



# Natural Language Processing

Info 159/259

Lecture 21: Information Extraction (April 9, 2020)

David Bamman, UC Berkeley

Investigating(SEC, Tesla)

**LEARN TO INVEST**

Don't know where to start? Sign up  
for the Investing Basics newsletter.



# SEC Probing Tesla CEO Musk's Tweets: Reports

By [Deborah DSouza](#) | Updated August 9, 2018 — 4:47 AM EDT

On Tuesday, Tesla Inc. ([TSLA](#)) CEO Elon Musk made the dramatic announcement that he was considering taking Tesla private for \$420 a share on Twitter. In an email sent to Tesla employees [posted on the company's official blog](#), Musk explained that he is mulling taking the firm private to protect it from short sellers and wild swings in stock prices. However, the email didn't provide any details regarding financing. (See also: [What if Tesla Goes Private?](#))



Withdraws-from(Sanders, US Presidential Race)

## ***Bernie Sanders Drops Out of 2020 Democratic Race for President***



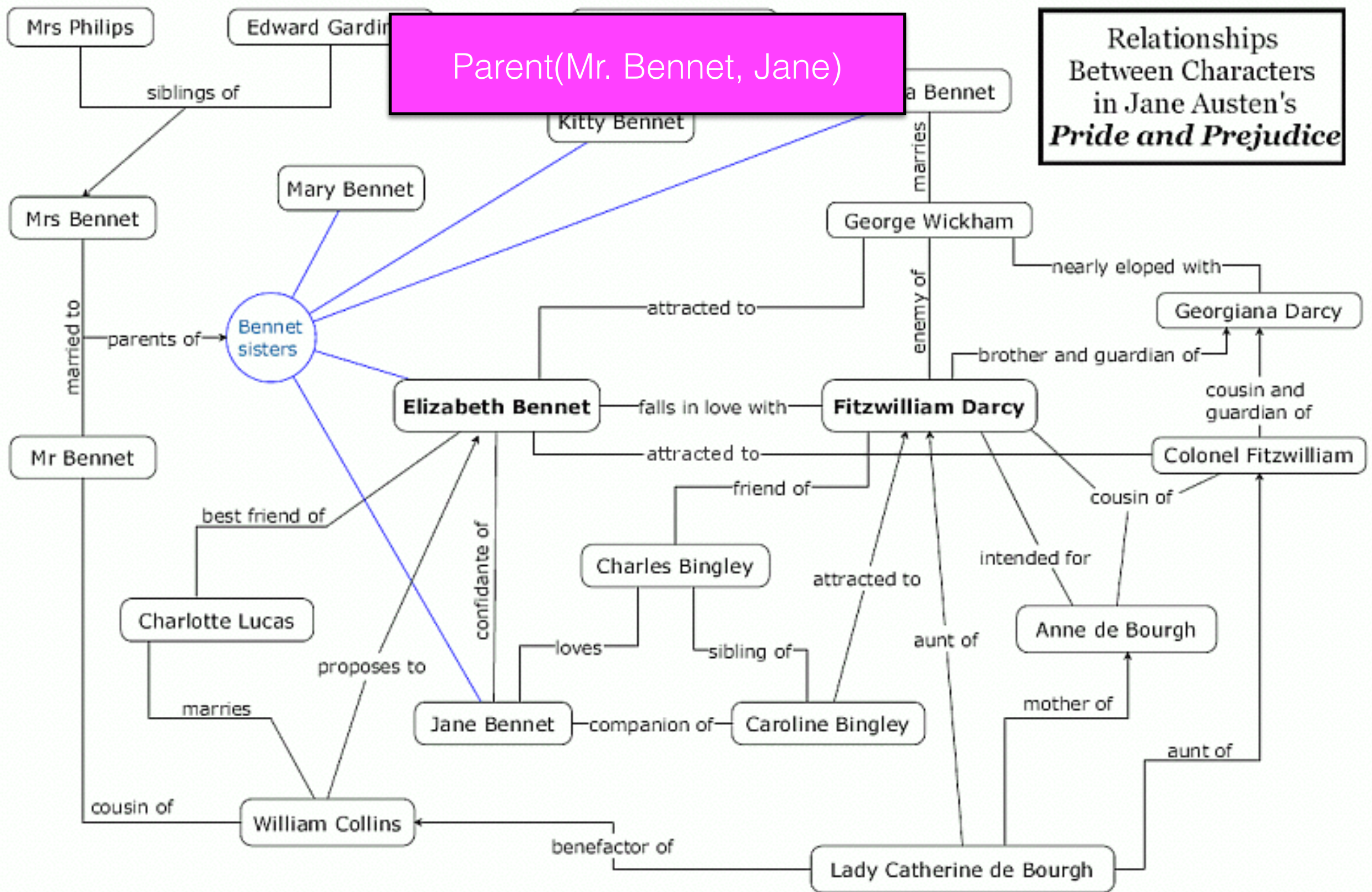
By Sydney Ember

April 8, 2020



4497

Senator [Bernie Sanders](#) of Vermont ended his presidential candidacy on Wednesday, concluding a quest that elevated him as a standard-bearer of American liberalism and clearing the way for a general election between the presumptive Democratic nominee, Joseph R. Biden Jr., and President Trump at a time of national crisis.



# Information extraction

- Named entity recognition
- Entity linking
- Relation extraction

# Named entity recognition

[tim cook]PER is the ceo of [apple]ORG

- Identifying spans of text that correspond to typed entities

# Named entity recognition

| Type                 | Tag | Sample Categories            | Example sentences                                          |
|----------------------|-----|------------------------------|------------------------------------------------------------|
| People               | PER | people, characters           | <b>Turing</b> is a giant of computer science.              |
| Organization         | ORG | companies, sports teams      | The <b>IPCC</b> warned about the cyclone.                  |
| Location             | LOC | regions, mountains, seas     | The <b>Mt. Sanitas</b> loop is in <b>Sunshine Canyon</b> . |
| Geo-Political Entity | GPE | countries, states, provinces | <b>Palo Alto</b> is raising the fees for parking.          |
| Facility             | FAC | bridges, buildings, airports | Consider the <b>Golden Gate Bridge</b> .                   |
| Vehicles             | VEH | planes, trains, automobiles  | It was a classic <b>Ford Falcon</b> .                      |

**Figure 17.1** A list of generic named entity types with the kinds of entities they refer to.

ACE NER categories (+weapon)

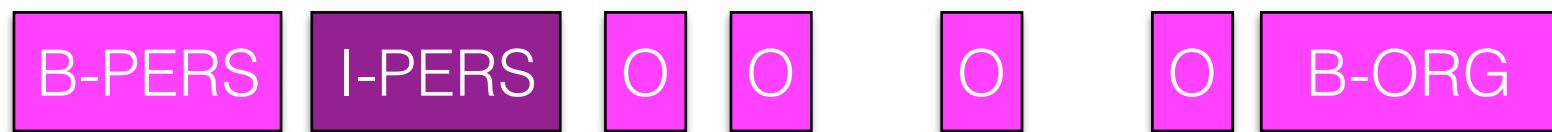
# Named entity recognition

- GENIA corpus of MEDLINE abstracts (biomedical)

We have shown that [interleukin-1]<sub>PROTEIN</sub> ([IL-1]<sub>PROTEIN</sub>) and [IL-2]<sub>PROTEIN</sub> control [IL-2 receptor alpha (IL-2R alpha) gene]<sub>DNA</sub> transcription in [CD4-CD8- murine T lymphocyte precursors]<sub>CELL LINE</sub>

|           |
|-----------|
| protein   |
| cell line |
| cell type |
| DNA       |
| RNA       |

# BIO notation



tim cook is the ceo of apple

- **B**eginning of entity
- **I**nside entity
- **O**utside entity

[tim cook]<sub>PER</sub> is the ceo of [apple]<sub>ORG</sub>

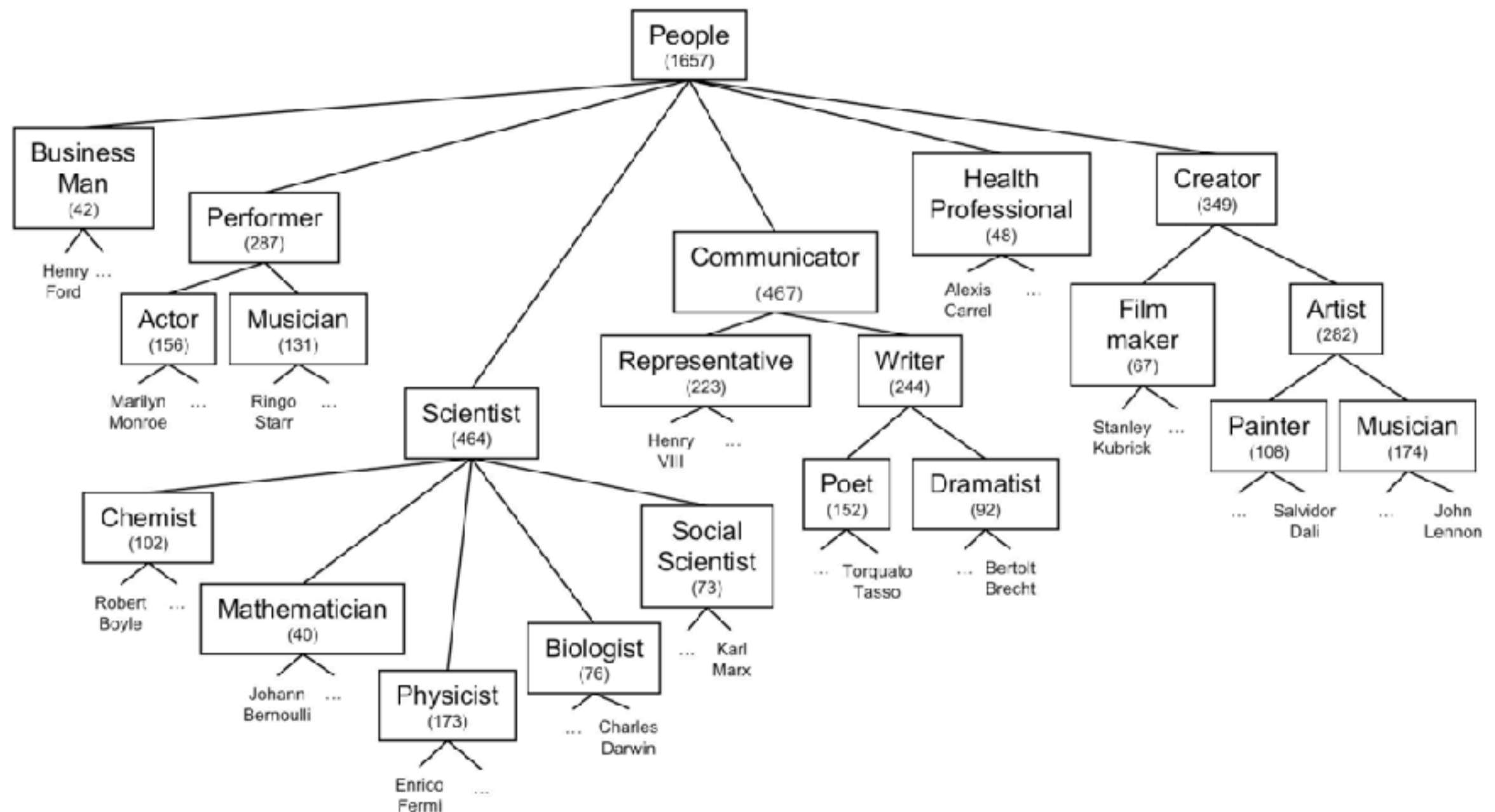
# Named entity recognition

B-PERS

B-PERS

After he saw Harry Tom went to the store

# Fine-grained NER



# Fine-grained NER

## WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations  
Display options for sense: (gloss) "an example sentence"

### Noun

- [S: \(n\) Brecht](#), **Bertolt Brecht** (German dramatist and poet who developed a style of epic theater (1898–1956))
  - [instance](#)
    - [S: \(n\) dramatist](#), [playwright](#) (someone who writes plays)
    - [S: \(n\) poet](#) (a writer of poems (the term is usually reserved for writers of good poetry))

# Entity recognition

|              |                                                                     |
|--------------|---------------------------------------------------------------------|
| Person       | ... named after [the daughter of a Mattel co-founder] ...           |
| Organization | [The Russian navy] said the submarine was equipped with 24 missiles |
| Location     | Fresh snow across [the upper Midwest] on Monday, closing schools    |
| GPE          | The [Russian] navy said the submarine was equipped with 24 missiles |
| Facility     | Fresh snow across the upper Midwest on Monday, closing [schools]    |
| Vehicle      | The Russian navy said [the submarine] was equipped with 24 missiles |
| Weapon       | The Russian navy said the submarine was equipped with [24 missiles] |

ACE entity categories

<https://www ldc.upenn.edu/sites/www ldc.upenn.edu/files/english-entities-guidelines-v6.6.pdf>

# Named entity recognition

- Most **named** entity recognition datasets have flat structure (i.e., non-hierarchical labels).
  - ✓ [The University of California]**ORG**
  - ✗ [The University of [California]**GPE**]**ORG**
- Mostly fine for **named** entities, but more problematic for general entities:

[[John]**PER**'s mother]**PER** said ...

# Nested NER

|       |       |       |          |       |       |        |            |
|-------|-------|-------|----------|-------|-------|--------|------------|
| named | after | the   | daughter | of    | a     | Mattel | co-founder |
| B-ORG |       |       |          |       |       |        |            |
|       |       |       |          |       | B-PER | I-PER  | I-PER      |
| B-PER |       | I-PER | I-PER    | I-PER | I-PER | I-PER  | I-PER      |

# Sequence labeling

$$x = \{x_1, \dots, x_n\}$$

$$y = \{y_1, \dots, y_n\}$$

- For a set of inputs  $x$  with  $n$  sequential time steps, one corresponding label  $y_i$  for each  $x_i$
- Model correlations in the labels  $y$ .

# Sequence labeling

- Feature-based models (MEMM, CRF)

identity of  $w_i$ , identity of neighboring words  
embeddings for  $w_i$ , embeddings for neighboring words  
part of speech of  $w_i$ , part of speech of neighboring words  
base-phrase syntactic chunk label of  $w_i$  and neighboring words  
presence of  $w_i$  in a **gazetteer**  
 $w_i$  contains a particular prefix (from all prefixes of length  $\leq 4$ )  
 $w_i$  contains a particular suffix (from all suffixes of length  $\leq 4$ )  
 $w_i$  is all upper case  
word shape of  $w_i$ , word shape of neighboring words  
short word shape of  $w_i$ , short word shape of neighboring words  
presence of hyphen

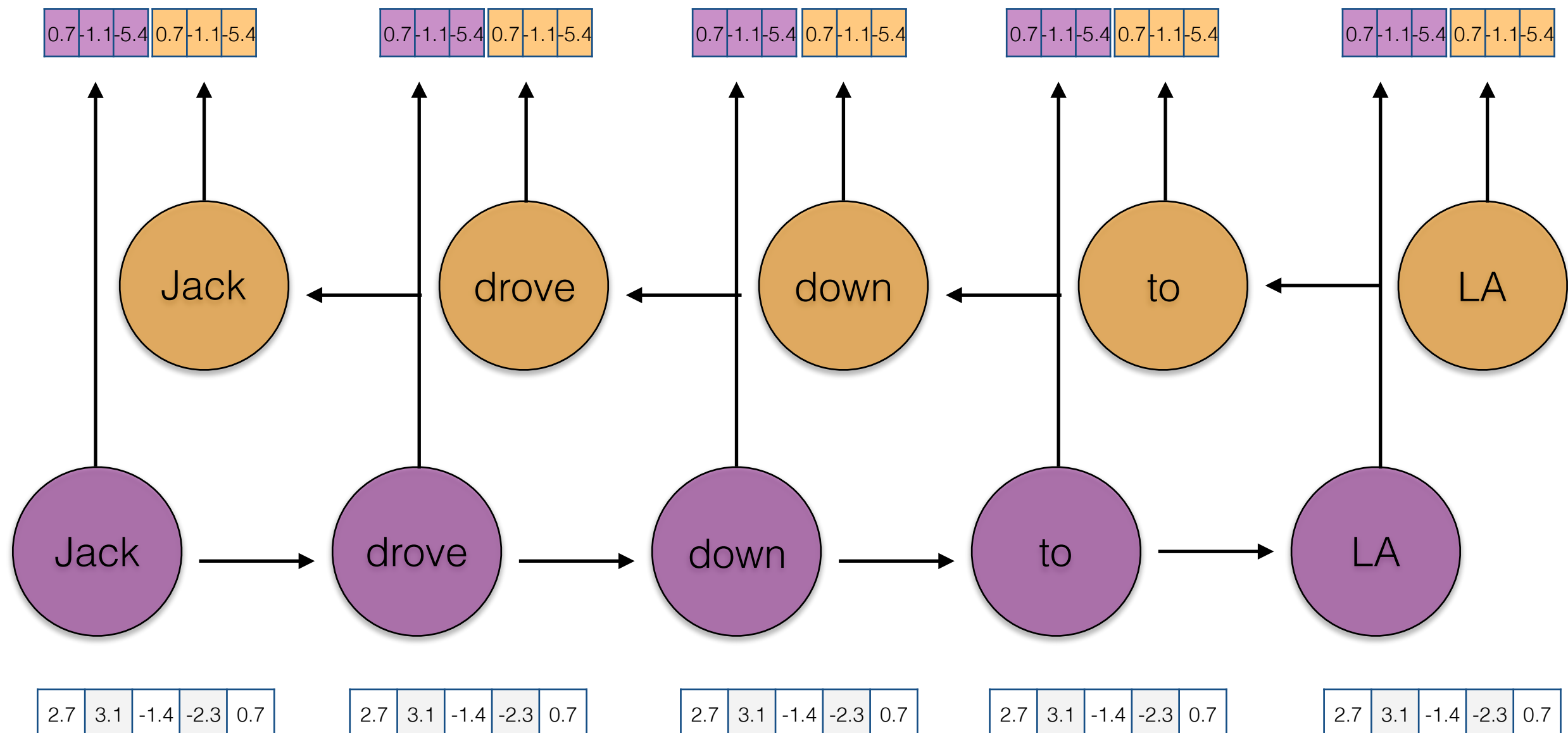
**Figure 17.5** Typical features for a feature-based NER system.

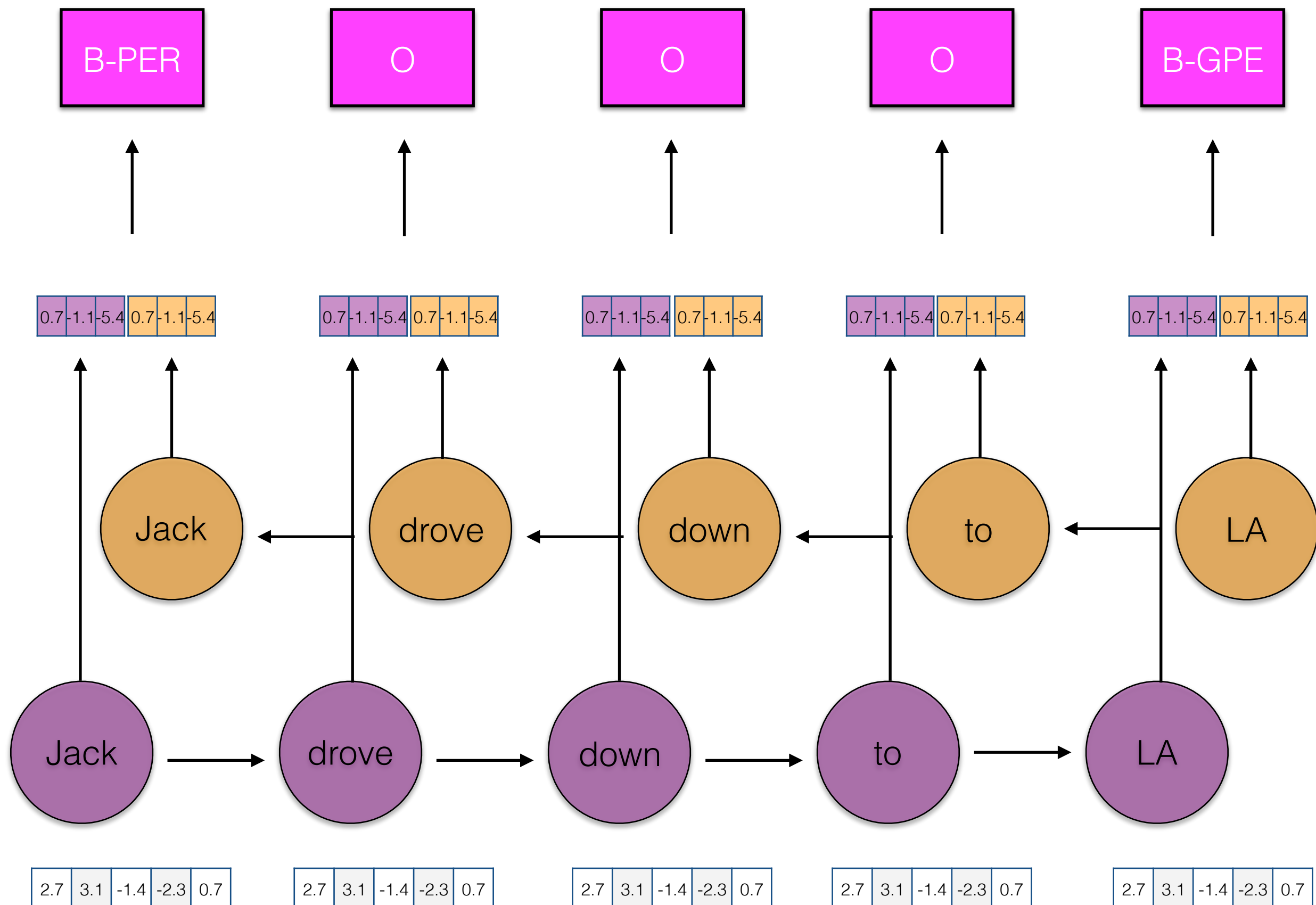
# Gazetteers

- List of place names; more generally, list of names of some typed category
- GeoNames (GEO), US SSN (PER), Getty Thesaurus of Geographic Placenames, Getty Thesaurus of Art and Architecture

Bun Crannich  
Dromore West  
Dromore  
Youghal Harbour  
Youghal Bay  
Youghal  
Eochail  
Yellow River  
Yellow Furze  
Woodville  
Wood View  
Woodtown House  
Woodstown  
Woodstock House  
Woodsgift House  
Woodrooff House  
Woodpark  
Woodmount  
Wood Lodge  
Woodlawn Station  
Woodlawn  
Woodlands Station  
Woodhouse  
Wood Hill  
Woodfort  
Woodford River  
Woodford  
Woodfield House  
Woodenbridge Junction Station  
Woodenbridge  
Woodbrook House  
Woodbrook  
Woodbine Hill  
Wingfield House  
Windy Harbour  
Windy Gap

# Bidirectional RNN

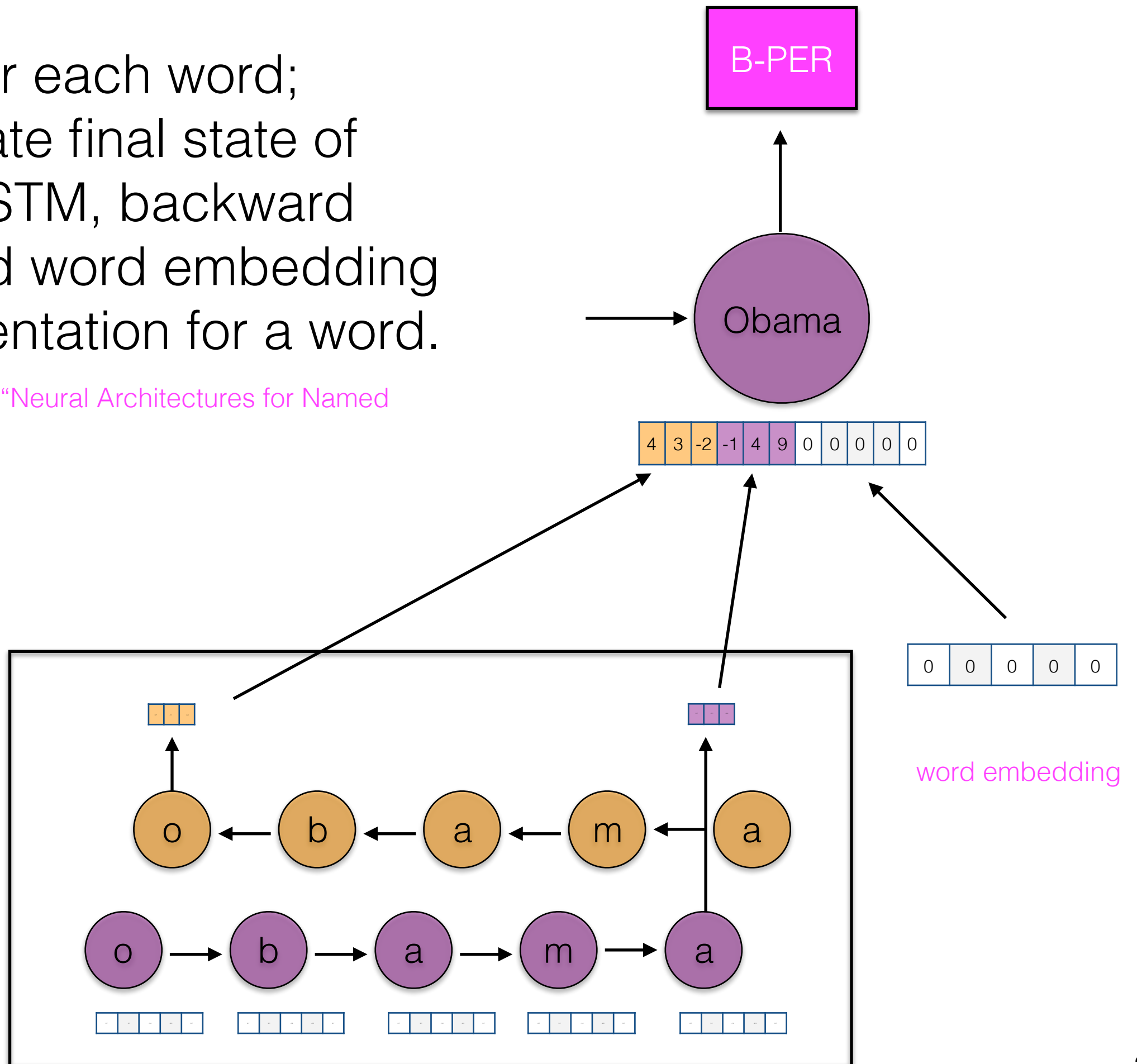




BiLSTM for each word;  
concatenate final state of  
forward LSTM, backward  
LSTM, and word embedding  
as representation for a word.

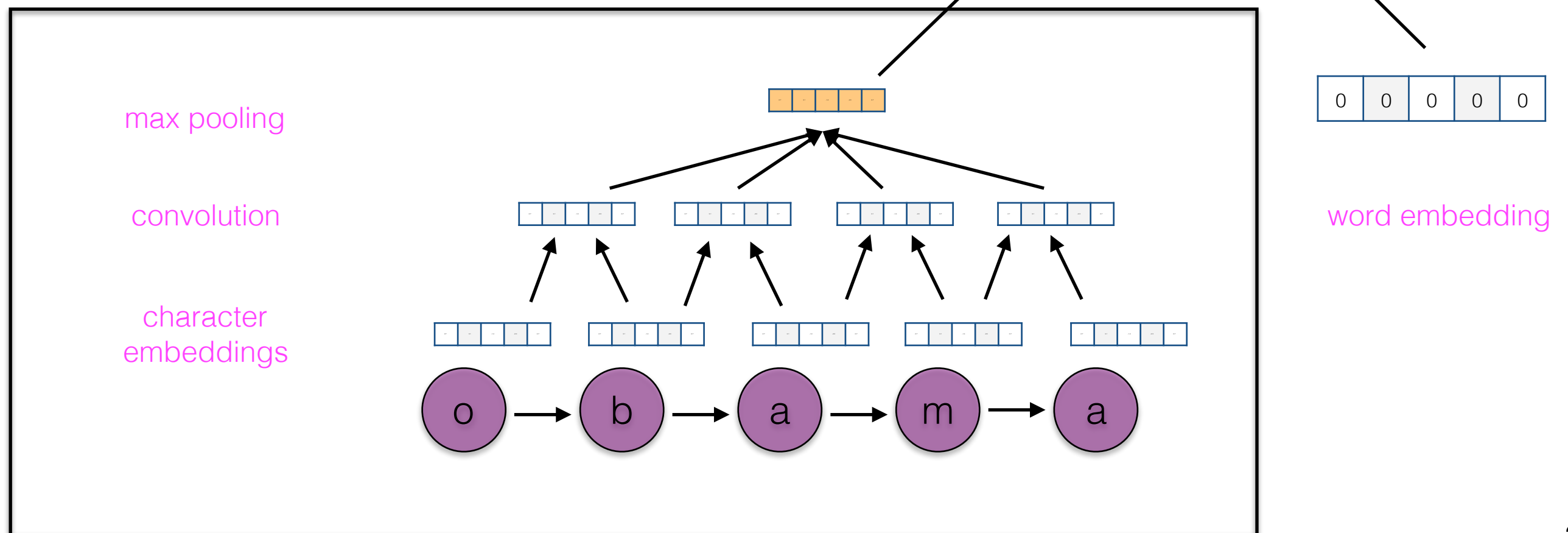
Lample et al. (2016), "Neural Architectures for Named  
Entity Recognition"

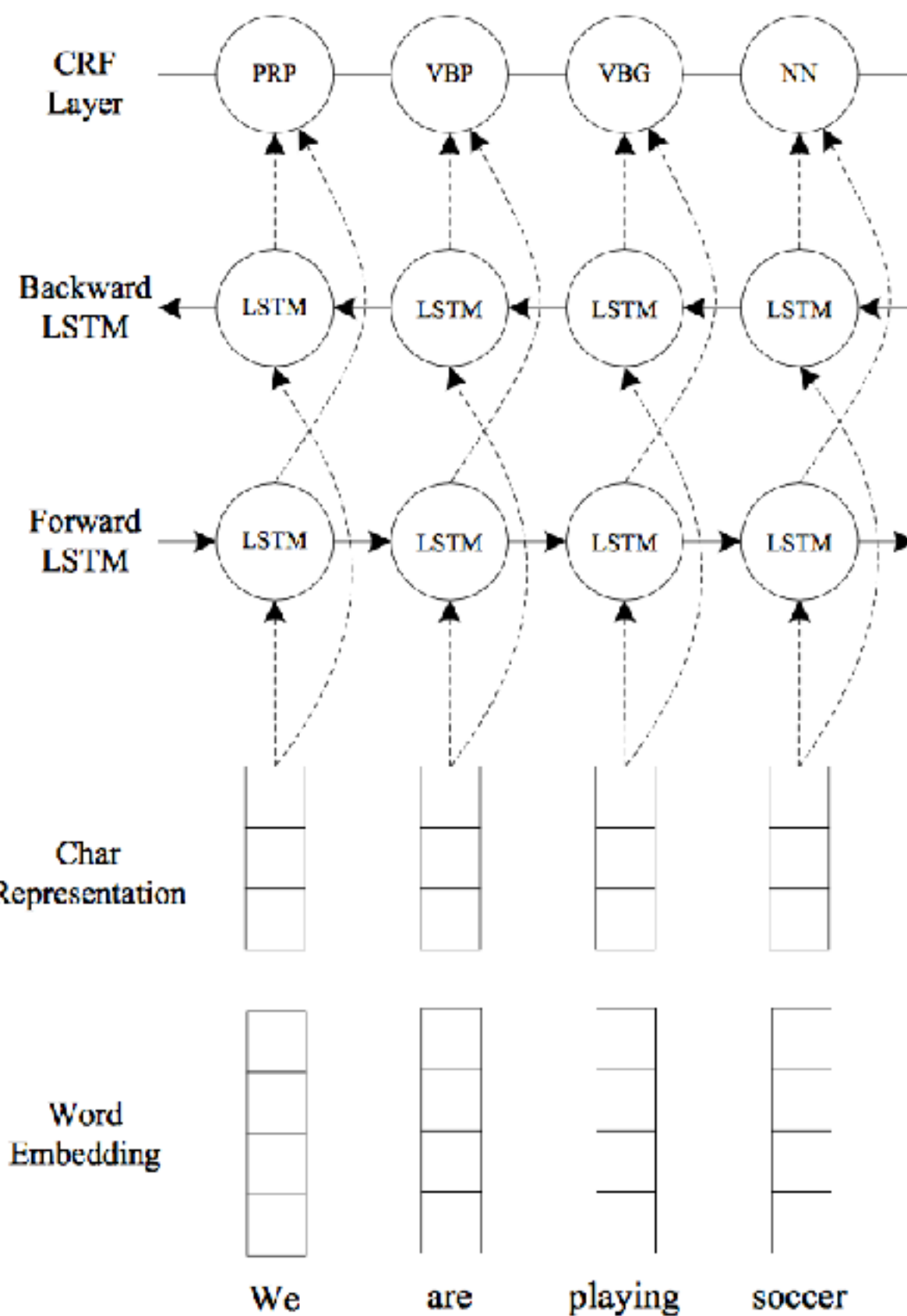
character BiLSTM



Character CNN for each word;  
concatenate character CNN  
output and word embedding  
as representation for a word.

Chu et al. (2016), "Named Entity Recognition with  
Bidirectional LSTM-CNNs"





Ma and Hovy (2016), "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF"

Some transitions (e.g. **O I-PER**) are impossible should never be predicted in NER (or any task that uses BIO notation)

| Model        | POS   |       | NER   |        |       |       |        |       |
|--------------|-------|-------|-------|--------|-------|-------|--------|-------|
|              | Dev   | Test  | Dev   |        |       | Test  |        |       |
|              | Acc.  | Acc.  | Prec. | Recall | F1    | Prec. | Recall | F1    |
| BRNN         | 96.56 | 96.76 | 92.04 | 89.13  | 90.56 | 87.05 | 83.88  | 85.44 |
| BLSTM        | 96.88 | 96.93 | 92.31 | 90.85  | 91.57 | 87.77 | 86.23  | 87.00 |
| BLSTM-CNN    | 97.34 | 97.33 | 92.52 | 93.64  | 93.07 | 88.53 | 90.21  | 89.36 |
| BRNN-CNN-CRF | 97.46 | 97.55 | 94.85 | 94.63  | 94.74 | 91.35 | 91.06  | 91.21 |

Ma and Hovy (2016), “End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF”

# Evaluation

- We evaluate NER with precision/recall/F1 over typed chunks.

# Evaluation

|               | 1     | 2     | 3  | 4   | 5     | 6  | 7     |
|---------------|-------|-------|----|-----|-------|----|-------|
|               | tim   | cook  | is | the | CEO   | of | Apple |
| <i>gold</i>   | B-PER | I-PER | O  | O   | O     | O  | B-ORG |
| <i>system</i> | B-PER | O     | O  | O   | B-PER | O  | B-ORG |

<start, end, type>

|           |     |
|-----------|-----|
| Precision | 1/3 |
| Recall    | 1/2 |

*gold*

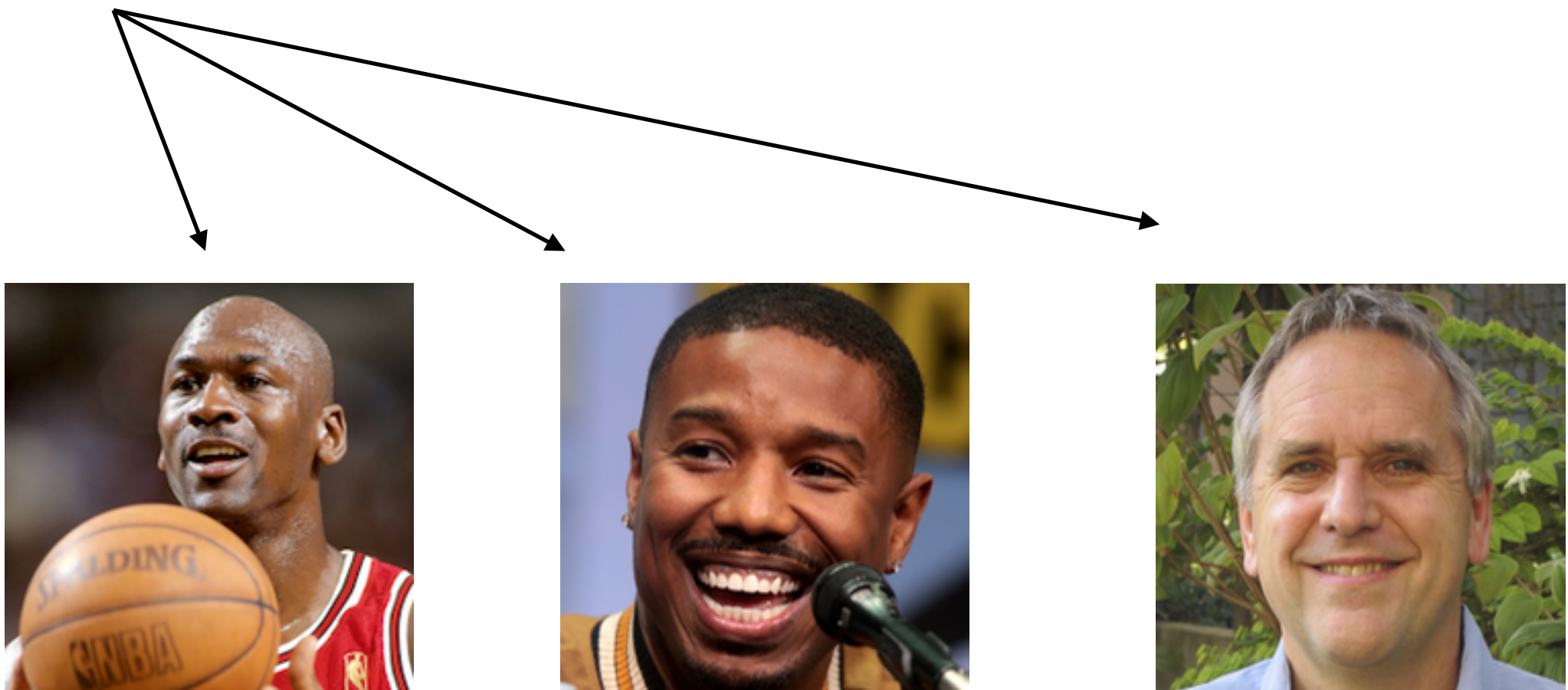
<1,2,PER>  
<7,7,ORG>

*system*

<1,1,PER>  
<5,5,PER>  
<7,7,ORG>

# Entity linking

|         |        |     |      |      |     |      |       |      |
|---------|--------|-----|------|------|-----|------|-------|------|
| Michael | Jordan | can | dunk | from | the | free | throw | line |
| B-PER   | I-PER  |     |      |      |     |      |       |      |



# Entity linking

- Task: Given a database of candidate referents, identify the correct referent for a mention in context.

| Text                                                                                                                                                       | True wikipedia page    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Hornets owner <b>Michael Jordan</b> thinks having one or two “super teams” is a detriment to the NBA because the other 28 teams “are going to be garbage.” | wiki/Michael_Jordan    |
| In 2001, <b>Michael Jordan</b> and others resigned from the Editorial Board of <i>Machine Learning</i> .                                                   | wiki/Michael_I._Jordan |
| The stars are aligning for leading man <b>Michael Jordan</b> , who just signed on for a new film, according to Variety.                                    | wiki/Michael_B._Jordan |
| <b>Michael Jordan</b> played in 1,072 regular-season games in his 15-season career                                                                         | wiki/Michael_Jordan    |

# Michael Jordan (disambiguation)

---

From Wikipedia, the free encyclopedia

**Michael Jordan** (born 1963) is an American basketball player.

**Michael** or **Mike Jordan** may also refer to:

## People [\[ edit \]](#)

---

### Sports [\[ edit \]](#)

- [Michael Jordan \(footballer\)](#) (born 1986), English goalkeeper
- [Mike Jordan \(racing driver\)](#) (born 1958), English racing driver
- [Mike Jordan \(baseball, born 1863\)](#) (1863–1940), baseball player
- [Mike Jordan \(cornerback\)](#) (born 1992), American football cornerback
- [Michael-Hakim Jordan](#) (born 1977), American professional basketball player
- [Michal Jordán](#) (born 1990), Czech ice hockey player

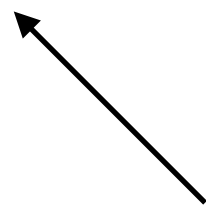
### Other people [\[ edit \]](#)

- [Michael B. Jordan](#) (born 1987), American actor
- [Michael Jordan \(insolvency baron\)](#) (born 1931), English businessman
- [Michael Jordan \(Irish politician\)](#), Irish Farmers' Party TD from Wexford, 1927–1932
- [Michael I. Jordan](#) (born 1956), American researcher in machine learning and artificial intelligence
- [Michael H. Jordan](#) (1936–2010), American executive for CBS, PepsiCo, Westinghouse
- [Michael Jordan \(mycologist\)](#), English mycologist

# Learning to rank

- Entity linking is often cast as a learning to rank problem: given a mention  $x$ , some set of candidate entities  $\mathcal{Y}(x)$  for that mention, and context  $c$ , select the **highest scoring** entity from that set.

$$\hat{y} = \arg \max_{y \in \mathcal{Y}(x)} \Psi(y, x, c)$$



Some scoring function  
over the mention  $x$ ,  
candidate  $y$ , and context  $c$

# Learning to rank

- We learn the parameters of the scoring function by minimizing the ranking loss

$$\ell(\hat{y}, y, x, c) = \max (0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1)$$

# Learning to rank

$$\ell(\hat{y}, y, x, c) = \max (0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1)$$

We suffer some loss if the predicted entity has a higher score than the true entity

$$\ell(\hat{y}, y, x, c) = \max (0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1)$$

You can't have a negative loss (if the true entity scores way higher than the predicted entity)

$$\ell(\hat{y}, y, x, c) = \max (0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1)$$

The true entity needs to score at least some constant margin better than the prediction; beyond that the higher score doesn't matter.

# Learning to rank

Some scoring function  
over the mention  $x$ ,  
candidate  $y$ , and context  $c$

$$\Psi(y, x, c)$$

| feature = $f(x, y, c)$                                   |
|----------------------------------------------------------|
| string similarity between $x$ and $y$                    |
| popularity of $y$                                        |
| NER $\text{type}(x) = \text{type}(y)$                    |
| cosine similarity between $c$ and Wikipedia page for $y$ |

$$\Psi(y, x, c) = f(x, y, c)^{\top} \beta$$

# Neural learning to rank

Parameters measuring the compatibility of the candidate and mention

Parameters measuring the compatibility of the candidate and context

$$\Psi(y, x, c) = v_y^T \Theta(x, y) x + v_y^T \Theta(y, c) c$$

Embedding  
for candidate

Embedding  
for mention

Embedding  
for context

# Learning to rank

- We learn the parameters of the scoring function by minimizing the ranking loss; take the derivative of the loss and backprop using SGD.

$$\ell(\hat{y}, y, x, c) = \max (0, \Psi(\hat{y}, x, c) - \Psi(y, x, c) + 1)$$

# Relation extraction

***The Big Sleep*** is a 1946 film noir directed by Howard Hawks,<sup>[2][3]</sup> the first film version of Raymond Chandler's 1939 novel of the same name. The film stars Humphrey Bogart as private detective Philip Marlowe and Lauren Bacall as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."<sup>[4]</sup> William Faulkner, Leigh Brackett and Jules Furthman co-wrote the screenplay.

| <i>subject</i> | <i>predicate</i> | <i>object</i>    |
|----------------|------------------|------------------|
| The Big Sleep  | directed_by      | Howard Hawks     |
| The Big Sleep  | stars            | Humphrey Bogart  |
| The Big Sleep  | stars            | Lauren Bacall    |
| The Big Sleep  | screenplay_by    | William Faulkner |
| The Big Sleep  | screenplay_by    | Leigh Brackett   |
| The Big Sleep  | screenplay_by    | Jules Furthman   |

# Relation extraction

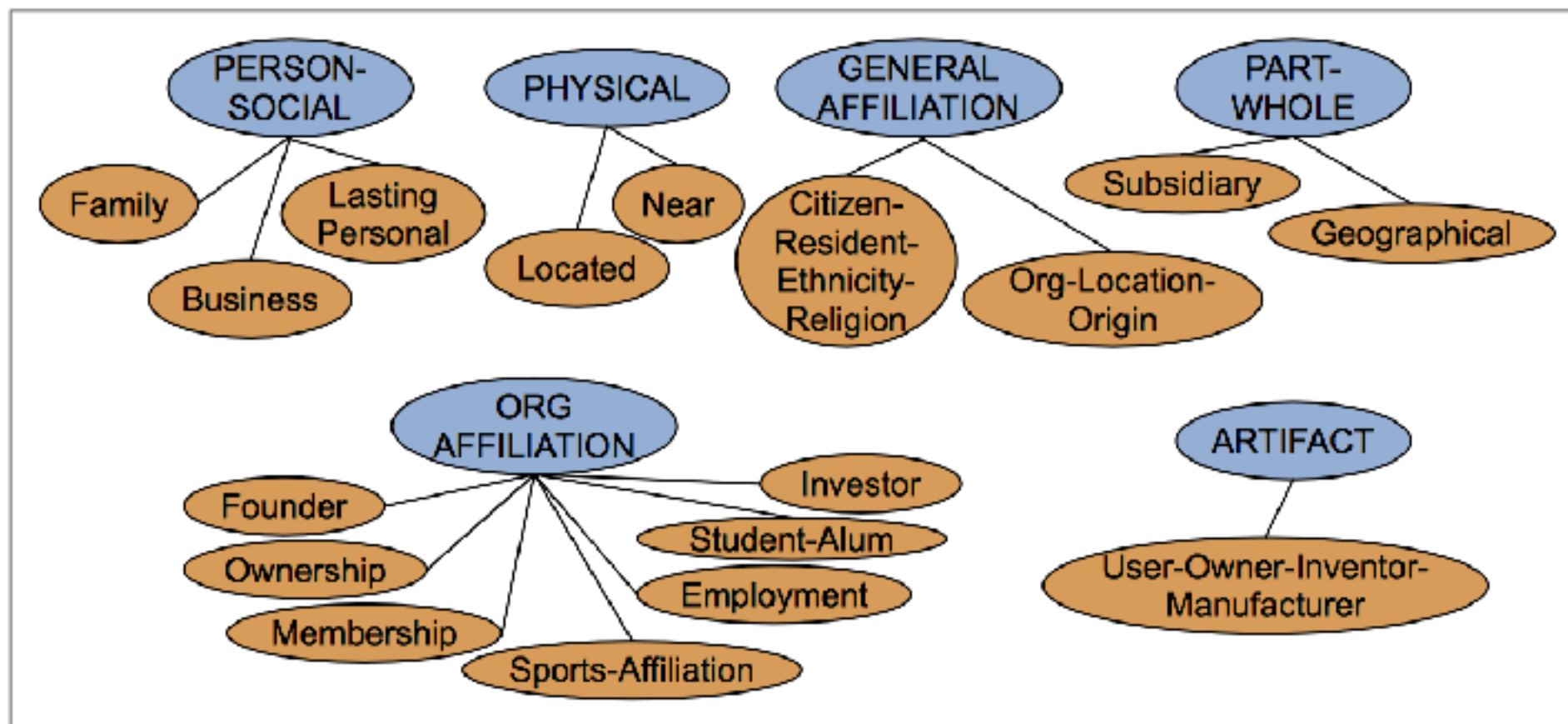


**Figure 17.9** The 17 relations used in the ACE relation extraction task.

ACE relations, SLP3

# Relation extraction

| Entity                  | Relation    | Entity                 |
|-------------------------|-------------|------------------------|
| Injury                  | disrupts    | Physiological Function |
| Bodily Location         | location-of | Biologic Function      |
| Anatomical Structure    | part-of     | Organism               |
| Pharmacologic Substance | causes      | Pathological Function  |
| Pharmacologic Substance | treats      | Pathologic Function    |

Unified Medical Language System (UMLS), SLP3

# Wikipedia Infoboxes

*The Big Sleep* is a 1946 film noir directed by Howard Hawks,<sup>[2][3]</sup> the first film version of Raymond Chandler's 1939 novel of the same name. The film stars Humphrey Bogart as private detective Philip Marlowe and Lauren Bacall as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."<sup>[4]</sup> William Faulkner, Leigh Brackett and Jules Furthman co-wrote the screenplay.

A remake starring Robert Mitchum as Philip Marlowe was released in 1978. This was the second film in three years featuring Mitchum as Marlowe. The remake was arguably more faithful to the novel, possibly due to fewer restrictions in 1978 on what could be portrayed on screen, however, it was far less successful than the original 1946 version. In 1997, the U.S. Library of Congress deemed the film "culturally, historically, or aesthetically significant," and added it to the National Film Registry.



Theatrical release lobby card

|                       |                                                                            |
|-----------------------|----------------------------------------------------------------------------|
| <b>Directed by</b>    | Howard Hawks                                                               |
| <b>Produced by</b>    | Howard Hawks                                                               |
| <b>Screenplay by</b>  | William Faulkner<br>Leigh Brackett<br>Jules Furthman                       |
| <b>Based on</b>       | <i>The Big Sleep</i><br>by Raymond Chandler                                |
| <b>Starring</b>       | Humphrey Bogart<br>Lauren Bacall                                           |
| <b>Music by</b>       | Max Steiner                                                                |
| <b>Cinematography</b> | Sidney Hickox                                                              |
| <b>Edited by</b>      | Christian Nyby                                                             |
| <b>Distributed by</b> | Warner Bros.                                                               |
| <b>Release date</b>   | August 23, 1946 (United States)                                            |
| <b>Running time</b>   | 114 minutes<br>(released cut)<br>116 minutes<br>(re-released original cut) |

# Regular expressions

- Regular expressions are precise ways of extracting high-precision relations
- “NP<sub>1</sub> is a film directed by NP<sub>2</sub>” → `directed_by`(NP<sub>1</sub>, NP<sub>2</sub>)
- “NP<sub>1</sub> was the director of NP<sub>2</sub>” → `directed_by`(NP<sub>2</sub>, NP<sub>1</sub>)

# Hearst patterns

| <i>pattern</i>                                     | <i>sentence</i>                                           |
|----------------------------------------------------|-----------------------------------------------------------|
| NP {, NP}* {,} (and or) other NP <sub>H</sub>      | temples, treasures, and other important civic buildings   |
| NP <sub>H</sub> such as {NP,}* {(or and)} NP       | red algae such as Gelidium                                |
| such NP <sub>H</sub> as {NP,}* {(or and)} NP       | such authors as Herrick, Goldsmith, and Shakespeare       |
| NP <sub>H</sub> {,} including {NP,}* {(or and)} NP | common-law countries, including Canada and England        |
| NP <sub>H</sub> {,} especially {NP}* {(or and)} NP | European countries, especially France, England, and Spain |

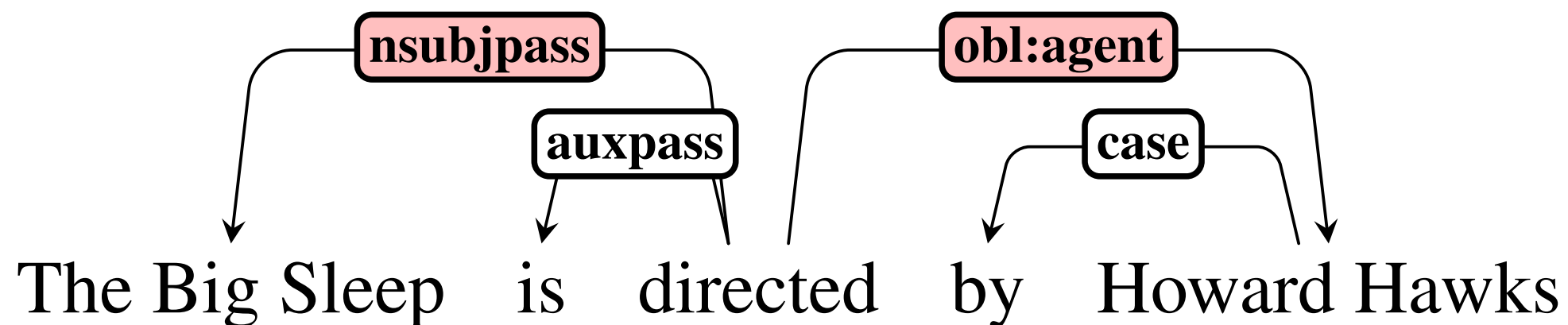
# Supervised relation extraction

[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>, the first film version of Raymond Chandler's 1939 novel of the same name.

| feature(m1, m2)               |
|-------------------------------|
| headwords of m1, m2           |
| bag of words in m1, m2        |
| bag of words between m1, m2   |
| named entity types of m1, m2  |
| syntactic path between m1, m2 |

# Supervised relation extraction

[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>, the first film version of Raymond Chandler's 1939 novel of the same name.



[The Big Sleep]<sub>m1</sub>  $\leftarrow$  *nsubjpass* directed  $\rightarrow$  *obl:agent* [Howard Hawks]<sub>m2</sub>,

$m1 \leftarrow$  *nsubjpass*  $\leftarrow$  directed  $\rightarrow$  *obl:agent*  $\rightarrow$   $m2$

# Supervised relation extraction

---

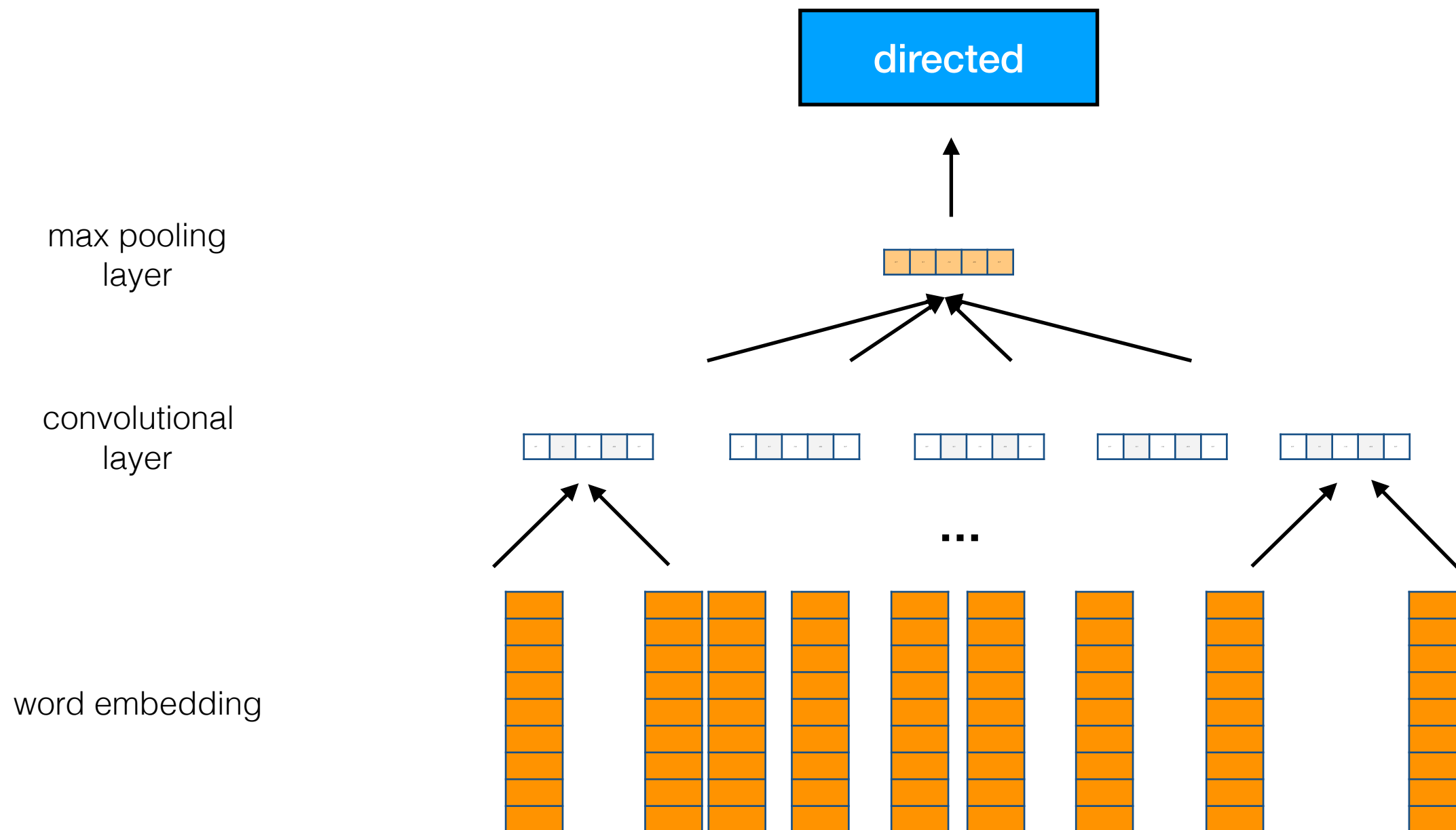
|                                                 |                                                                                                                         |
|-------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| 1. <i>George Bush traveled to France</i>        | <i>George Bush</i> $\leftarrow$ <i>traveled</i> $\rightarrow$ <i>France</i><br>NSUBJ OBL                                |
| 2. <i>Ahab traveled to Nantucket</i>            | <i>Ahab</i> $\leftarrow$ <i>traveled</i> $\rightarrow$ <i>Nantucket</i><br>NSUBJ OBL                                    |
| 3. <i>George Bush will travel to France</i>     | <i>George Bush</i> $\leftarrow$ <i>travel</i> $\rightarrow$ <i>France</i><br>NSUBJ OBL                                  |
| 4. <i>George Bush wants to travel to France</i> | <i>George Bush</i> $\leftarrow$ <i>wants</i> $\rightarrow$ <i>travel</i> $\rightarrow$ <i>France</i><br>NSUBJ XCOMP OBL |
| 5. <i>Ahab traveled to a city in France</i>     | <i>Ahab</i> $\leftarrow$ <i>traveled</i> $\rightarrow$ <i>city</i> $\rightarrow$ <i>France</i><br>NSUBJ OBL NMOD        |
| 6. <i>We await Ahab's visit to France</i>       | <i>Ahab</i> $\leftarrow$ <i>visit</i> $\rightarrow$ <i>France</i><br>NMOD:POSS NMOD                                     |

---

# Supervised relation extraction

```
function FINDRELATIONS(words) returns relations  
  
  relations  $\leftarrow$  nil  
  entities  $\leftarrow$  FINDENTITIES(words)  
  forall entity pairs  $\langle e1, e2 \rangle$  in entities do  
    if RELATED?(e1, e2)  
      relations  $\leftarrow$  relations + CLASSIFYRELATION(e1, e2)
```

**Figure 17.13** Finding and classifying the relations among entities in a text.



[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>

We don't know which entities we're classifying!

**directed(Howard Hawks, The Big Sleep)**  
**genre(The Big Sleep, Film Noir)**  
**year\_of\_release(The Big Sleep, 1946)**

# Neural RE

- To solve this, we'll add positional embeddings to our representation of each word — the distance from each word  $w$  in the sentence to  $m1$  and  $m2$

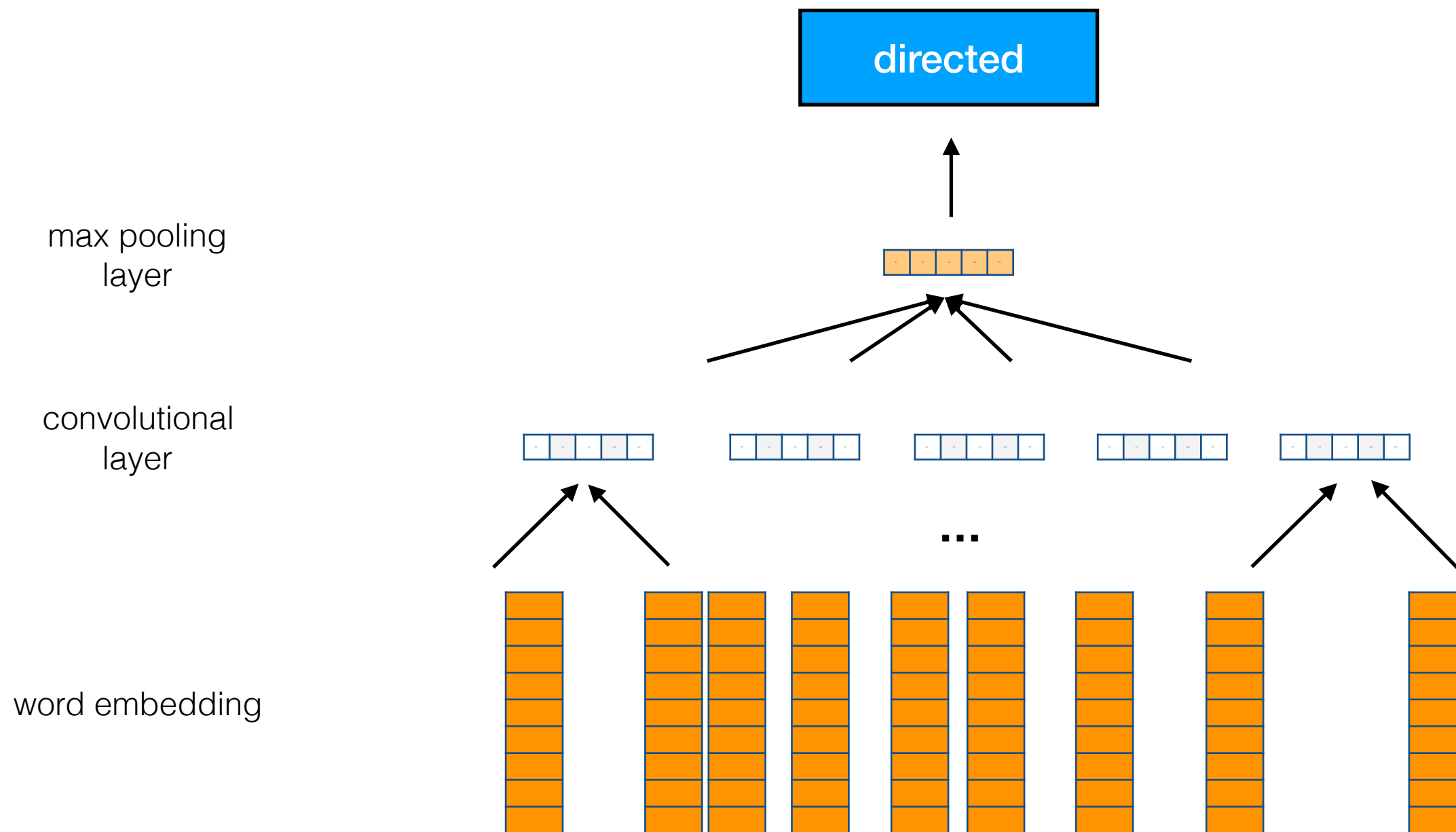
|              |                 |    |    |      |      |      |          |    |                |
|--------------|-----------------|----|----|------|------|------|----------|----|----------------|
| dist from m1 | 0               | 1  | 2  | 3    | 4    | 5    | 6        | 7  | 8              |
| dist from m2 | -8              | -7 | -6 | -5   | -4   | -3   | -2       | -1 | 0              |
|              | [The Big Sleep] | is | a  | 1946 | film | noir | directed | by | [Howard Hawks] |

- 0 here uniquely identifies the head and tail of the relation; other position indicate how close the word is (maybe closer words matter more)

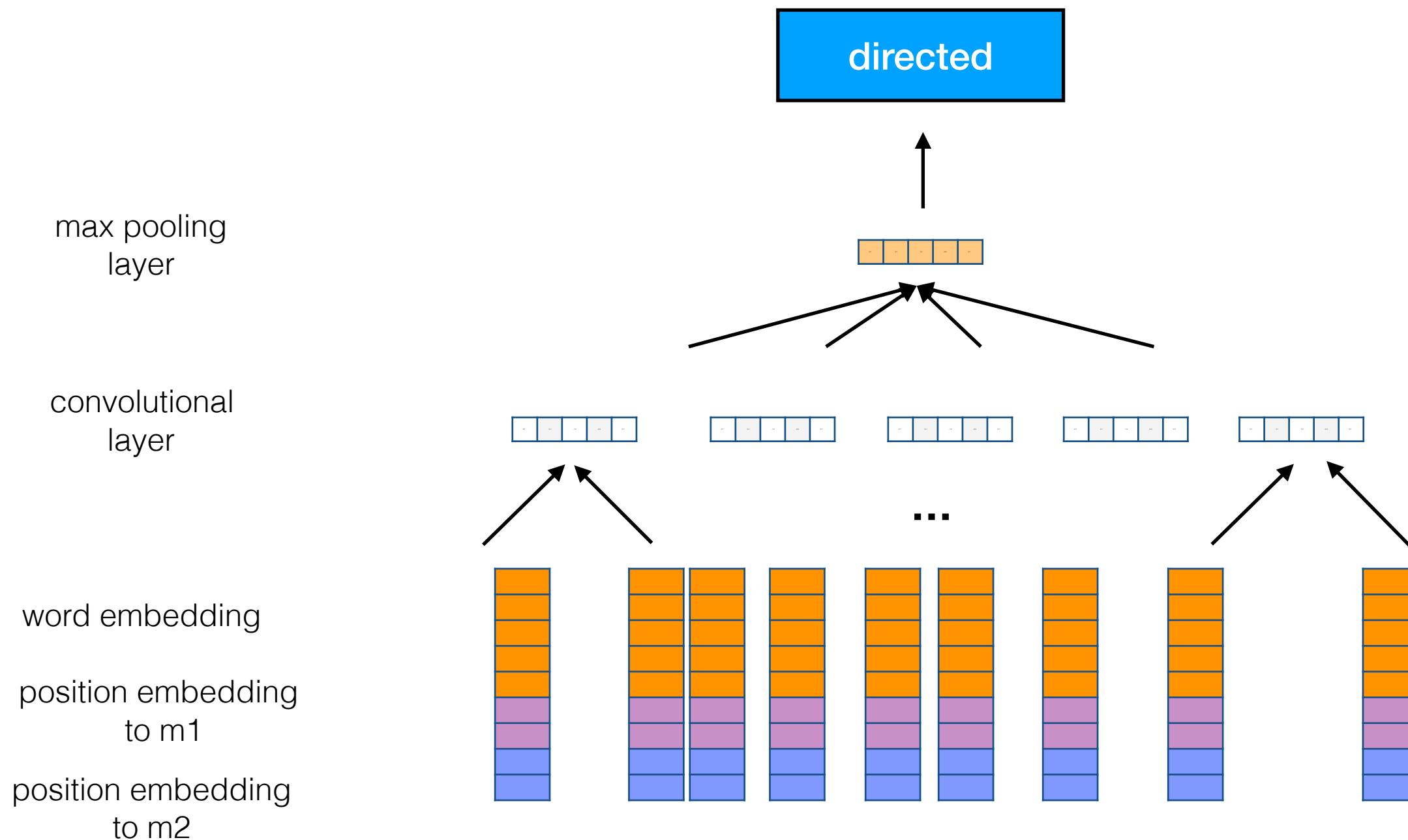
# Neural RE

Each position then has an embedding

|    |      |      |      |      |      |      |
|----|------|------|------|------|------|------|
| -4 | 2    | -0.5 | 1.1  | 0.3  | 0.4  | -0.5 |
| -3 | -1.4 | 0.4  | -0.2 | -0.9 | 0.5  | 0.9  |
| -2 | -1.1 | -0.2 | -0.5 | 0.2  | -0.8 | 0    |
| -1 | 0.7  | -0.3 | 1.5  | -0.3 | -0.4 | 0.1  |
| 0  | -0.8 | 1.2  | 1    | -0.7 | -1   | -0.4 |
| 1  | 0    | 0.3  | -0.3 | -0.9 | 0.2  | 1.4  |
| 2  | 0.8  | 0.8  | -0.4 | -1.4 | 1.2  | -0.9 |
| 3  | 1.6  | 0.4  | -1.1 | 0.7  | 0.1  | 1.6  |
| 4  | 1.2  | -0.2 | 1.3  | -0.4 | 0.3  | -1.0 |



[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>



[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>

# Distant supervision

- It's uncommon to have labeled data in the form of <sentence, relation> pairs

| <i>sentence</i>                                                                                                                                                        | <i>relations</i>                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|
| [The Big Sleep] <sub>m1</sub> is a 1946 film noir directed by [Howard Hawks] <sub>m2</sub> , the first film version of Raymond Chandler's 1939 novel of the same name. | directed_by(The Big Sleep, Howard Hawks) |

# Distant supervision

- More common to have knowledge base data about entities and their relations that's separate from text.
- We know the text likely expresses the relations somewhere, but not **exactly where**.

# Wikipedia Infoboxes

*The Big Sleep* is a 1946 film noir directed by Howard Hawks,<sup>[2][3]</sup> the first film version of Raymond Chandler's 1939 novel of the same name. The film stars Humphrey Bogart as private detective Philip Marlowe and Lauren Bacall as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."<sup>[4]</sup> William Faulkner, Leigh Brackett and Jules Furthman co-wrote the screenplay.

A remake starring Robert Mitchum as Philip Marlowe was released in 1978. This was the second film in three years featuring Mitchum as Marlowe. The remake was arguably more faithful to the novel, possibly due to fewer restrictions in 1978 on what could be portrayed on screen, however, it was far less successful than the original 1946 version. In 1997, the U.S. Library of Congress deemed the film "culturally, historically, or aesthetically significant," and added it to the National Film Registry.



Theatrical release lobby card

|                       |                                                                            |
|-----------------------|----------------------------------------------------------------------------|
| <b>Directed by</b>    | Howard Hawks                                                               |
| <b>Produced by</b>    | Howard Hawks                                                               |
| <b>Screenplay by</b>  | William Faulkner<br>Leigh Brackett<br>Jules Furthman                       |
| <b>Based on</b>       | <i>The Big Sleep</i><br>by Raymond Chandler                                |
| <b>Starring</b>       | Humphrey Bogart<br>Lauren Bacall                                           |
| <b>Music by</b>       | Max Steiner                                                                |
| <b>Cinematography</b> | Sidney Hickox                                                              |
| <b>Edited by</b>      | Christian Nyby                                                             |
| <b>Distributed by</b> | Warner Bros.                                                               |
| <b>Release date</b>   | August 23, 1946 (United States)                                            |
| <b>Running time</b>   | 114 minutes<br>(released cut)<br>116 minutes<br>(re-released original cut) |

| Relation name                           | Size    | Example                                 |
|-----------------------------------------|---------|-----------------------------------------|
| /people/person/nationality              | 281,107 | John Dugard, South Africa               |
| /location/location/contains             | 253,223 | Belgium, Nijlen                         |
| /people/person/profession               | 208,888 | Dusa McDuff, Mathematician              |
| /people/person/place_of_birth           | 105,799 | Edwin Hubble, Marshfield                |
| /dining/restaurant/cuisine              | 86,213  | MacAyo's Mexican Kitchen, Mexican       |
| /business/business_chain/location       | 66,529  | Apple Inc., Apple Inc., South Park, NC  |
| /biology/organism_classification_rank   | 42,806  | Scorpaeniformes, Order                  |
| /film/film/genre                        | 40,658  | Where the Sidewalk Ends, Film noir      |
| /film/film/language                     | 31,103  | Enter the Phoenix, Cantonese            |
| /biology/organism_higher_classification | 30,052  | Calopteryx, Calopterygidae              |
| /film/film/country                      | 27,217  | Turtle Diary, United States             |
| /film/writer/film                       | 23,856  | Irving Shulman, Rebel Without a Cause   |
| /film/director/film                     | 23,539  | Michael Mann, Collateral                |
| /film/producer/film                     | 22,079  | Diane Eskenazi, Aladdin                 |
| /people/deceased_person/place_of_death  | 18,814  | John W. Kern, Asheville                 |
| /music/artist/origin                    | 18,619  | The Octopus Project, Austin             |
| /people/person/religion                 | 17,582  | Joseph Chartrand, Catholicism           |
| /book/author/works_written              | 17,278  | Paul Auster, Travels in the Scriptorium |
| /soccer/football_position/players       | 17,244  | Midfielder, Chen Tao                    |
| /people/deceased_person/cause_of_death  | 16,709  | Richard Daintree, Tuberculosis          |
| /book/book/genre                        | 16,431  | Pony Soldiers, Science fiction          |
| /film/film/music                        | 14,070  | Stavisky, Stephen Sondheim              |
| /business/company/industry              | 13,805  | ATS Medical, Health care                |

Table 2: The 23 largest Freebase relations we use, with their size and an instance of each relation.

# Distant supervision

mayor(Maynard Jackson, Atlanta)

Elected mayor of Atlanta in 1973, Maynard Jackson...

Atlanta's airport will be renamed to honor Maynard Jackson, the city's first Black mayor

Born in Dallas, Texas in 1938, Maynard Holbrook Jackson, Jr. moved to Atlanta when he was 8.

mayor(Fiorello LaGuardia, New York)

Fiorello LaGuardia was Mayor of New York for three terms...

Fiorello LaGuardia, then serving on the New York City Board of Aldermen...

# Distant supervision

- For feature-based models, we can represent the tuple  $\langle m1, m2 \rangle$  by aggregating together the representations from all the sentences they appear in

# Distant supervision

[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>, the first film version of Raymond Chandler's 1939 novel of the same name.

[Howard Hawks]<sub>m2</sub> directed the [The Big Sleep]<sub>m1</sub>

| feature(m1, m2)                                          | value (e.g., normalized over all sentences) |
|----------------------------------------------------------|---------------------------------------------|
| “directed” between m1, m2                                | 0.37                                        |
| “by” between m1, m2                                      | 0.42                                        |
| m1 ← <i>nsubjpass</i> ← directed → <i>obl:agent</i> → m2 | 0.13                                        |
| m2 ← <i>nsubj</i> ← directed → <i>obj</i> → m2           | 0.08                                        |

# Distant supervision

- Discovering Hearst patterns from distant supervision using WordNet (Snow et al. 2005)

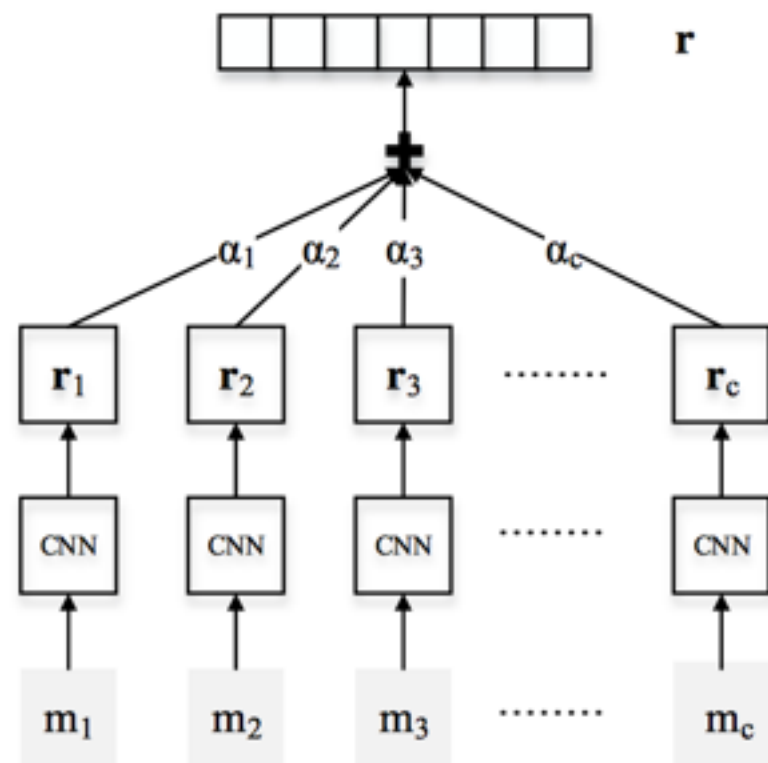
| <i>pattern</i>            | <i>sentence</i>                   |
|---------------------------|-----------------------------------|
| NP <sub>H</sub> like NP   | Many hormones like leptin...      |
| NP <sub>H</sub> called NP | a markup language called XHTML    |
| NP is a NP <sub>H</sub>   | Ruby is a programming language... |
| NP, a NP <sub>H</sub>     | IBM, a company with a long...     |

# Multiple Instance Learning

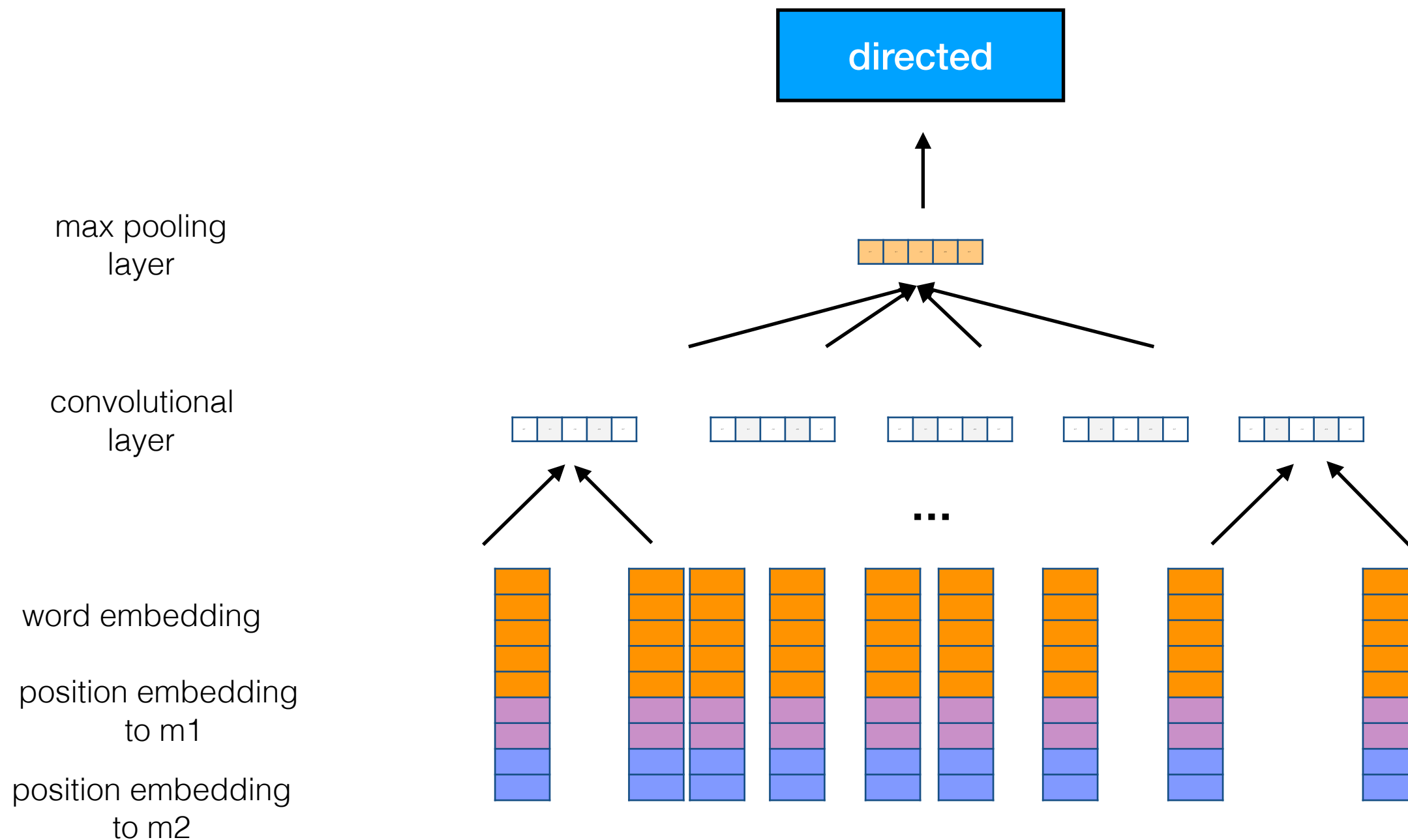
- Labels are assigned to a set of sentences, each containing the pair of entities  $m_1$  and  $m_2$ ; not all of those sentences express the relation between  $m_1$  and  $m_2$ .

# Attention

- Let's incorporate structure (and parameters) into a network that captures which **sentences** in the input we should be **attending** to (and which we can ignore).

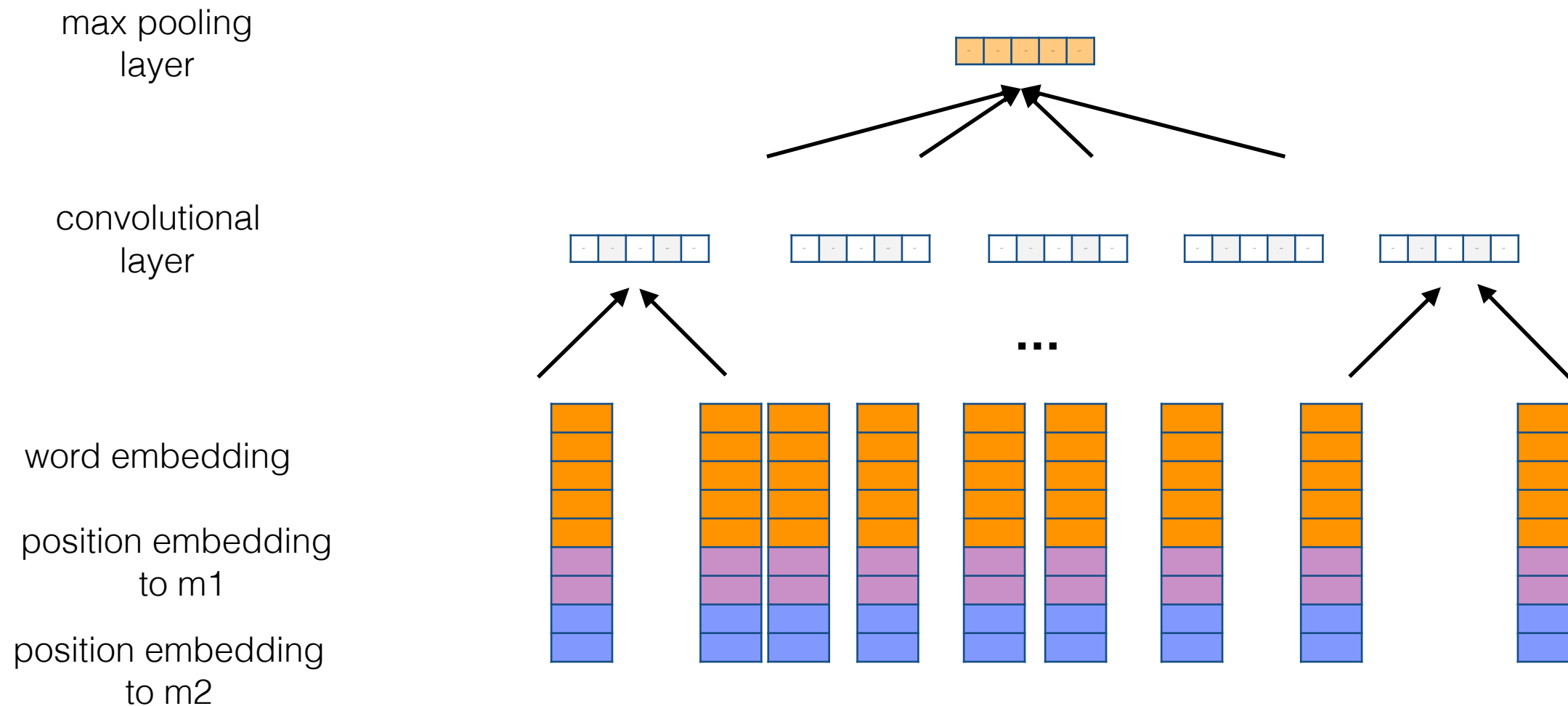


Lin et al (2016), "Neural Relation Extraction with Selective Attention over Instances" (ACL)



[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>

Now we just have an  
encoding of a sentence



[The Big Sleep]<sub>m1</sub> is a 1946 film noir directed by [Howard Hawks]<sub>m2</sub>

sentence  
encoding

[The Big Sleep]<sub>m1</sub> is a  
1946 film noir  
directed by [Howard  
Hawks]<sub>m2</sub>

|     |     |      |      |     |
|-----|-----|------|------|-----|
| 2.7 | 3.1 | -1.4 | -2.3 | 0.7 |
|-----|-----|------|------|-----|

[Howard Hawks]<sub>m2</sub>  
directed [The Big  
Sleep]<sub>m1</sub>

|     |     |      |      |     |
|-----|-----|------|------|-----|
| 2.7 | 3.1 | -1.4 | -2.3 | 0.7 |
|-----|-----|------|------|-----|

After [The Big  
Sleep]<sub>m1</sub> [Howard  
Hawks]<sub>m2</sub> married Dee  
Hartford

|     |     |      |      |     |
|-----|-----|------|------|-----|
| 2.7 | 3.1 | -1.4 | -2.3 | 0.7 |
|-----|-----|------|------|-----|

$$x_1 a_1 + x_2 a_2 + x_3 a_3$$

weighted sum

|     |     |      |      |     |
|-----|-----|------|------|-----|
| 2.7 | 3.1 | -1.4 | -2.3 | 0.7 |
|-----|-----|------|------|-----|

directed

# Information Extraction

- Named entity recognition
- Entity linking
- Relation extraction
- Template filling
- Event detection
- Event coreference
- Extra-propositional information (veridicality, hedging)