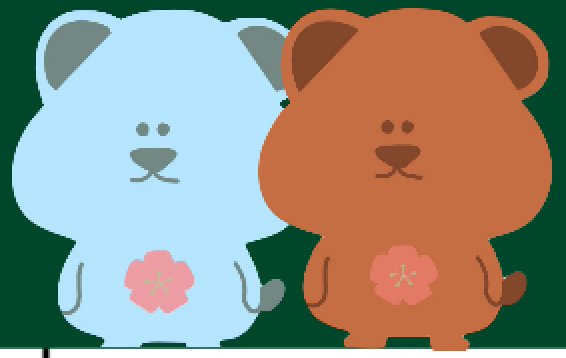
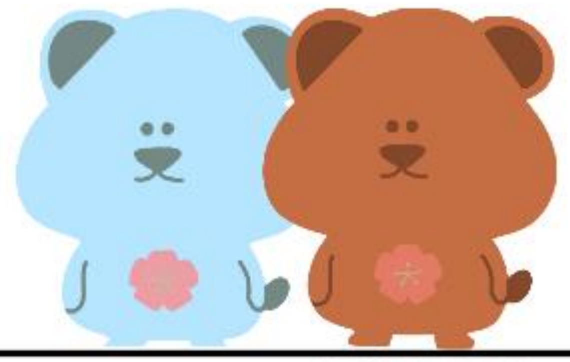


Coreference Resolution

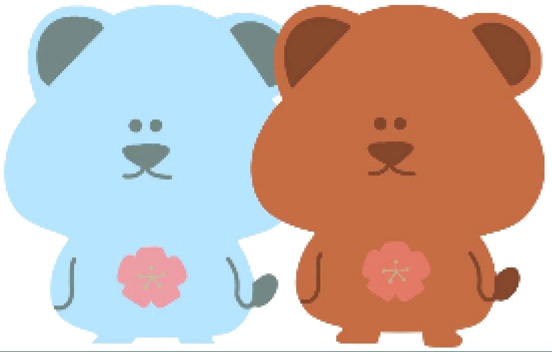
김경민, 주연우

EWHA EURON

목차



Introduction



#01 Introduction

#1 Coreference Resolution

=상호 참조 해결, 동일 지시어 찾기

문장 속 등장하는 지시어 중 같은 것을 지칭하는 지시어 찾기

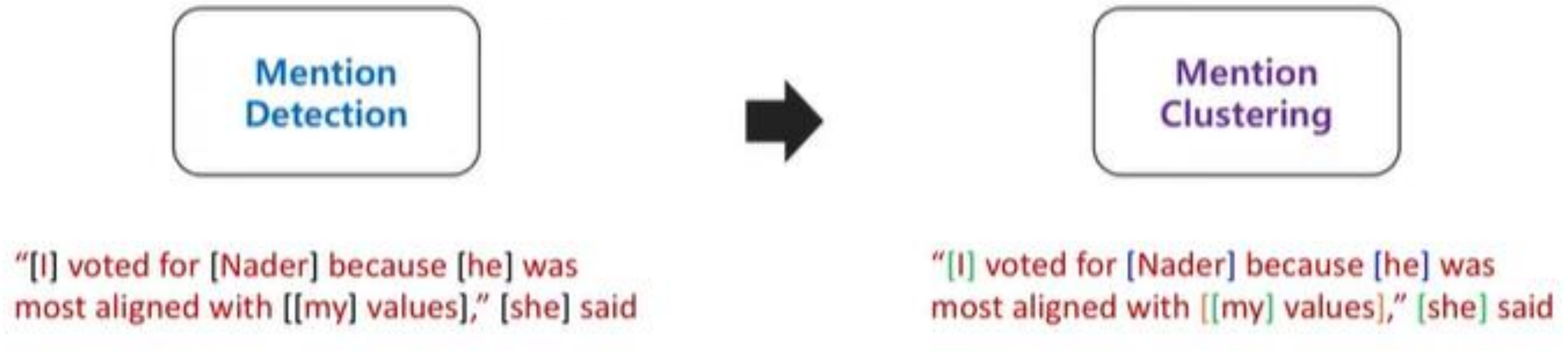
Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

#01 Introduction

#1 Coreference Resolution

Mention: 어떤 개체를 지시하고 있는 텍스트의 범위

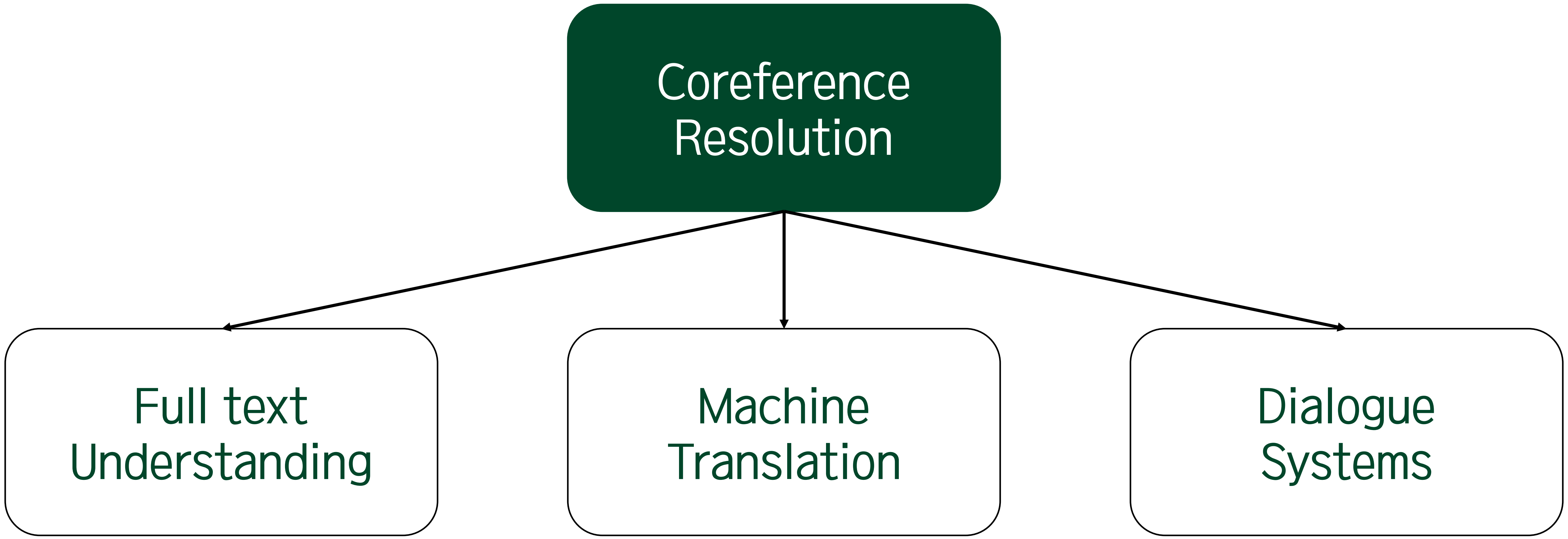


Mention을 찾는 Detection을 수행하고, 같은 Mention끼리 Clustering 수행

#01 Introduction

#2 적용 분야

Coreference Resolution은 여러 문제를 푸는 데 적용 가능하고 성능 향상에 도움이 됨



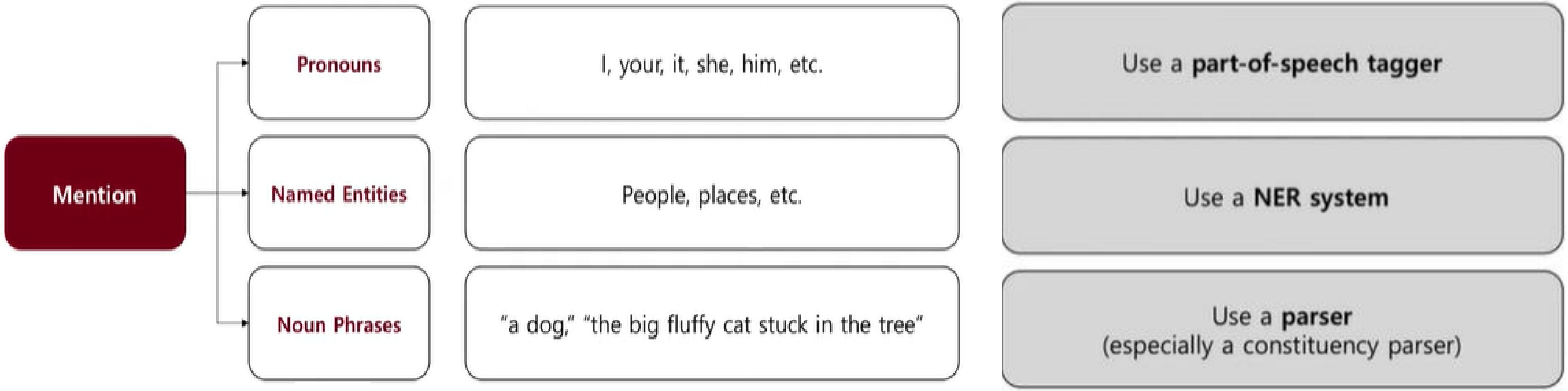
Mention Detection

#02 Mention Detection

#1 Mention의 종류

Mention의 종류는 3가지로 분류

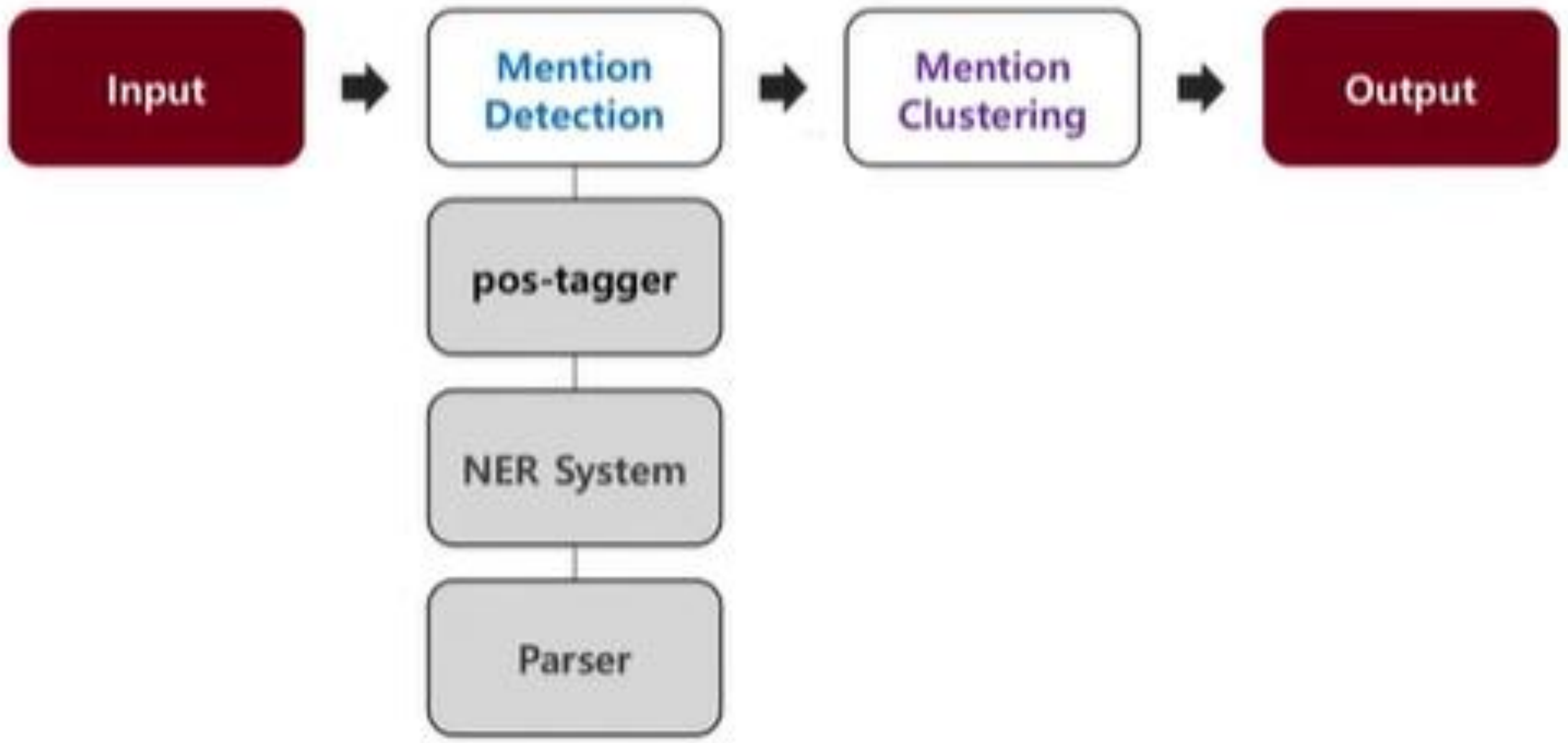
Mention의 종류에 따라 다른 NLP 시스템을 활용하여 Mention Detection 수행



#02 Mention Detection

#2 Mention Detection

2 step의 파이프라인 시스템 사용 → End-to-End 사용

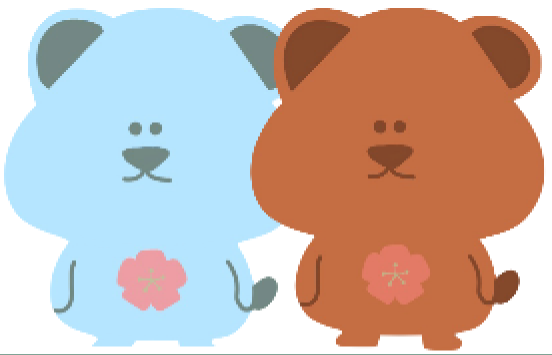


[2 step의 파이프라인 시스템]

[End-to-End]

EWHA EURON

Linguistics of Coreference



#03 Linguistics of Coreference

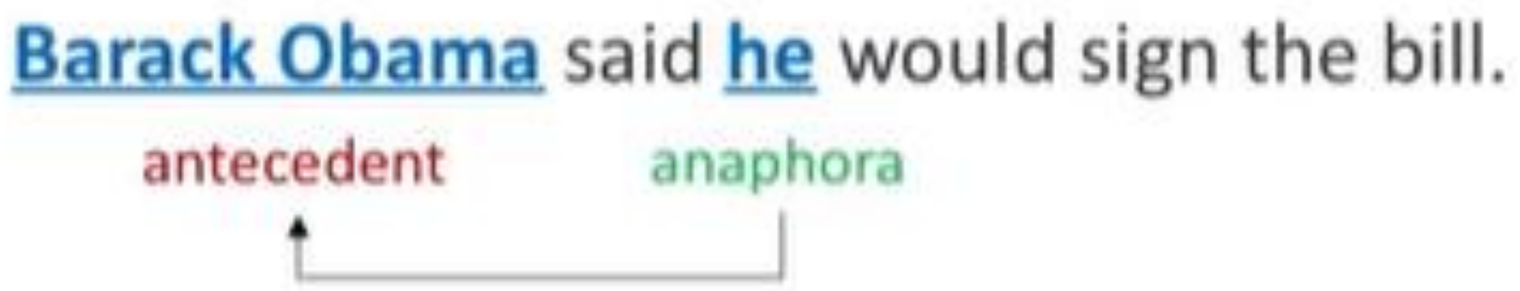
#1 Anaphora

=전방 조응 문맥적 지시로 선행 명사구와 동일 지시 관계를 나타내는 것

선행 어구의 대용으로서 대명사 등을 쓰는 일

⇨ 문장 내의 어떤 단어들은 독자적으로 참조 할 수 없다

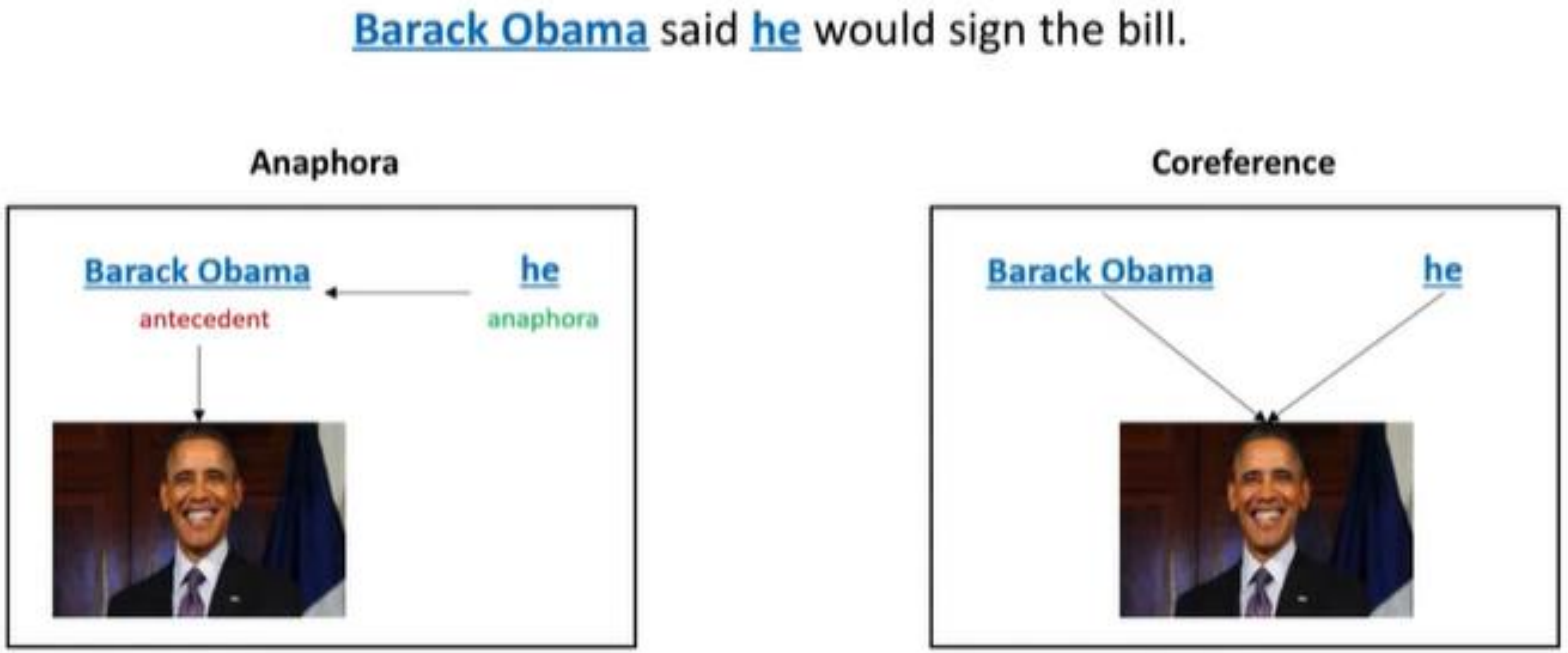
Anaphora에 해당하는 단어들은 antecedent를 참조 해야만 제대로 해석 할 수 있다



EWHA EURON

#03 Linguistics of Coreference

#2 Anaphora vs. Coreference

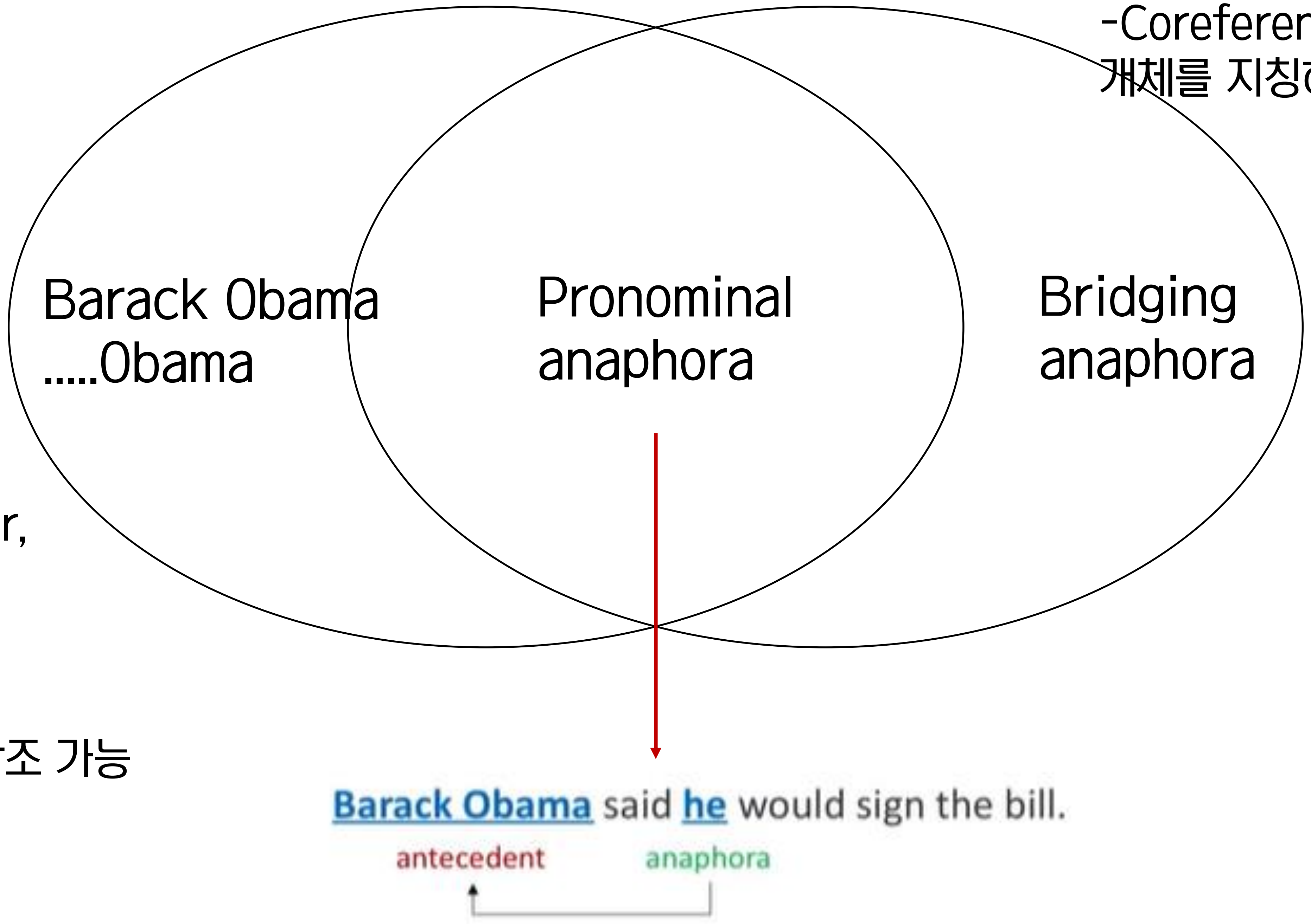


#03 Linguistics of Coreference

#2 Anaphora vs. Coreference

We went to see a concert last night. The tickets were really expensive.

-The tickets은 앞서 등장한 a concert에 대한 참조 없이 설명 불가
-Coreference를 적용하면 두 단어는 다른 개체를 지칭하기 때문에 관계 파악 어려움



Barack Obama was born in Hawaii in 1961 and 48 years later, Obama was elected president of The United States.

-뒤에 언급된 Obama는 독립적으로 참조 가능
Coreference만 적용 가능

#03 Linguistics of Coreference

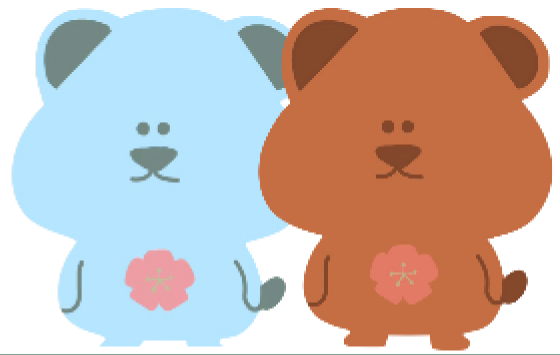
#3 Cataphora

=후방 조응 문맥적 지시로 후행하는 명사구와 동일 지시 관계를 나타내는 표현

후속 어구를 지시하는 어구의 사용

대부분의 antecedent는 anaphora 앞에 등장하지만 항상 그런 것은 아니다

Coreference Models



#04 Coreference Models

4 Kinds of Coreference Models

#1 Rule-based

#2 Mention Pair

#3 Mention Ranking

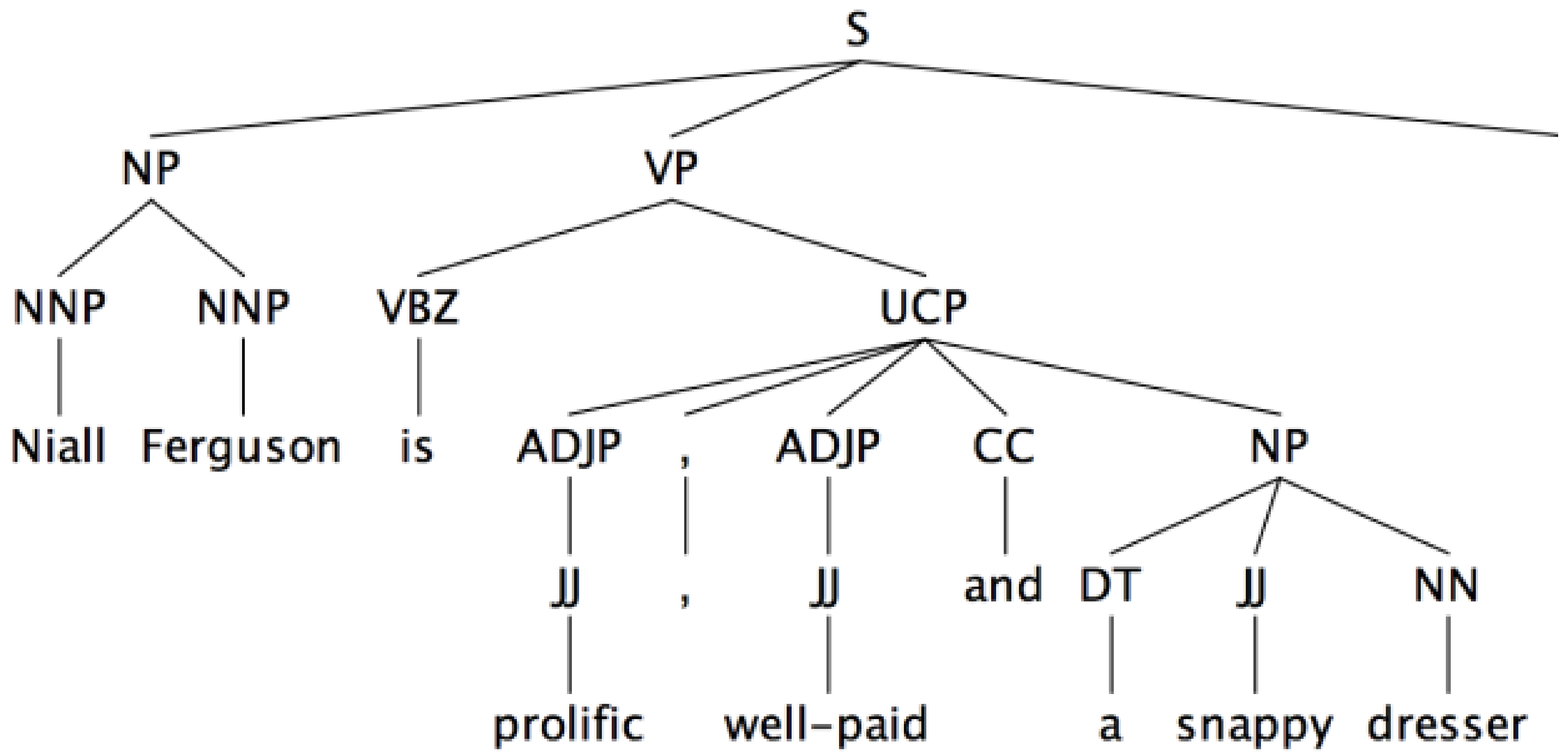
#4 Clustering

#04 Coreference Models

Rule-based Model

Hobb's Naïve Algorithm

- 전통적으로 대명사의 anaphora를 resolution
- 알고리즘의 아홉 단계 반복 -> 선행사를 찾아가는 과정 반복



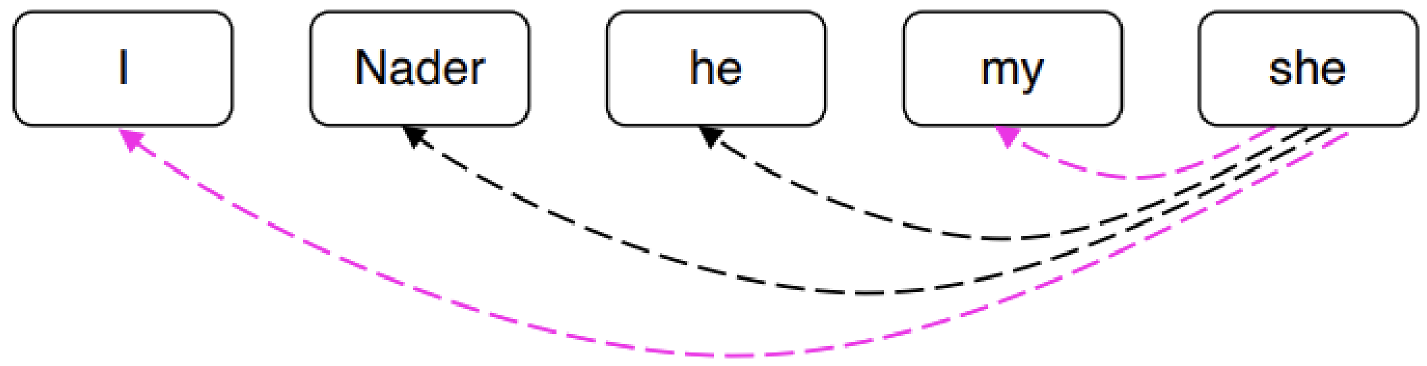
#04 Coreference Models

Mention Pair Model

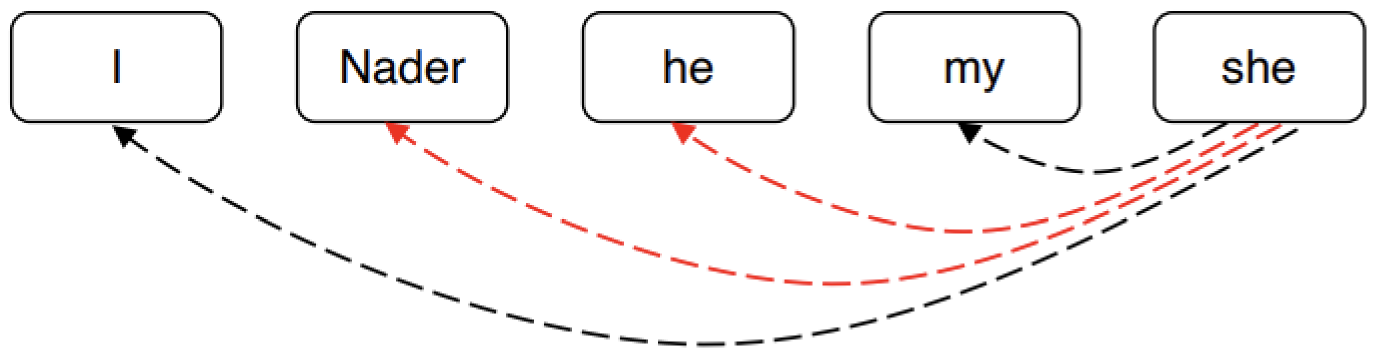
Mention Pair

- 두 mention (mention pair) 간의 coreference가 있는가?
- 가능한 모든 mention pair의 cross-entropy loss를 줄이는 방향으로 학습
- Mention들 간의 coreference link는 판단 불가
- 긴 문장에서 상대적으로 거리가 먼 단어가 선행사일 가능성은 낮음

$$p(m_i, m_j)$$



Positive examples: want $p(m_i, m_j)$ to be near 1



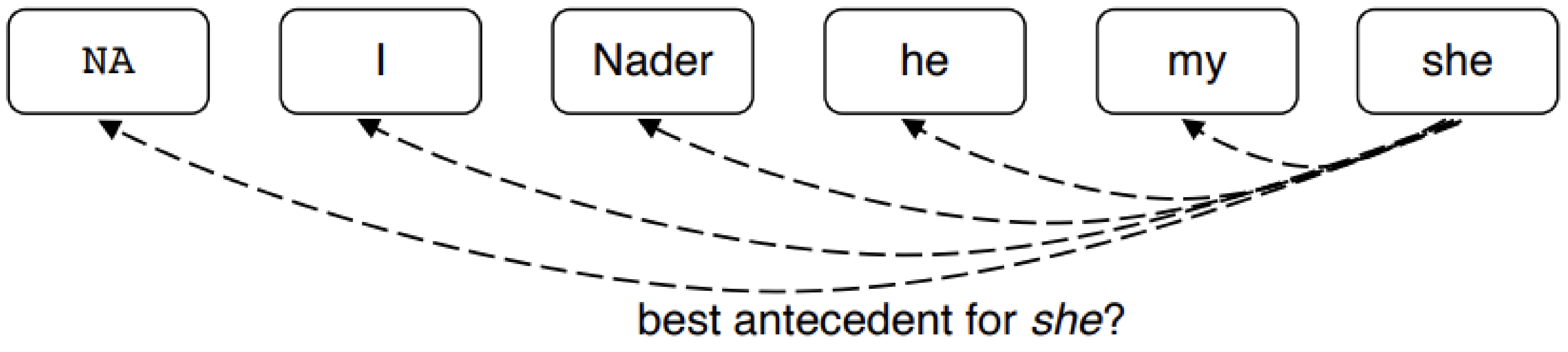
Negative examples: want $p(m_i, m_j)$ to be near 0

#04 Coreference Models

Mention Ranking Model

Mention Ranking

- Mention 하나 당 한 개의 best 선행사(선행사일 가능성이 가장 높은 다른 mention)를 찾는 방식
- Dummy mention <NA>를 사용해서 선행사가 없을 수 있는 mention 처리
- 하나의 선행사만을 연결시켜주지만, mention pair model과 비슷한 결과
- Clustering 과정이 없지만, coreference가 있는 단어끼리 연결 good



#04 Coreference Models

How do we compute the probabilities?

#1 Non-neural statistical classifier

#2 Simple neural network

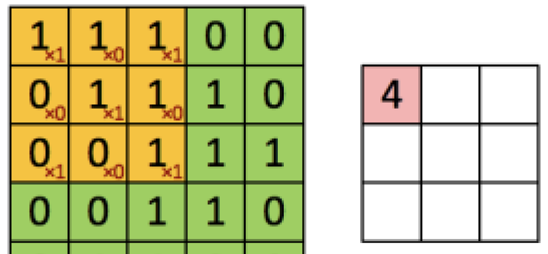
#3 More advanced model using LSTMs, attention

#04 Coreference Models

Convolutional Networks for Language

- Compute vectors of every possible word subsequence of a certain length
- Convolution is classically used to extract features from images

$$(f * g)[n] = \sum_{m=-M}^{M} f[n-m]g[m].$$



Image

Convolved Feature

From Stanford UFLDL wiki

1D convolution for text with padding

∅	0.0	0.0	0.0	0.0
tentative	0.2	0.1	−0.3	0.4
deal	0.5	0.2	−0.3	−0.1
reached	−0.1	−0.3	−0.2	0.4
to	0.3	−0.3	0.1	0.1
keep	0.2	−0.3	0.4	0.2
government	0.1	0.2	−0.1	−0.1
open	−0.4	−0.4	0.2	0.3
∅	0.0	0.0	0.0	0.0

∅,t,d	−0.6
t,d,r	−1.0
d,r,t	−0.5
r,t,k	−3.6
t,k,g	−0.2
k,g,o	0.3
g,o,∅	−0.5

Apply a **filter** (or **kernel**) of size 3

3	1	2	−3
−1	2	1	−3
1	1	−1	1

conv1d, padded with ave pooling over time

∅	0.0	0.0	0.0	0.0
tentative	0.2	0.1	−0.3	0.4
deal	0.5	0.2	−0.3	−0.1
reached	−0.1	−0.3	−0.2	0.4
to	0.3	−0.3	0.1	0.1
keep	0.2	−0.3	0.4	0.2
government	0.1	0.2	−0.1	−0.1
open	−0.4	−0.4	0.2	0.3
∅	0.0	0.0	0.0	0.0

∅,t,d	−0.6	0.2	1.4
t,d,r	−1.0	1.6	−1.0
d,r,t	−0.5	−0.1	0.8
r,t,k	−3.6	0.3	0.3
t,k,g	−0.2	0.1	1.2
k,g,o	0.3	0.6	0.9
g,o,∅	−0.5	−0.9	0.1
ave p	−0.87	0.26	0.53

Apply 3 **filters** of size 3

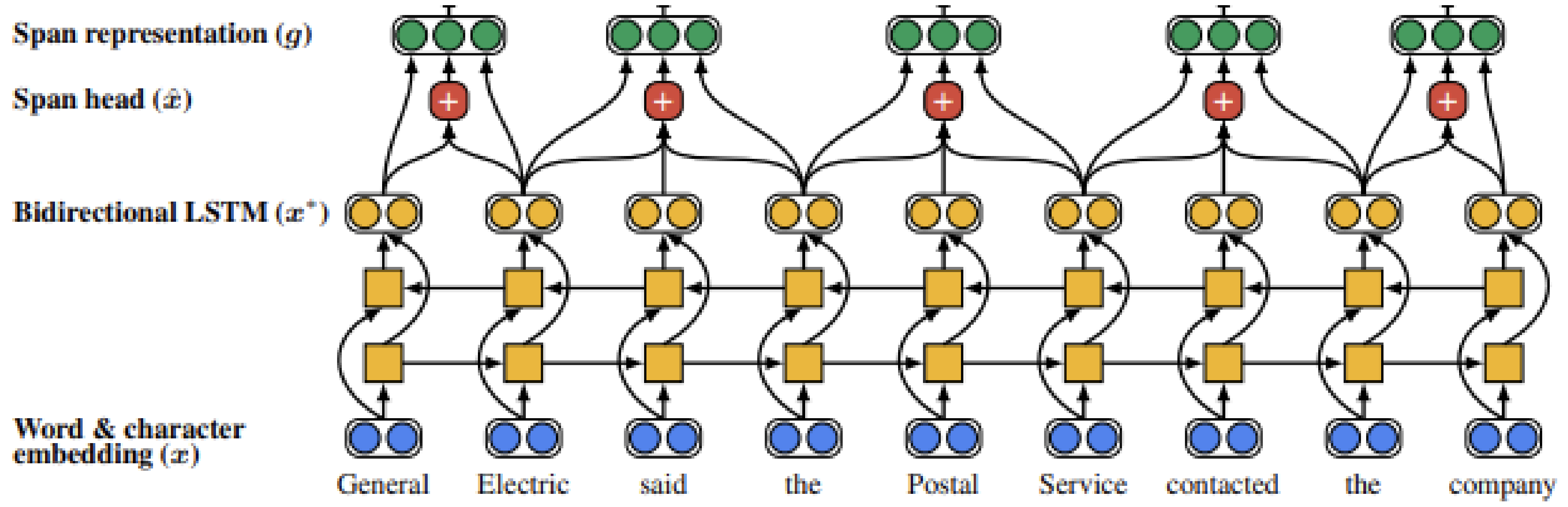
3	1	2	−3
−1	2	1	−3
1	1	−1	1

1	0	0	1
1	0	−1	−1
0	1	0	1

1	−1	2	−1
1	0	−1	3
0	2	2	1

#04 Coreference Models

End-to-end Neural Coref Model



Kenton Lee et al. (EMNLP 2017)

- Mention Ranking Model
- LSTM & Attention
  - Attention – mention 내의 어떤 단어가 중요 단어인지 학습
- End-to-End model -> No mention detection step
  - 특정 길이 이상의 test span -> mention 후보로 가정
- 모든 span of text에 대해서 coreferent score을 계산하는 것은 비효율적 -> Pruning

#04 Coreference Models

BERT-based Coref

Mention Ranking

- Pretrained transformers can learn long-distance semantic dependencies in text

#1. SpanBERT

- Pretrains BERT models to be better at span-based prediction task

#2. BERT-QA for coref

- Treat coreference like a deep QA task

Performance & Conclusion

#05 Performance & Conclusion

#1 System Performance

Model	English	Chinese
Lee et al. (2010)	~55	~50
Chen & Ng (2012) [CoNLL 2012 Chinese winner]	54.5	57.6
Fernandes (2012) [CoNLL 2012 English winner]	60.7	51.6
Wiseman et al. (2015)	63.3	—
Clark & Manning (2016)	65.4	63.7
Lee et al. (2017)	67.2	—
Joshi et al. (2019)	79.6	—
Wu et al. (2019) [CorefQA]	79.9	—
Xu et al. (2020)	80.2	
Wu et al. (2020) [CorefQA + SpanBERT-large]	83.1	

EWHA EURON

#05 Performance & Conclusion

#2 Conclusion

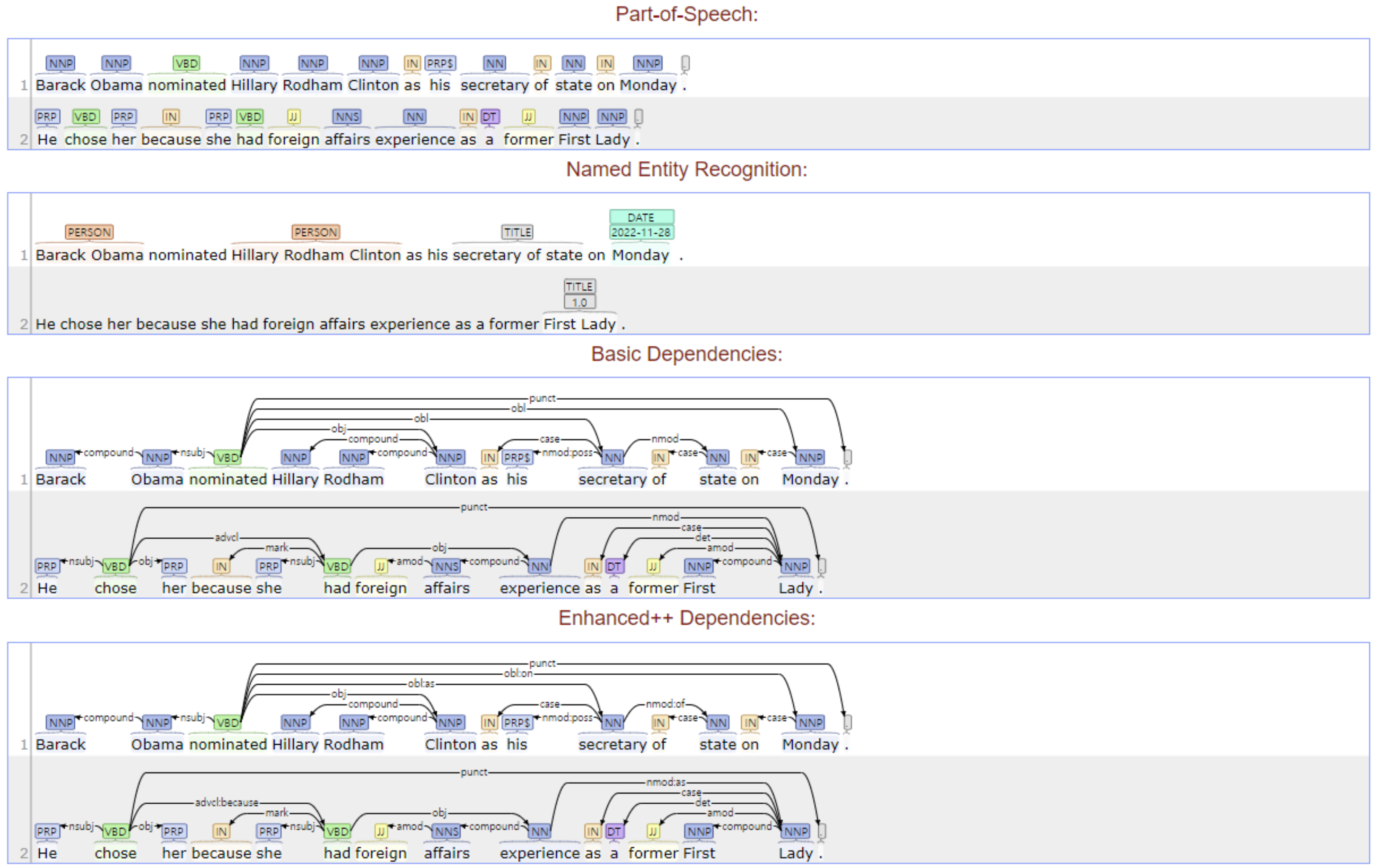
Coreference is useful, challenging, and linguistically interesting task

- 다양한 종류의 Coreference resolution system 존재

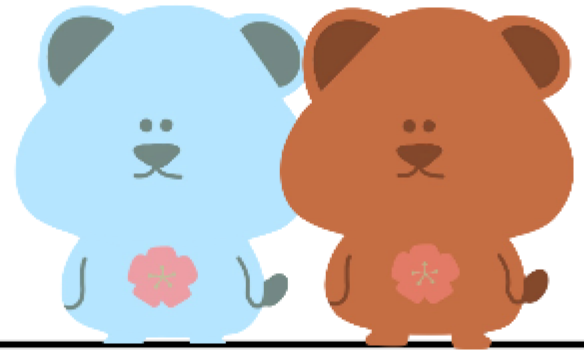
Systems are getting better rapidly, largely due to better neural models

- 전반적으로 아직 실수가 많이 있다

http://corenlp.run/
https://huggingface.co/coref/



THANK YOU



EWHA EURON