Challenges in semantics

- Examples from the Natural Questions data set (Kwiatkowski et al. 2019)
- Questions typed into the Google search engine + a Wikipedia page that Google returned as a result
- Annotators identify a paragraph in the Wikipedia page that contains the answer (or say the answer isn't in the page), and identify the answer

Challenges in text understanding: semantic role realization

- Who appoints the chairman of union public service commission in India?

- *As per Art. 316,* ***the Chairman*** *and other members of Union Public Service Commission* ***shall be appointed by the President****. In case the office of the Chairman becomes vacant his duties shall be performed by one of the other members of the Commission as the President may appoint for the purpose.*

Challenges in text understanding: coreference resolution (pronouns)

- Who won Wimbledon in 2019?

- *Simona Halep is a female tennis player.* ***She won Wimbledon in 2019****.*

Challenges in text understanding: coreference resolution beyond pronouns

- Who wrote the song love is in the air?
- *"**Love Is in the Air"** is a 1977 disco song sung by John Paul Young. **The song** **was written by** **George Young** and Harry Vanda and it became a worldwide hit in 1978, peaking at No. 3 on the Australian charts and No. 5 in the UK Singles Chart.*

Challenges in text understanding: bridging

- Who won America's Got Talent season 11?
- The 11th season of America's Got Talent, an American talent show competition, began broadcasting in the United States during 2016. **Grace VanderWaal was announced as the winner** on September 14, 2016.

Challenges in text understanding: bridging

- Who sang the national anthem at the first game of 2017 world series?

- Game 1 of the 2017 World Series: The ceremonial first pitch was thrown out by members of former Dodger Jackie Robin-son's family, including his widow Rachel. The game marked the 45th anniversary of Robin-son's death. **Keith Williams Jr., a gospel singer, performed "The Star-Spangled Banner", the national anthem.**

Challenges in text understanding: predicate entailment

- How many seats in University of Michigan stadium?
- Michigan Stadium, nicknamed “The Big House”, is the football stadium for the University of Michigan in Ann Arbor, Michigan. It is the largest stadium in the United States and the second largest stadium in the world. **Its official capacity is 107,601**.
- N seats in X <-> X’s capacity is N

Addressing the challenges

- Lexical semantics: word senses and relations between word meanings
- Semantic roles
- Coreference

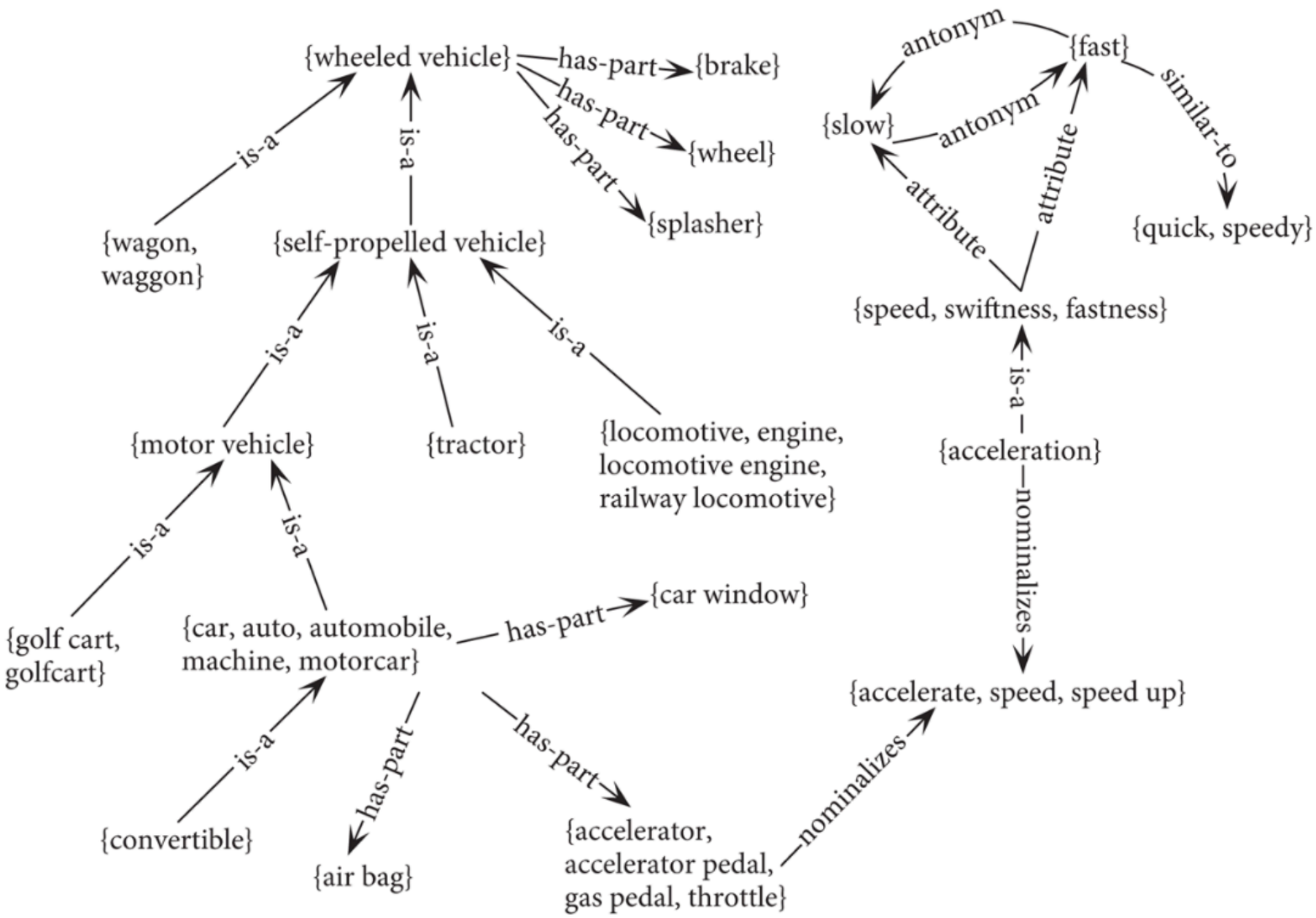
Addressing the challenges

- Pipeline approach: dedicated systems for each component of the problem
  - Rule-based
  - Machine-learning based
- End-to-end neural network training
- Fine-tuning a pre-trained sentence encoder

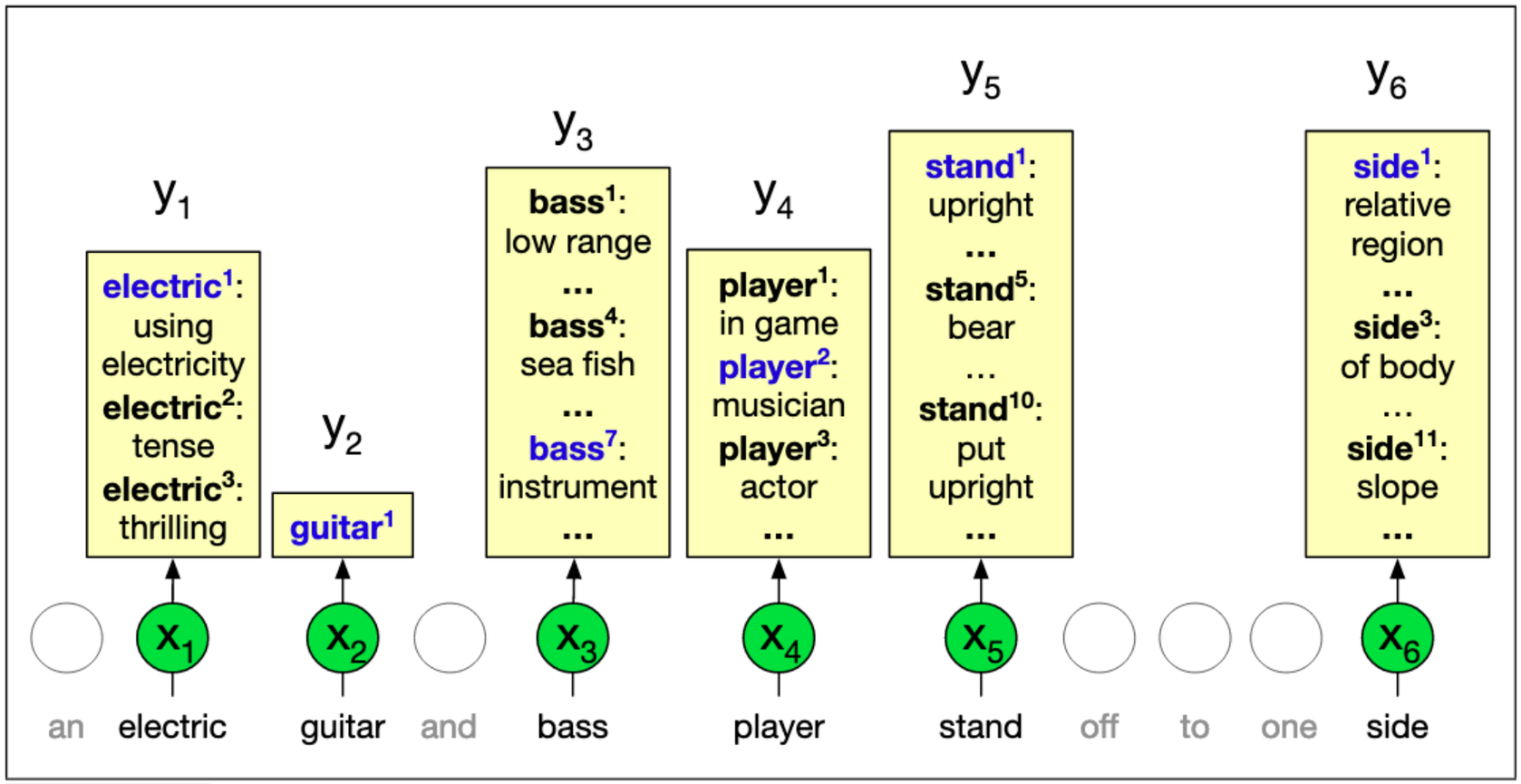
Expert-created lexical semantic resources: WordNet

- Which words are synonyms? (*couch = sofa*)
    - If a question is about couches, we can answer with a fact about sofas
- Hyponyms (*a mango is a kind of fruit*)
    - If we know something about fruits, we can extend that knowledge to mangos
- What senses does a word have?
    - Synonymy is a relation between senses! (“synsets”)

WordNet as a graph of synsets

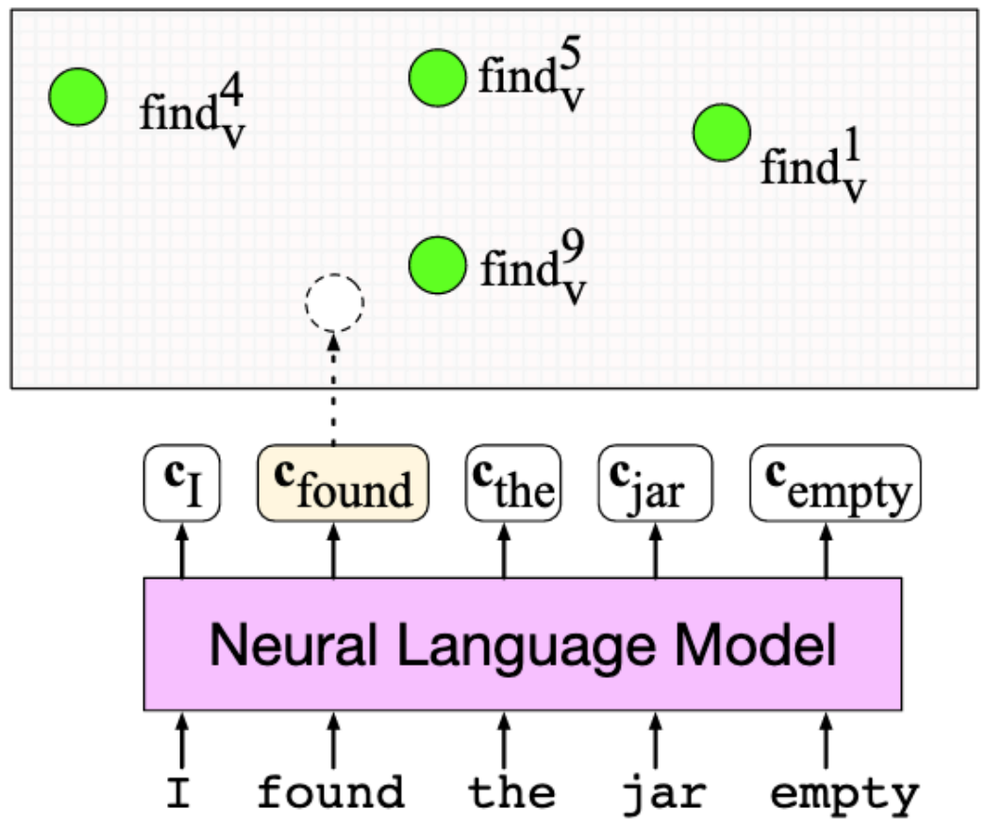


The word sense disambiguation (WSD) task



WSD with contextualized word embeddings

Contextualized embeddings from sense-annotated corpus



- Fall back on hypernym embeddings if the specific sense isn't in the training corpus

Alternate evaluation of contextualized embeddings: the words-in-context task

F	There's a lot of trash on the *bed* of the river — I keep a glass of water next to my *bed* when I sleep
F	*Justify* the margins — The end *justifies* the means
T	*Air* pollution — Open a window and let in some *air*
T	The expanded *window* will give us time to catch the thieves — You have a two-hour *window* of clear weather to finish working on the lawn

Table 1: Sample positive (T) and negative (F) pairs from the WiC dataset (target word in *italics*).

	MLP
Contextualized word-based models	
Context2vec	57.9 ± 0.9
ElMo_1	56.4 ± 0.6
ElMo_3	57.2 ± 0.8
BERT_{base}	$\mathbf{60.2} \pm 0.4$
BERT_{large}	57.4 ± 1.0

(Pilehvar & Camacho-Collados, 2018)

Semantic roles

- The same event can be expressed in a range of ways:
  - XYZ corporation bought the stock.
  - They sold the stock to XYZ corporation.
  - The stock was bought by XYZ corporation.
  - The purchase of the stock by XYZ corporation…
  - The stock purchase by XYZ corporation...

Standard inventory of thematic roles

- What do the *kicker, spiller, breaker* have in common?

Thematic Role	Example
AGENT	*The waiter* spilled the soup.
EXPERIENCER	*John* has a headache.
FORCE	*The wind* blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke *the ice*...
RESULT	The city built a *regulation-size baseball diamond*...
CONTENT	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	I flew in *from Boston*.
GOAL	I drove *to Portland*.

PropBank

Arg0-PAG: *causer of increase* (vnrole: 45.4-agent)
Arg1-PPT: *thing increasing* (vnrole: 45.4-patient, 45.6-1-patient)
Arg2-EXT: *amount increased by, EXT or MNR* (vnrole: 45.6-1-extent)
Arg3-DIR: *start point*
Arg4-GOL: *end point*

The Polish government increased home electricity charges by 150%.

Arg0: The Polish government
Rel: increased
Arg1: home electricity charges
Arg2: by 150%

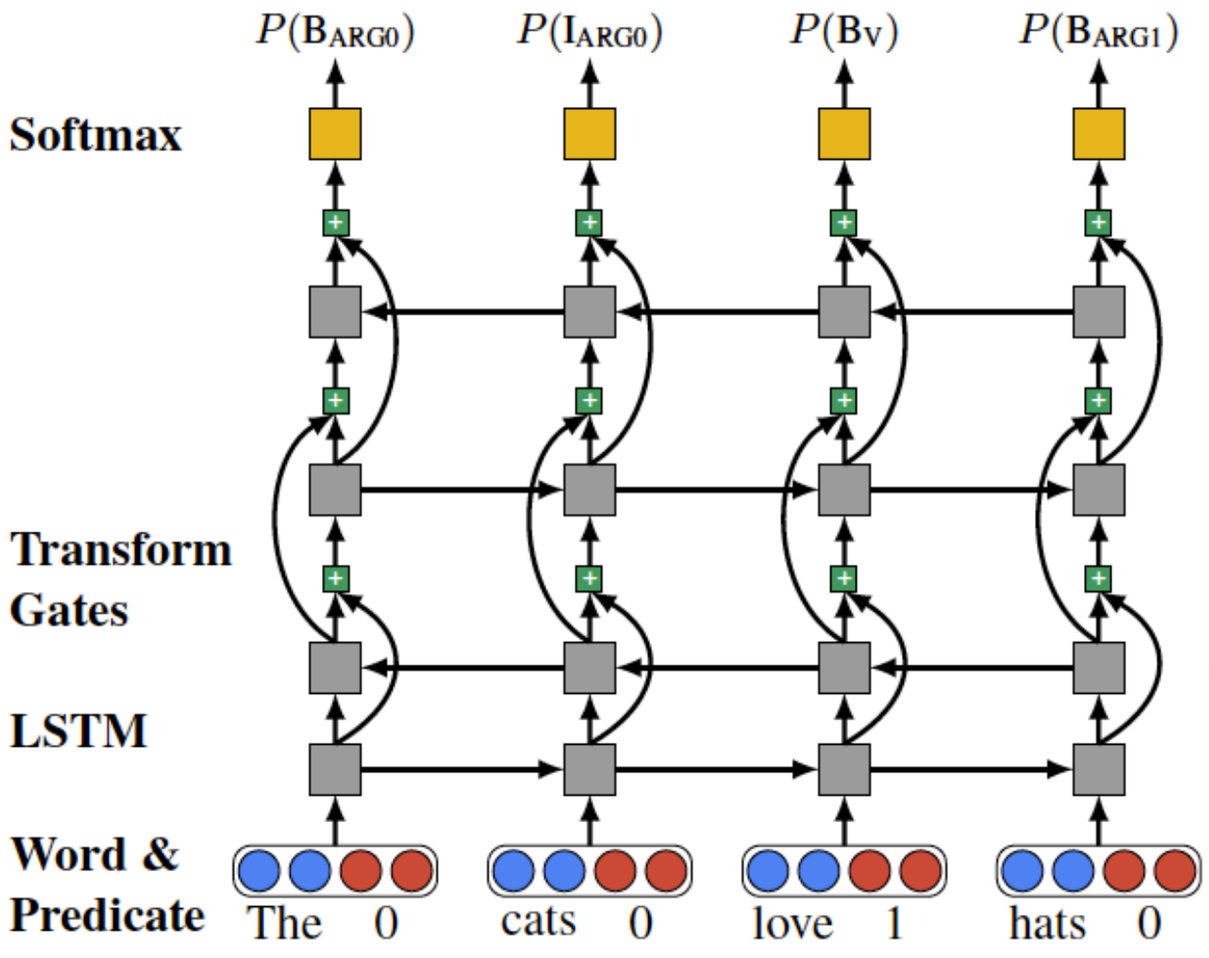
Same analysis for “home electricity charges were increased by 150% by the Polish government”!

http://verbs.colorado.edu/propbank/framesets-english-aliases/increase.html

SRL as span prediction

b-arg0	i-arg0	i-arg0	o	b-arg1	i-arg1	i-arg1
The	Polish	government	increased	home	electricity	charges

BiLSTM approach to SRL



(He et al. 2017)

BiLSTM approach to SRL

- Constraint search (using A*); instead of taking the argmax at each point, try to enforce constraints, e.g.:
  - General BIO violations: B-ARG0 I-ARG1 is an invalid sequence
  - Each core role (e.g. ARG0) should appear at most once per predicate
  - Roles should match syntactic constituents from a parser

(He et al. 2017)

Challenges in text understanding: coreference resolution (pronouns)

- Who won Wimbledon in 2019?

- *Simona Halep is a female tennis player.* ***She won Wimbledon in 2019****.*

Some factors that affect coreference

- Agreement:
    - Maryam has a theorem. She is exciting. (she=Maryam, not the theorem)
- Syntactic constraints (“binding theory”):
    - Janet bought **her** a bottle of fish sauce.
- World knowledge (Winograd schemas):
    - The trophy doesn't fit in the brown suitcase because **it**'s too big.

More (and less) than pronouns

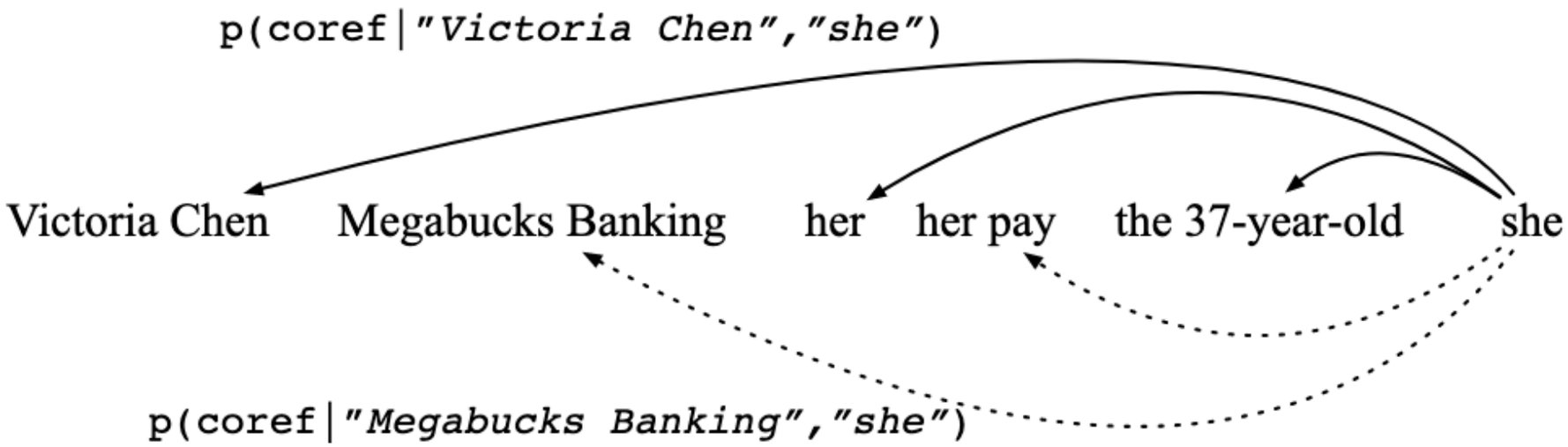
- Some pronouns don't refer to any entity
  - **It** rained all day yesterday.
  - You can make **it** in Hollywood.
- Coreference with full NPs:
  - **"Love Is in the Air"** is a 1977 disco song sung by John Paul Young. **The song** was written by George Young and Harry Vanda
- In many languages, like Italian, pronouns can (or must) be omitted:
  - **Giovanni** andò a far visita a degli amici. Per via **ϕ** comprò del vino.
  - (**Giovanni** went to visit some friends. On the way **he** bought some wine.)

How do we come up with the correct coreference chains?

[Victoria Chen]^1_a, CFO of [Megabucks Banking]^2_a, saw [[her]^1_b pay]^3_a jump to $2.3 million, as [the 38-year-old]^1_c also became [[the company]^2_b's president. It is widely known that [she]^1_d came to [Megabucks]^2_c from rival [Lotsabucks]^4_a.

1. {*Victoria Chen, her, the 38-year-old, She*}
2. {*Megabucks Banking, the company, Megabucks*}
3. { *her pay*}
4. { *Lotsabucks*}

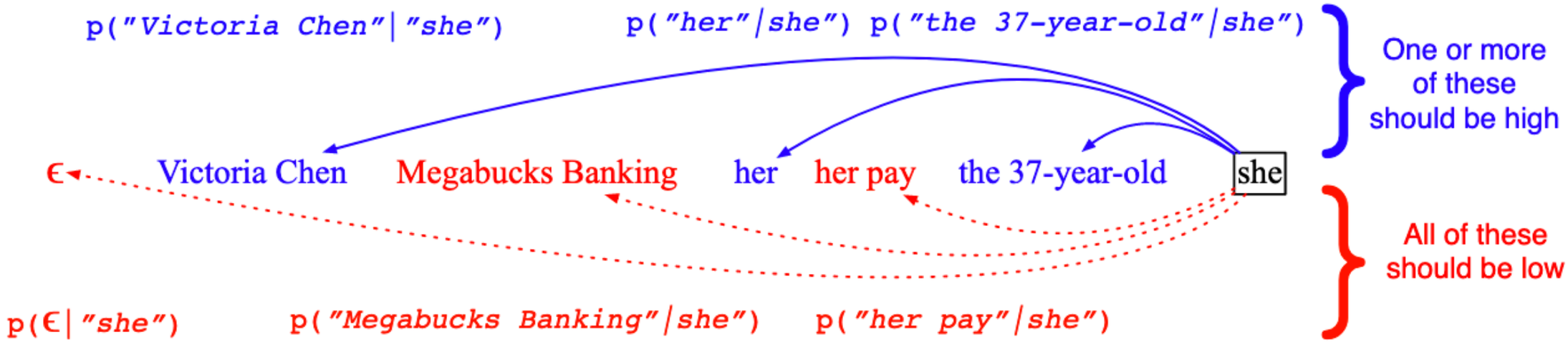
Mention-pair: do these two NPs refer to the same thing?



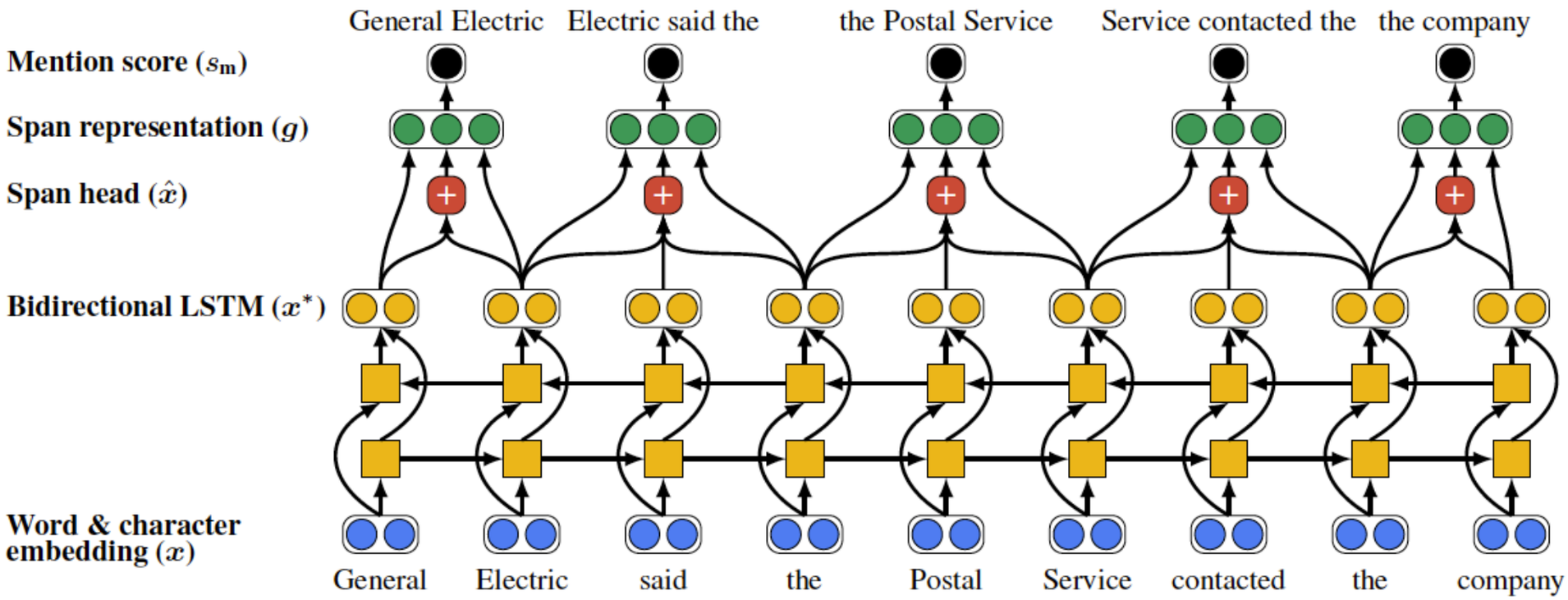
- Parse the sentence to identify noun phrases and their head words

- Train a simple classifier to make the mention-pair decision based on features of the two words: distance, head word, agreement features (singular, feminine…)

Mention ranking: a probability distribution over all previous mentions



End-to-end coreference resolution (without parsing first)

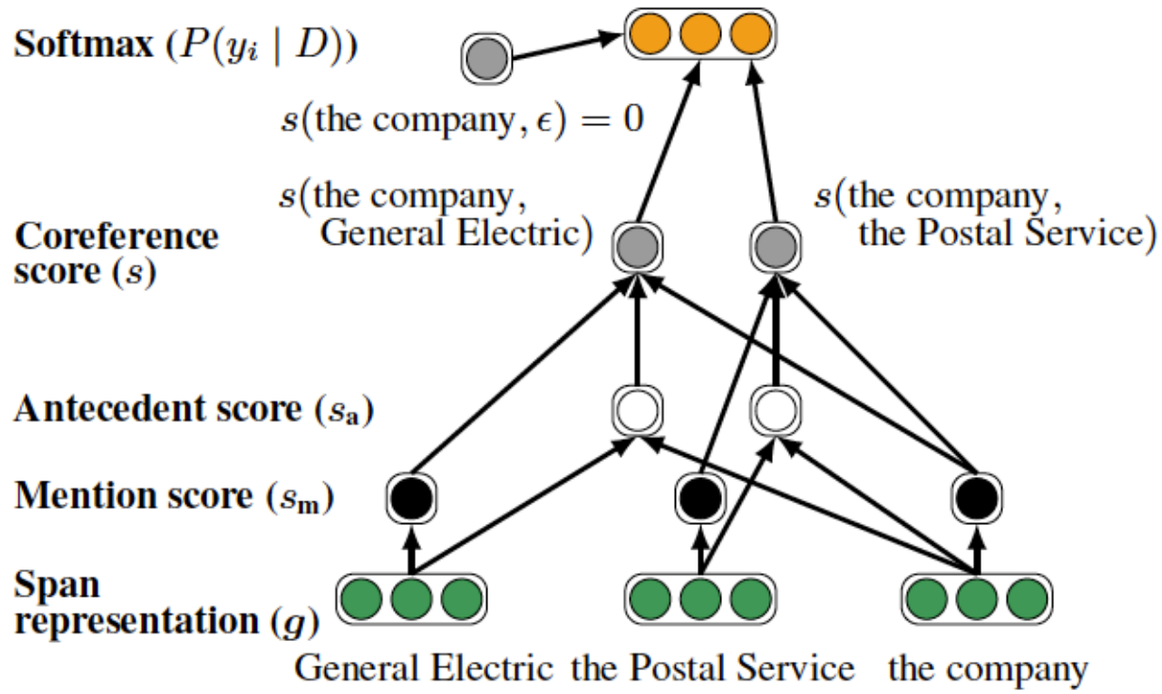


- Span representation $\mathbf{g}_i$: attention-weighted average of contextualized word embeddings
- Mention score: $\mathbf{w}_m \cdot \text{FFNN}_m(\mathbf{g}_i)$
- Score **all** spans of words, not just certified NPs

(Lee et al 2017)

End-to-end coreference resolution (without parsing first)

Mention ranking for "the company":

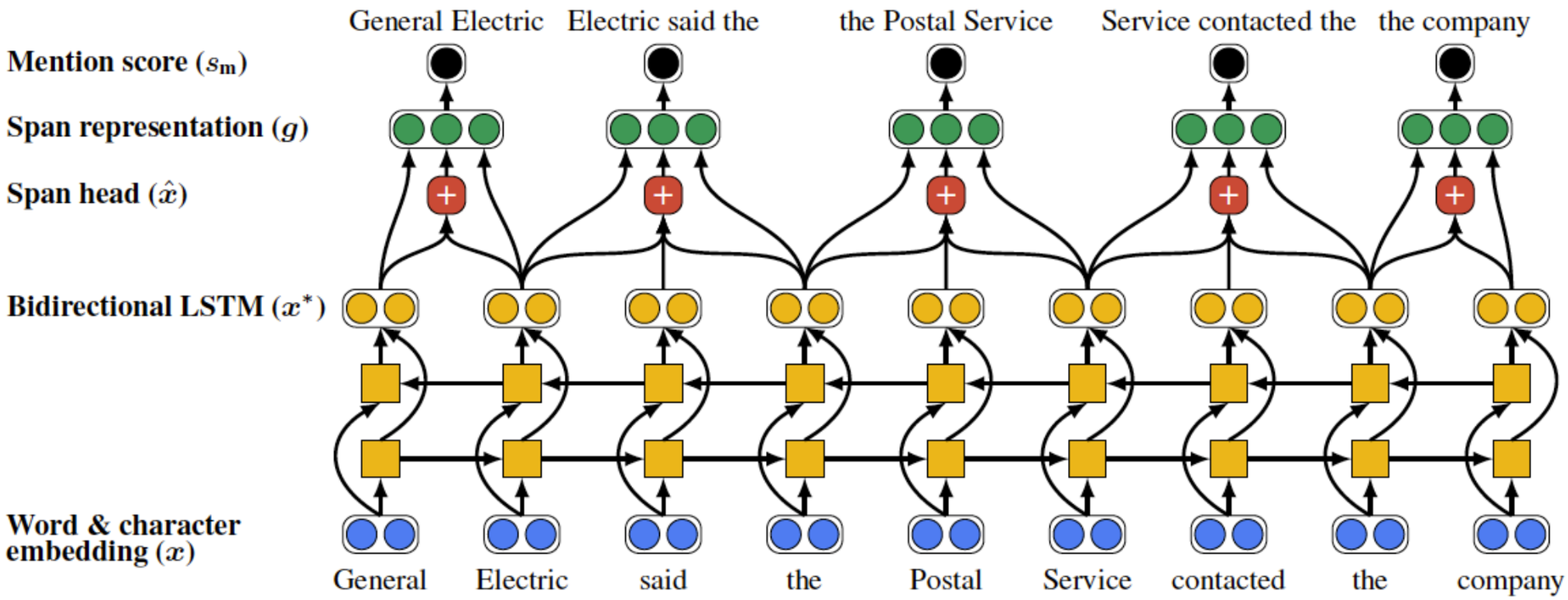


- Antecedent score: $s_a(i,j) = \mathbf{w}_a \cdot \text{FFNN}_a([\mathbf{g}_i, \mathbf{g}_j, \mathbf{g}_i \circ \mathbf{g}_j, \phi(i,j)])$

- Coreference score: sum the mention scores of both spans and their pairwise antecedent score

(Lee et al 2017)

End-to-end coreference resolution: some recent development



- Add ELMo word representations to the input layer (Peters et al 2018)
- Completely replace the BiLSTM encoder with BERT (Joshi et al 2019)