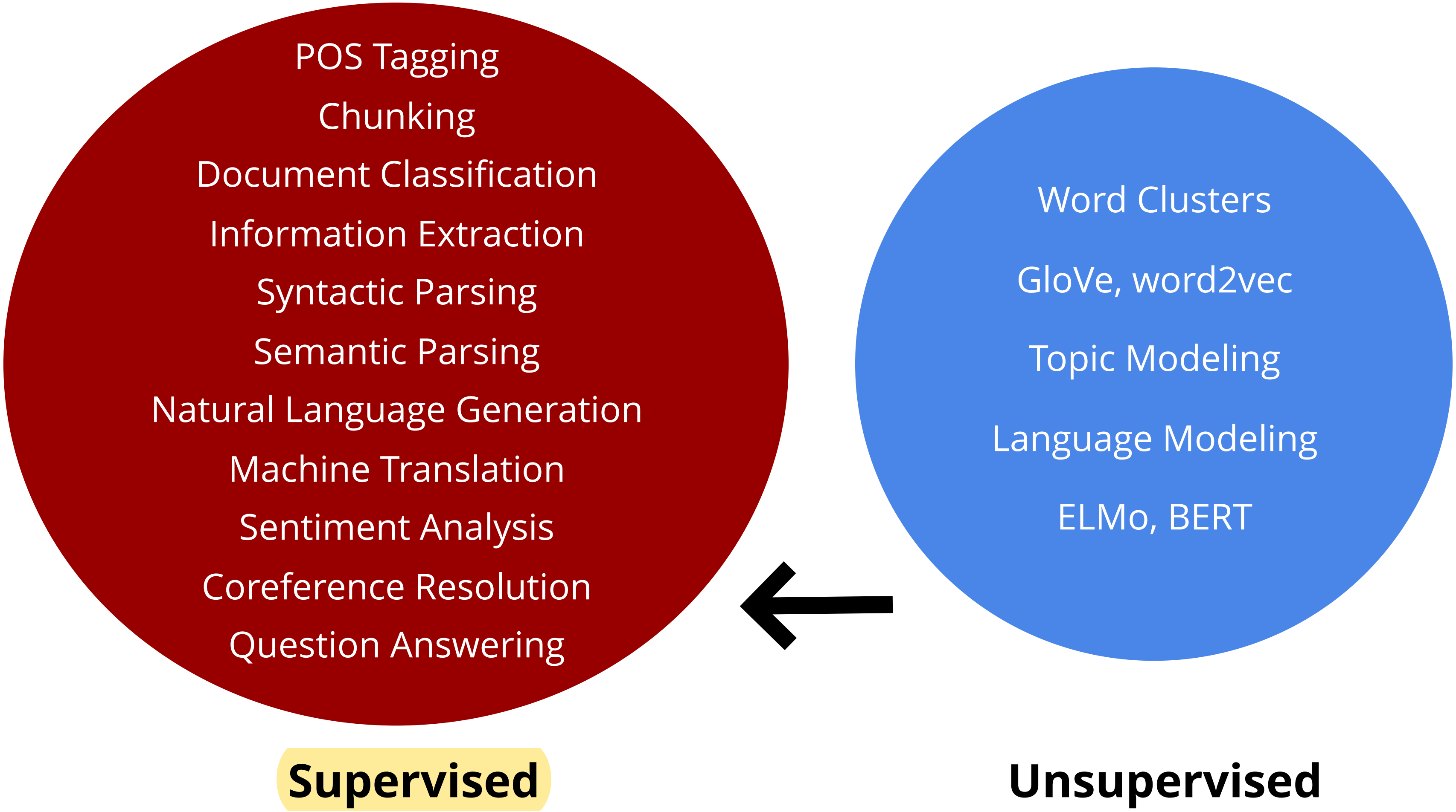
50.040
Natural Language Processing

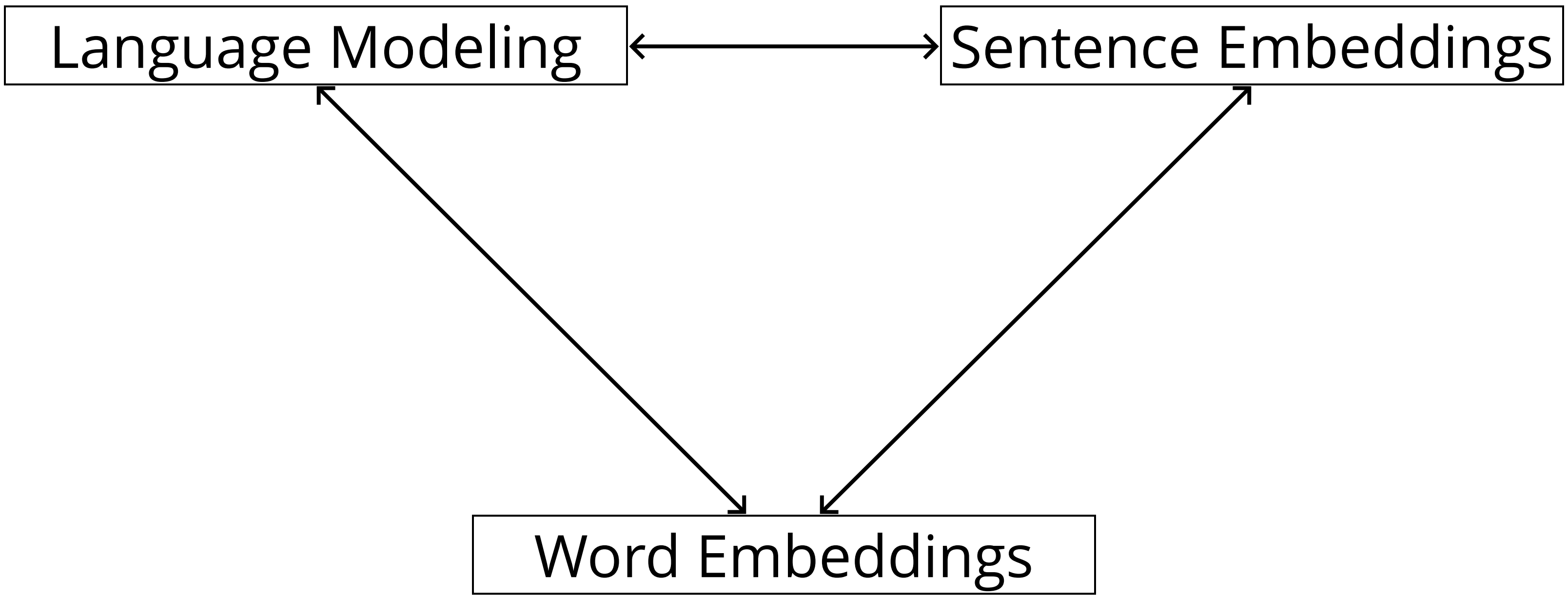
Lu, Wei

SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Tasks in NLP


POS Tagging
Chunking
Document Classification
Information Extraction
Syntactic Parsing
Semantic Parsing
Natural Language Generation
Machine Translation
Sentiment Analysis
Coreference Resolution
Question Answering
Word Clusters
GloVe, word2vec
Topic Modeling
Language Modeling
ELMo, BERT
Supervised
Unsupervised

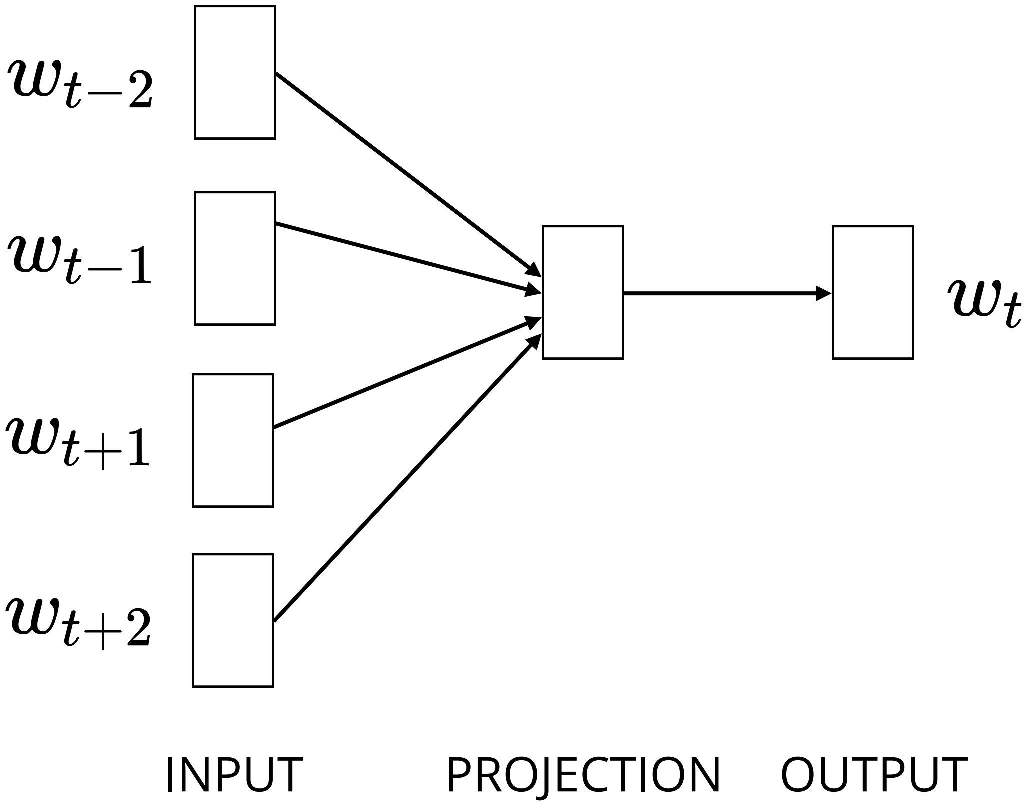
Three Tasks



These three tasks are closely related!

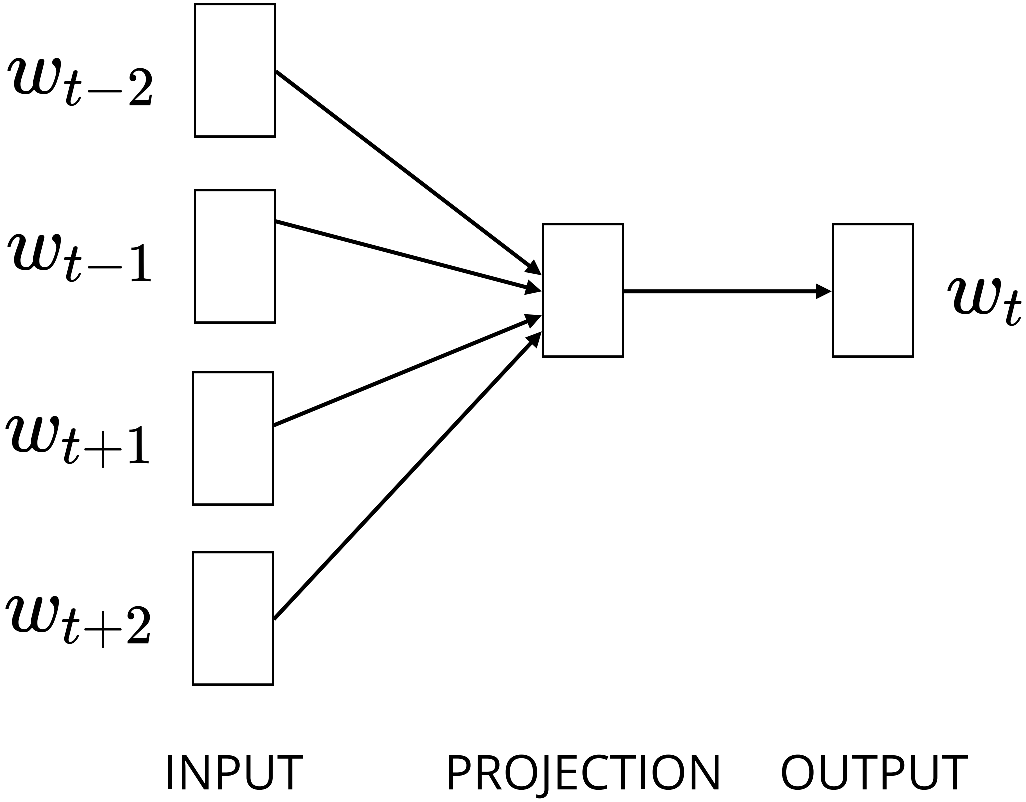
Word Embedding

Mikolov et al. (2013)



CBOW

Word Embedding



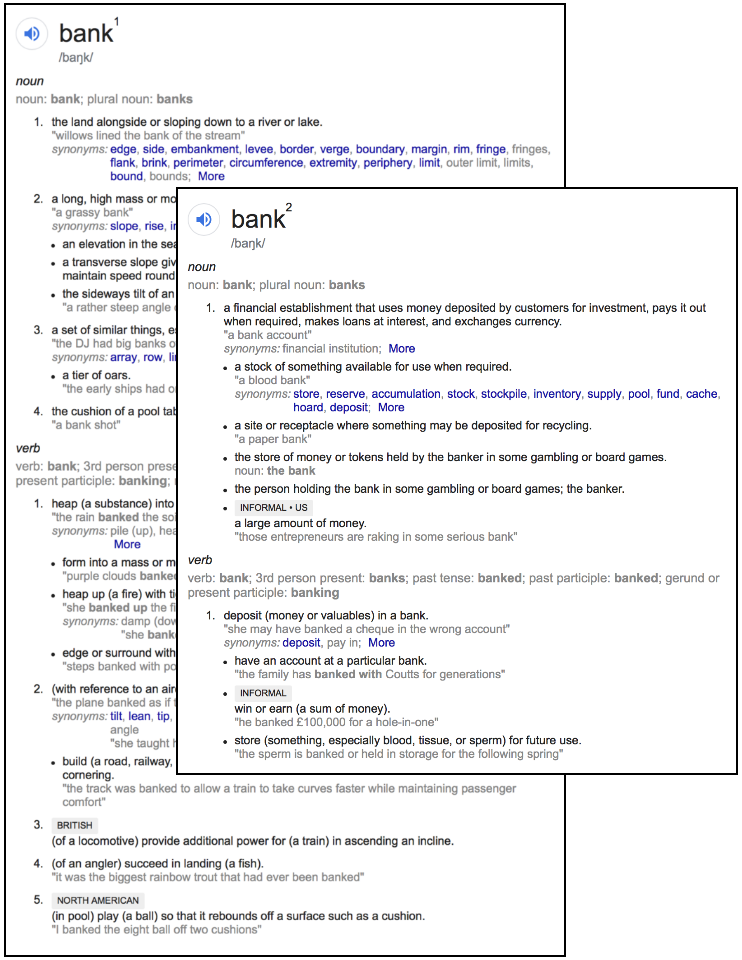
One observation:

It learns for each word token a vector.

A word may have multiple senses.

Word Senses

A word may have multiple senses.



bank[1]
/baŋk/
noun
noun: **bank**; plural noun: **banks**

1. the land alongside or sloping down to a river or lake.
"willows lined the bank of the stream"
synonyms: edge, side, embankment, levee, border, verge, boundary, margin, rim, fringe, fringes, flank, brink, perimeter, circumference, extremity, periphery, limit, outer limit, limits, bound, bounds; More
2. a long, high mass or mo
"a grassy bank"
synonyms: slope, rise, i
- an elevation in the sea
- a transverse slope giv maintain speed round
- the sideways tilt of an "a rather steep angle
3. a set of similar things, e
"the DJ had big banks o
synonyms: array, row, li
- a tier of oars.
"the early ships had o
4. the cushion of a pool tab
"a bank shot"

verb
verb: **bank**; 3rd person prese present participle: **banking**;

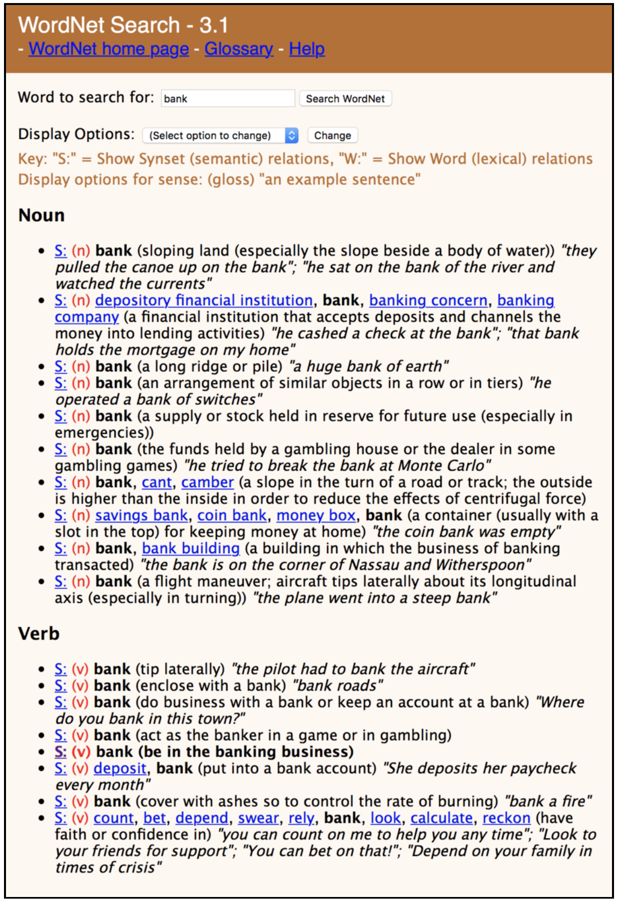
1. heap (a substance) into
"the rain **banked** the so
synonyms: pile (up), hea More
- form into a mass or m
"purple clouds **banke**
- heap up (a fire) with ti
"she **banked up** the fi
synonyms: damp (dow
"she **bank**
- edge or surround with
"steps banked with po
2. (with reference to an air
"the plane banked as if
synonyms: tilt, lean, tip, angle
"she taught
- build (a road, railway, cornering.
"the track was banked to allow a train to take curves faster while maintaining passenger comfort"
3. BRITISH
(of a locomotive) provide additional power for (a train) in ascending an incline.
4. (of an angler) succeed in landing (a fish).
"it was the biggest rainbow trout that had ever been banked"
5. NORTH AMERICAN
(in pool) play (a ball) so that it rebounds off a surface such as a cushion.
"I banked the eight ball off two cushions"

bank[2]
/baŋk/
noun
noun: **bank**; plural noun: **banks**

1. a financial establishment that uses money deposited by customers for investment, pays it out when required, makes loans at interest, and exchanges currency.
"a bank account"
synonyms: financial institution; More
- a stock of something available for use when required.
"a blood bank"
synonyms: store, reserve, accumulation, stock, stockpile, inventory, supply, pool, fund, cache, hoard, deposit; More
- a site or receptacle where something may be deposited for recycling.
"a paper bank"
- the store of money or tokens held by the banker in some gambling or board games.
noun: **the bank**
- the person holding the bank in some gambling or board games; the banker.
- INFORMAL • US
a large amount of money.
"those entrepreneurs are raking in some serious bank"

verb
verb: **bank**; 3rd person present: **banks**; past tense: **banked**; past participle: **banked**; gerund or present participle: **banking**

1. deposit (money or valuables) in a bank.
"she may have banked a cheque in the wrong account"
synonyms: deposit, pay in; More
- have an account at a particular bank.
"the family has **banked with** Coutts for generations"
- INFORMAL
win or earn (a sum of money).
"he banked £100,000 for a hole-in-one"
- store (something, especially blood, tissue, or sperm) for future use.
"the sperm is banked or held in storage for the following spring"



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: bank Search WordNet

Display Options: (Select option to change) Change

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

Noun

- S: (n) **bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
- S: (n) depository financial institution, **bank**, banking concern, banking company (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
- S: (n) **bank** (a long ridge or pile) *"a huge bank of earth"*
- S: (n) **bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- S: (n) **bank** (a supply or stock held in reserve for future use (especially in emergencies))
- S: (n) **bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- S: (n) **bank**, cant, camber (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- S: (n) savings bank, coin bank, money box, **bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- S: (n) **bank**, bank building (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*
- S: (n) **bank** (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning)) *"the plane went into a steep bank"*

Verb

- S: (v) **bank** (tip laterally) *"the pilot had to bank the aircraft"*
- S: (v) **bank** (enclose with a bank) *"bank roads"*
- S: (v) **bank** (do business with a bank or keep an account at a bank) *"Where do you bank in this town?"*
- S: (v) **bank** (act as the banker in a game or in gambling)
- **S: (v) bank (be in the banking business)**
- S: (v) deposit, **bank** (put into a bank account) *"She deposits her paycheck every month"*
- S: (v) **bank** (cover with ashes so to control the rate of burning) *"bank a fire"*
- S: (v) count, bet, depend, swear, rely, **bank**, look, calculate, reckon (have faith or confidence in) *"you can count on me to help you any time"; "Look to your friends for support"; "You can bet on that!"; "Depend on your family in times of crisis"*

Question

How do we automatically model multiple senses of a word?

Need a way to *dynamically* capture the specific ***context.***

Language Model

$$p(x_1, x_2, \ldots, x_m) = \prod_{i=1,\ldots,m} p(x_i | \underbrace{x_1, \ldots, x_{i-1}}_{\text{Context}})$$

We discussed:

1) n-gram language model

2) a neural language model (Bengio et al, 2003)

Language Model

$$p(x_1, x_2, \ldots, x_m) = \prod_{i=1,\ldots,m} p(x_i | \underbrace{x_1, \ldots, x_{i-1}}_{\text{Context}})$$

We discussed:

1) n-gram language model

2) a neural language model (Bengio et al, 2003)

Both models rely on the Markov (independence) assumption.

Language Model

$$p(x_1, x_2, \dots, x_m) = \prod_{i=1,\dots,m} p(x_i | \underbrace{x_1, \dots, x_{i-1}}_{\text{Context}})$$

We discussed:

1) n-gram language model

2) a neural language model (Bengio et al, 2003)

Both models rely on the Markov (independence) assumption.

Can we remove this assumption now
(since we use word embeddings) ?

Language Model

$$p(x_1, x_2, \ldots, x_m) = \prod_{i=1,\ldots,m} p(x_i | \underbrace{x_1, \ldots, x_{i-1})}_{\text{Context}}$$

We discussed:

1) n-gram language model

2) a neural language model (Bengio et al, 2003)

Both models rely on the Markov (independence) assumption.

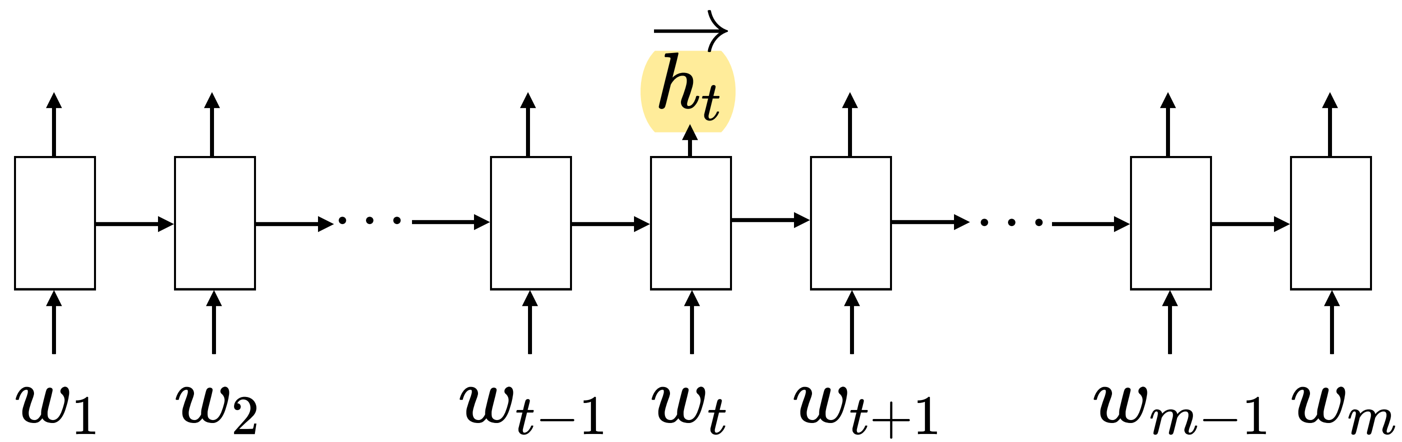
Need a way to *dynamically* capture the specific ***context***.

Contextual Embedding

LSTM

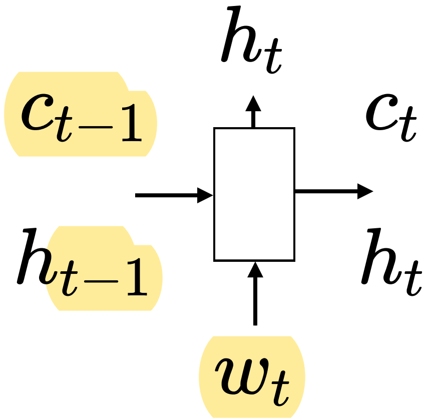
(Long Short-term Memory)

Hochreiter & Schmidhuber (1997)

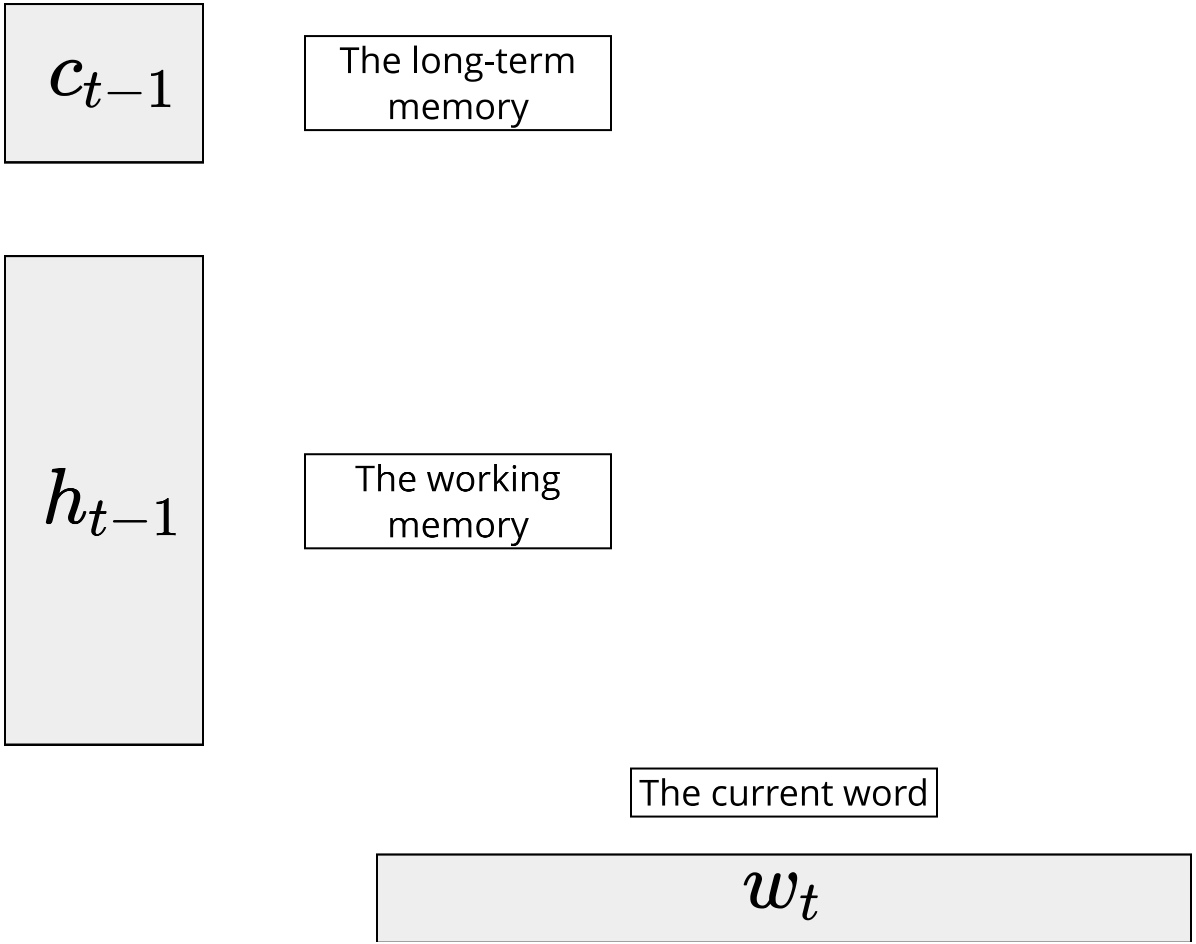


$$\overrightarrow{h_t} = \overrightarrow{\mathrm{LSTM}}(w_1, w_2, \ldots, w_t)$$

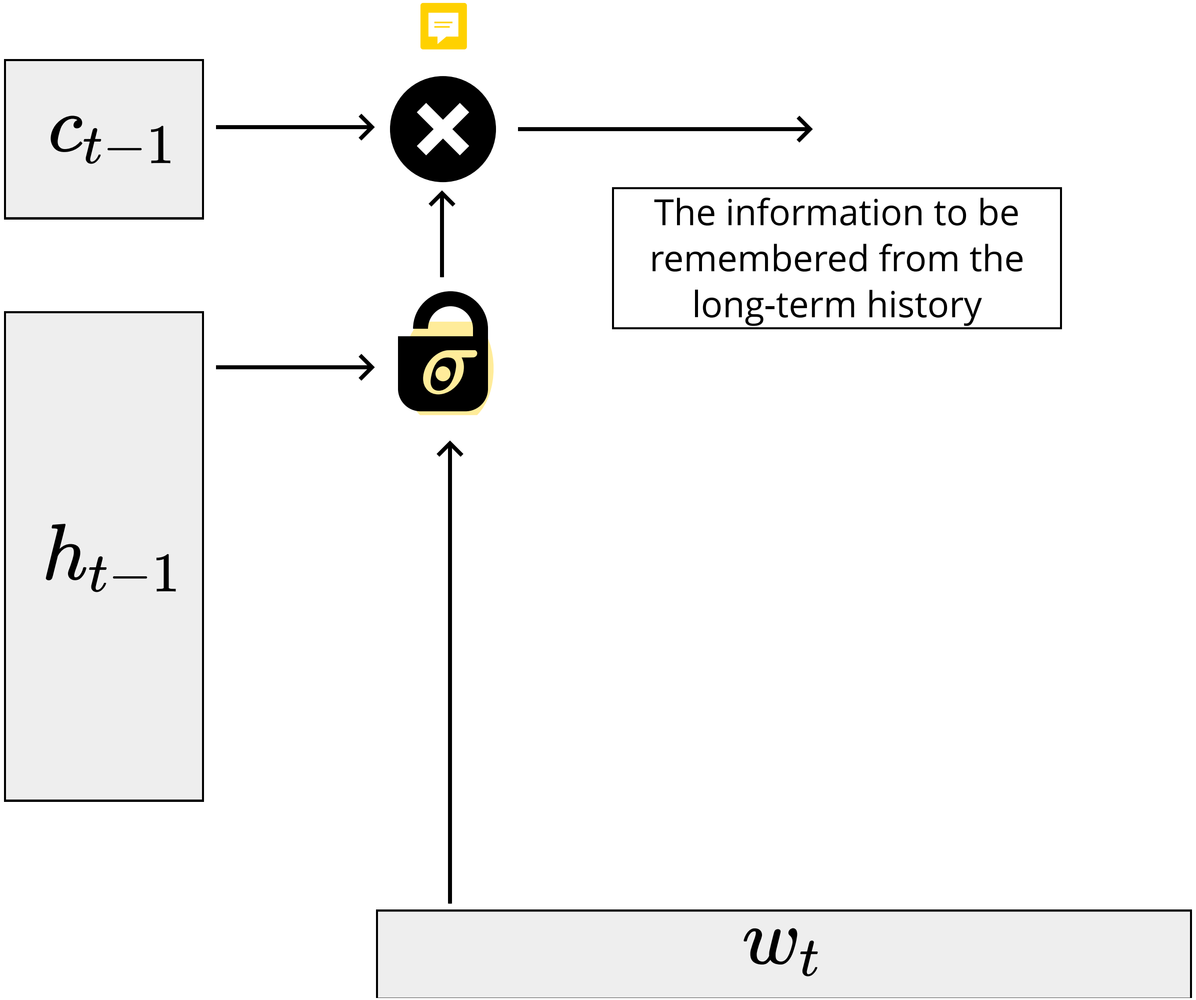
LSTM



LSTM

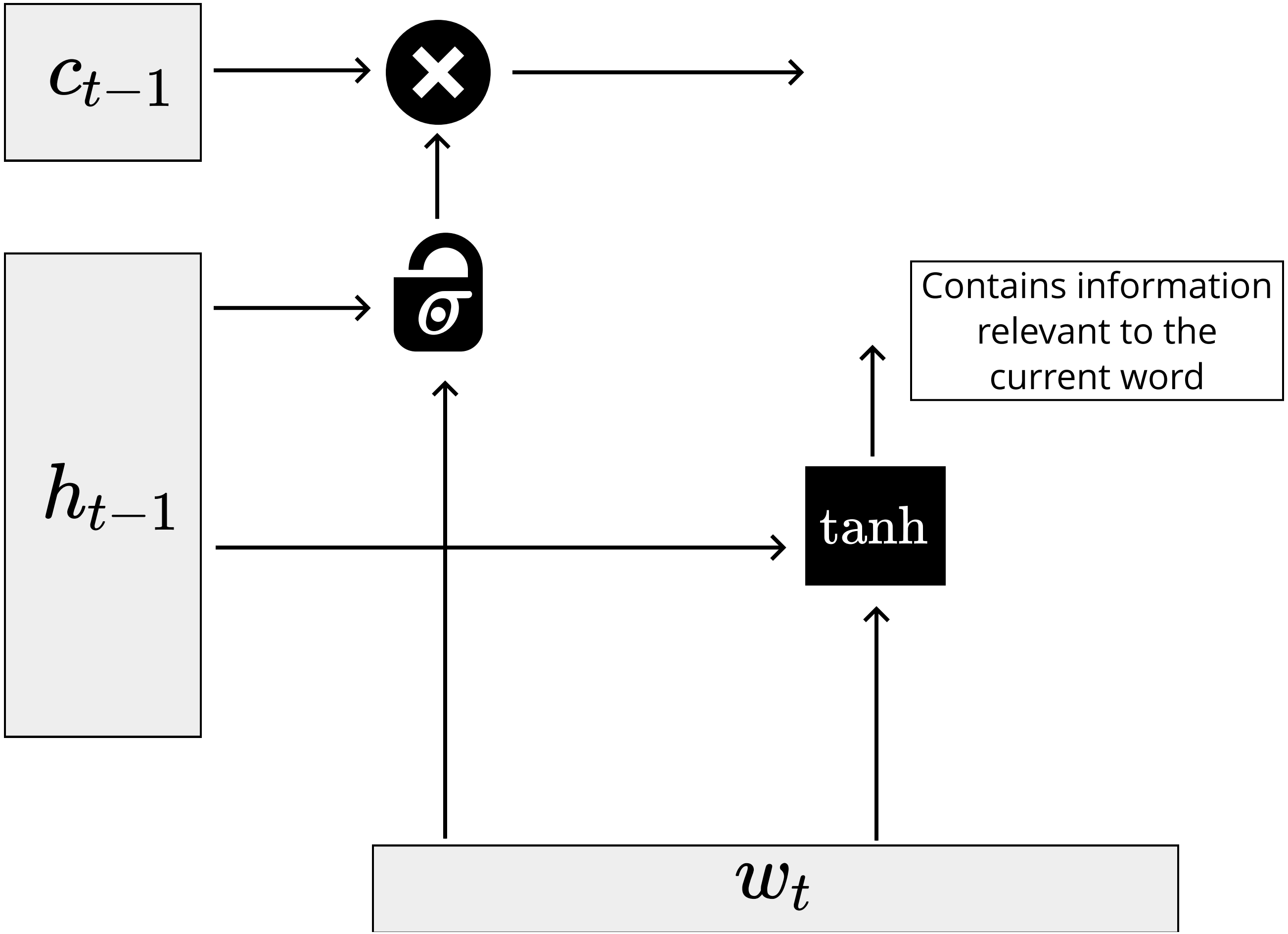

c_{t-1}
The long-term memory
h_{t-1}
The working memory
The current word
w_t

LSTM

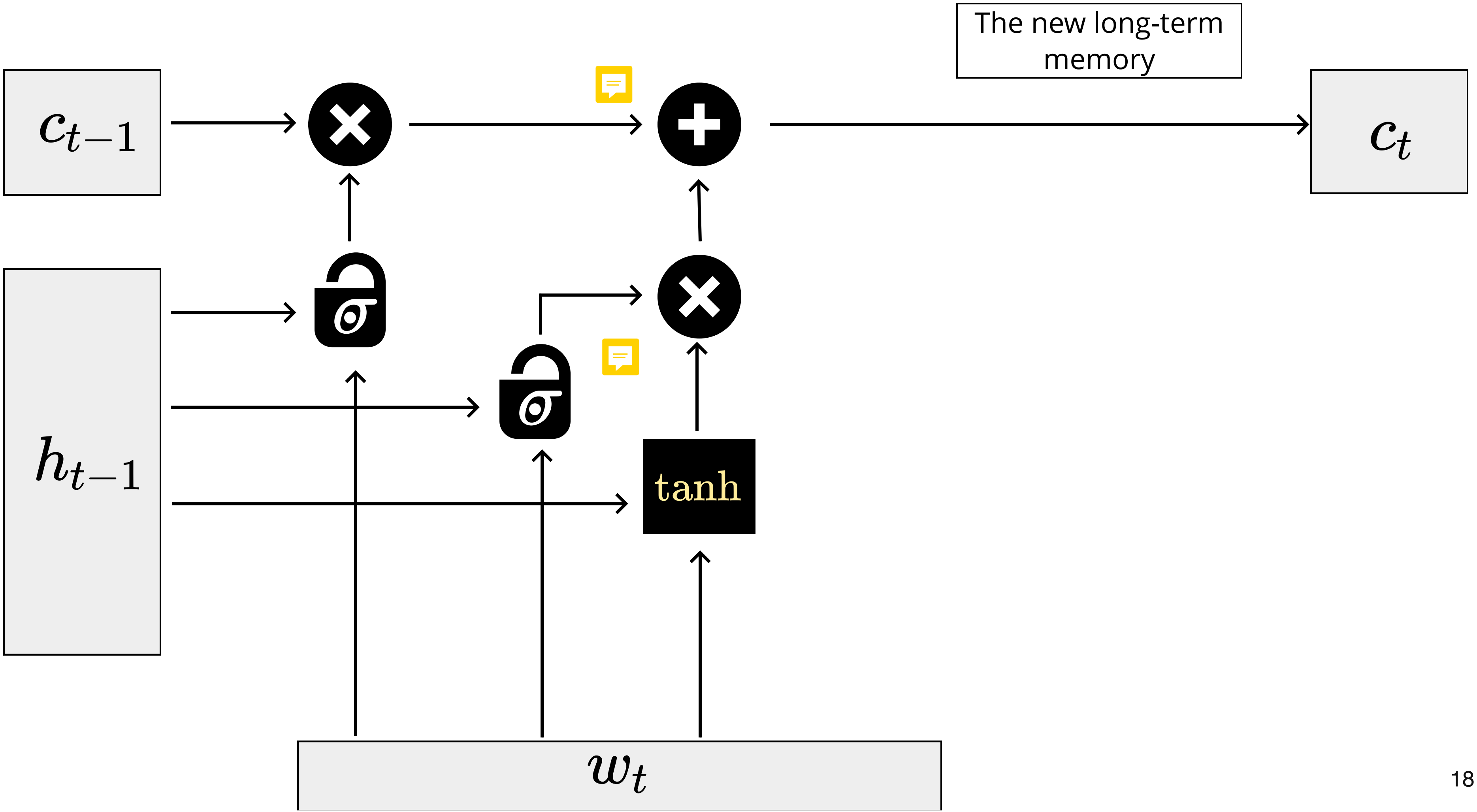


c_{t-1}
h_{t-1}
The information to be remembered from the long-term history
w_t

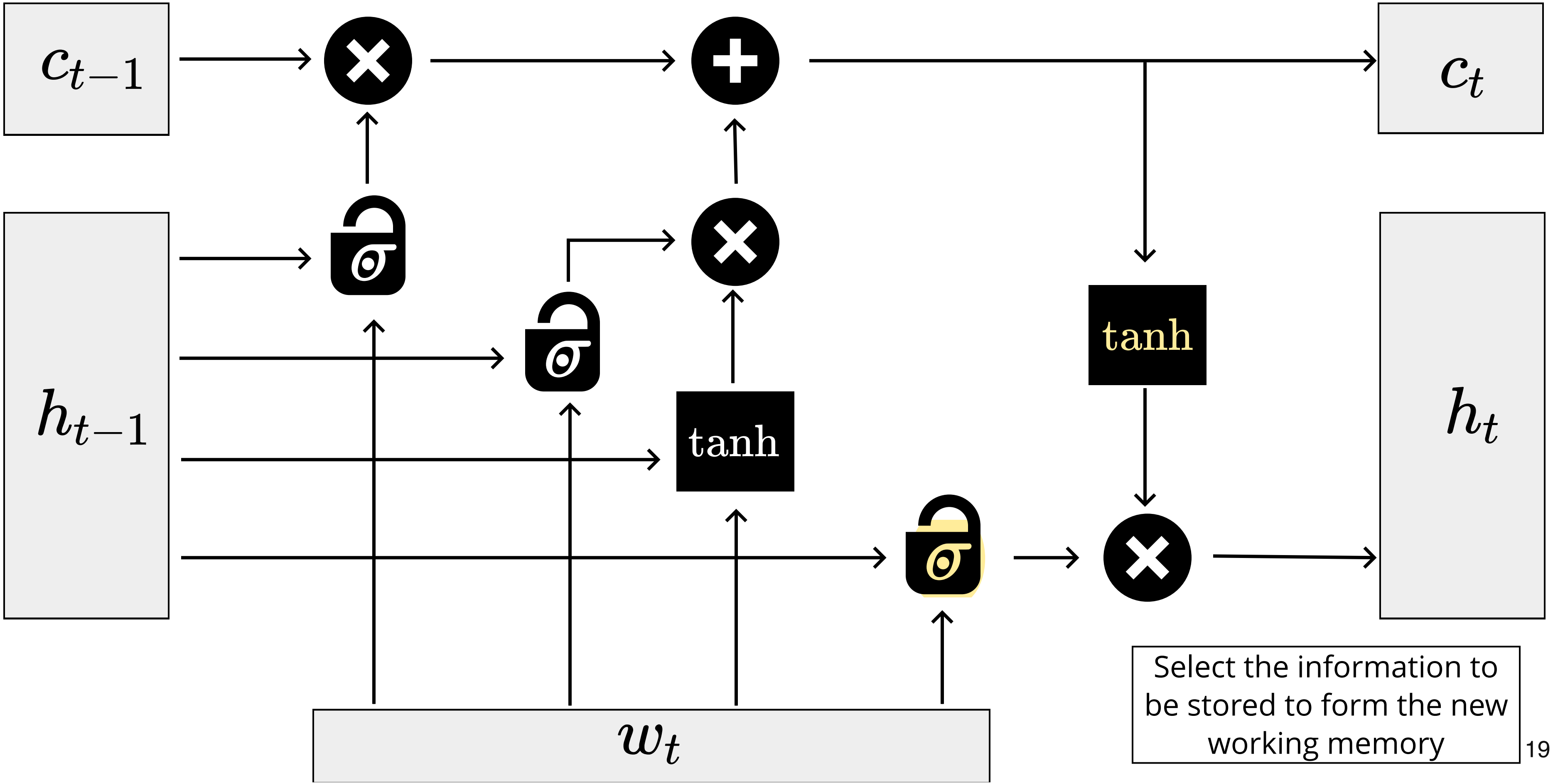
LSTM


c_{t-1}
h_{t-1}
Contains information relevant to the current word
tanh
w_t

LSTM

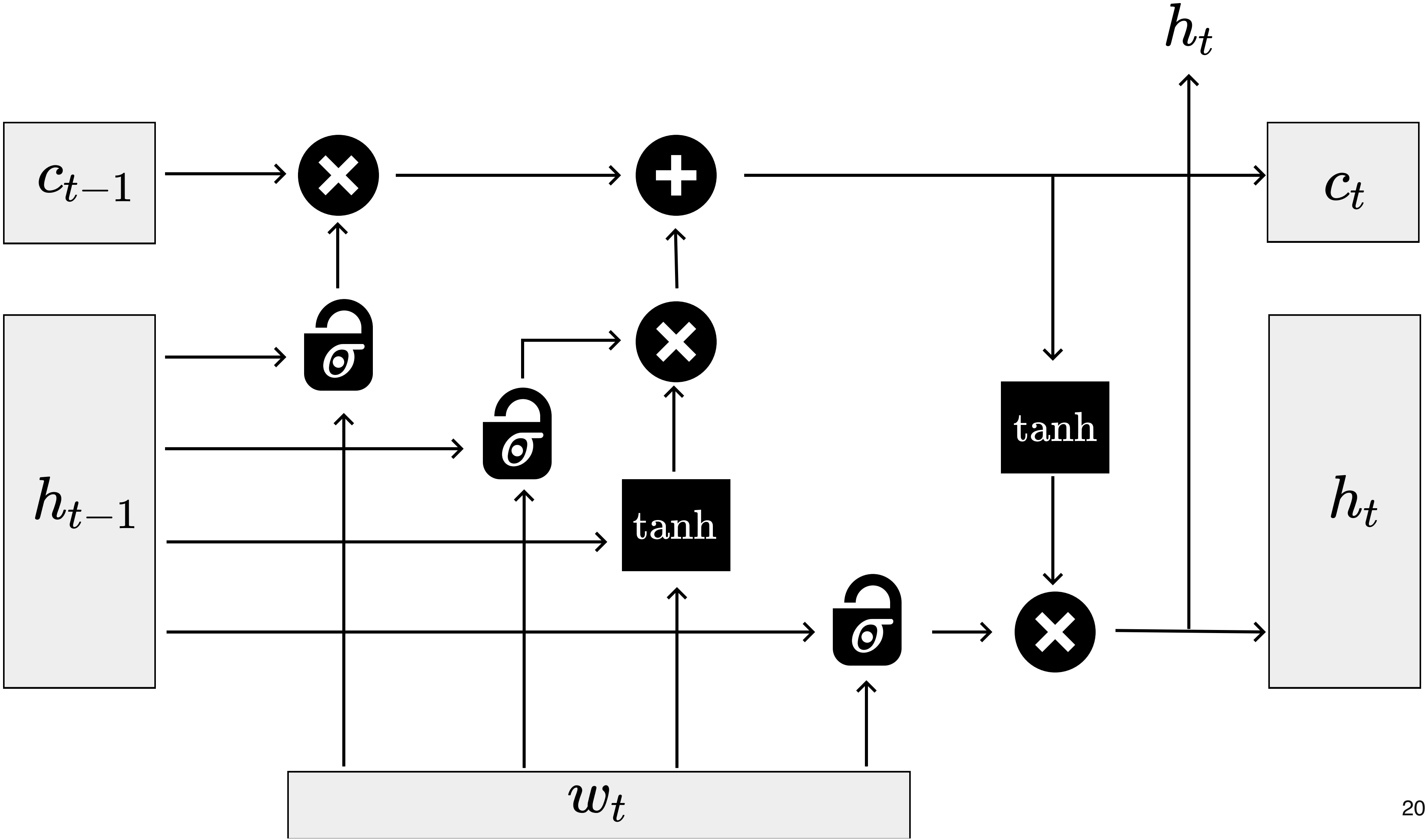


LSTM

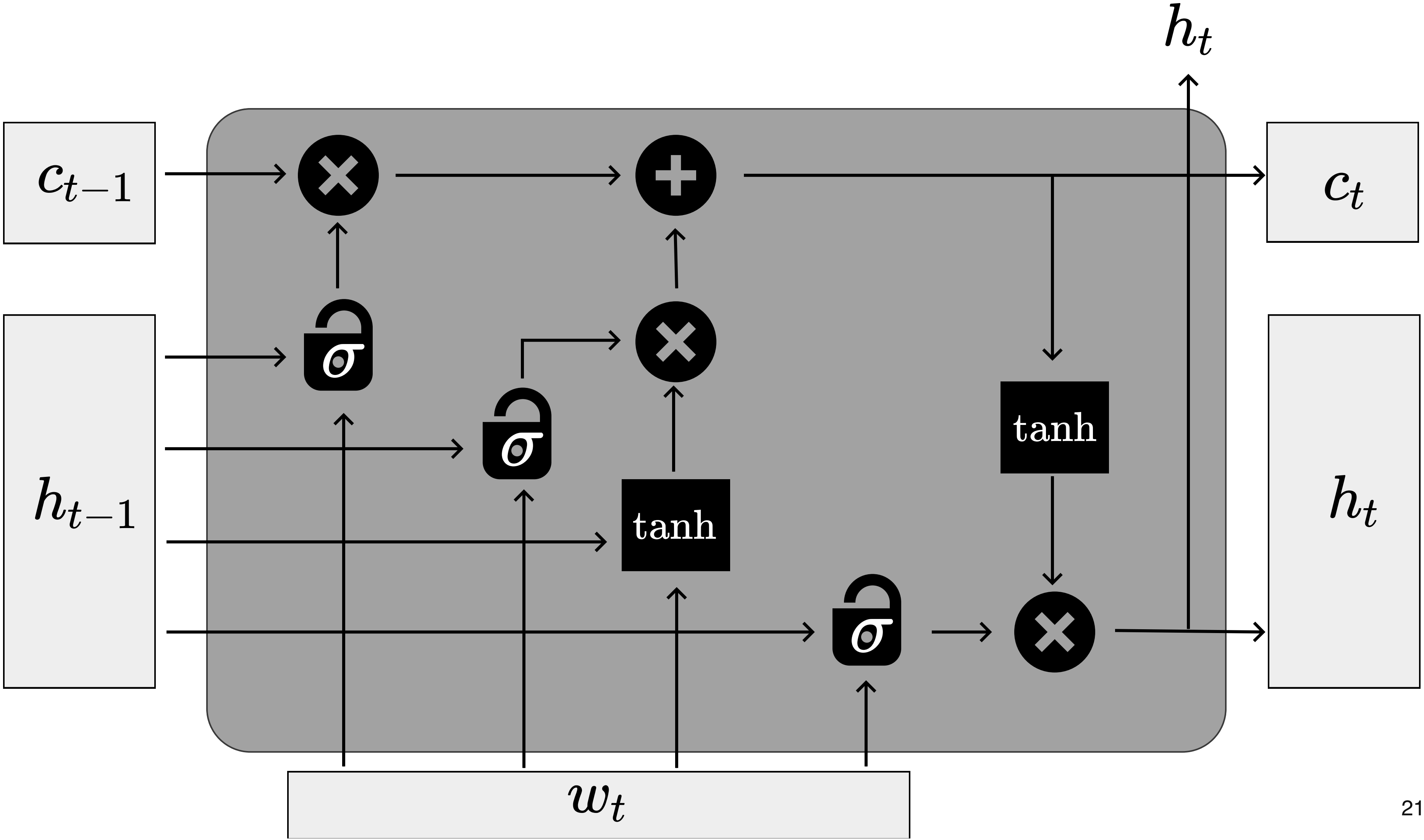


Select the information to be stored to form the new working memory

LSTM

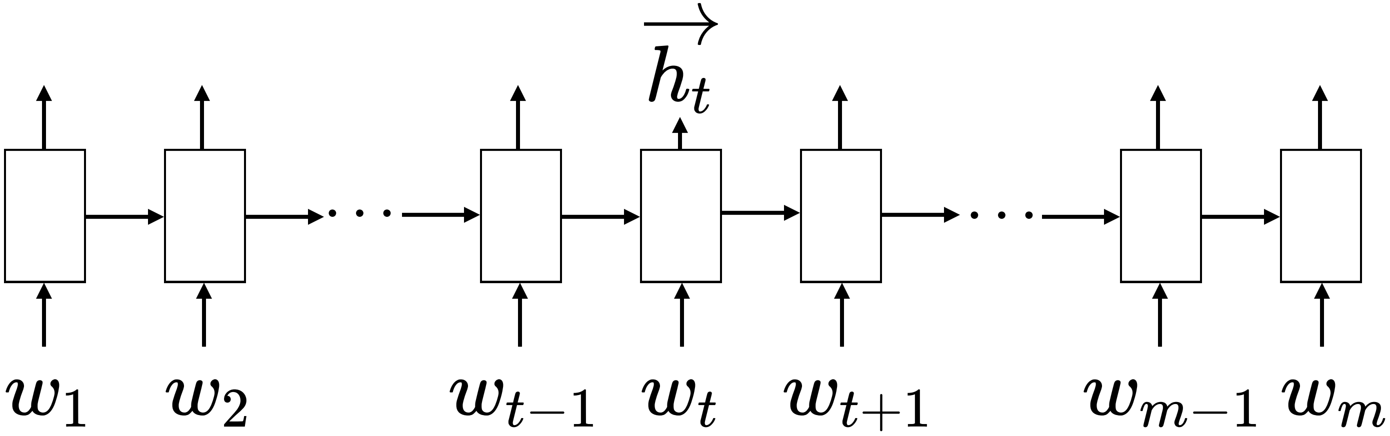


LSTM



LSTM Language Model

Sundermeyer et al. (2012)



$$\overrightarrow{h_t} = \overrightarrow{\text{LSTM}}(w_1, w_2, \ldots, w_t)$$

Individual probability at each position:

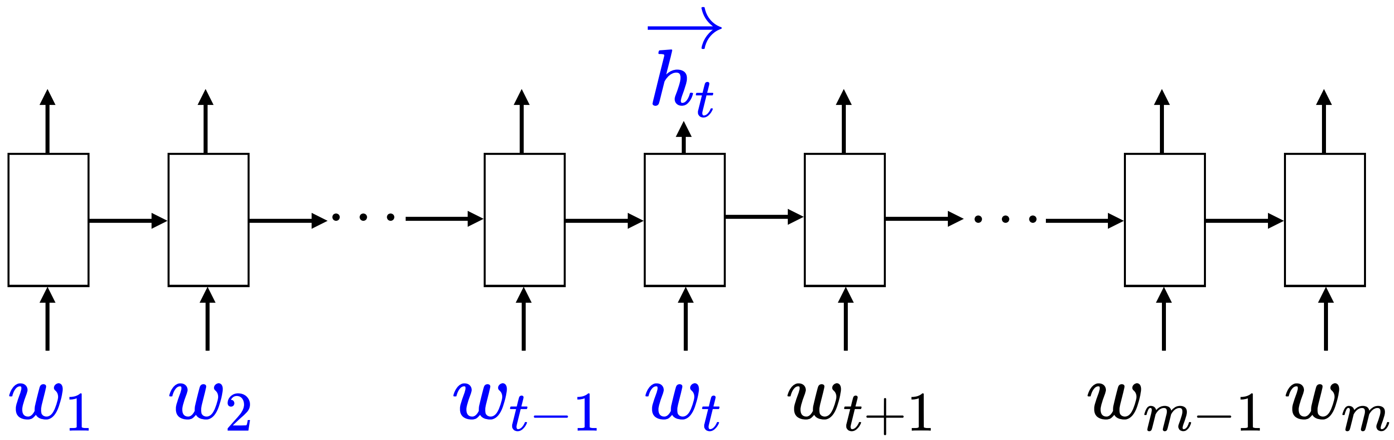
1. project $\overrightarrow{h_t}$ to a $|V|$ dimensional space

2. apply softmax on top of the vector

3. get the probability of generating the desired word w_{t+1}

LSTM Language Model

Sundermeyer et al. (2012)



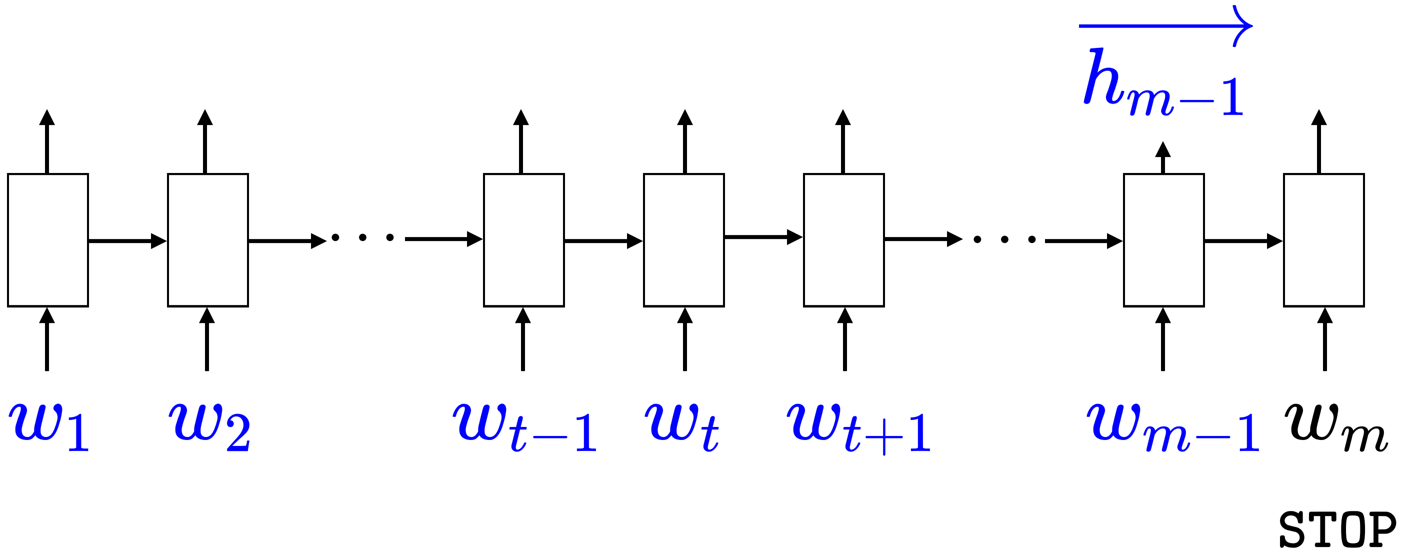
$$\overrightarrow{h_t} = \overrightarrow{\mathrm{LSTM}}(w_1, w_2, \ldots, w_t)$$

Context Embedding!

A function over a sequence of word embeddings.

LSTM Language Model

Sundermeyer et al. (2012)



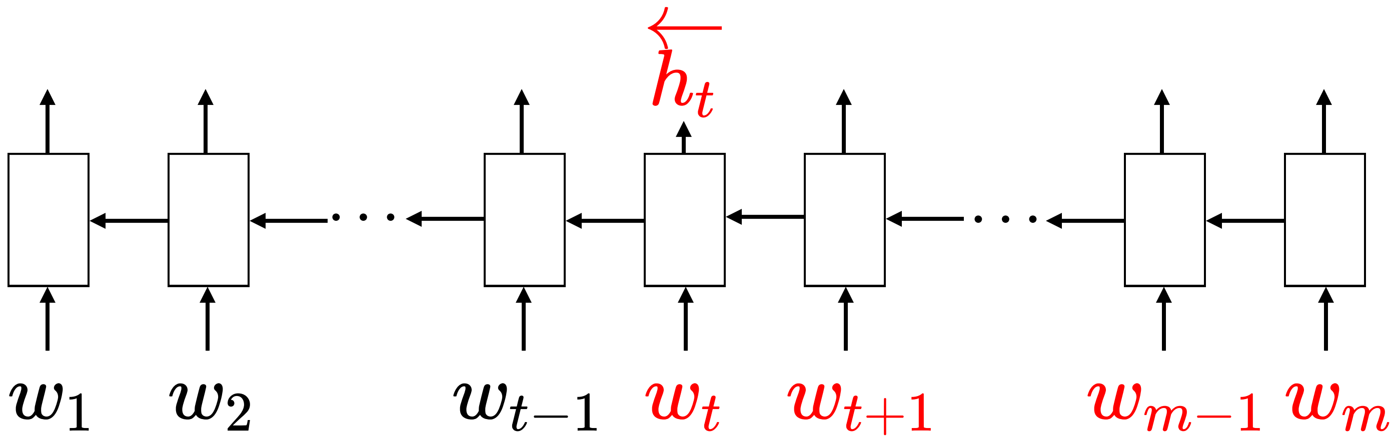
$$\overrightarrow{h_{m-1}} = \overrightarrow{\text{LSTM}}(w_1, w_2, \ldots, w_{m-1})$$

Sentence Embedding!

It's essentially a special context embedding.

LSTM Language Model

Sundermeyer et al. (2012)



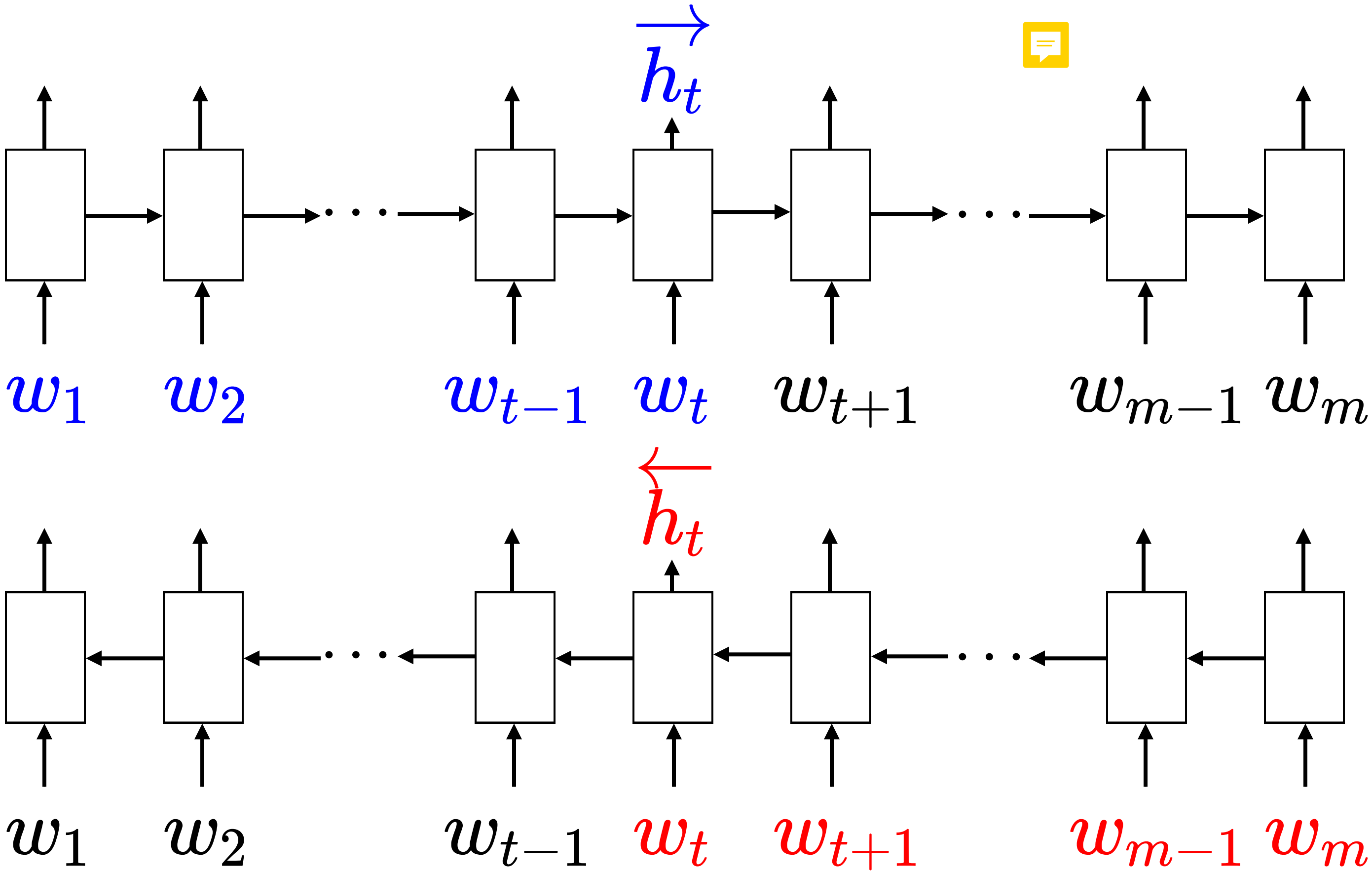
$$\overleftarrow{h_t} = \overleftarrow{\text{LSTM}}(w_m, w_{m-1}, \ldots, w_t)$$

Contextual Embedding!

A function over the sequence of word embeddings.

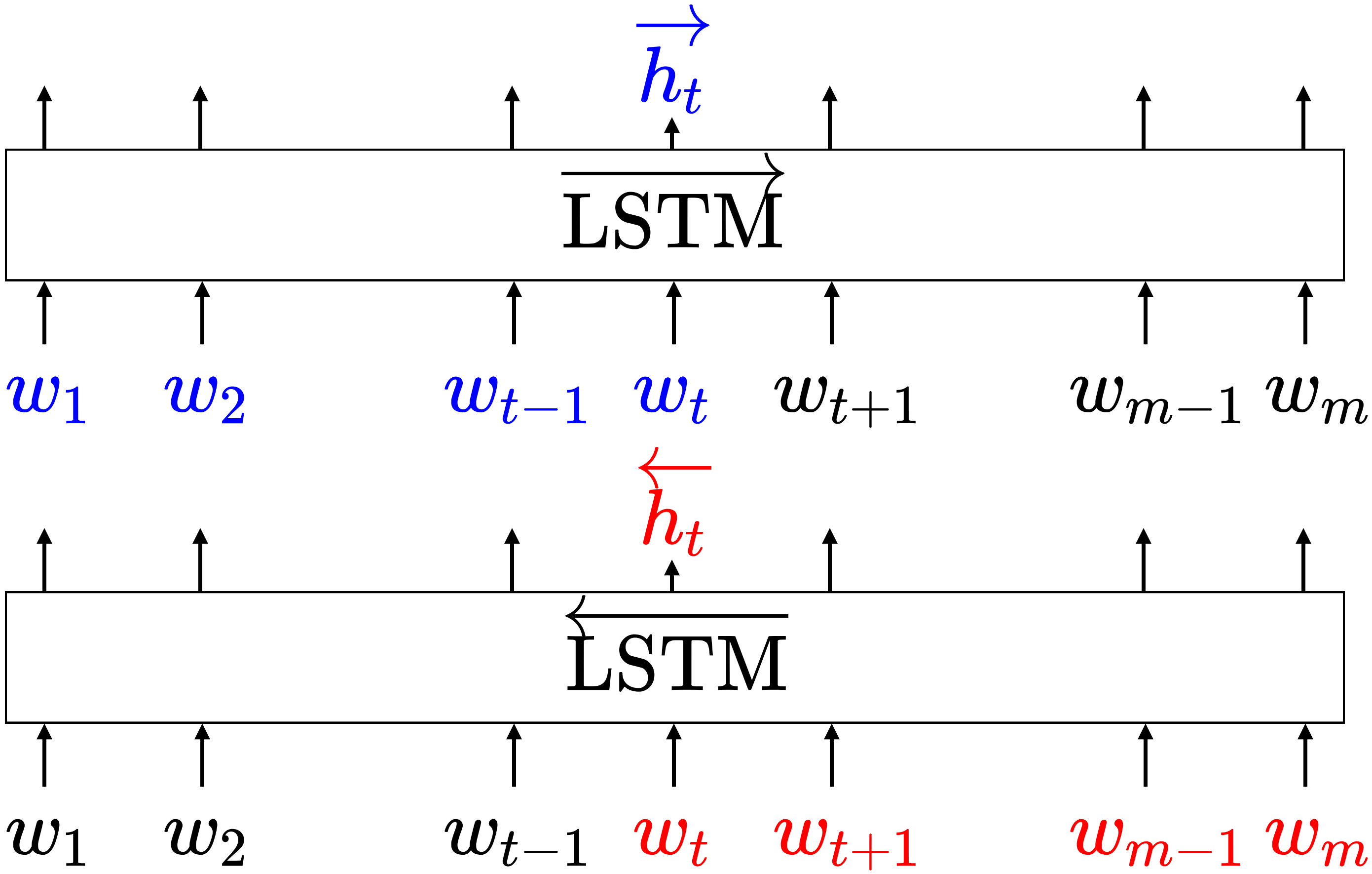
Bidirectional LSTM

Context Embedding $[\overrightarrow{h_t}; \overleftarrow{h_t}]$



Bidirectional LSTM

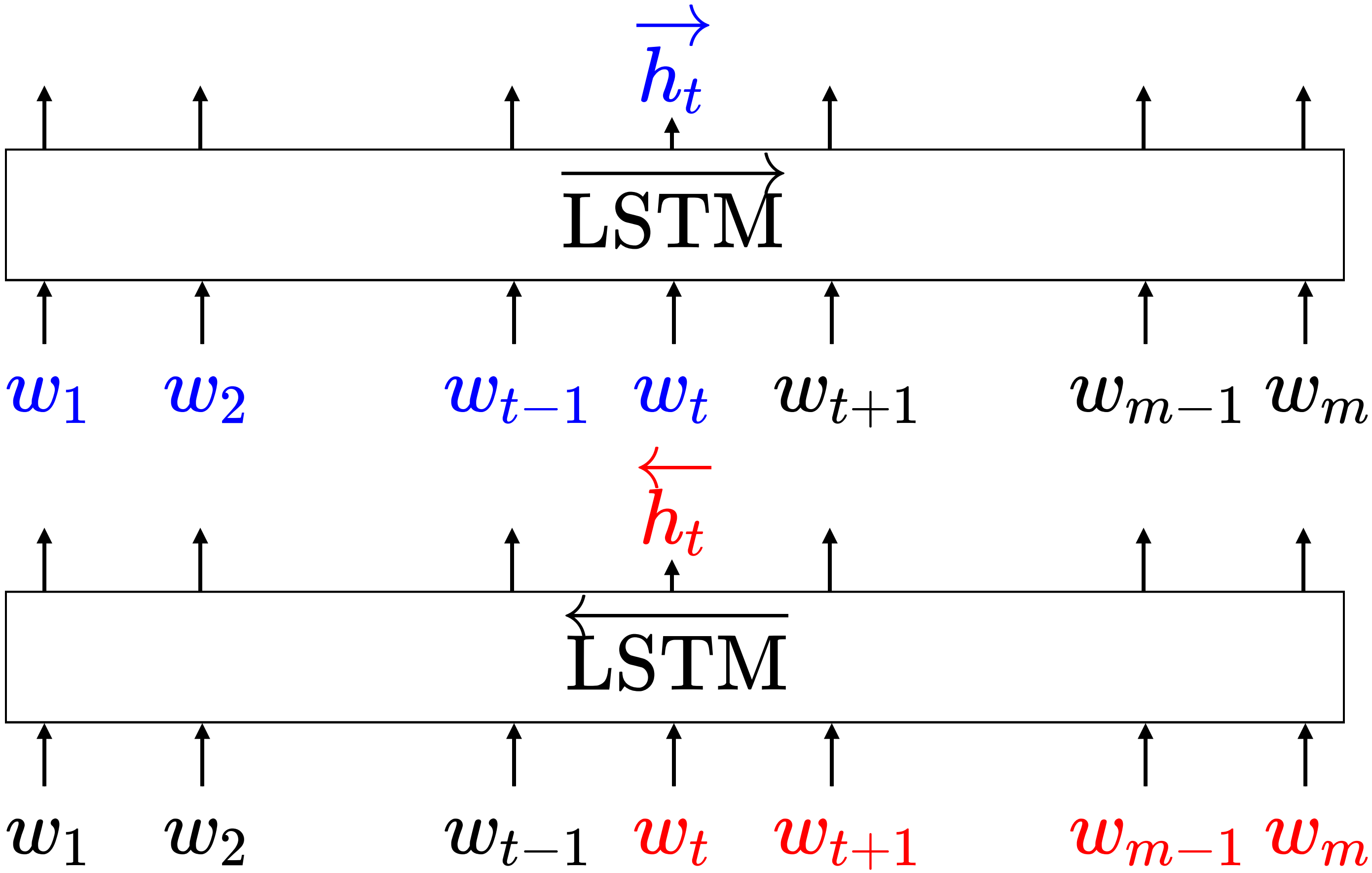
Context Embedding $[\overrightarrow{h_t}; \overleftarrow{h_t}]$



Bidirectional LSTM

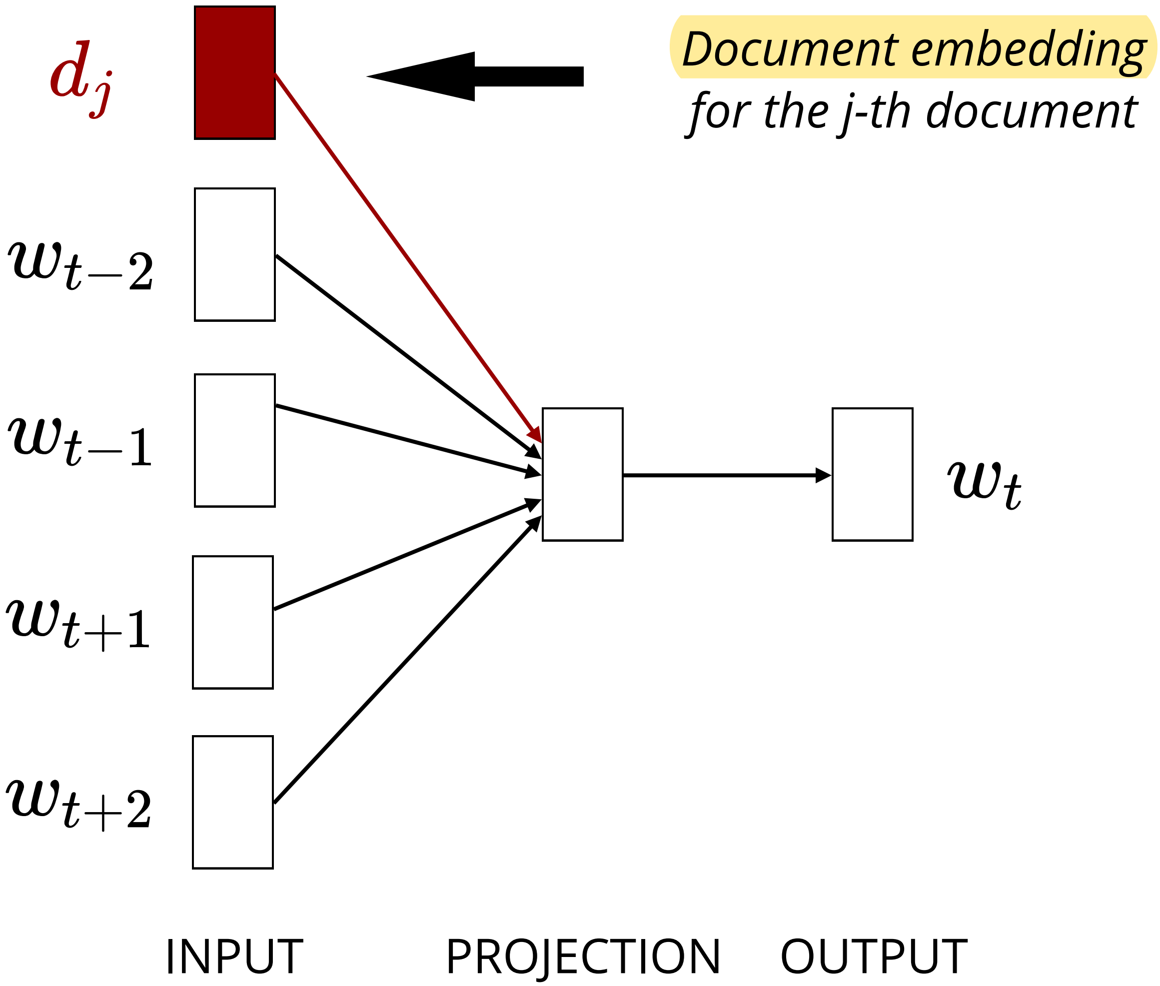
Context Embedding

$$[\overrightarrow{h_t}; \overleftarrow{h_t}] = \mathbf{h}_t$$



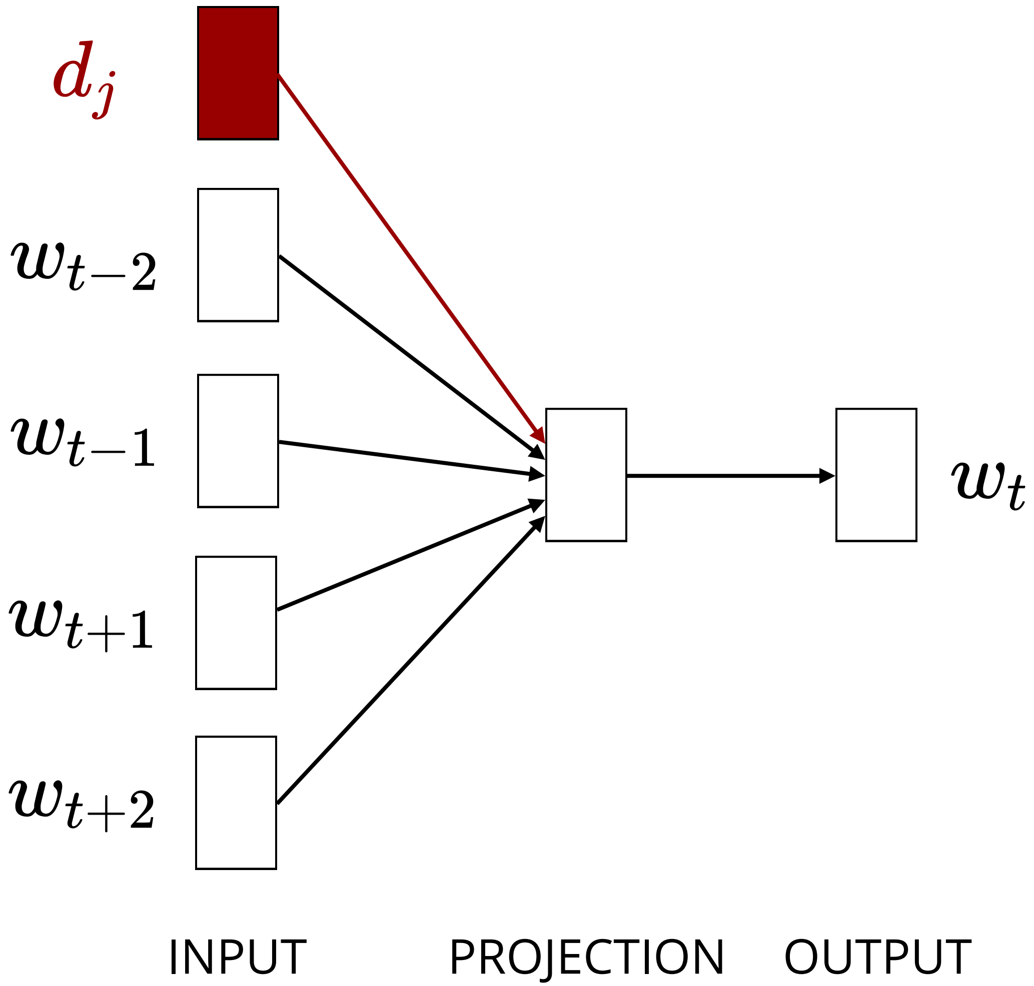
Doc2Vec

Le & Mikolov (2014)



PV-DM

Doc2Vec

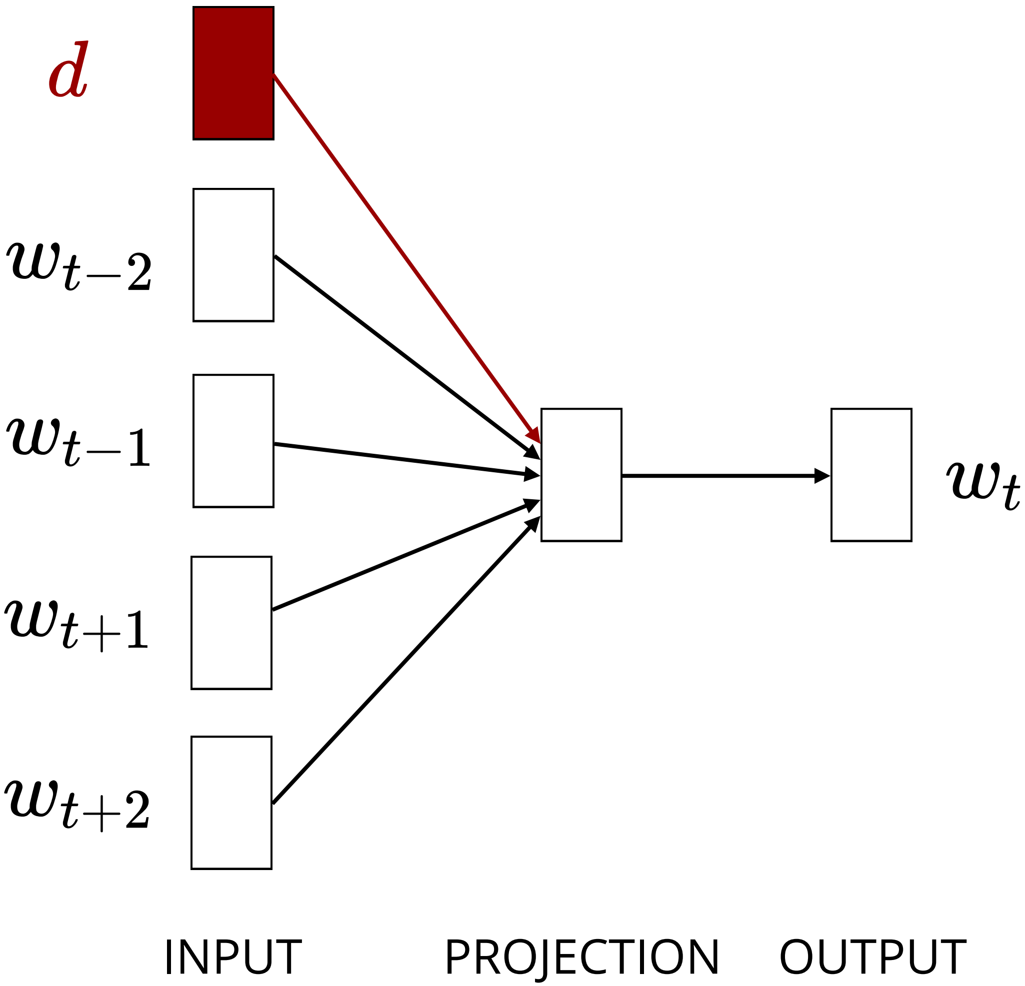


Training

Almost the same as CBOW, except that there is an additional document vector from the input layer

Concatenation was preferred (over sum/averaging) when constructing the projection layer

Doc2Vec



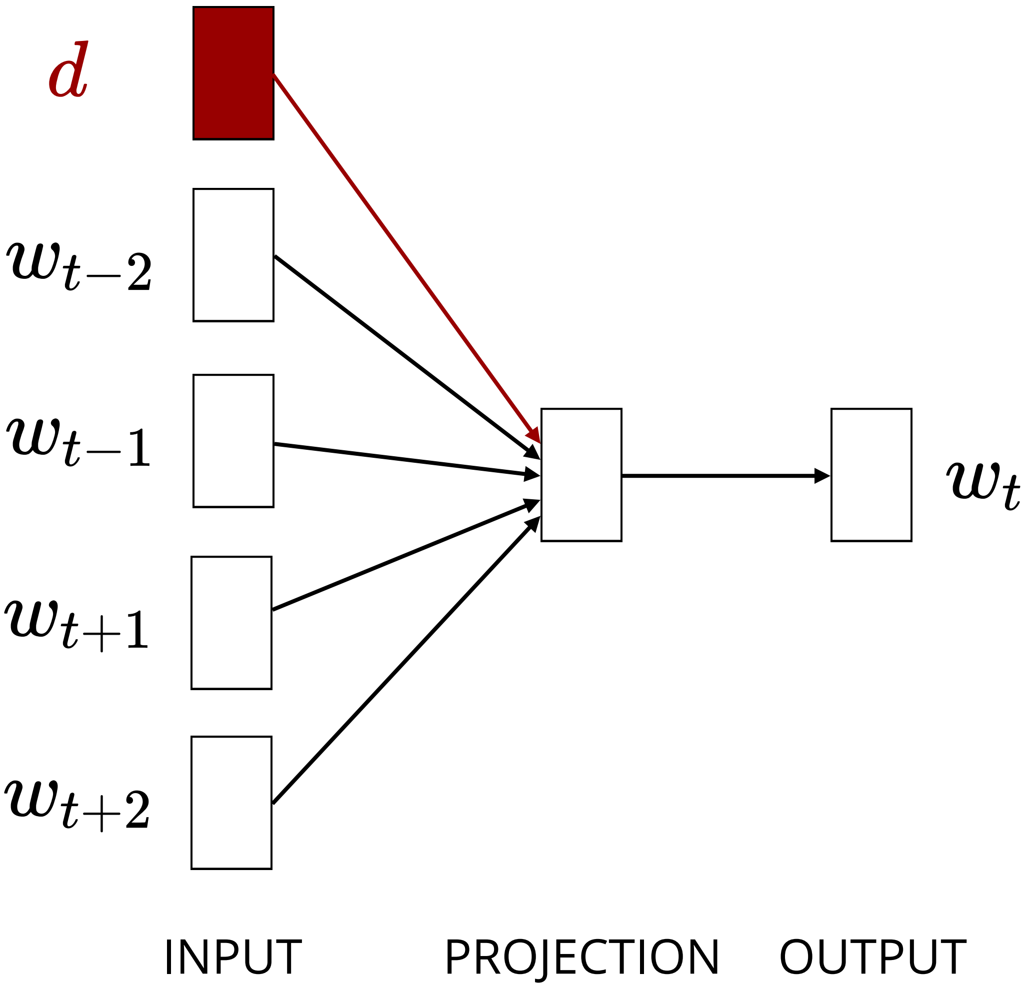
Inference

Fix all the learned parameters (including the learned word embeddings), fine tune the document embedding

Doc2Vec

$$\min_d f(d, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, w_t)$$

Note that this is a simplified optimization problem!!

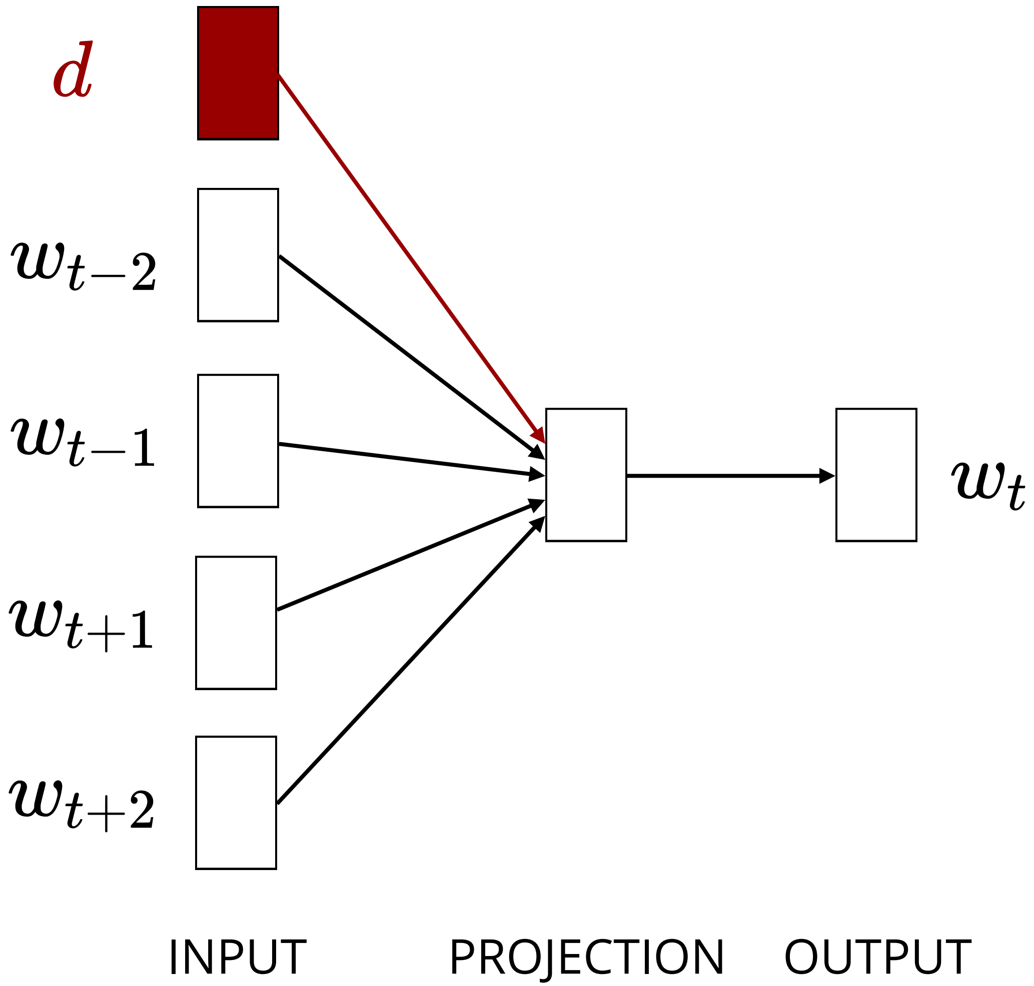


This is a process for finding an analytical solution to a minimization problem that involves words in the document!

Doc2Vec

$$\min_d f(d, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, w_t)$$

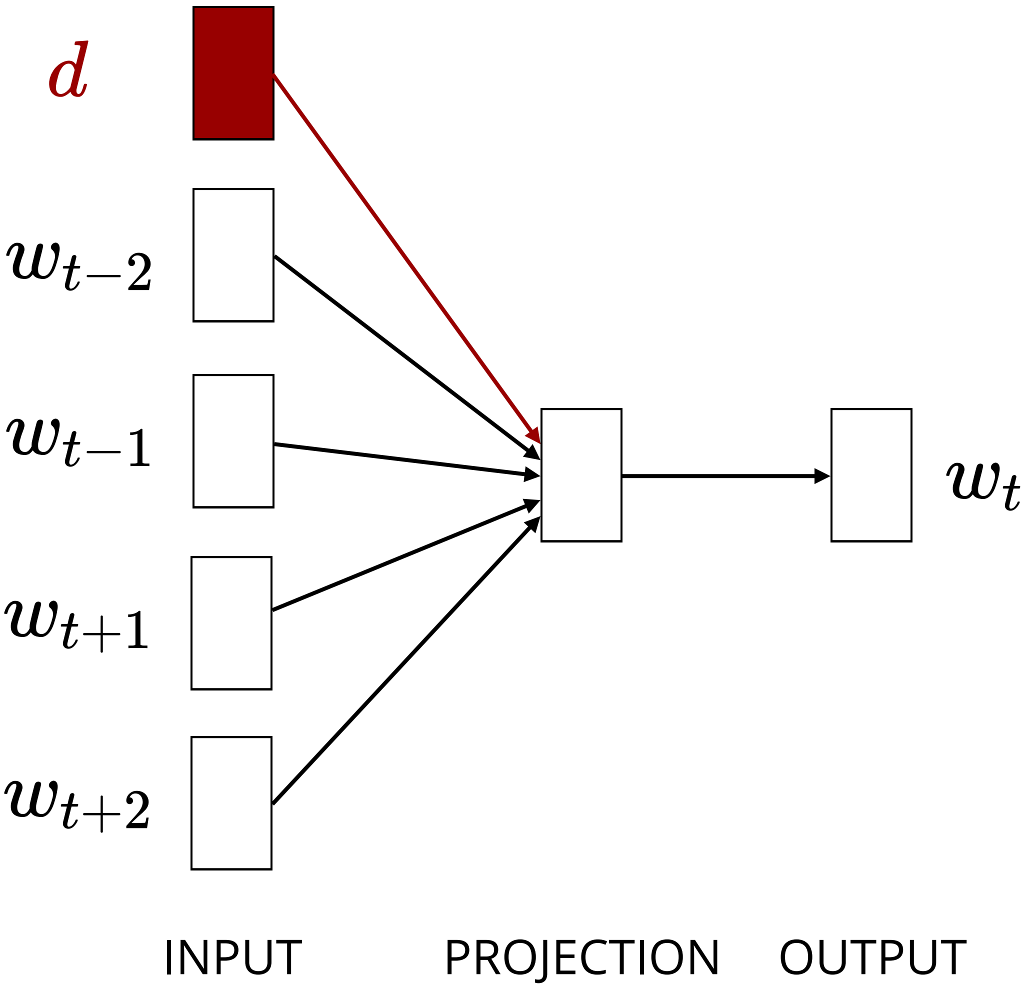
Note that this is a simplified optimization problem!!



The sentence embedding is again a context embedding - a function of the word embeddings.

Doc2Vec

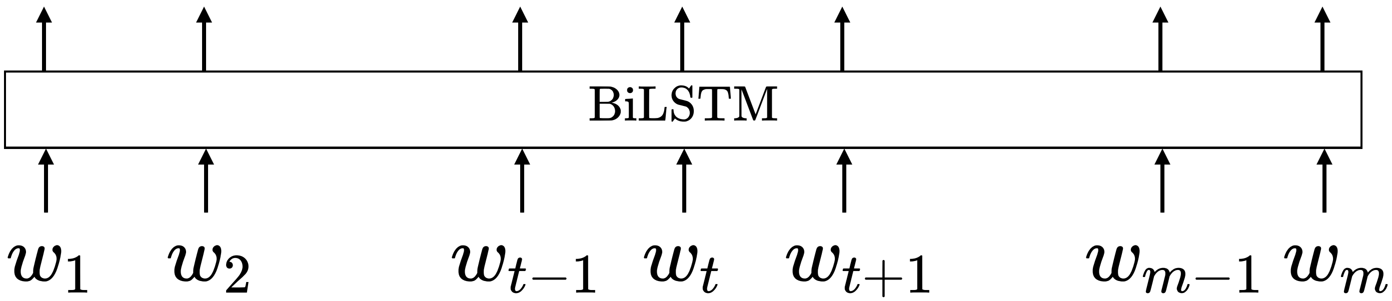
$$d = g(w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, w_t)$$



The sentence embedding is again a context embedding - a function of the word embeddings.

Deep Contextual Embedding

ELMo
(Embeddings from Language Models)
Peters et al. (2018)

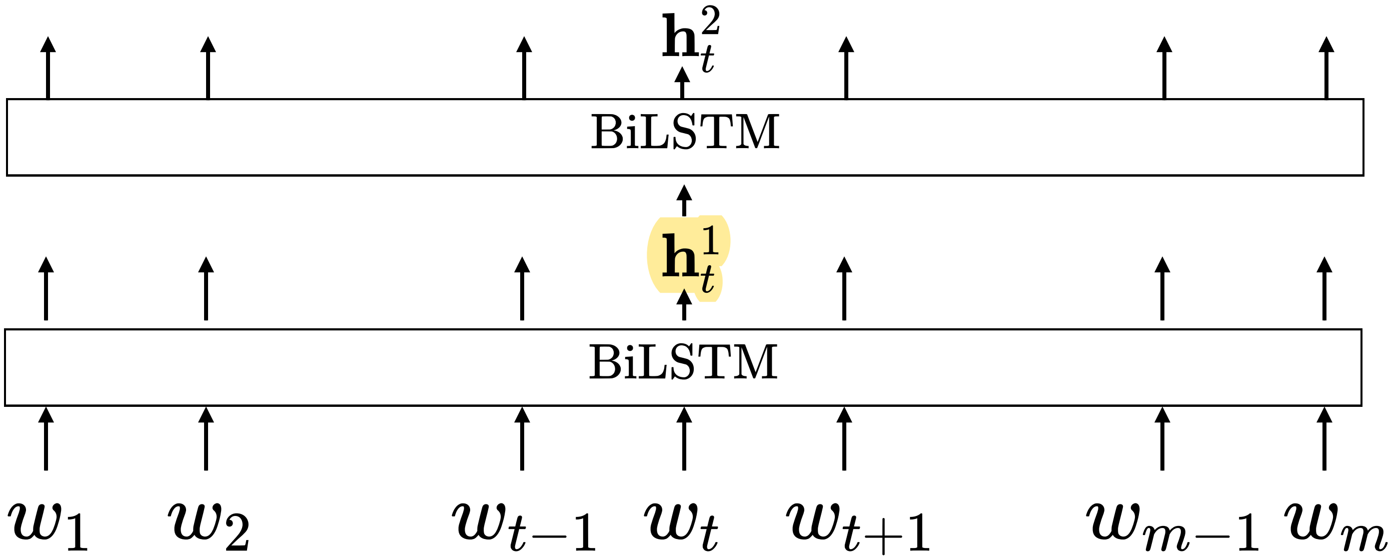


$$\sum_{t=1,\ldots,m} \Big(\log p(w_t \mid w_1, \ldots, w_{t-1}; \Theta_x, \overrightarrow{\Theta}_{LSTM}, \Theta_s) + \log p(w_t \mid w_{t+1}, \ldots, w_m; \Theta_x, \overleftarrow{\Theta}_{LSTM}, \Theta_s) \Big)$$

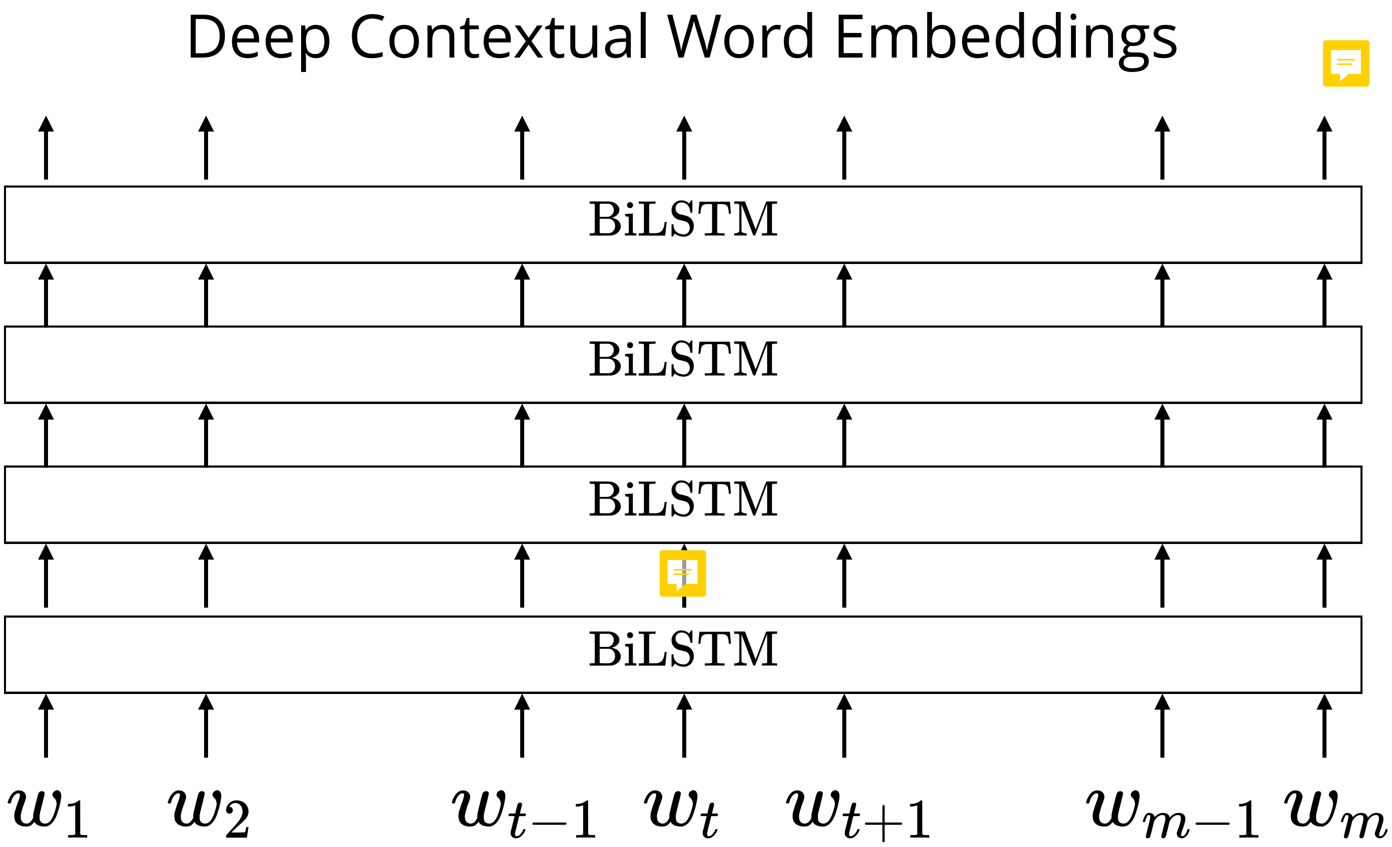
Tie the parameters for both the token representation (Θ_x), and softmax layer (Θ_s).

Separate parameters for the LSTMs in each direction.

ELMo
(Embeddings from Language Models)



ELMo
(Embeddings from Language Models)



ELMo
(Embeddings from Language Models)

$$R_t = \{\mathbf{h}_t^j \mid j = 0, \ldots, L\}$$

$$\mathbf{ELMo}_t = E(R_t, \Theta_e) = \gamma \sum_{j=0}^{L} s^j \mathbf{h}_t^j$$

Task-specific hyperparameters

$\mathbf{h}_t^L$

Layer L BiLSTM

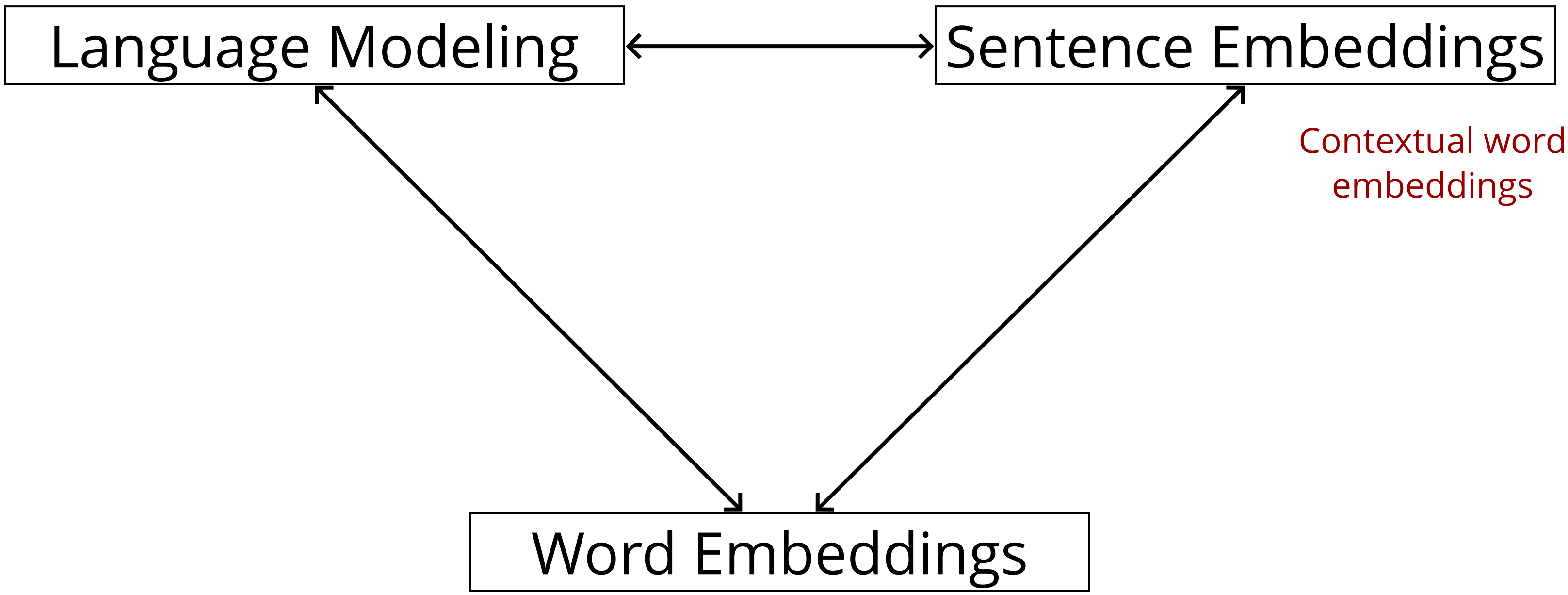
⋮

$\mathbf{h}_t^1$

Layer 1 BiLSTM

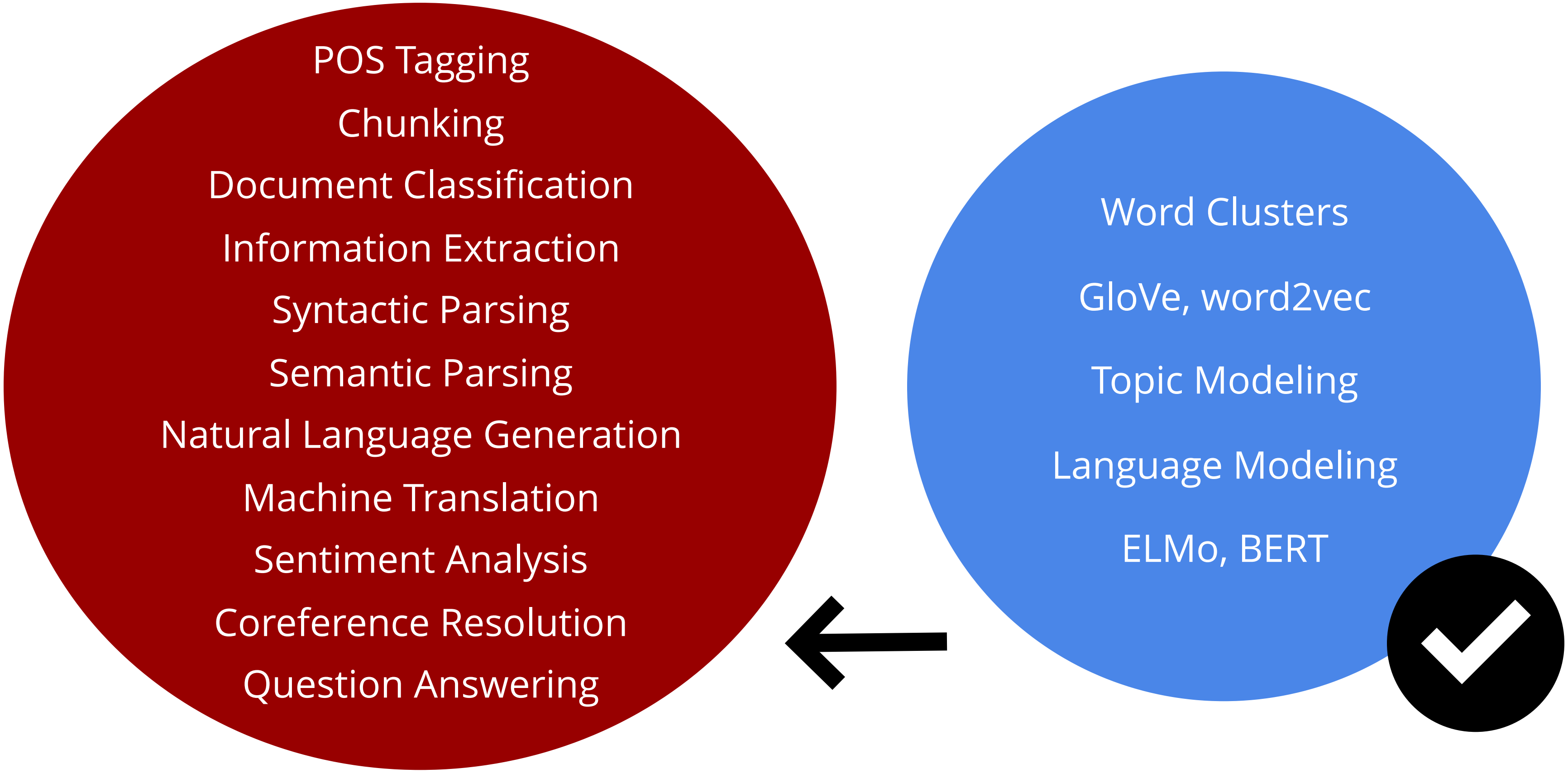
Layer 0 w_t $(\mathbf{h}_t^0)$

Three Tasks



These three tasks are closely related!

Tasks in NLP

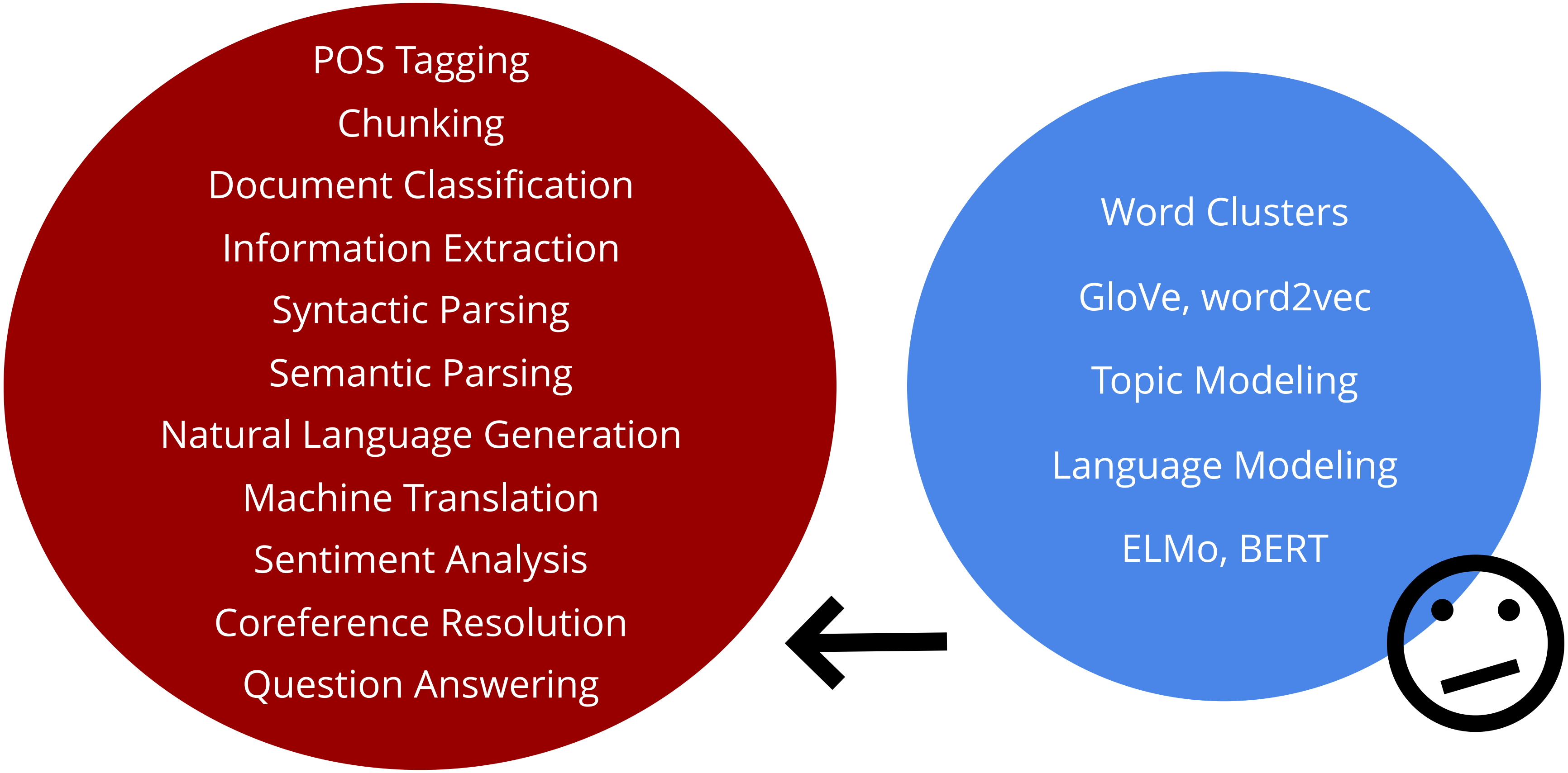


Supervised **Unsupervised**

Question

What about the unsupervised HMM that we learned in the ML class?

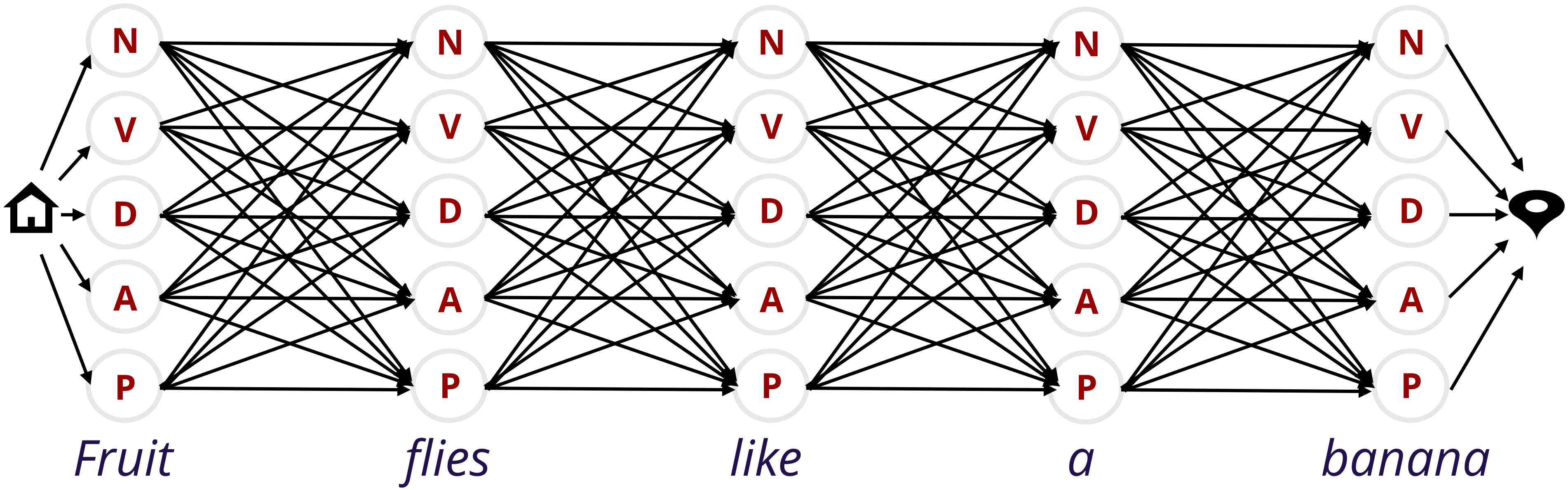
Tasks in NLP



Supervised **Unsupervised**

Hidden Markov Model
Supervised Learning

$$p(x_1, \ldots, x_m, \overset{\texttt{START}}{\overset{\downarrow}{y_0}}, y_1, \ldots, y_m, \overset{\texttt{STOP}}{\overset{\downarrow}{y_{m+1}}})$$

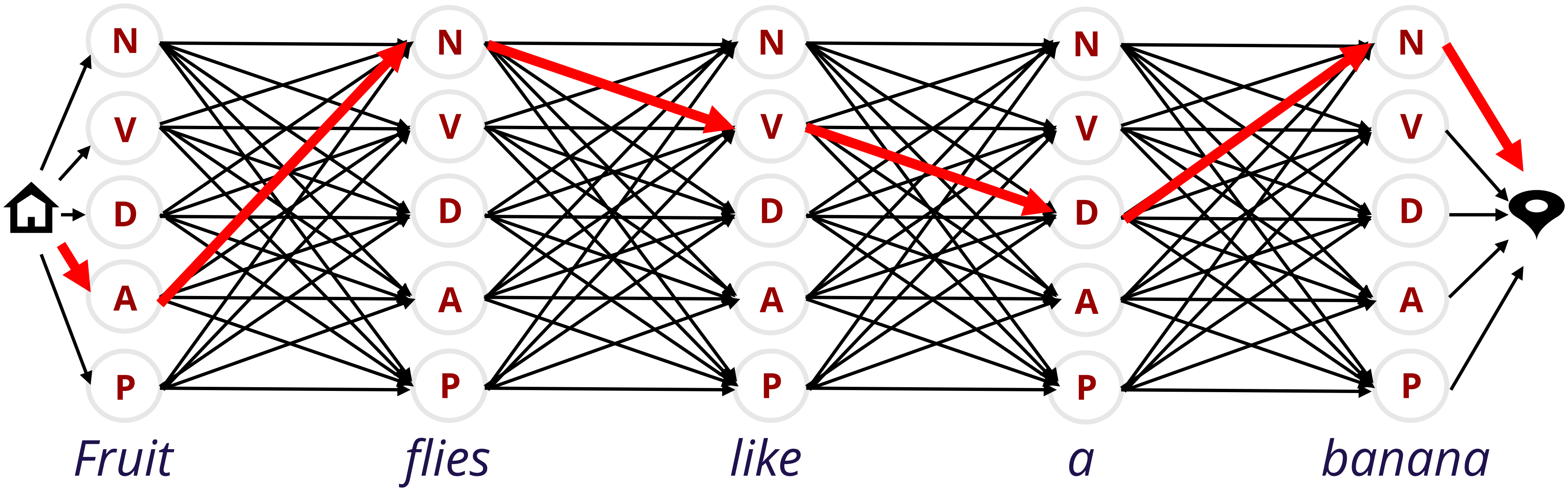


Hidden Markov Model

Decoding

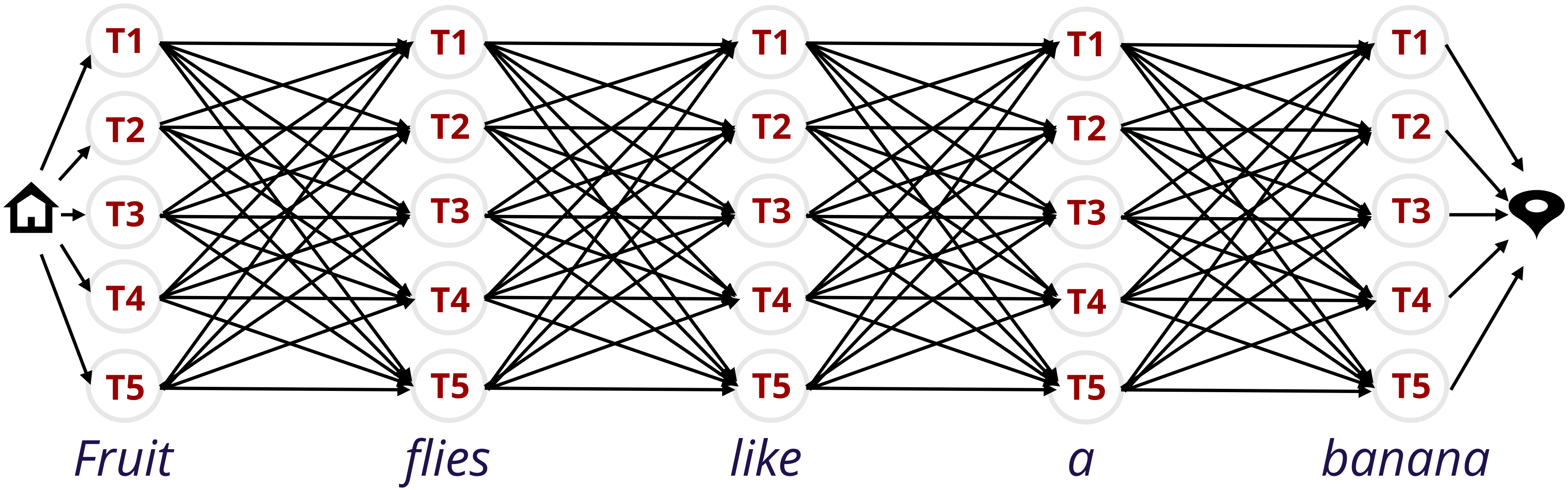
$$y_1^*, \ldots, y_m^*$$

$$= \arg\max_{y_1,\ldots,y_m} p(x_1, \ldots, x_m, y_0, y_1, \ldots, y_m, y_{m+1})$$



Hidden Markov Model
Unsupervised Learning

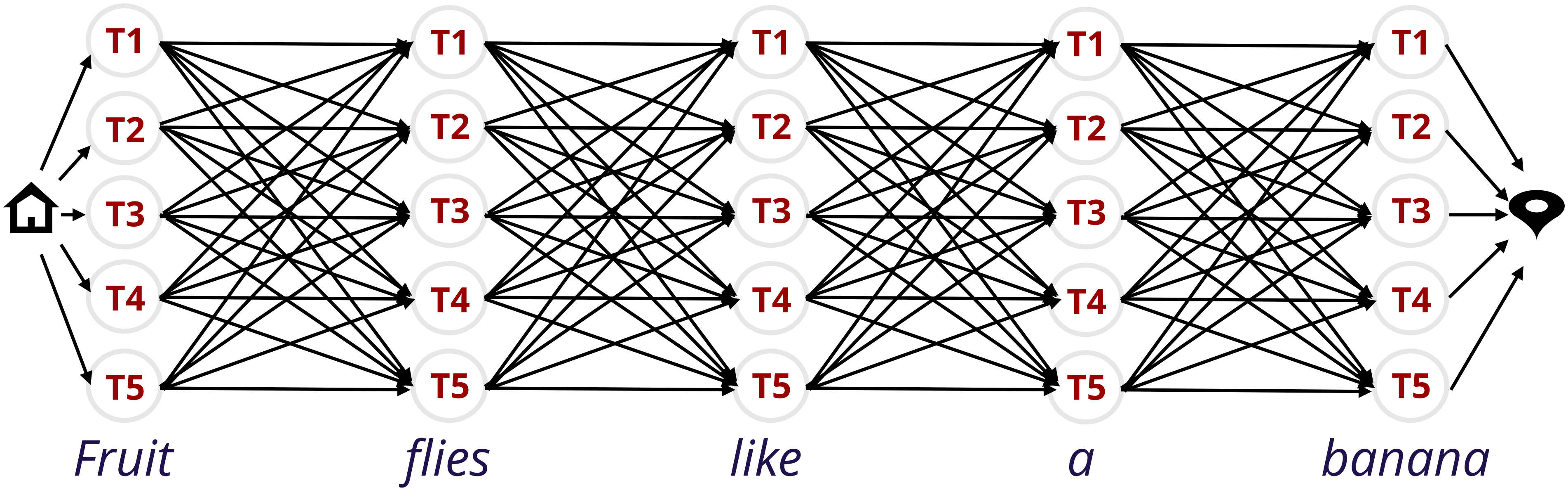
$$p(x_1, \ldots, x_m)$$



Hidden Markov Model
Unsupervised Learning

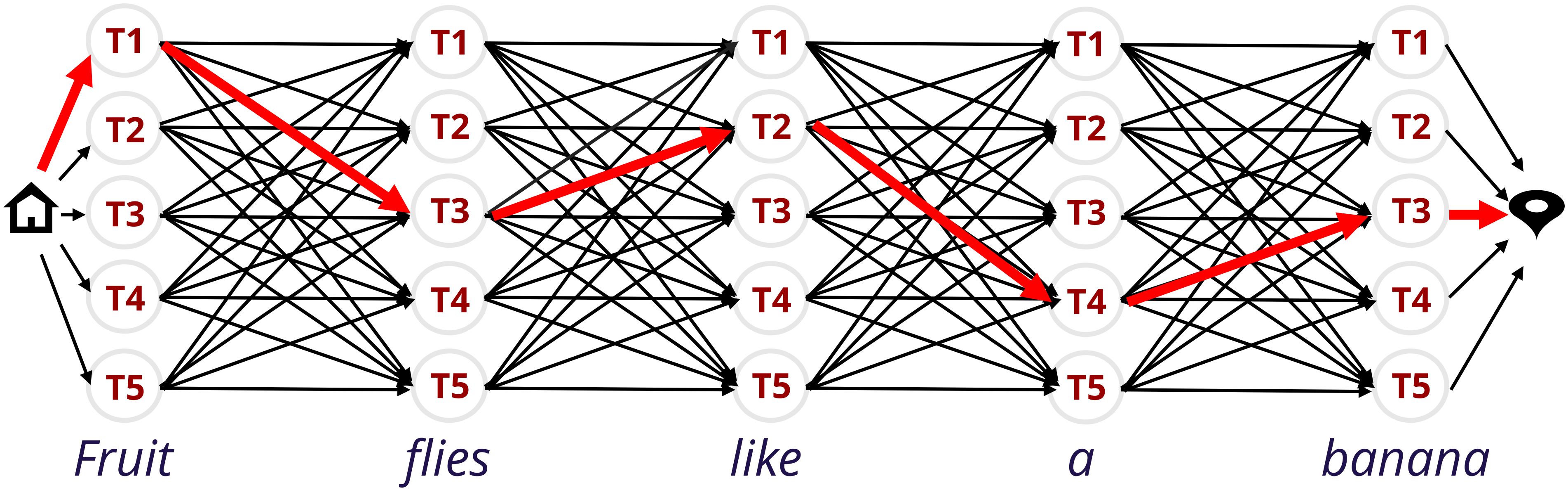
$$p(x_1, \ldots, x_m)$$

$$= \sum_{x_1, \ldots, x_m, y_1, \ldots, y_m} p(x_1, \ldots, x_m, y_0, y_1, \ldots, y_m, y_{m+1})$$

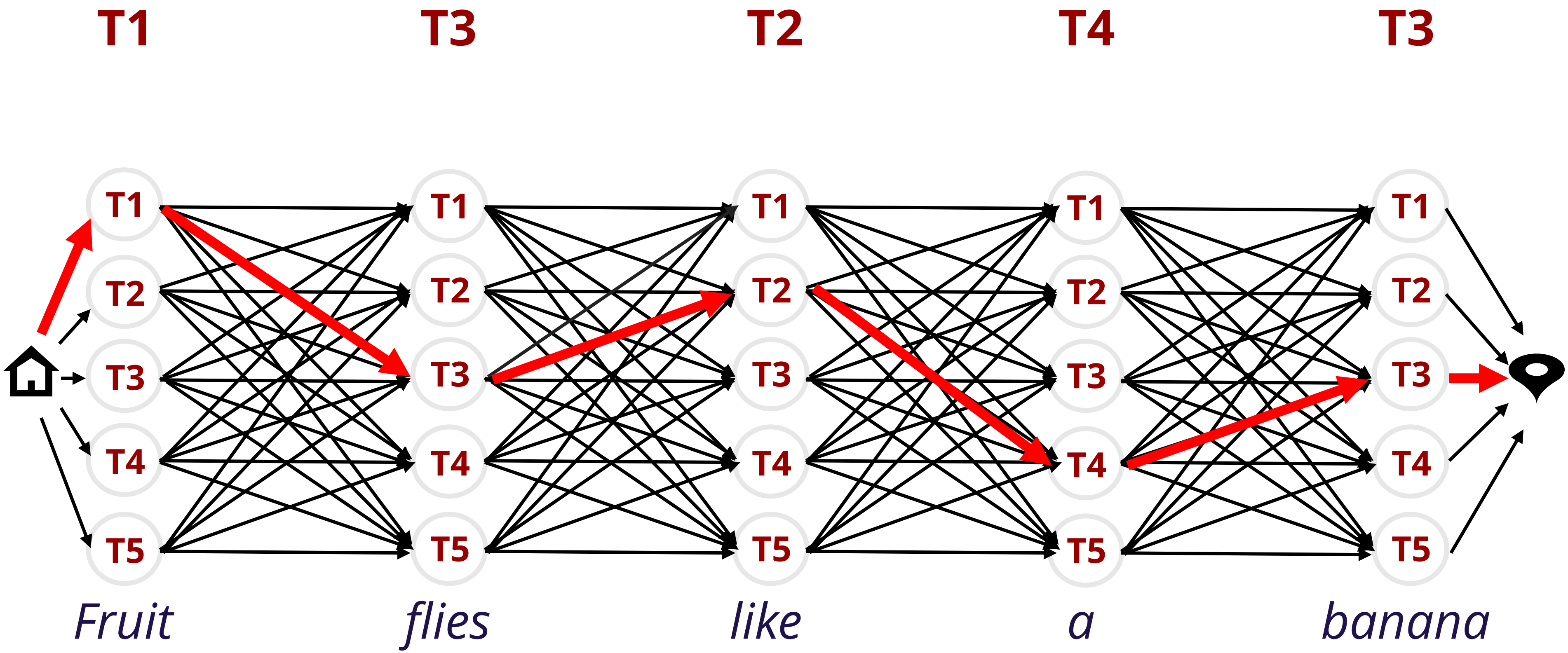


Hidden Markov Model
Decoding

$$y_1^*, \ldots, y_m^*$$
$$= \arg\max_{y_1, \ldots, y_m} p(x_1, \ldots, x_m, y_0, y_1, \ldots, y_m, y_{m+1})$$

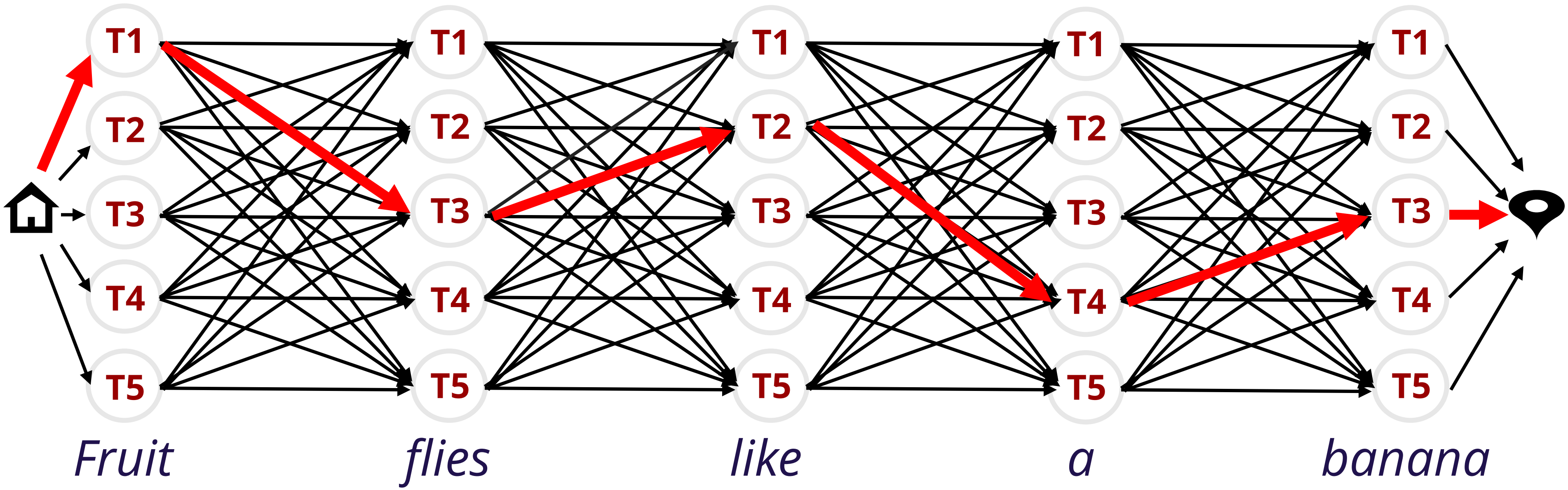


Hidden Markov Model
Decoding



Hidden Markov Model
Decoding

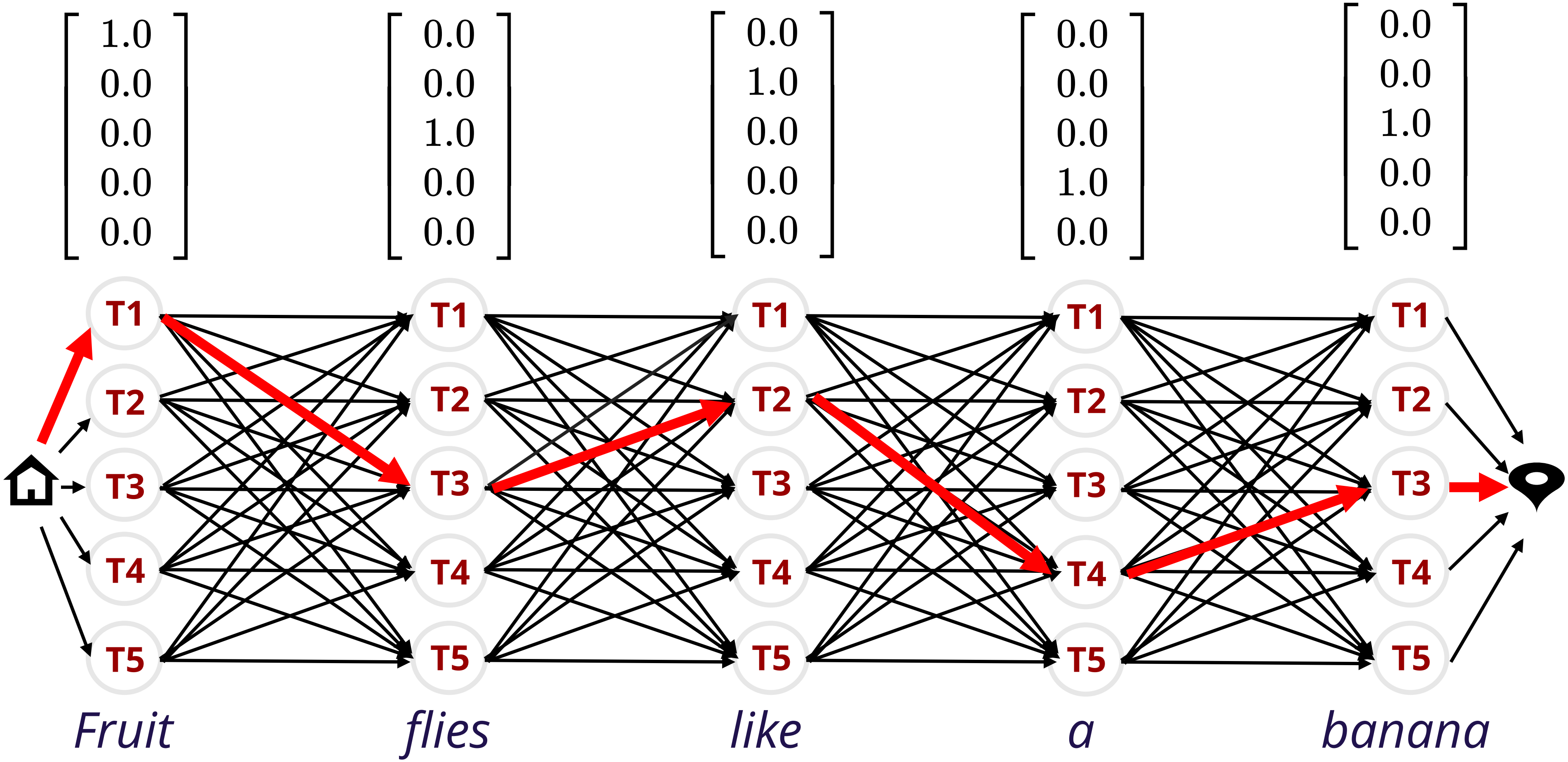
$$\begin{array}{c} \mathbf{T1} \\ \mathbf{T2} \\ \mathbf{T3} \\ \mathbf{T4} \\ \mathbf{T5} \end{array} \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{bmatrix} \quad \begin{bmatrix} 0.0 \\ 0.0 \\ 1.0 \\ 0.0 \\ 0.0 \end{bmatrix} \quad \begin{bmatrix} 0.0 \\ 1.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{bmatrix} \quad \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \\ 0.0 \end{bmatrix} \quad \begin{bmatrix} 0.0 \\ 0.0 \\ 1.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$



Contextual Embedding

Low-Dimensional One-hot Vectors

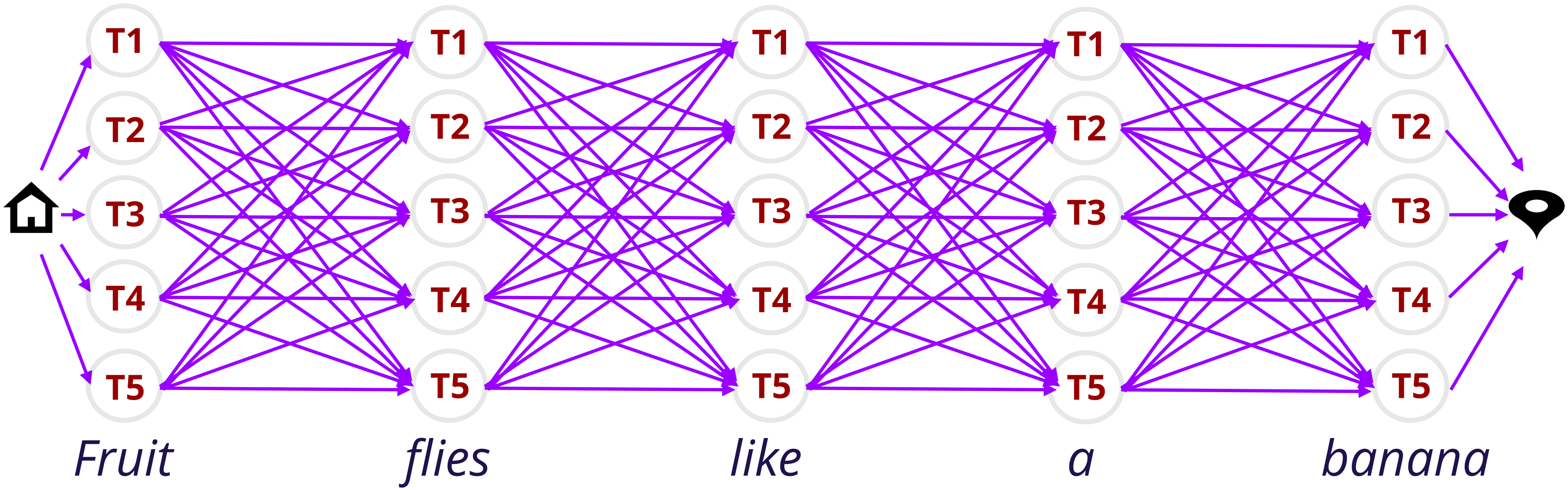
Similar to Brown Clusters (Brown et al. 1992)



Contextual Embedding

Use Forward-backward Algorithm to find probabilities for all tags at each position

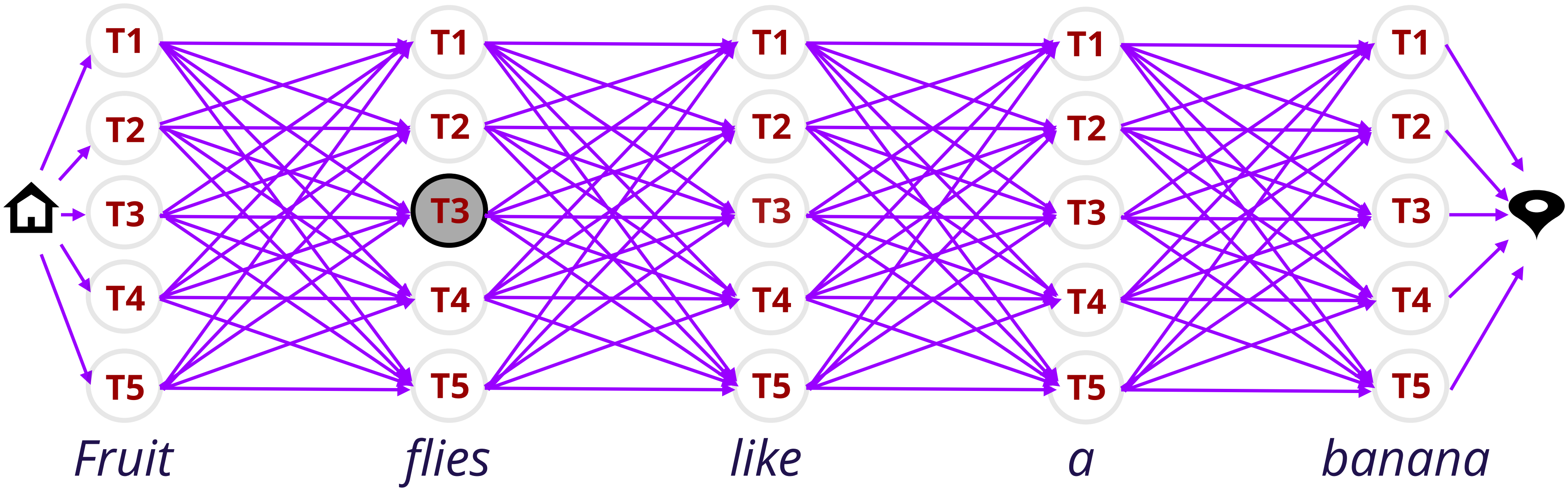
$$\begin{bmatrix} 0.76 \\ 0.11 \\ 0.03 \\ 0.05 \\ 0.05 \end{bmatrix} \quad \begin{bmatrix} 0.08 \\ 0.02 \\ 0.87 \\ 0.02 \\ 0.01 \end{bmatrix} \quad \begin{bmatrix} 0.01 \\ 0.91 \\ 0.02 \\ 0.03 \\ 0.03 \end{bmatrix} \quad \begin{bmatrix} 0.01 \\ 0.01 \\ 0.01 \\ 0.92 \\ 0.05 \end{bmatrix} \quad \begin{bmatrix} 0.01 \\ 0.02 \\ 0.89 \\ 0.04 \\ 0.04 \end{bmatrix}$$



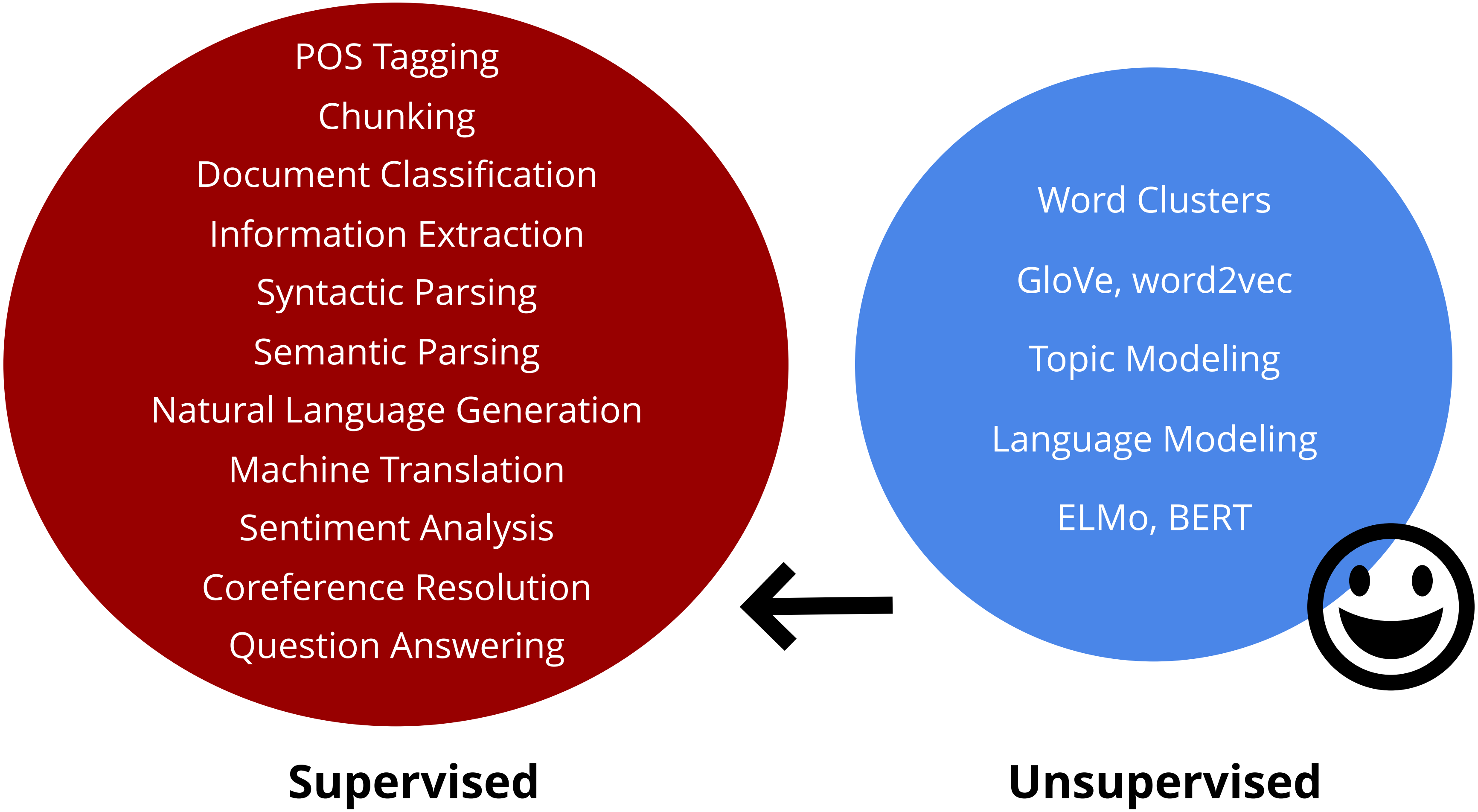
Contextual Embedding

Use Forward-backward Algorithm to find probabilities for all tags at each position

$$\begin{bmatrix} 0.08 \\ 0.02 \\ \color{red}{0.87} \\ 0.02 \\ 0.01 \end{bmatrix} \longleftarrow \frac{\alpha_{T_3}(2)\beta_{T_3}(2)}{Z}$$



Tasks in NLP


POS Tagging
Chunking
Document Classification
Information Extraction
Syntactic Parsing
Semantic Parsing
Natural Language Generation
Machine Translation
Sentiment Analysis
Coreference Resolution
Question Answering
Word Clusters
GloVe, word2vec
Topic Modeling
Language Modeling
ELMo, BERT
Supervised
Unsupervised

Tasks in NLP

POS Tagging
Chunking
Document Classification
Information Extraction
Syntactic Parsing
Semantic Parsing
Natural Language Generation
Machine Translation
Sentiment Analysis
Coreference Resolution
Question Answering

Supervised

Next Class