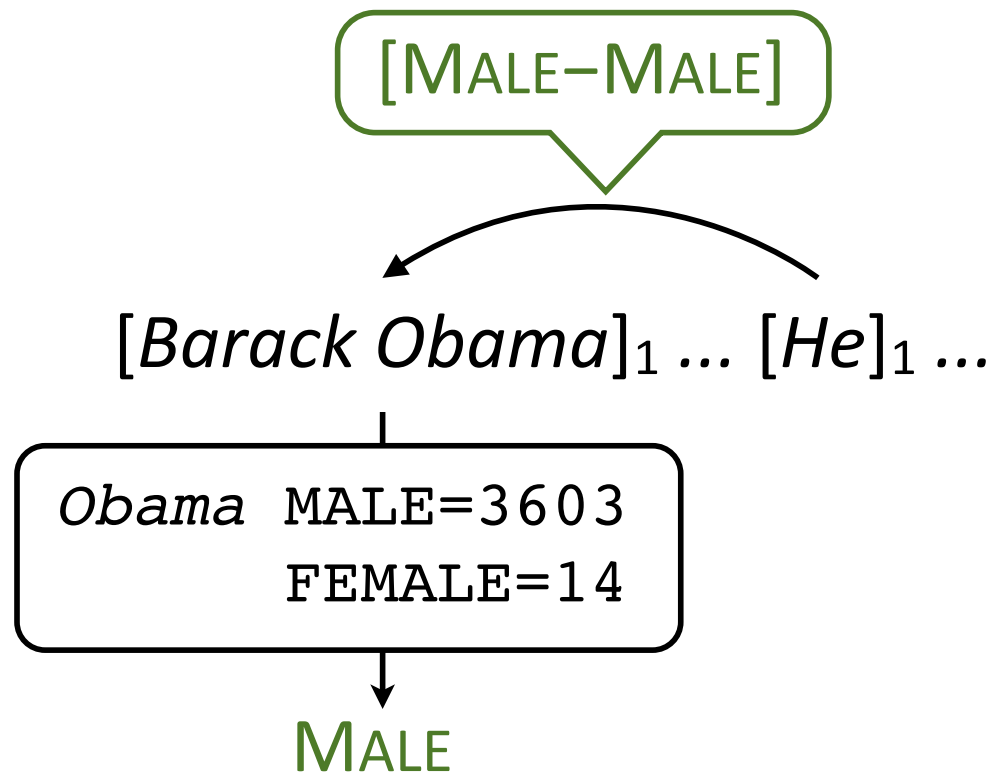Bergsma and Lin (2006)

Pronoun Agreement

[MALE–MALE]

[*Barack Obama*]$_1$ ... [*He*]$_1$ ...

Bergsma and Lin (2006)

Pronoun Agreement

[MALE–MALE]

[*Barack Obama*]₁ ... [*He*]₁ ...

Obama MALE=3603
FEMALE=14

MALE

Bergsma and Lin (2006)

Pronoun Agreement

[MALE–MALE]

[*Barack Obama*]₁ ... [*He*]₁ ...

Obama MALE=3603
FEMALE=14

MALE   MALE

Bergsma and Lin (2006)
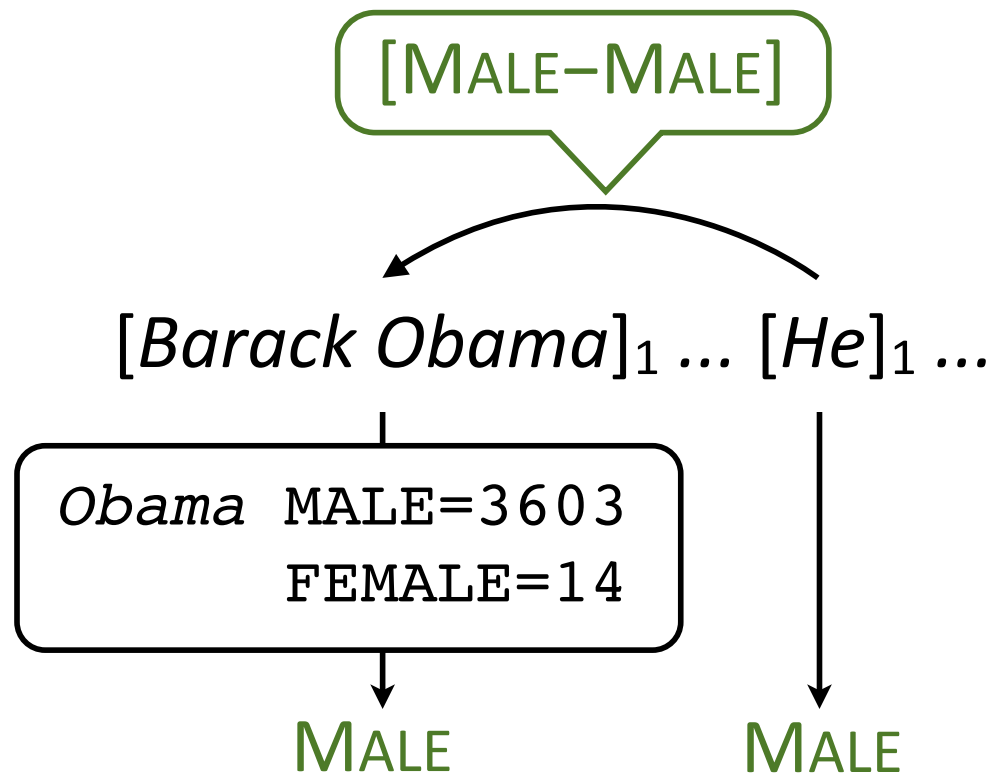
Pronoun Agreement

$[Barack\ Obama]_1 ... [He]_1 ...$

Pronoun Agreement

[*Obama–He*]

[*Barack Obama*]$_1$ ... [*He*]$_1$ ...
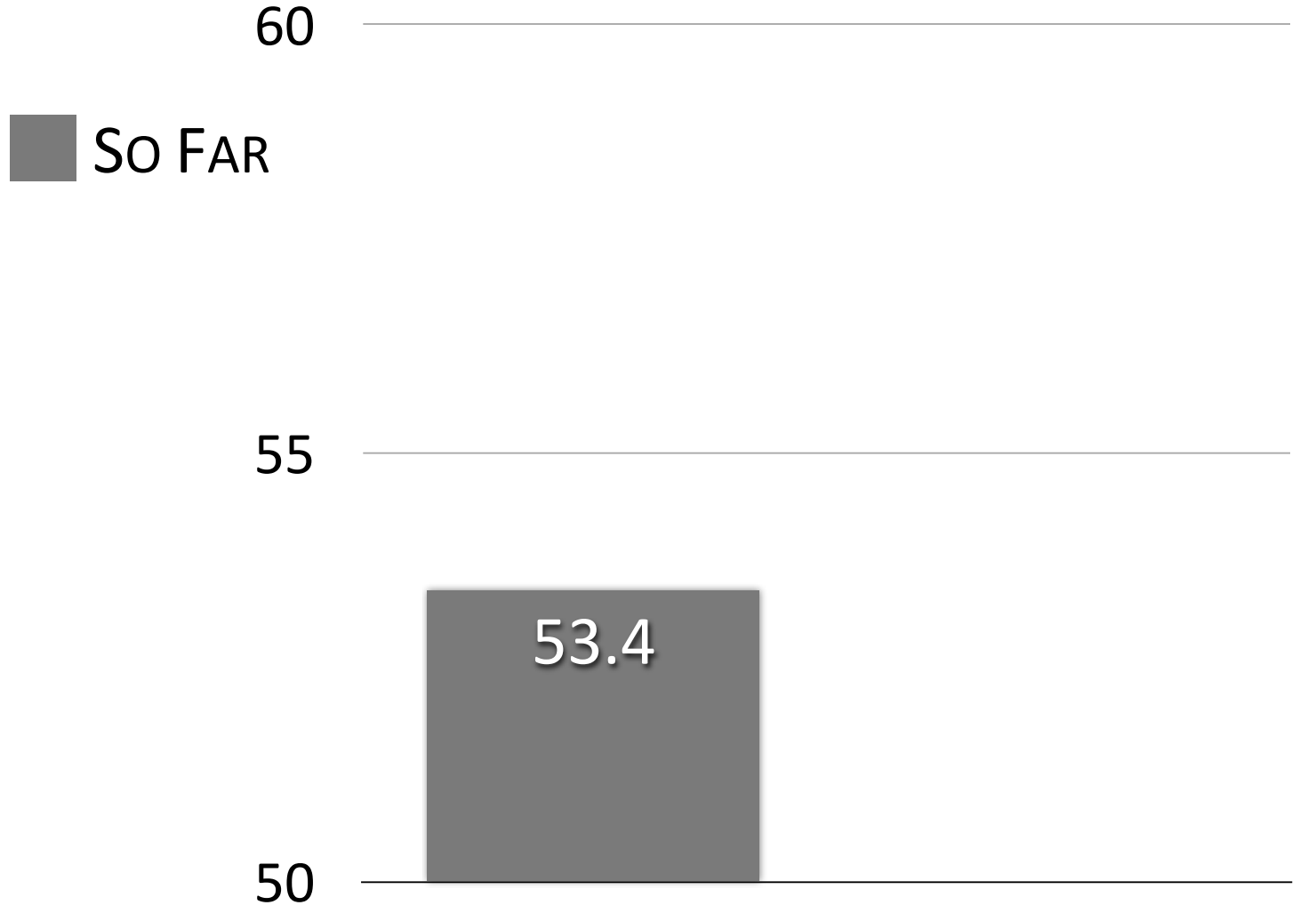
Pronoun Agreement

So Far

53.4

(Dev set CoNLL-F1, predicted mentions)

Pronoun Agreement

So Far
Gender/number

60

57.9

55

53.4

50

(Dev set CoNLL-F1, predicted mentions)

Pronoun Agreement

So Far
Indicators

60

58.2

55

53.4

50
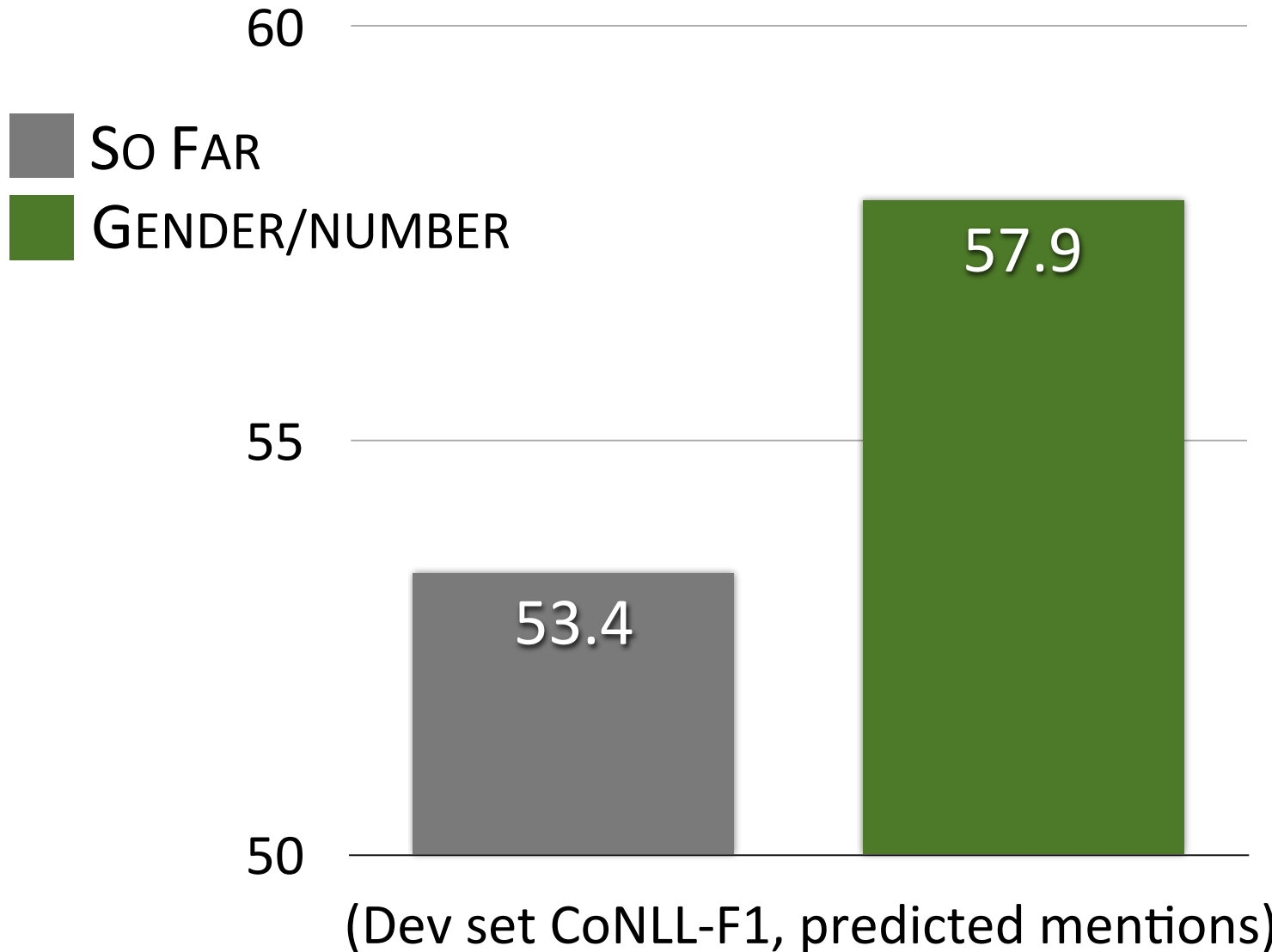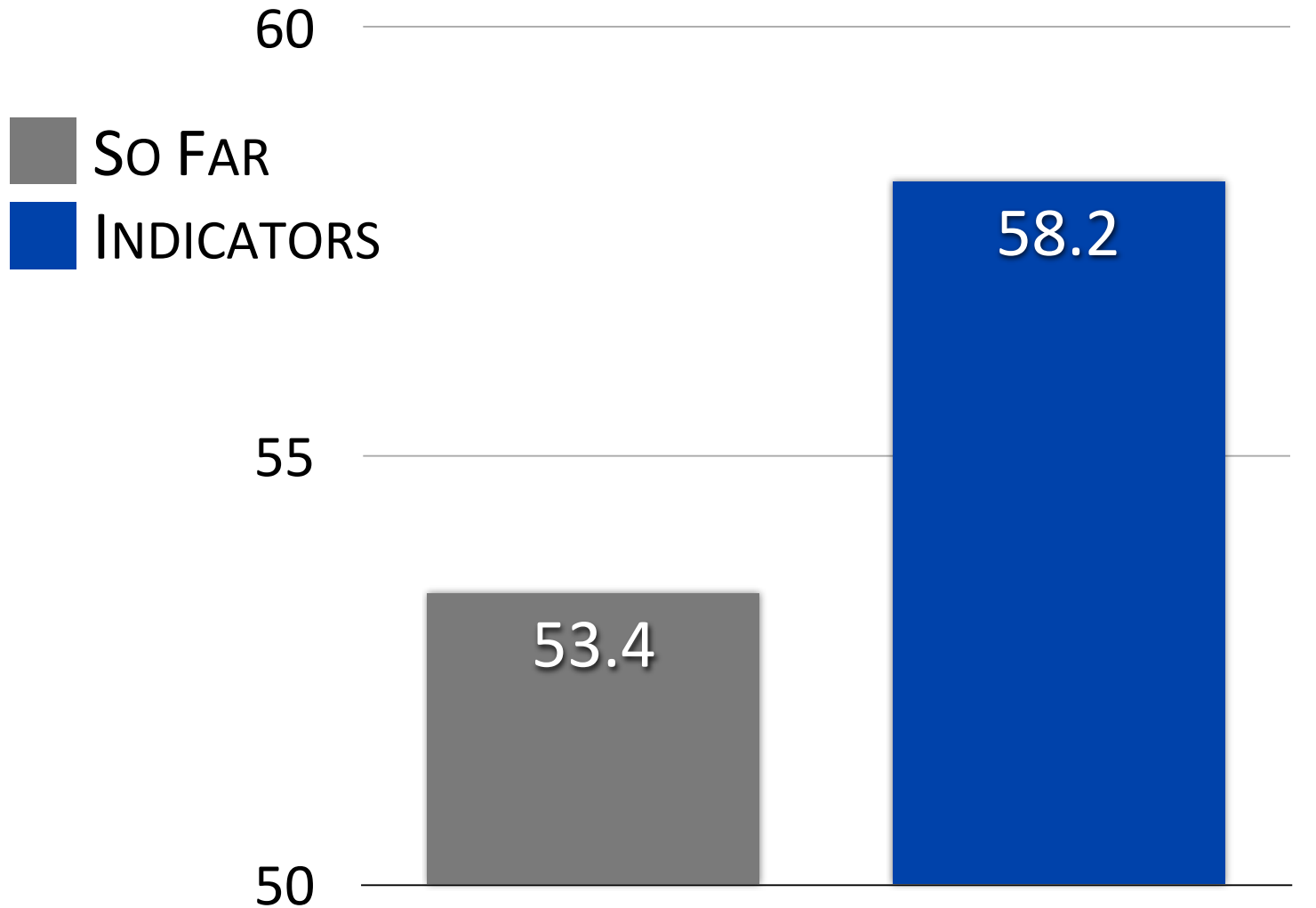
(Dev set CoNLL-F1, predicted mentions)

SURFACE Information

▸ Features conjoin surface-level mention attributes

SURFACE Information

▶ Features conjoin surface-level mention attributes

<s>[*President Barack Obama*] *signed the bill … Afterwards* [*he*] *said …*

SURFACE Information

▶ Features conjoin surface-level mention attributes

<s>[*President* Barack *Obama*] *signed the bill ... Afterwards* [*he*] said ...

SURFACE Information

▶ Features conjoin surface-level mention attributes

<s>[*President* Barack **Obama**] *signed* the bill ... *Afterwards* [*he*] said ...

SURFACE Information

▸ Features conjoin surface-level mention attributes

<s>[*President* Barack *Obama*] *signed* the bill ... *Afterwards* [*he*] said ...

LENGTH = 3   LENGTH = 1

SURFACE Information

▶ Features conjoin surface-level mention attributes

<s>[*President* Barack **Obama**] *signed* the bill ... *Afterwards* [*he*] said ...

LENGTH = 3   LENGTH = 1

PROPER   PRONOUN

SURFACE Information

▶ Features conjoin surface-level mention attributes

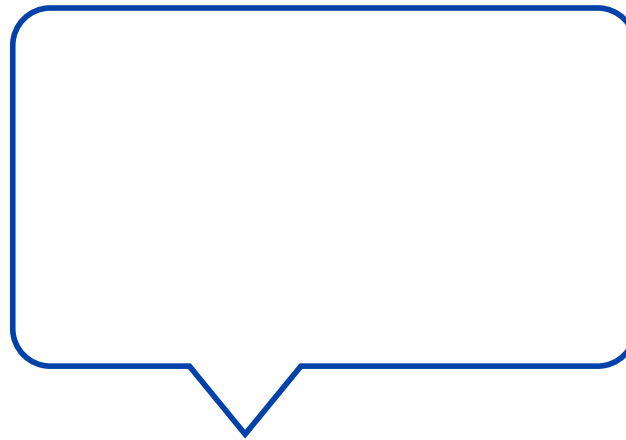<s> [*President* Barack **Obama**] *signed* the bill ... *Afterwards* [*he*] said ...

LENGTH = 3   LENGTH = 1

PROPER   PRONOUN

SURFACE Information

▶ Features conjoin surface-level mention attributes

> Mention distance
> Sentence distance

<s> [*President* Barack **Obama**] *signed* the bill ... *Afterwards* [*he*] said ...

LENGTH = 3   LENGTH = 1

PROPER   PRONOUN

SURFACE Information

▸ Features conjoin surface-level mention attributes

Mention distance
Sentence distance
Head match
Exact match

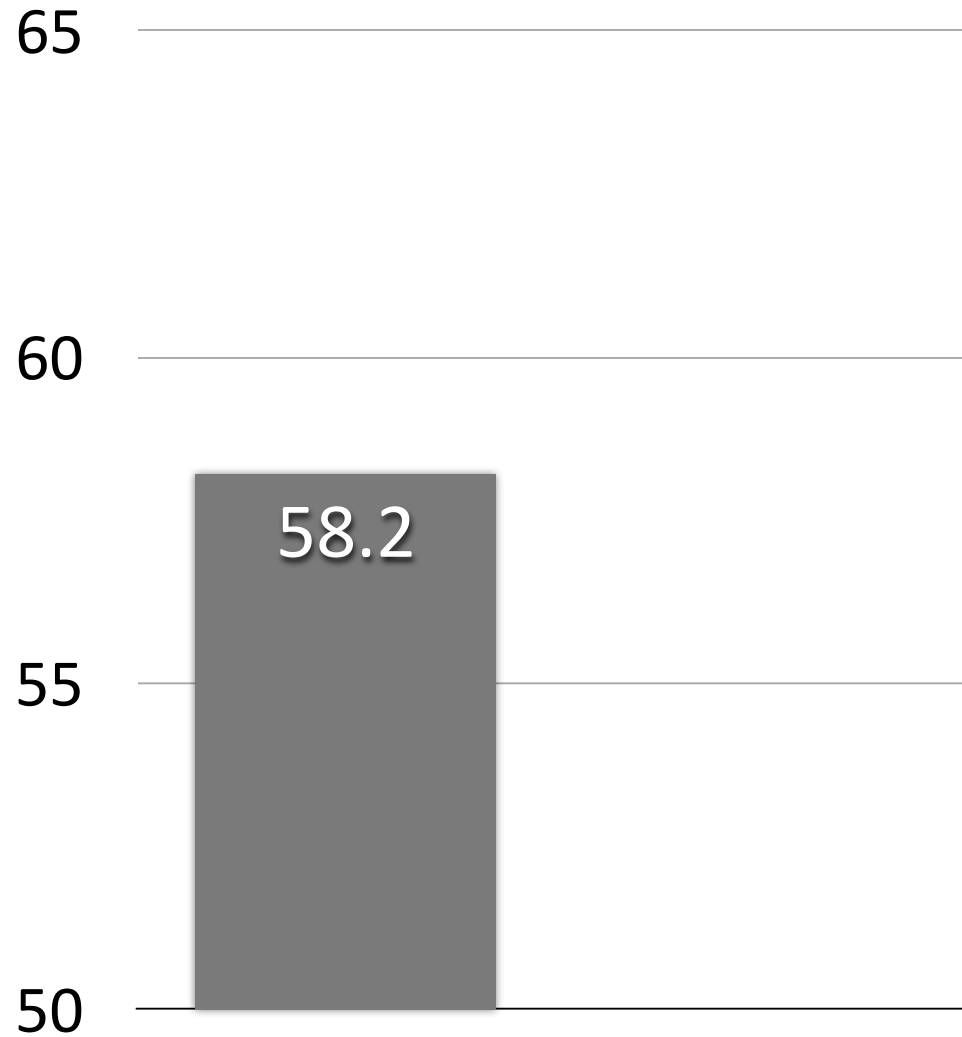<s> [*President* Barack *Obama*] *signed* the bill ... *Afterwards* [*he*] said ...

LENGTH = 3   LENGTH = 1
PROPER   PRONOUN

Surface Accuracy
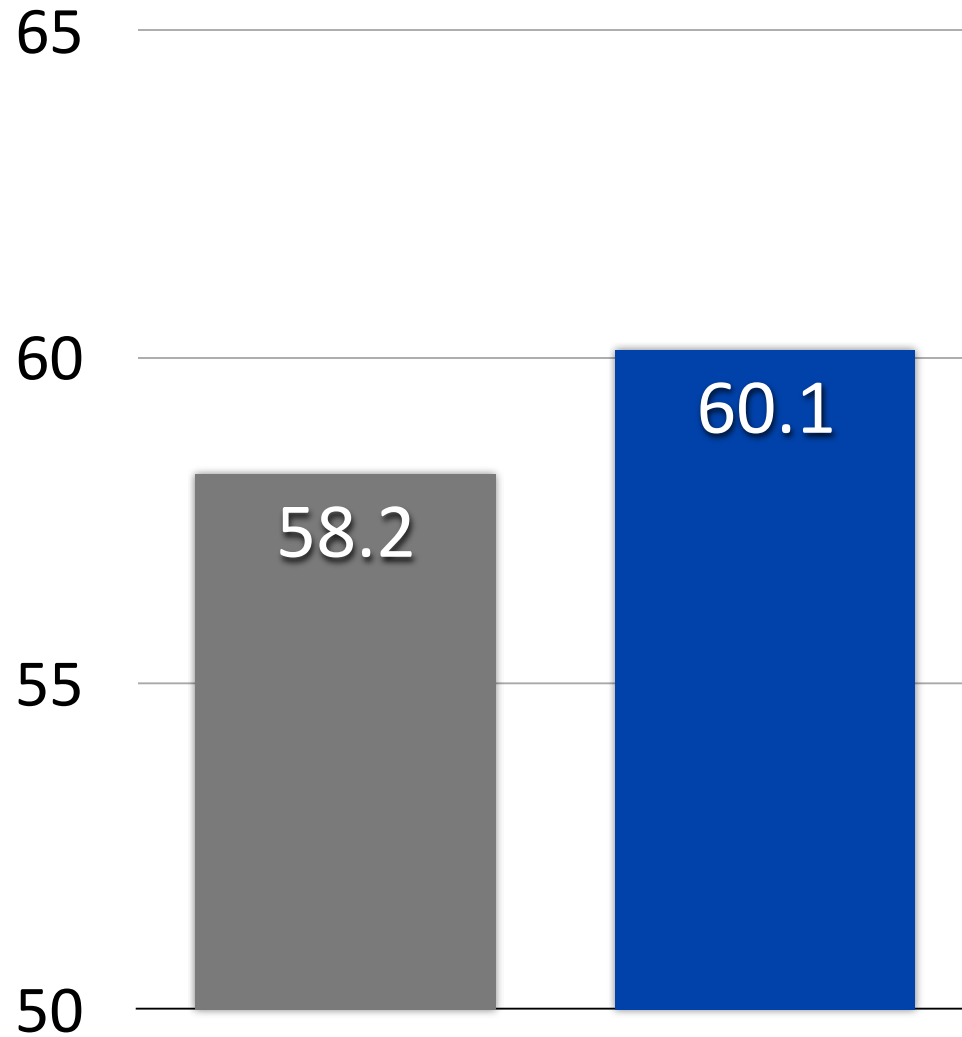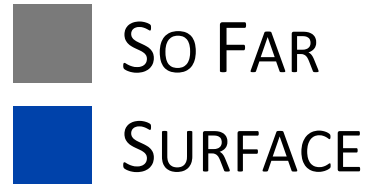

(Dev set CoNLL-F1, predicted mentions)

SURFACE Accuracy

(Dev set CoNLL-F1, predicted mentions)

Error Analysis

Error Analysis

Anaphoric pronouns

Obama ⟵ *he*

Error Analysis

Anaphoric pronouns

Obama ⟵ *he*

72.0%

Error Analysis

Anaphoric pronouns

Obama ⟵ *he*

Referring: head match

the U.S. president ⟵ *president*

72.0%

Error Analysis

Anaphoric pronouns

Obama ← *he*

Referring: head match

the U.S. president ← *president*

72.0%

82.7%

Error Analysis

Anaphoric pronouns

Obama ⟵ *he*

72.0%
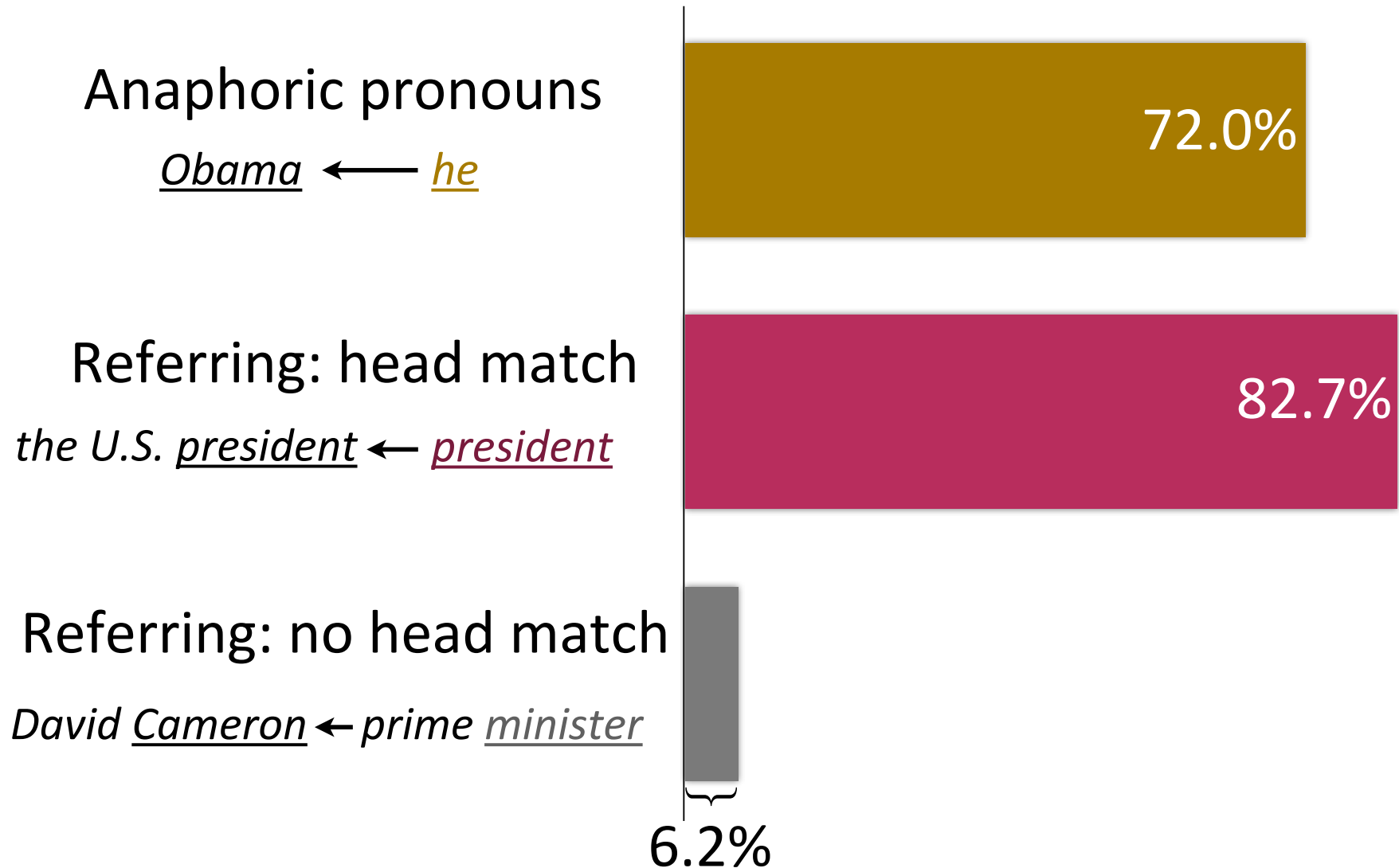
Referring: head match

the U.S. president ⟵ *president*

82.7%

Referring: no head match

David Cameron ⟵ *prime minister*

Error Analysis

Anaphoric pronouns

Obama ⟵ *he*

72.0%

Referring: head match

the U.S. president ⟵ *president*

82.7%

Referring: no head match

David Cameron ⟵ *prime minister*

6.2%

Compatibility

Compatibility

[*David Cameron*]$_1$ ... [*The prime minister*]$_1$ ...

Compatibility

[*David Cameron*]$_1$ ... [*The prime minister*]$_1$ ...

Soon et al. (2001)
inter alia

Compatibility

[*David Cameron*]$_1$ ... [*The prime minister*]$_1$ ...

▶ Number, gender

Soon et al. (2001)
inter alia

Compatibility

[*David Cameron*]$_1$ ... [*The prime minister*]$_1$ ...

- Number, gender
- Named entity type

Soon et al. (2001) inter alia

Compatibility

[*David Cameron*]$_1$ ... [*The prime minister*]$_1$ ...

- Number, gender
- Named entity type
- Unsupervised clustering labels
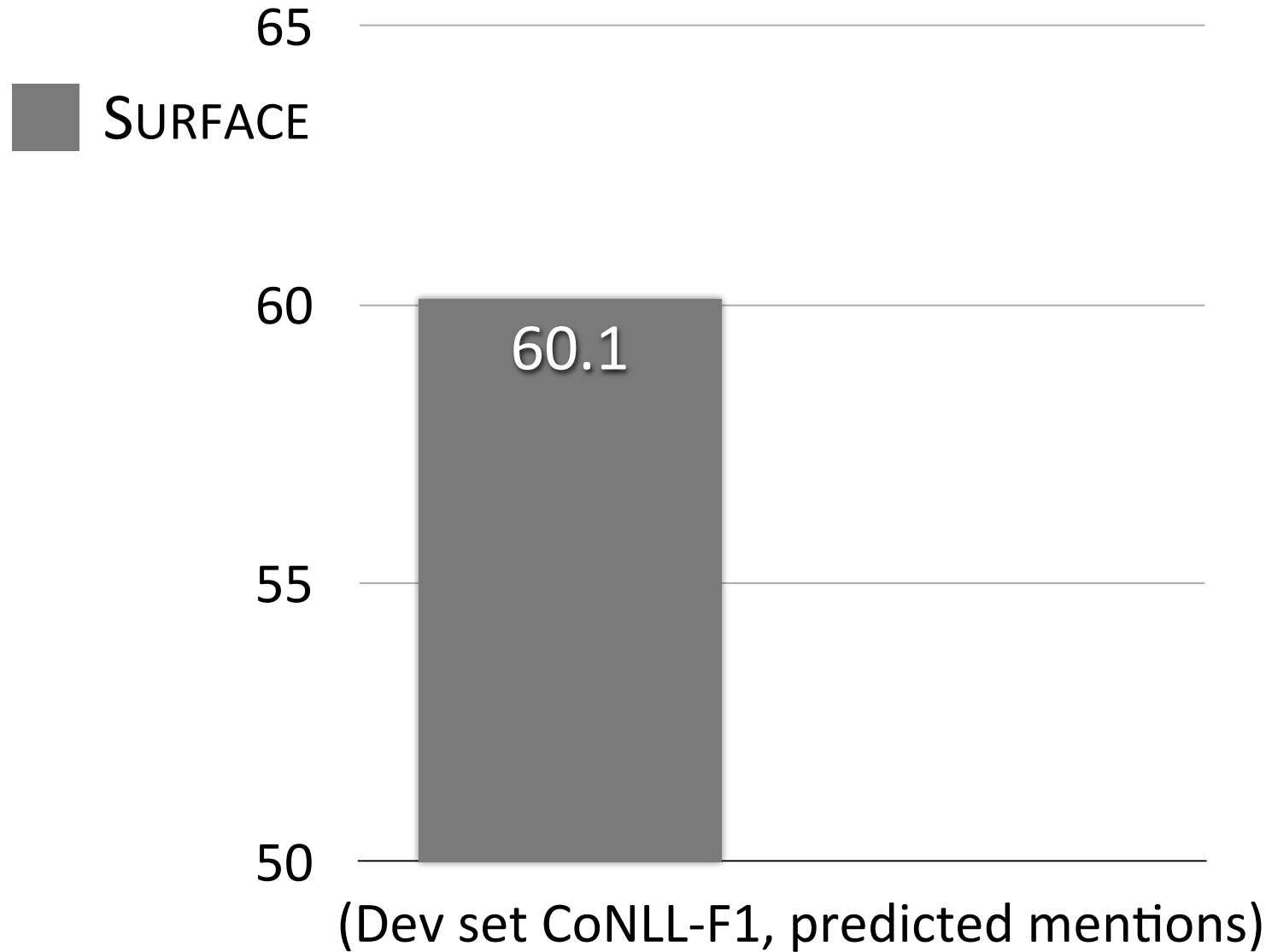
Soon et al. (2001) inter alia

Compatibility

[*David Cameron*]$_1$ ... [*The prime minister*]$_1$ ...

- Number, gender
- Named entity type
- Unsupervised clustering labels
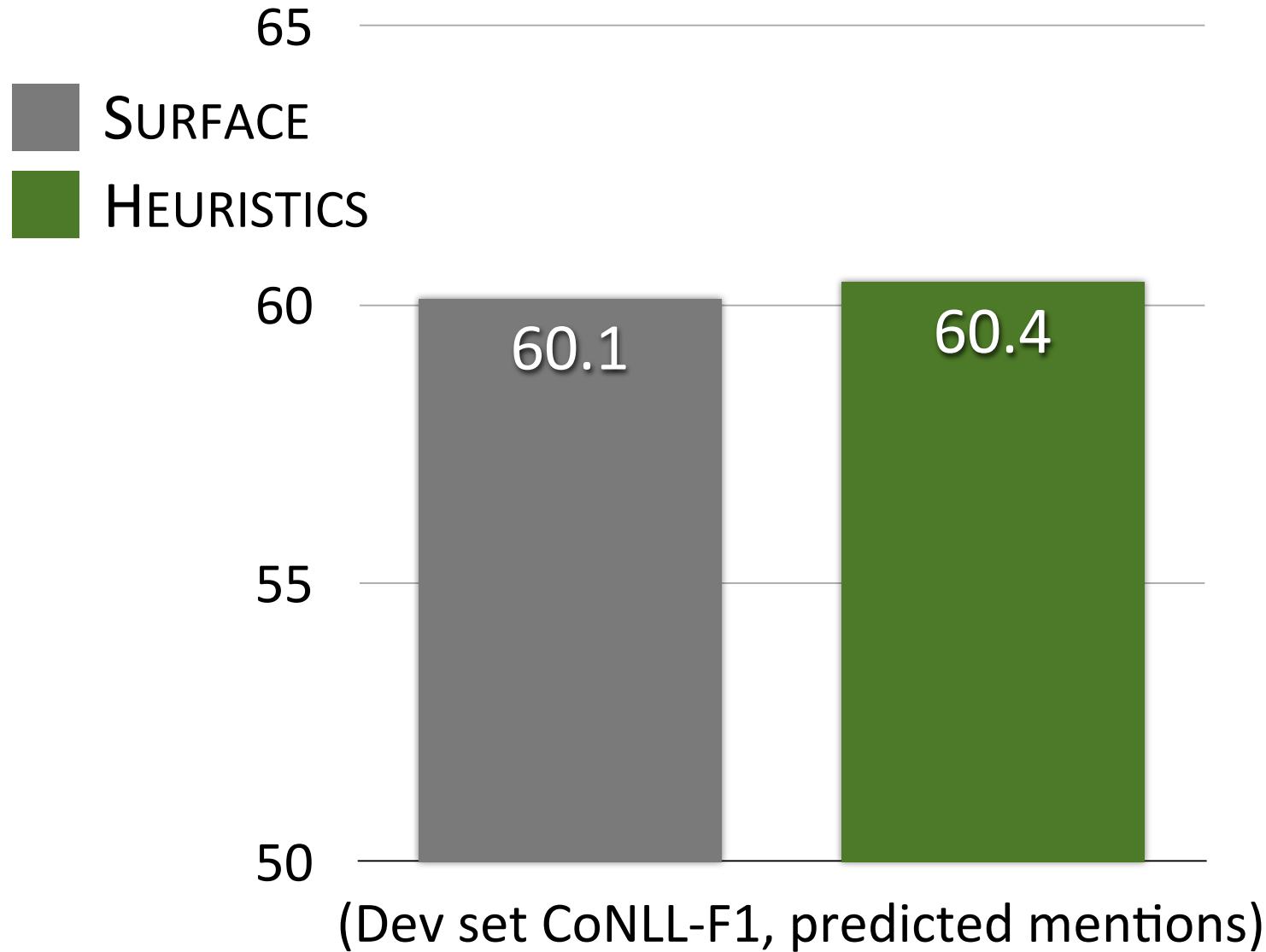- WordNet hypernymy / synonymy

Soon et al. (2001) inter alia

SURFACE Accuracy

60.1

(Dev set CoNLL-F1, predicted mentions)

SURFACE Accuracy

Legend:
- **SURFACE** (gray)
- **HEURISTICS** (green)

	SURFACE	HEURISTICS
Value	60.1	60.4

(Dev set CoNLL-F1, predicted mentions)

What works?

What works?

▶ Importing external information
with sophisticated heuristics

{
Ponzetto and Strube (2006)
Rahman and Ng (2011)
Bansal and Klein (2012)

What works?

▸ Importing external information with sophisticated heuristics $\left\{\begin{array}{l} \text{Ponzetto and Strube (2006)} \\ \text{Rahman and Ng (2011)} \\ \text{Bansal and Klein (2012)} \end{array}\right.$

▸ We can support additional heuristic features, including number and gender information (Bergsma and Lin, 2006)

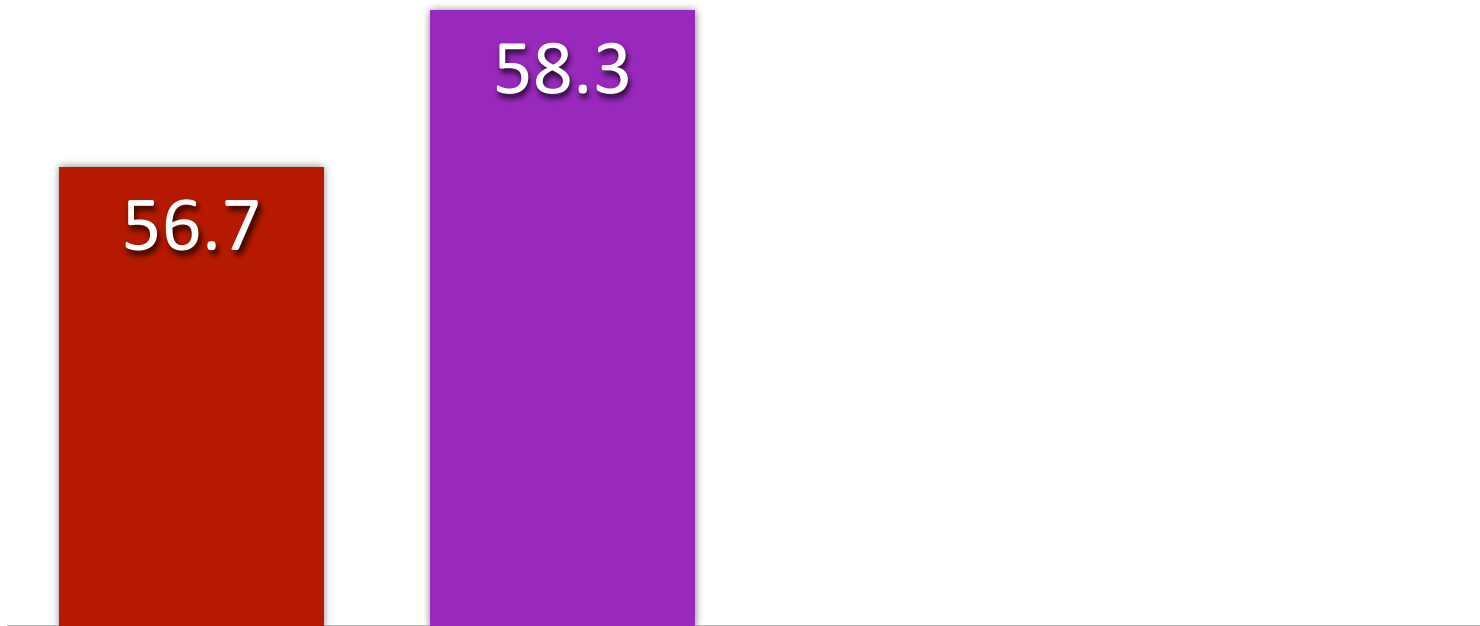# Test Set Accuracy

■ STANFORD (Lee et al., 2011)

56.7

(Test set CoNLL-F1, predicted mentions)

Test Set Accuracy

■ STANFORD (Lee et al., 2011)

■ IMS (Björkelund and Farkas, 2012)

56.7

58.3

(Test set CoNLL-F1, predicted mentions)

Test Set Accuracy

- ■ (red) Stanford (Lee et al., 2011)
- ■ (purple) IMS (Björkelund and Farkas, 2012)
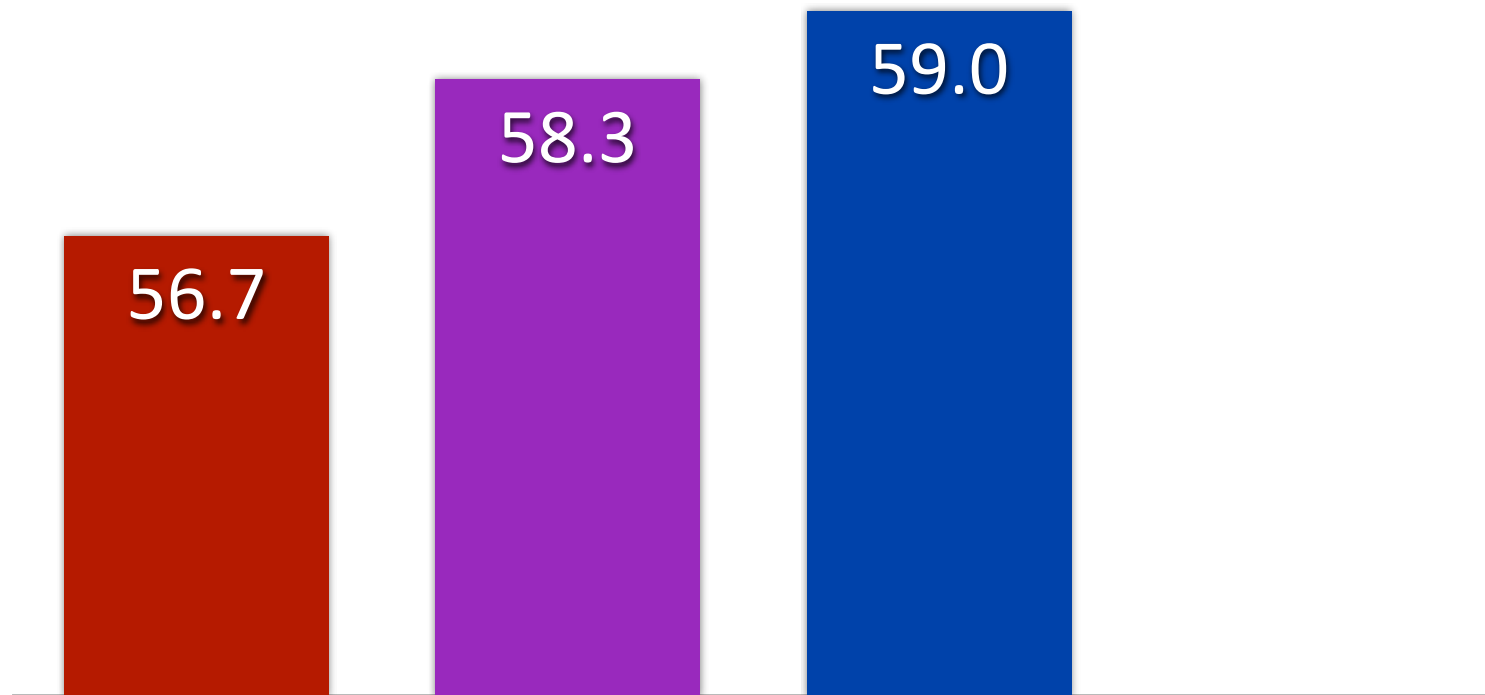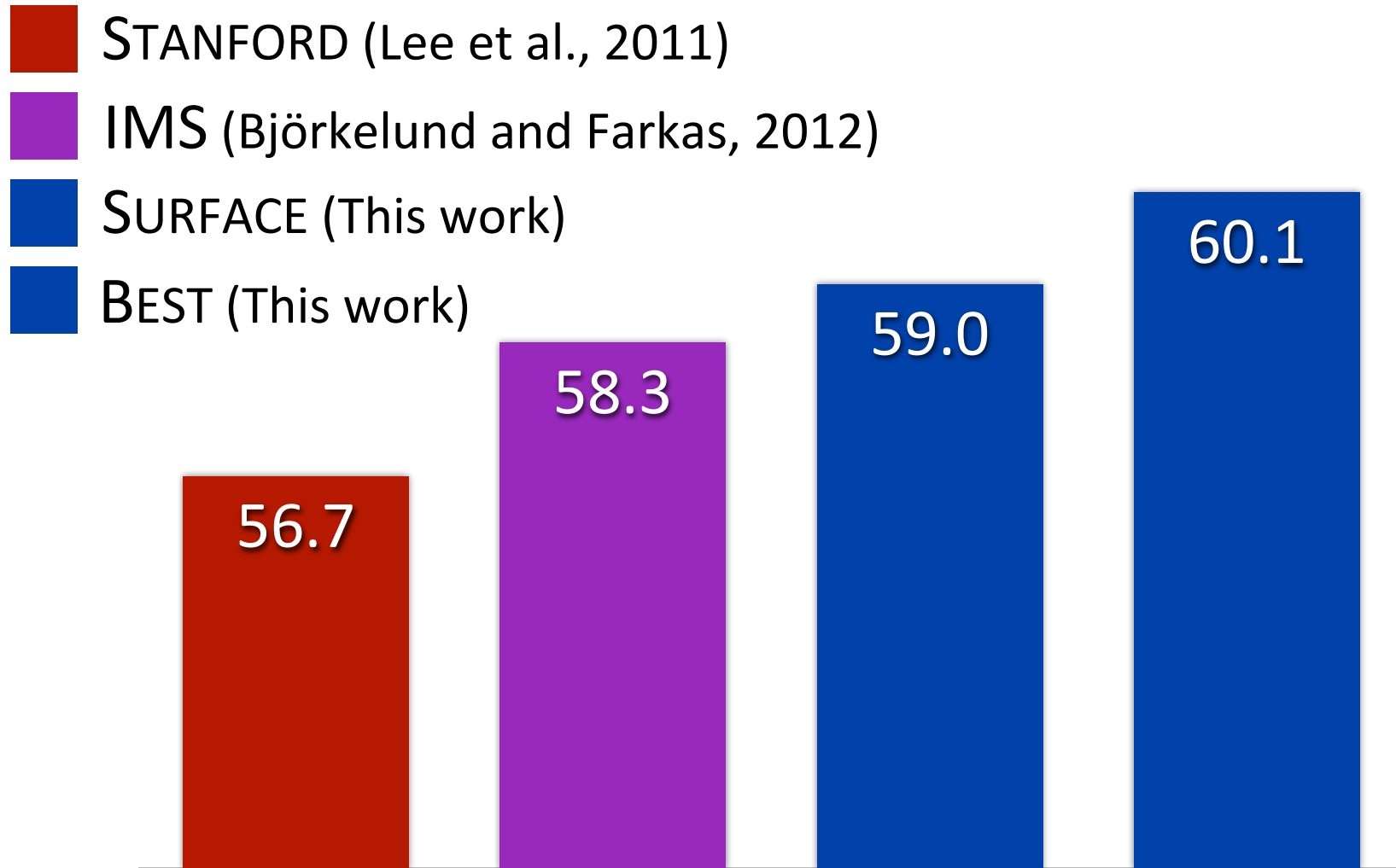- ■ (blue) Surface (This work)

56.7   58.3   59.0

(Test set CoNLL-F1, predicted mentions)

Test Set Accuracy

- STANFORD (Lee et al., 2011)
- IMS (Björkelund and Farkas, 2012)
- SURFACE (This work)
- BEST (This work)

56.7   58.3   59.0   60.1

(Test set CoNLL-F1, predicted mentions)

Conclusion

Conclusion

▶ Surface lexical features capture a wide range of linguistic phenomena in a unified way

Conclusion

▶ Surface lexical features capture a wide range of linguistic phenomena in a unified way

▶ Semantic errors require heavy-duty information from other knowledge sources

Conclusion

▶ Surface lexical features capture a wide range of linguistic phenomena in a unified way

▶ Semantic errors require heavy-duty information from other knowledge sources

▶ Extensible system that achieves state-of-the-art performance

Conclusion

▶ The Berkeley Coreference Resolution System: `http://nlp.cs.berkeley.edu`

    ▶ Full end-to-end system (accepts raw text as input)

Conclusion

▸ The Berkeley Coreference Resolution System:
`http://nlp.cs.berkeley.edu`

  ▸ Full end-to-end system (accepts raw text as input)

Thank you!

Stacking
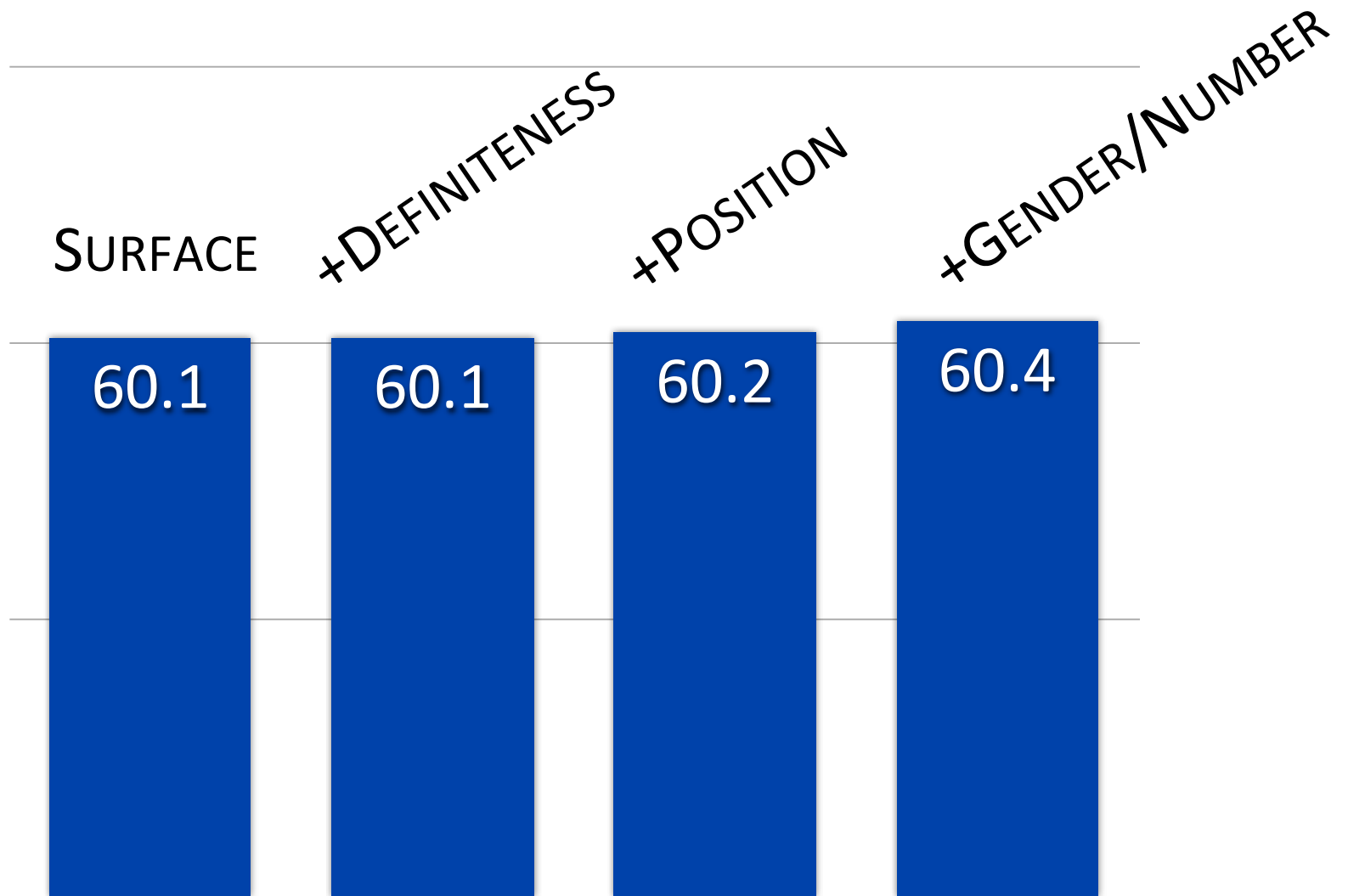
SURFACE	+DEFINITENESS	+POSITION	+GENDER/NUMBER
60.1	60.1	60.2	60.4

SURFACE Accuracy

- ⬛ (grey) SURFACE
- ⬛ (green) HEURISTICS
- ⬛ (blue) INDICATORS

60.1   60.4   60.2

Predicted mentions

SURFACE Accuracy

Legend:
- **SURFACE** (gray)
- **HEURISTICS** (green)
- **INDICATORS** (blue)

Predicted mentions
- SURFACE: 60.1
- HEURISTICS: 60.4
- INDICATORS: 60.2
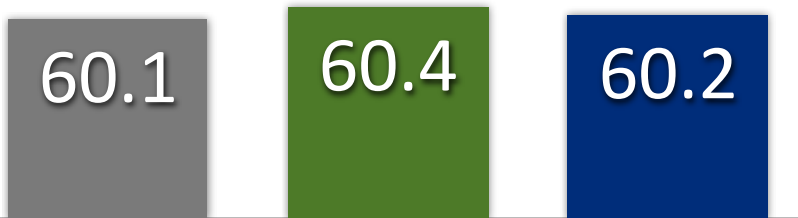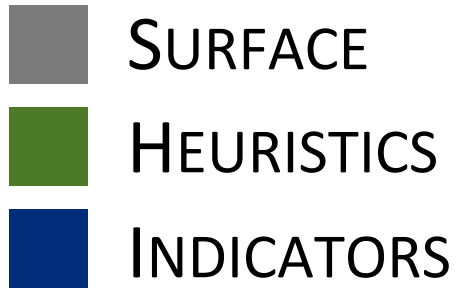
Gold mentions
- SURFACE: 75.1
- HEURISTICS: 76.7
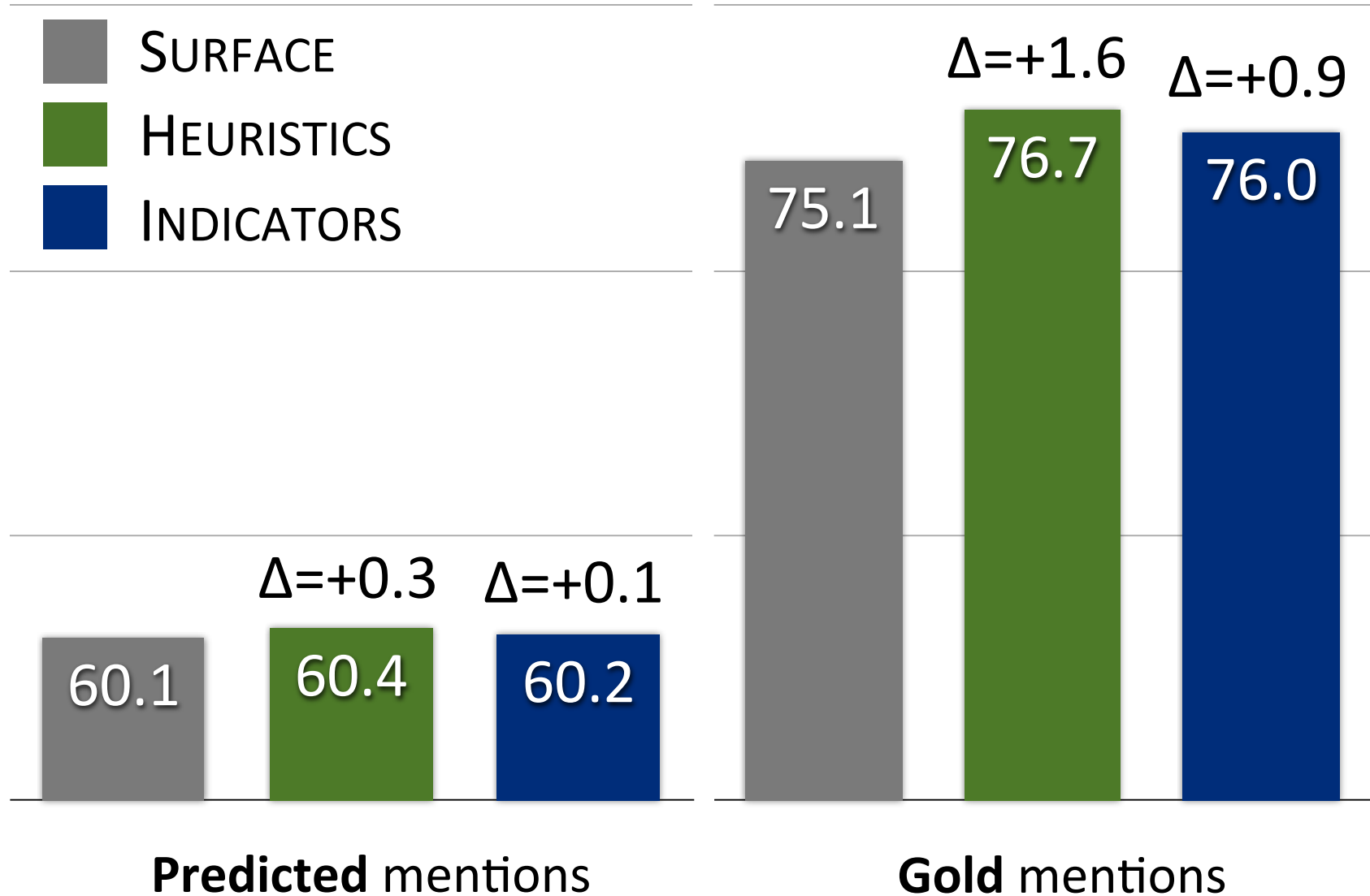- INDICATORS: 76.0
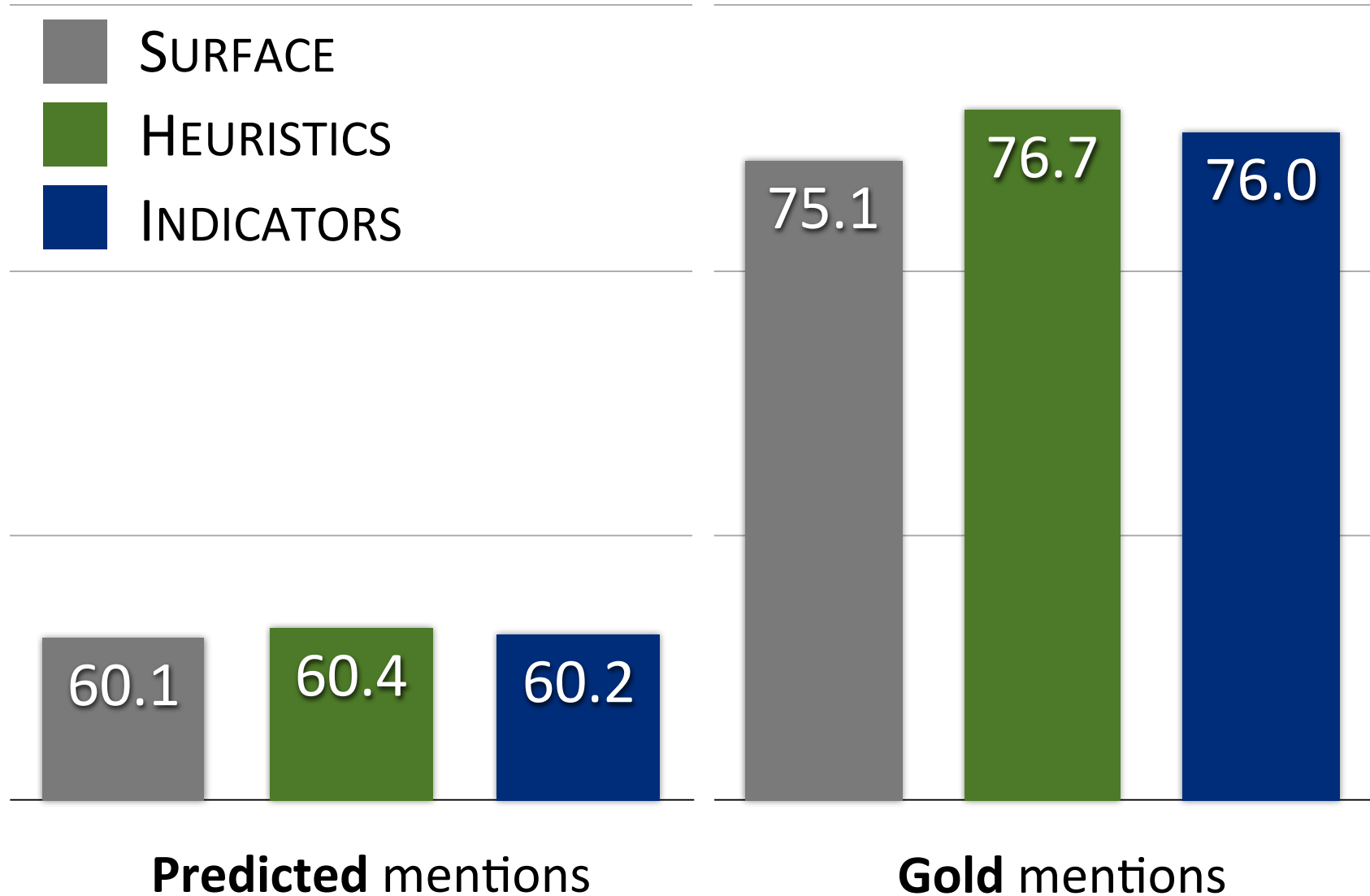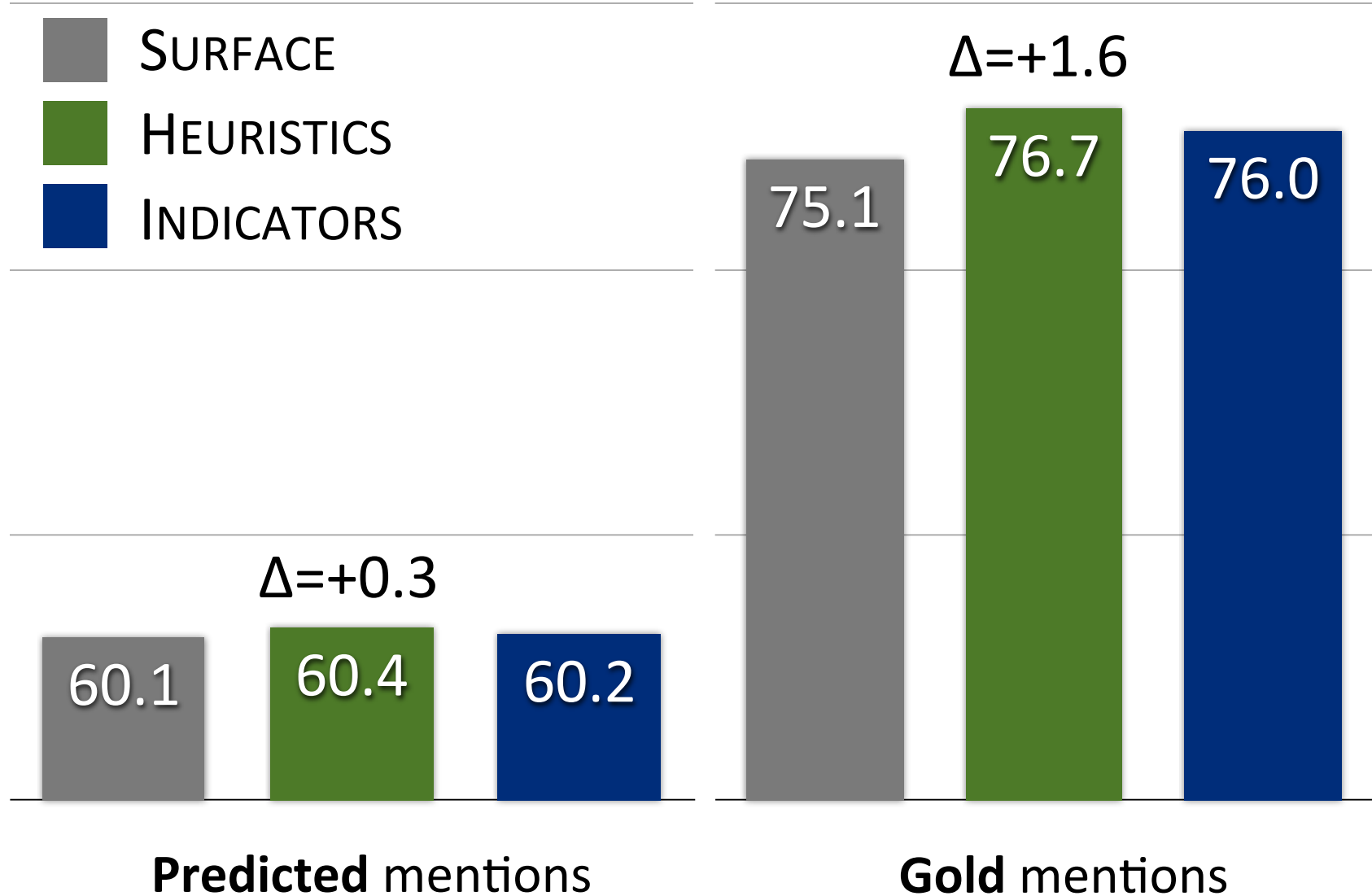
SURFACE Accuracy

Gold vs. System Mentions

New

[*the president*]

Gold vs. System Mentions

New

[*the president*]
PERSON

Gold vs. System Mentions

[*Barack Obama*]
PERSON
1

New

[*the GOP*]
ORG
2

[*the president*]
PERSON

[*the Capitol*]
LOC
3
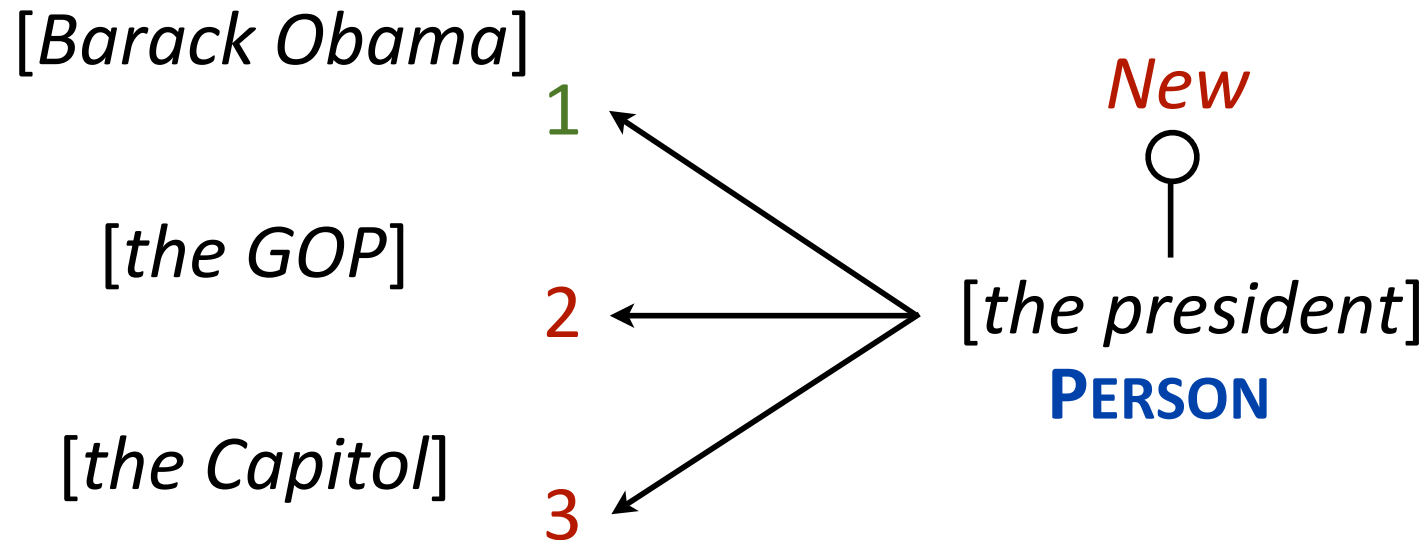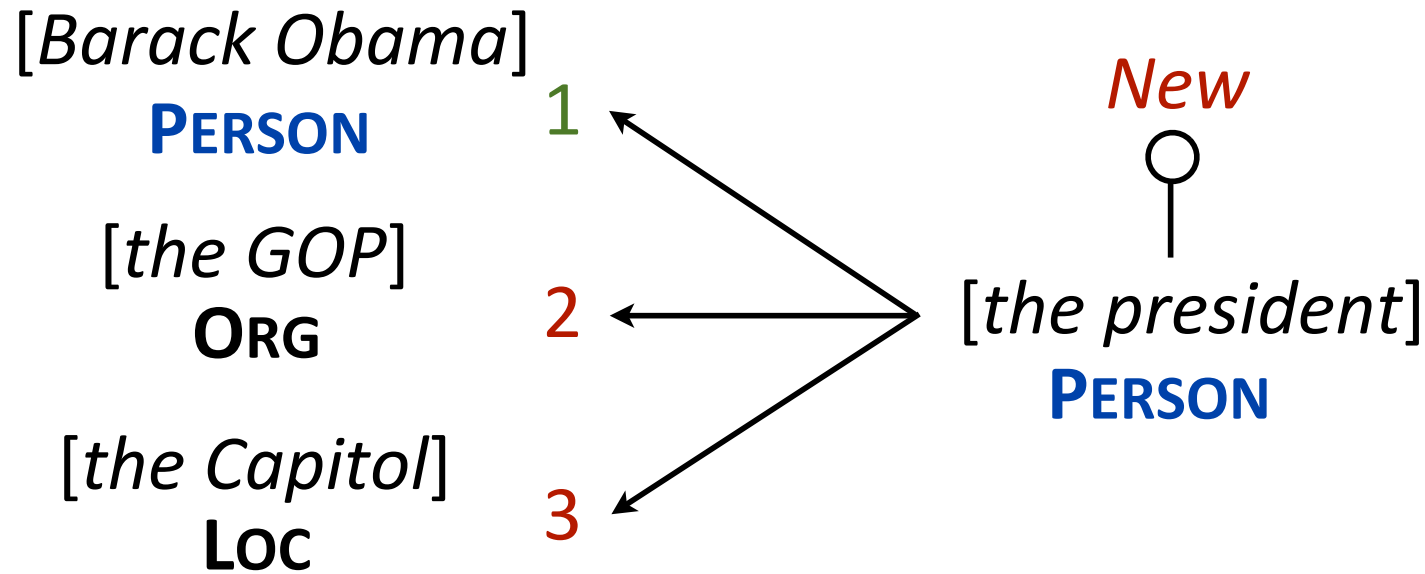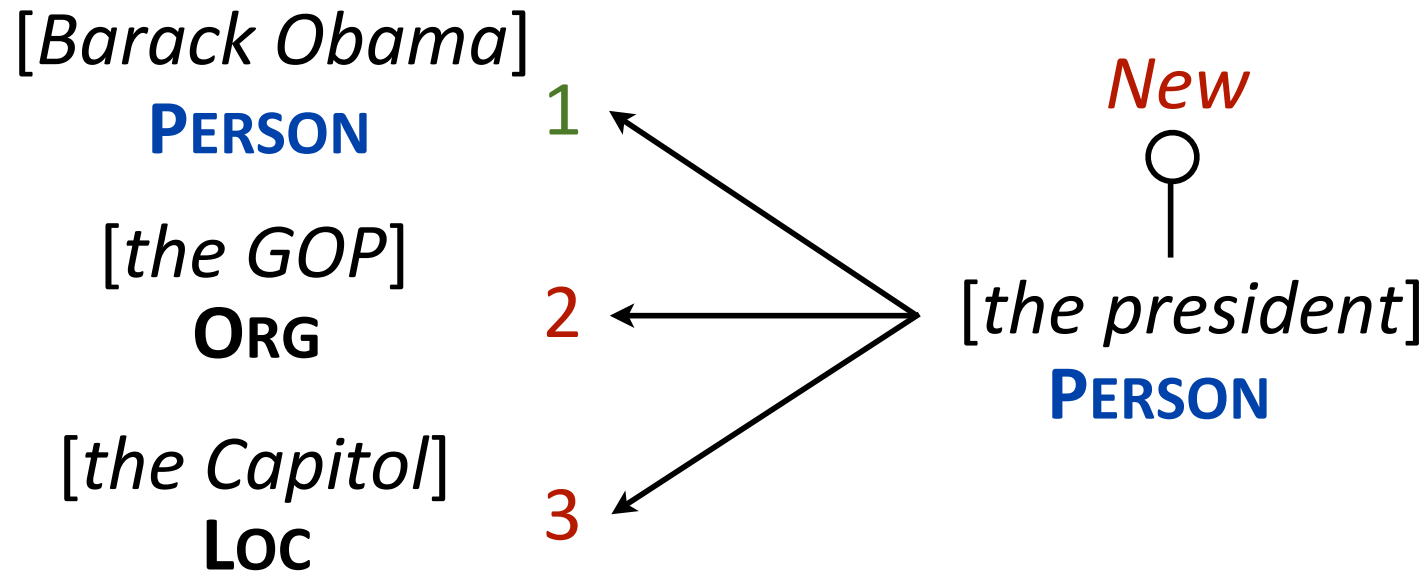
▶ 3x more mentions when singletons are included

Gold vs. System Mentions

[*Barack Obama*]
PERSON   1

[*the GOP*]
ORG   2

[*the Capitol*]
LOC   3   *New*

[*the advisor*]   ○
PERSON   4   [*the president*]
                                              PERSON
[*the Chief of Staff*]
PERSON   5

[*somebody*]
PERSON   6

[*the party*]
ORG   7

Gold vs. System Mentions

[*Barack Obama*]
PERSON   1

[*the GOP*]
ORG   2

[*the Capitol*]
LOC   3

[*the advisor*]
PERSON   4   [*the president*]
                                      PERSON

[*the Chief of Staff*]
PERSON   5

[*somebody*]
PERSON   6

[*the party*]
ORG   7

New

UPHILL BATTLE