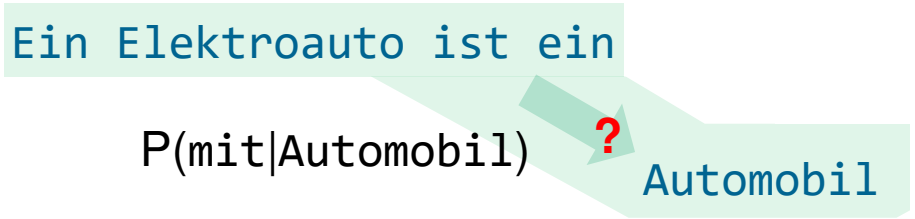
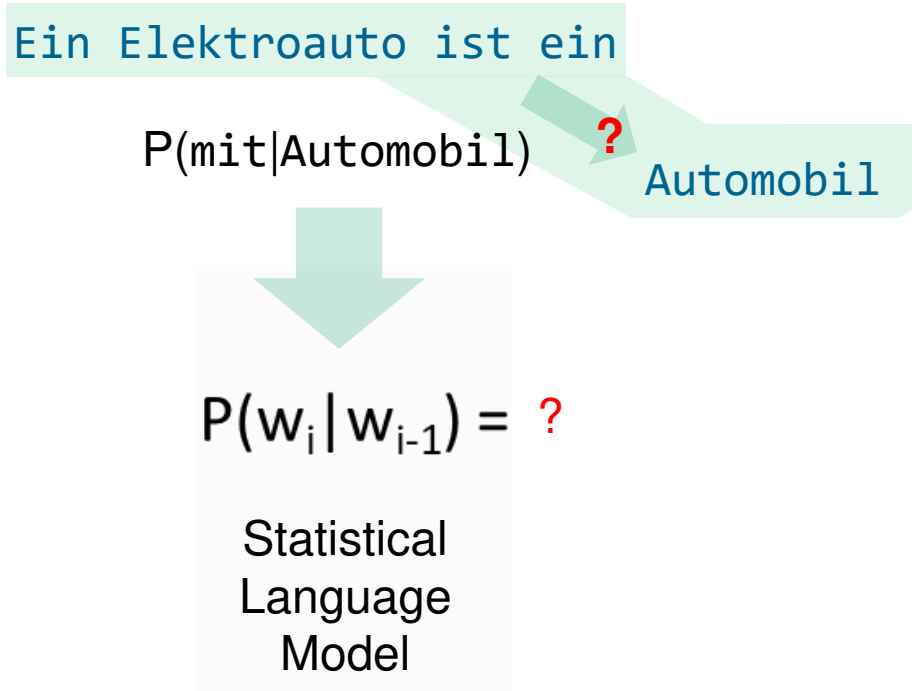
Statistical Language Model: How likely is the next word?



Statistical Language Model: How likely is the next word?



Statistical Language Model: How likely is the next word?

Ein Elektroauto ist ein

P(mit|Automobil) ? Automobil

$$P(w_i \mid w_{i-1}) = \frac{P(w_i w_{i-1})}{P(w_{i-1})} = ?$$

Statistical Language Model

n-gram Probabilities

Statistical Language Model: How likely is the next word?

Ein Elektroauto ist ein

P(mit|Automobil) ? Automobil

$$P(w_i \mid w_{i-1}) = \frac{P(w_i w_{i-1})}{P(w_{i-1})} = \frac{\frac{|w_i w_{i-1}|}{|corpus|}}{\frac{|w_{i-1}|}{|corpus|}} = \frac{|w_i w_{i-1}|}{|w_{i-1}|}$$

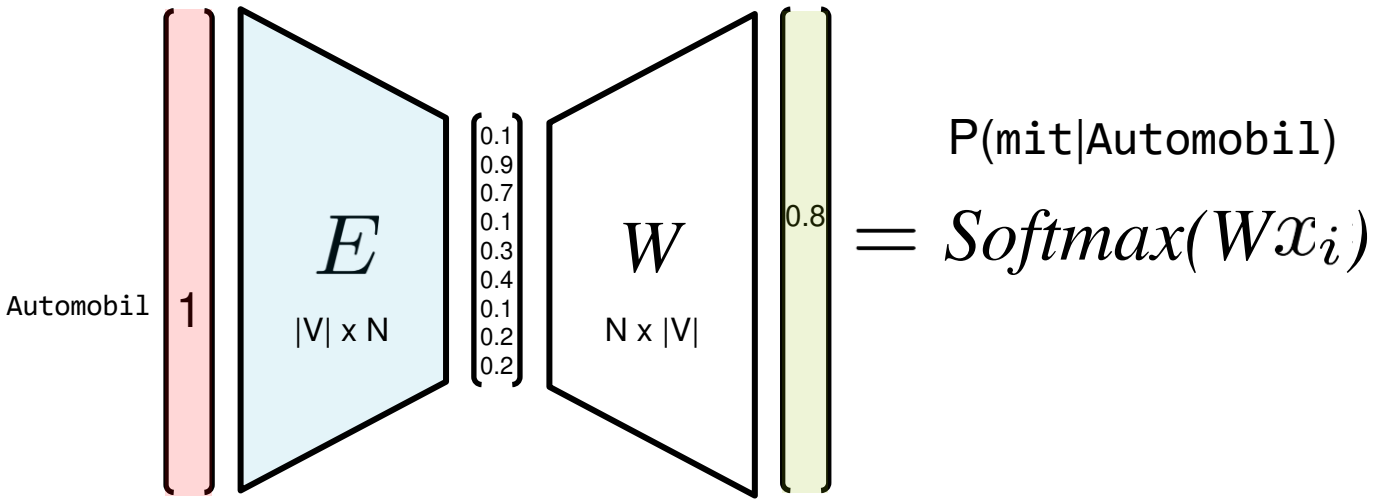
Statistical Language Model

n-gram Probabilities

MLE

Statistical Language Model: How likely is the next word?

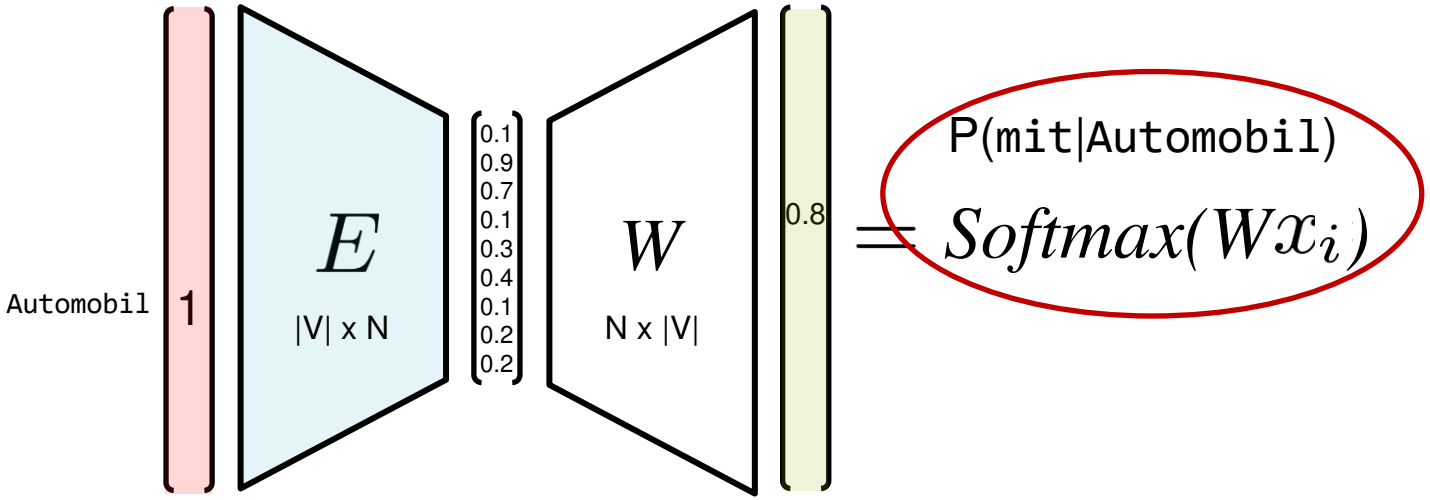
`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`



$$\text{P(mit|Automobil)} = Softmax(Wx_i)$$

Statistical Language Model: How likely is the next word?

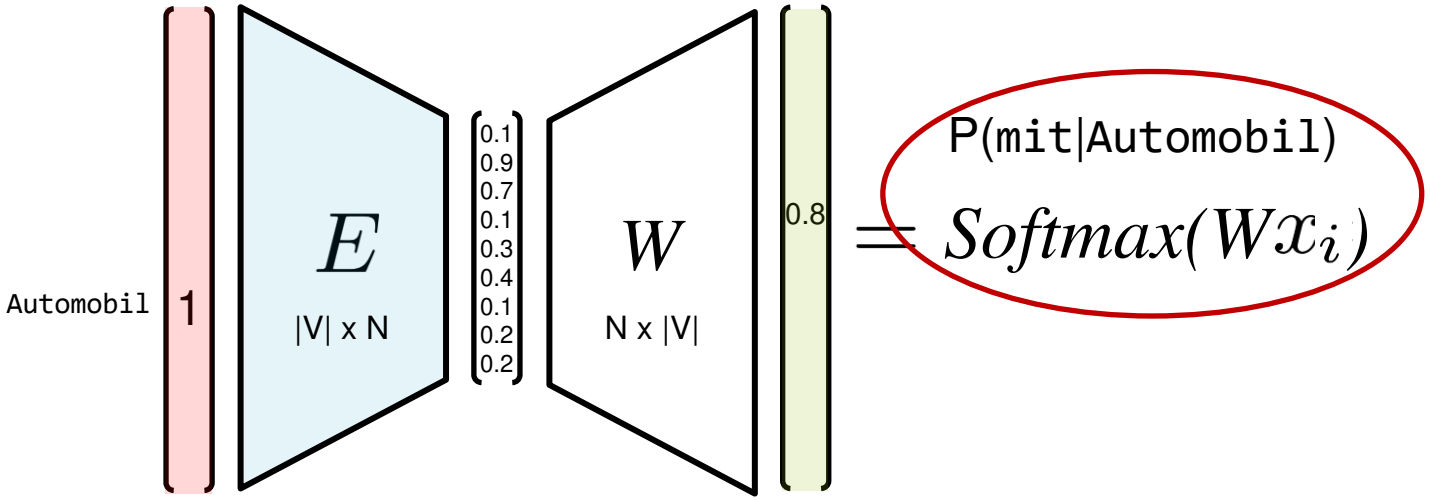
`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`



Same computational issues as in computing word embedding. Solutions are:

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

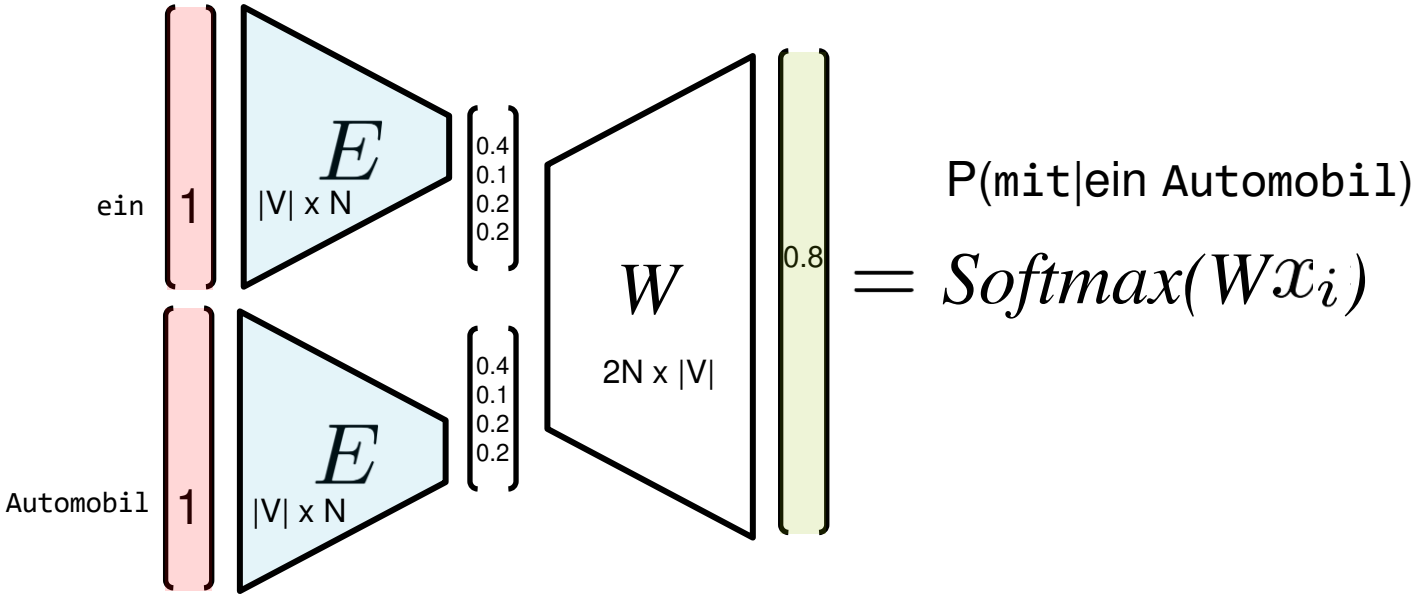


Same computational issues as in computing word embedding. Solutions are:

- Hierarchical Softmax
- “moving to binary classifiers”
- Negative Sampling etc.

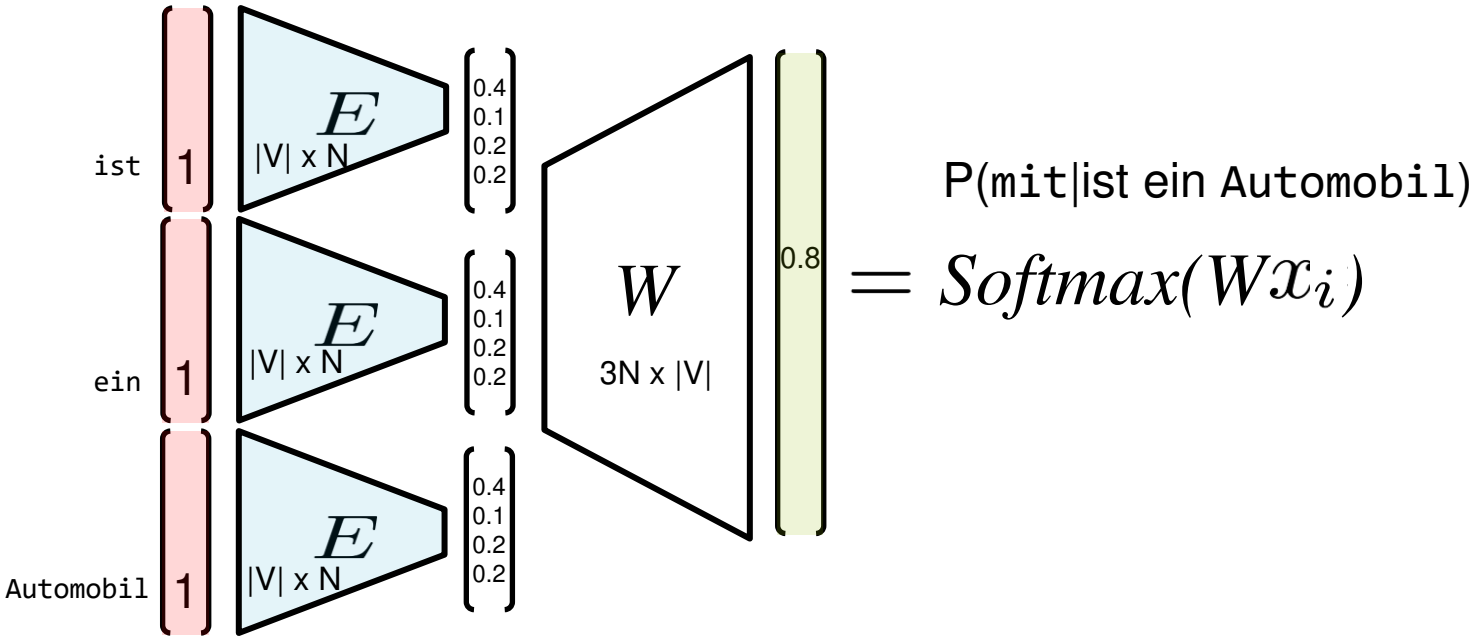
Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`



Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

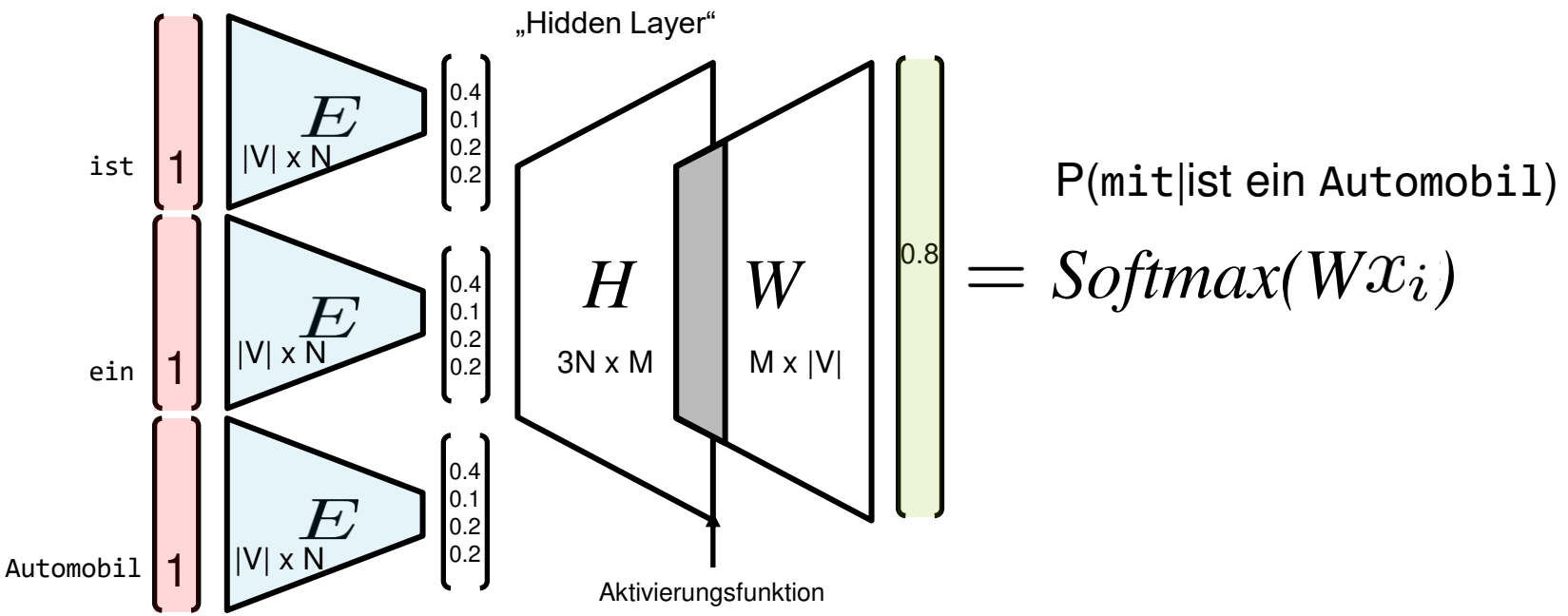


P(mit|ist ein Automobil)

$$= Softmax(Wx_i)$$

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

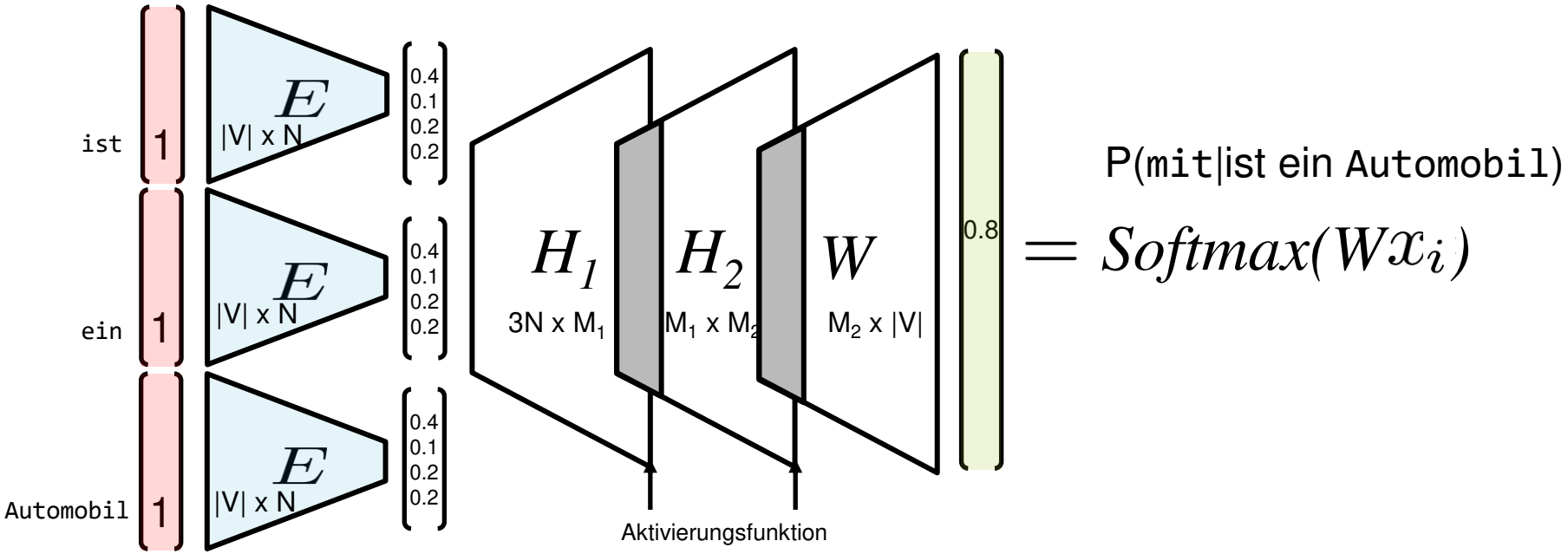


P(mit|ist ein Automobil)

$$= Softmax(Wx_i)$$

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

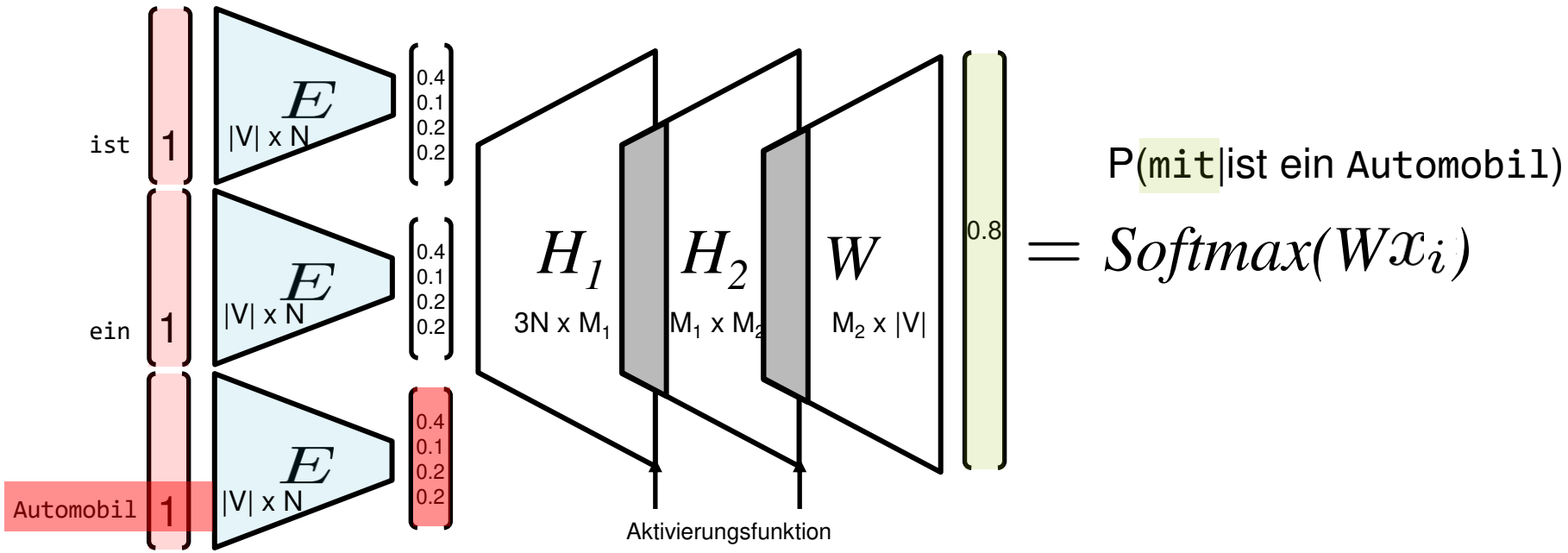


P(mit|ist ein Automobil)

$$= Softmax(W x_i)$$

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

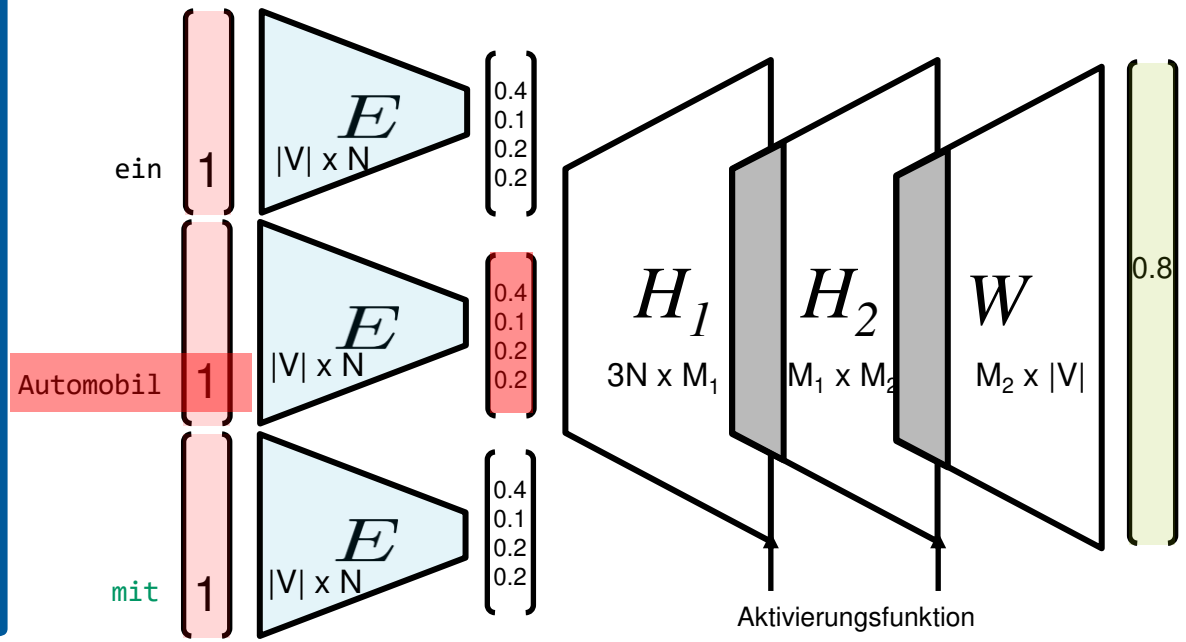


P(mit|ist ein Automobil)

$$= Softmax(W x_i)$$

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

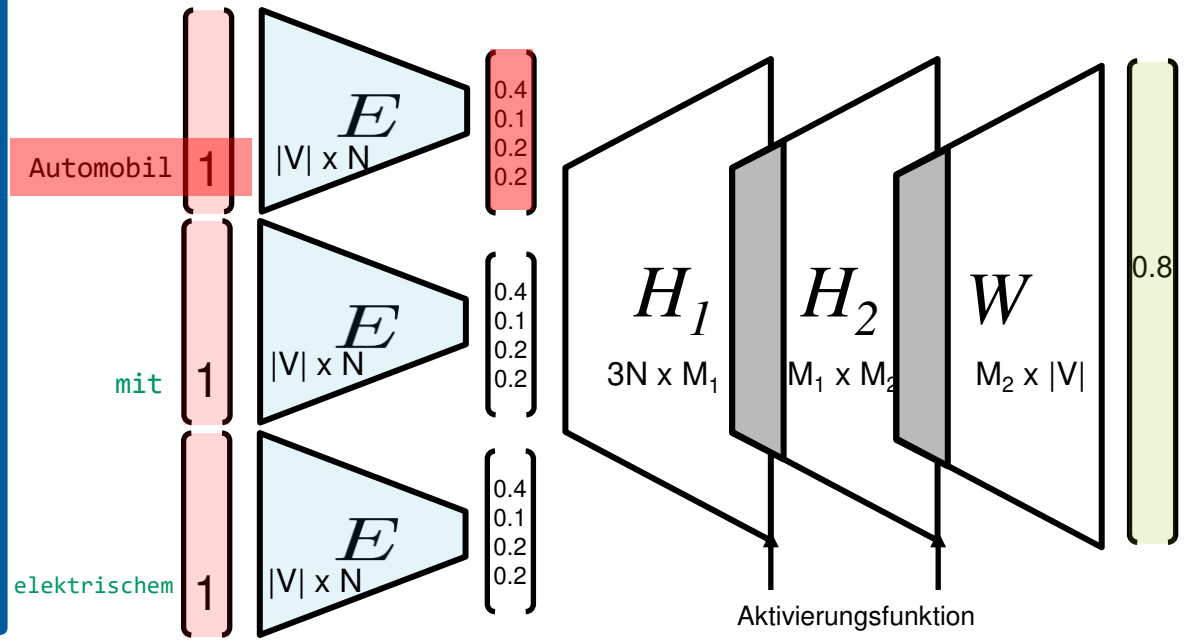


$$P(\texttt{elektrischem} \mid \text{ein Automobil } \texttt{mit}) = Softmax(Wx_i)$$

Vorrangegangene Vorhersage

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`



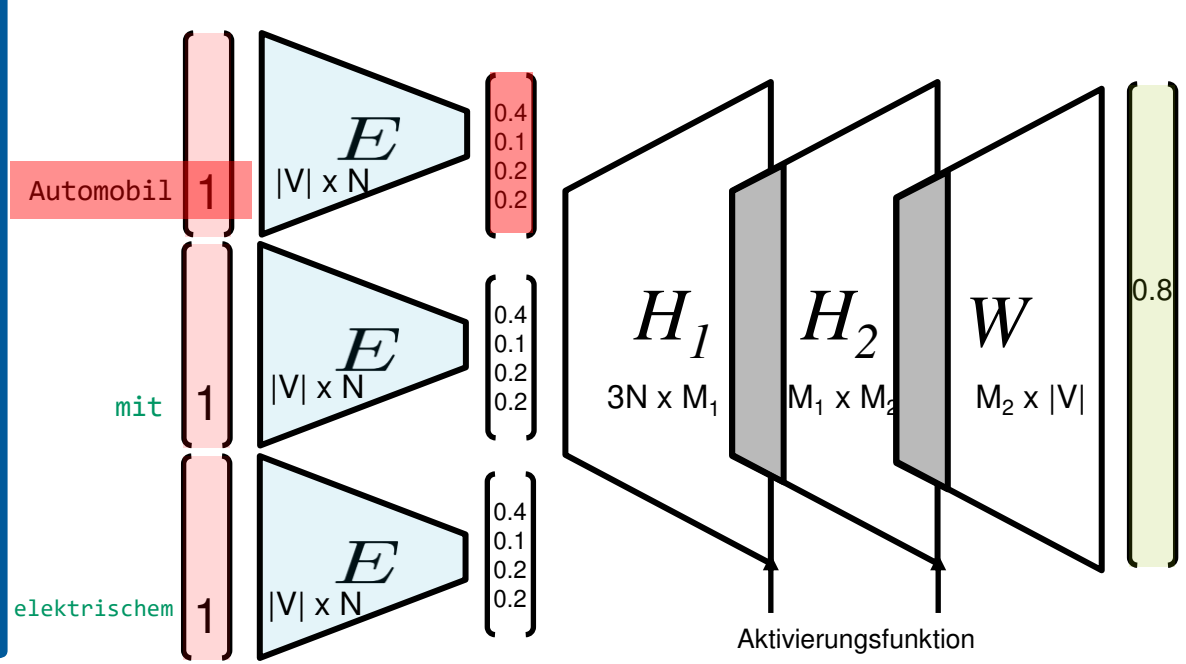
$$P(\texttt{Antrib} \mid \texttt{Automobil mit elektrischem}) = Softmax(W x_i)$$

Vorrangegangene Vorhersage

Vor-Vorrangegangene Vorhersage

Statistical Language Model: How likely is the next word?

`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`



$$= Softmax(W x_i)$$

History is n=3, as similar as with n-gram SLMs

$$P(\texttt{Antrib} \mid \texttt{Automobil mit elektrischem})$$

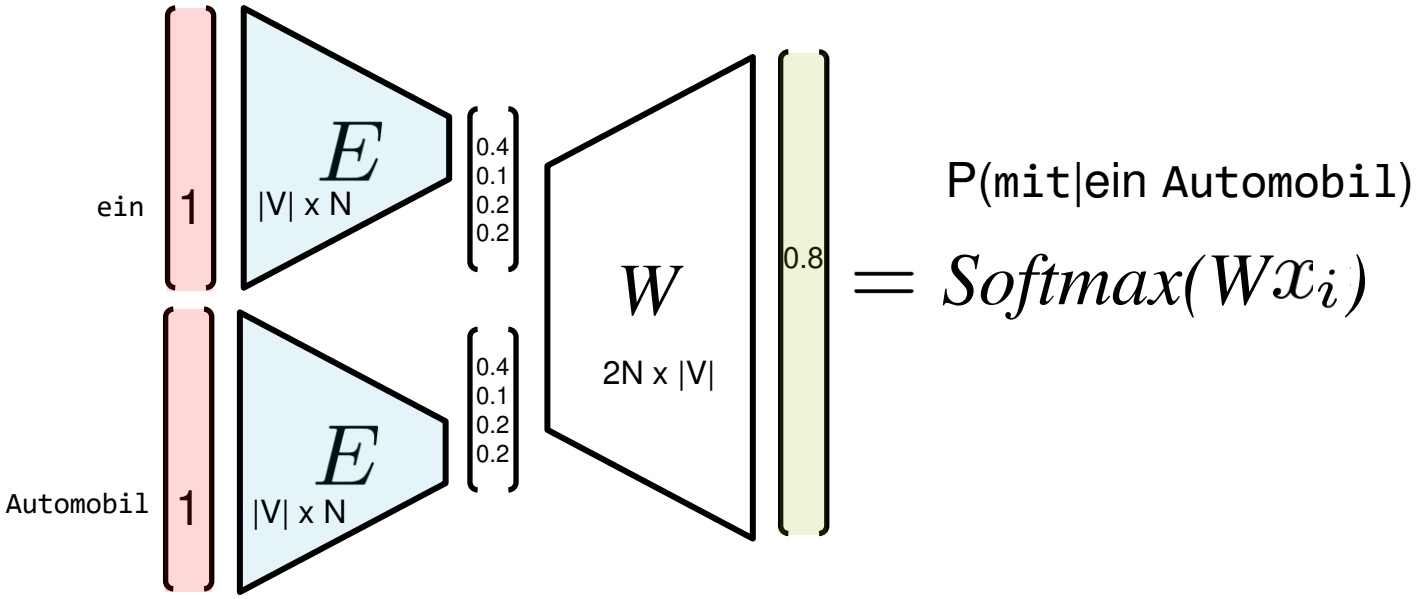
Vor-Vorrangegangene Vorhersage

Vorrangegangene Vorhersage

Statistical Language Model: How likely is the next word?

input output

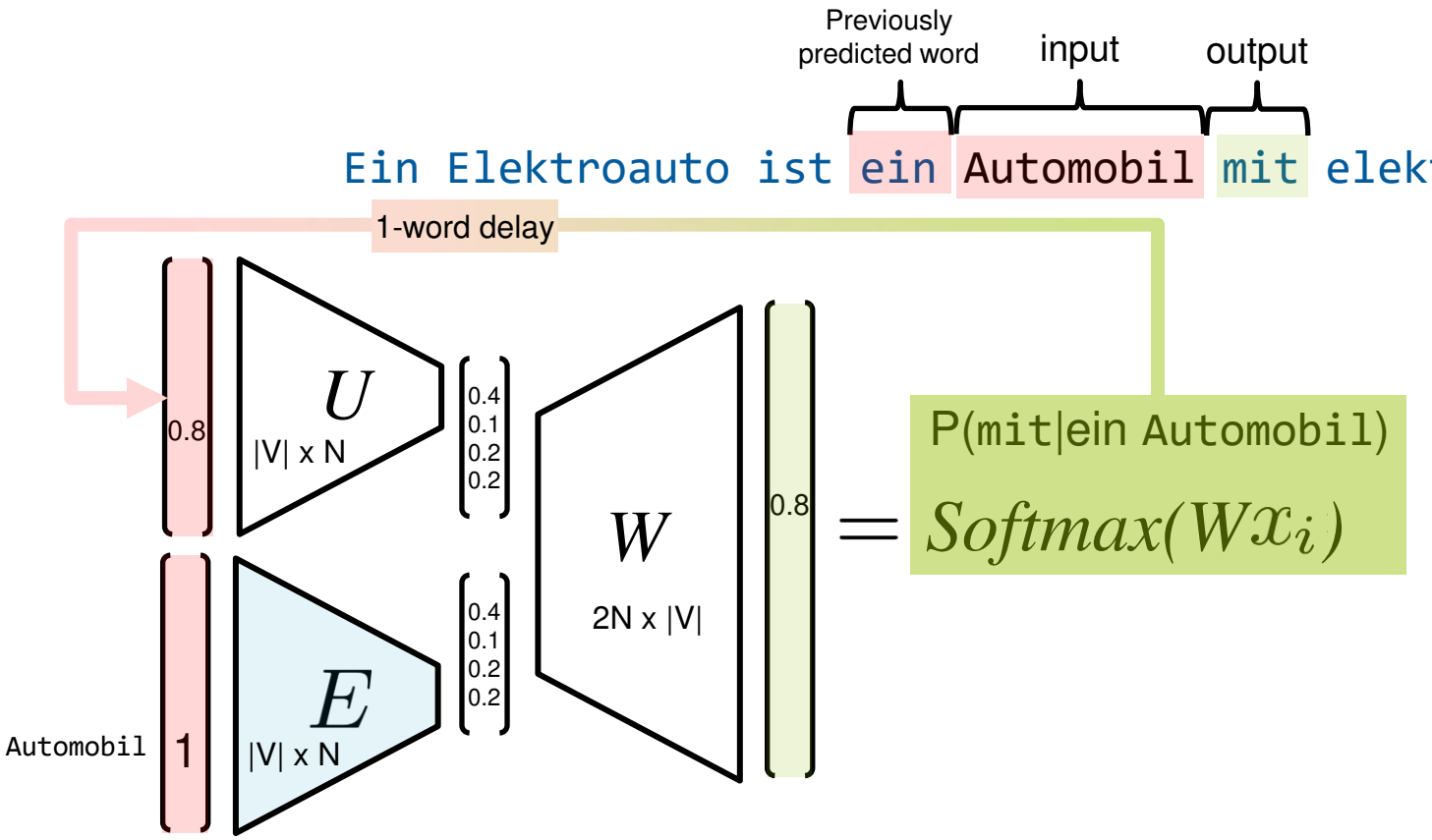
`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`



$$\text{P(mit|ein Automobil)} = Softmax(Wx_i)$$

Remember: we need to approximate the softmax or use some “tricks”.

Statistical Language Model: How likely is the next word?



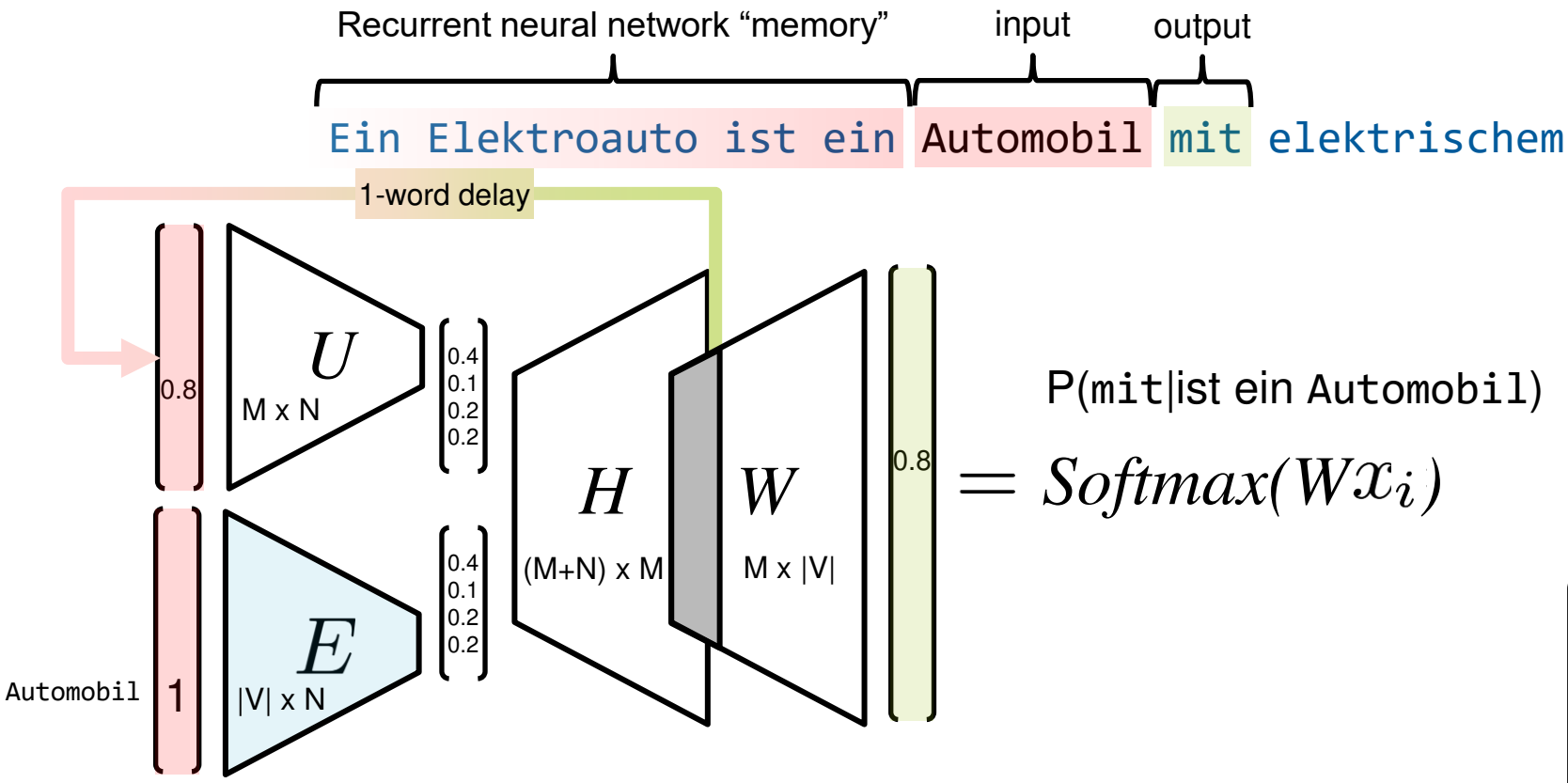
Remember: we need to approximate the softmax or use some “tricks”.

Minimal Jordan RNN Equation:

$$h_t = \sigma_h(W_h x_t + U_h y_{t-1} + b_h)$$

$$y_t = \sigma_y(W_y h_t + b_y)$$

Statistical Language Model: How likely is the next word?



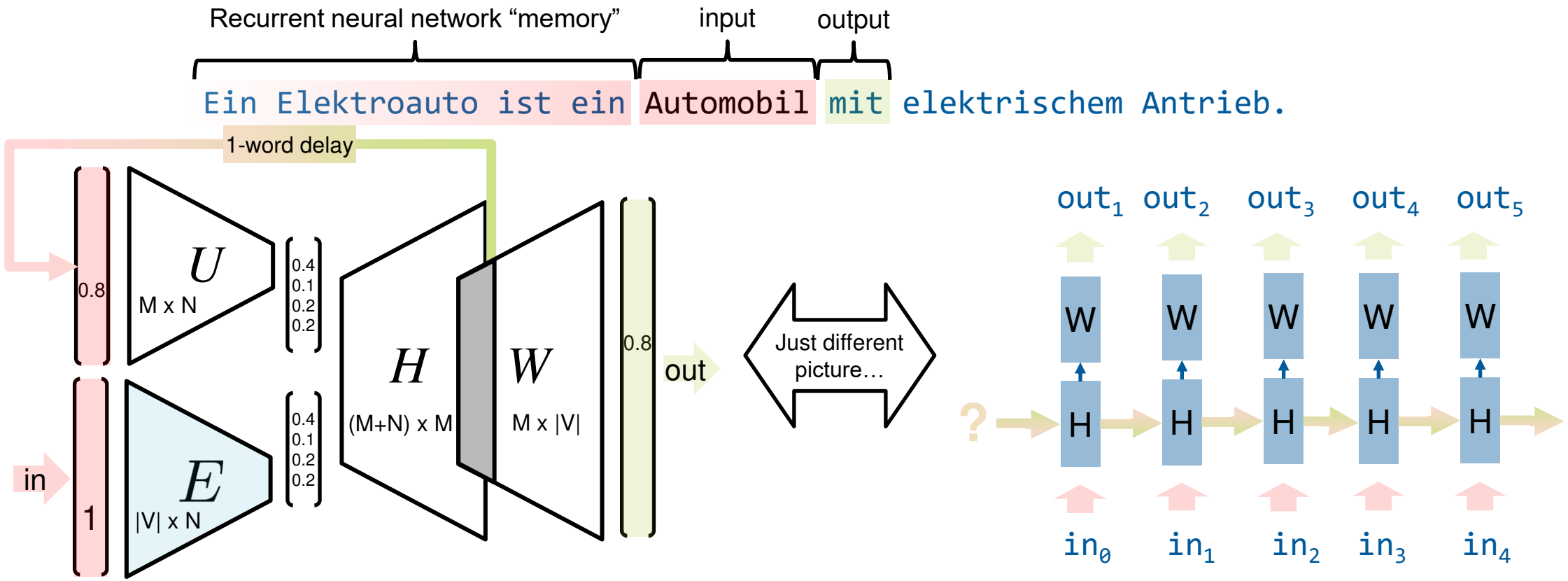
$$P(\texttt{mit}|\text{ist ein }\texttt{Automobil}) = Softmax(Wx_i)$$

Remember: we need to approximate the softmax or use some “tricks”.

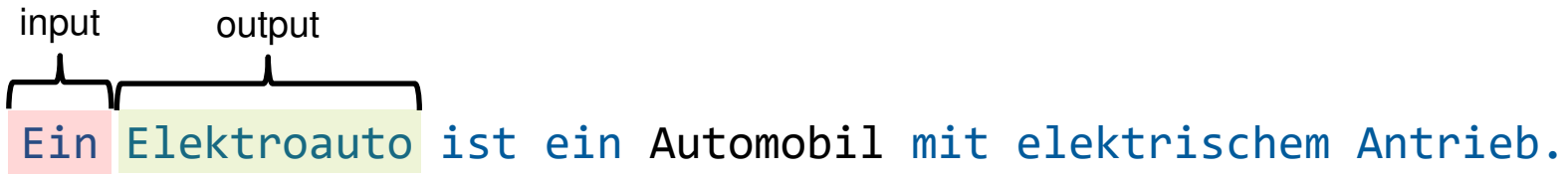
Minimal Elman RNN Equation:

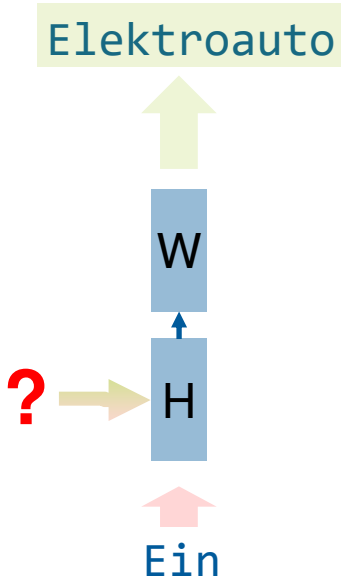
$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

Statistical Language Model: How likely is the next word?

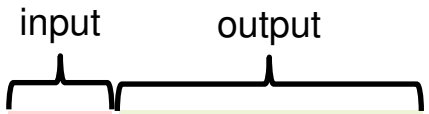


Statistical Language Model: How likely is the next word?



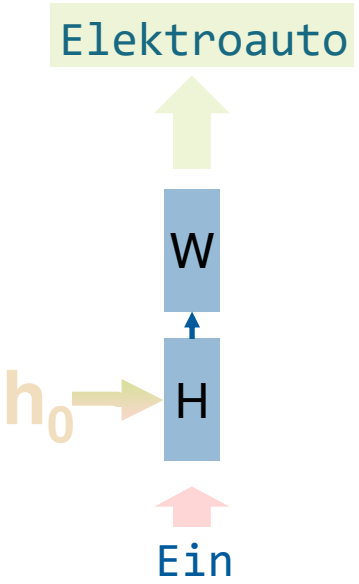


Statistical Language Model: How likely is the next word?

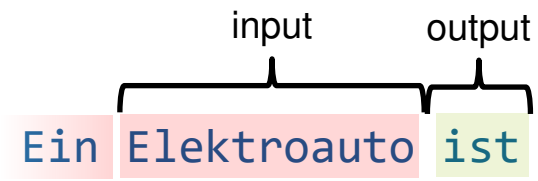


`Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.`

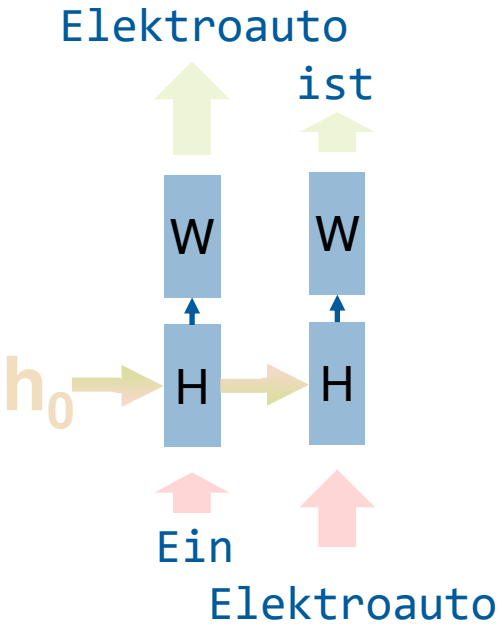
Initialization of the RNN is a real problem. Typically, it's just 0. But even if it's 0, when should we "reset" the RNN, e.g. after each sentence, document, corpus?



Statistical Language Model: How likely is the next word?



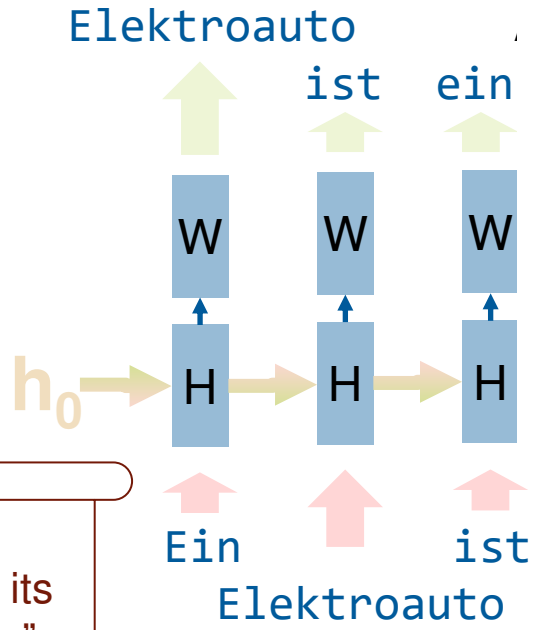
Note: The RNN predicts one output for each new input.



Statistical Language Model: How likely is the next word?

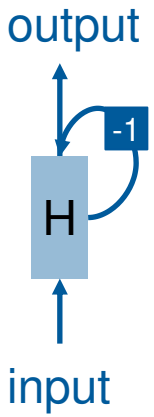
input output

```
Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.
```

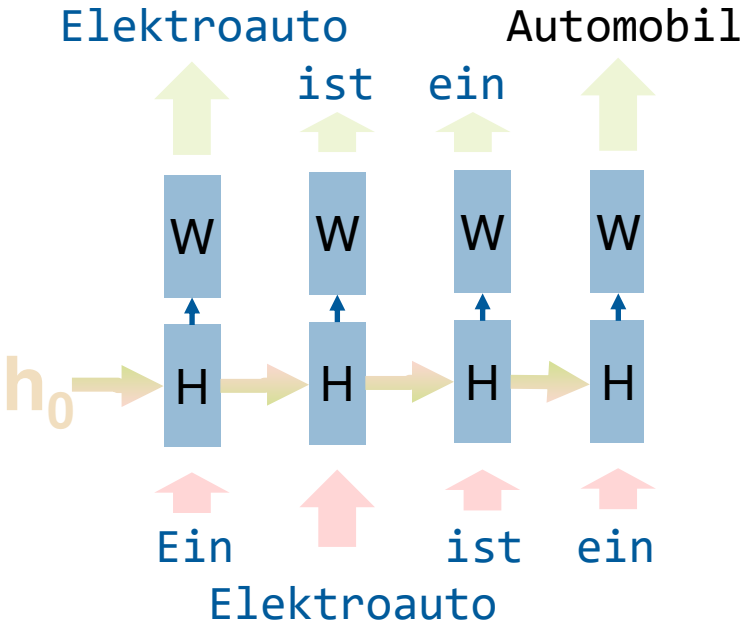


Remember: The word is given by its "word embedding"

Statistical Language Model: How likely is the next word?



Note: The RNN is inferenced with an "RNN iterator"

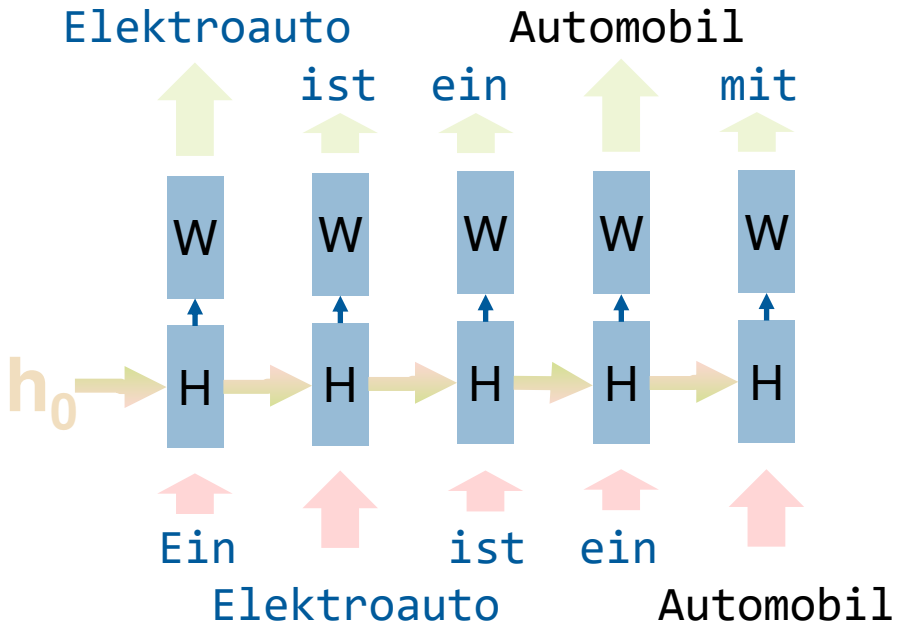


Statistical Language Model: How likely is the next word?

input output

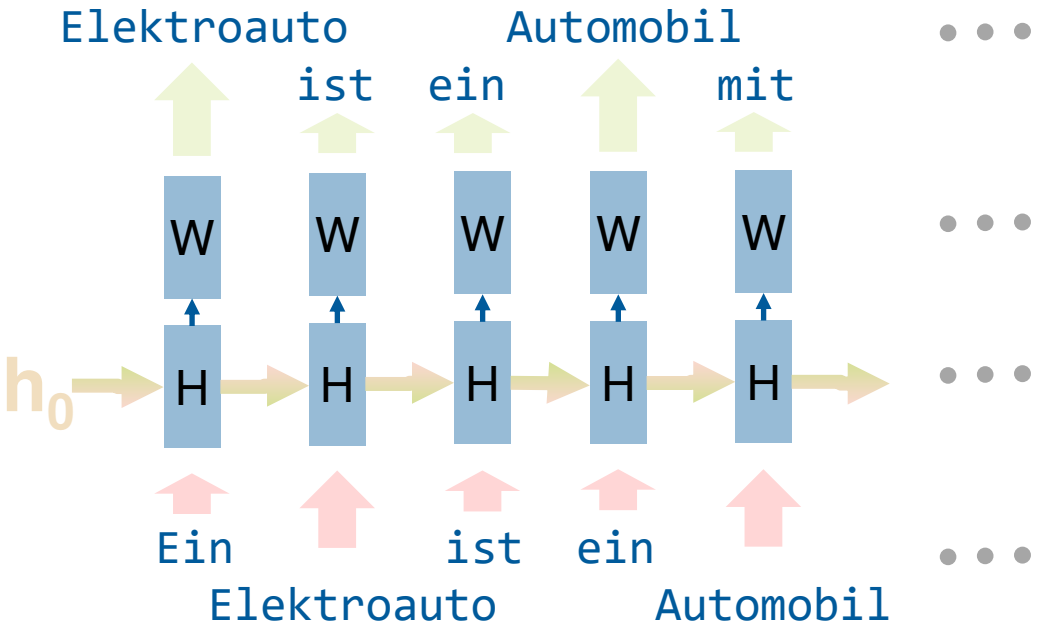
Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.

Note: The RNN is needs to store some values between inference steps.

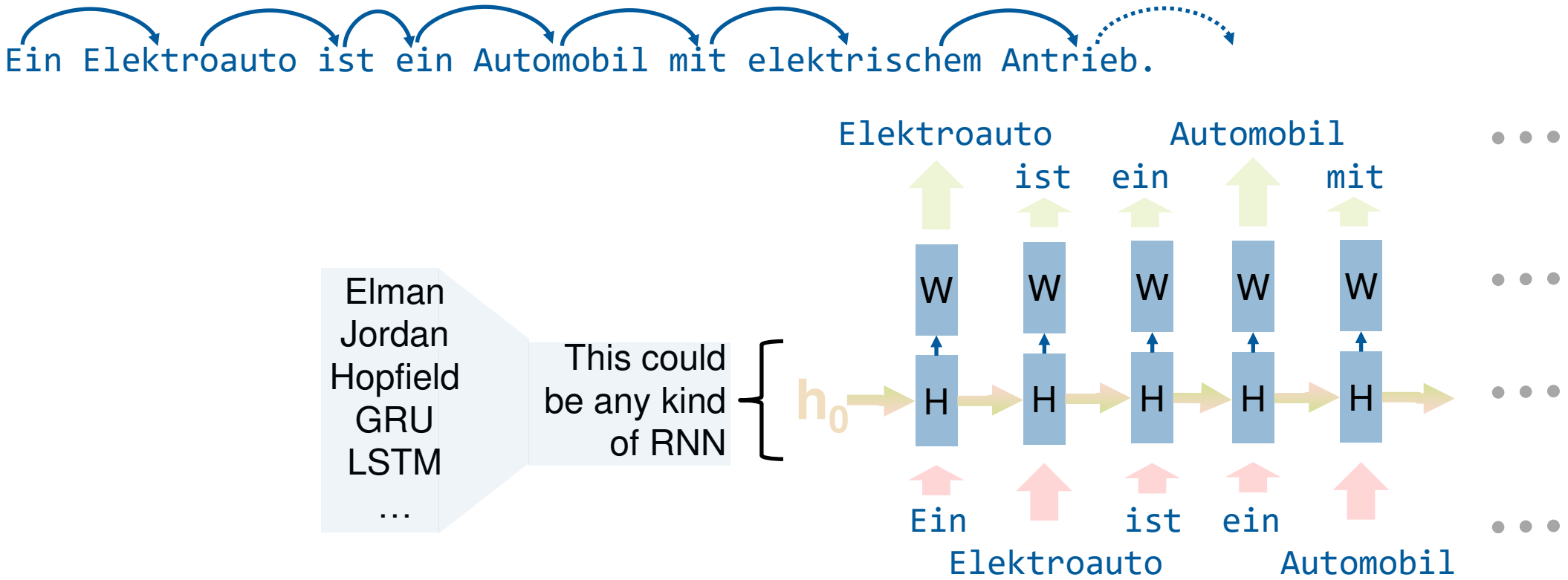


Statistical Language Model: How likely is the next word?

Note: The input sequence length is the same as the output sequence length.



Statistical Language Model: How likely is the next word?



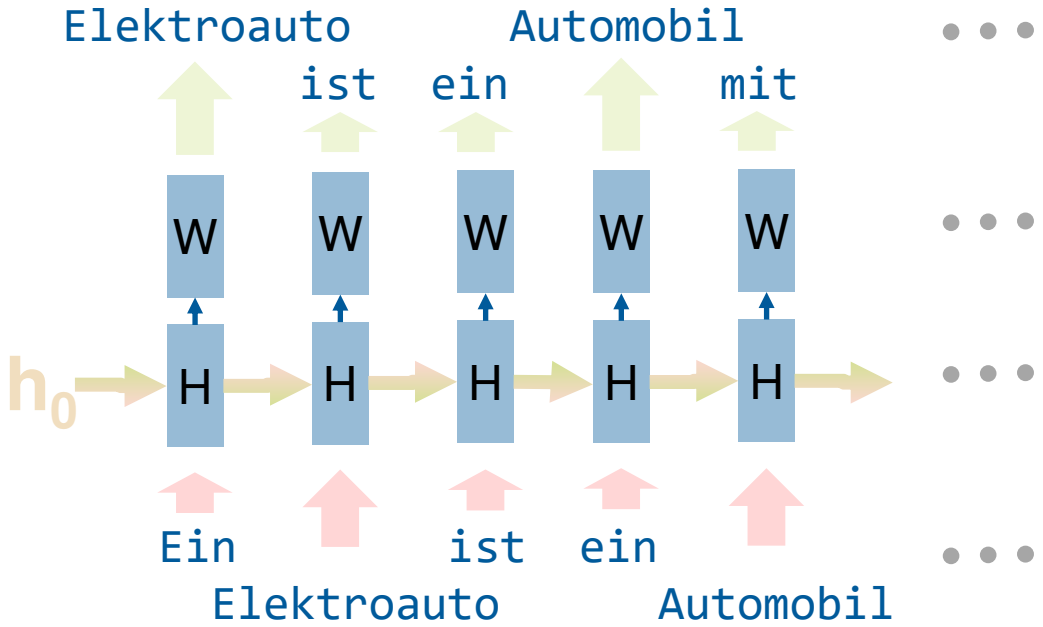
Statistical Language Model: How likely is the next word?

Remember the goal:

$P(w_i|w_0 \ldots w_{i-1})$

Elman
Jordan
Hopfield
GRU
LSTM
…

This could be any kind of RNN

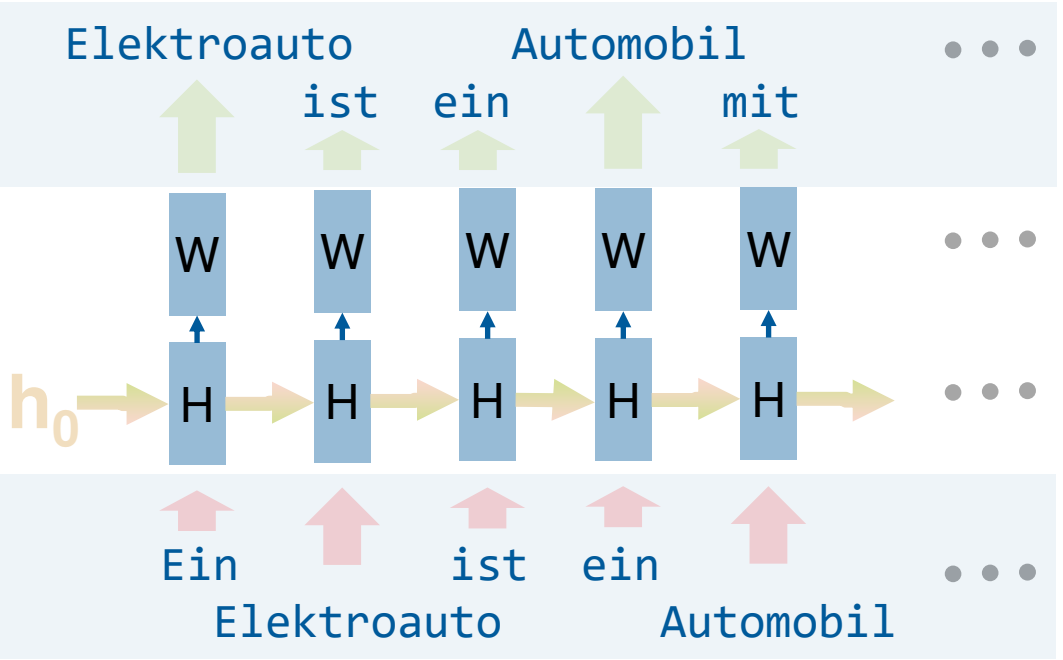


Statistical Language Model: How likely is the next-next word?

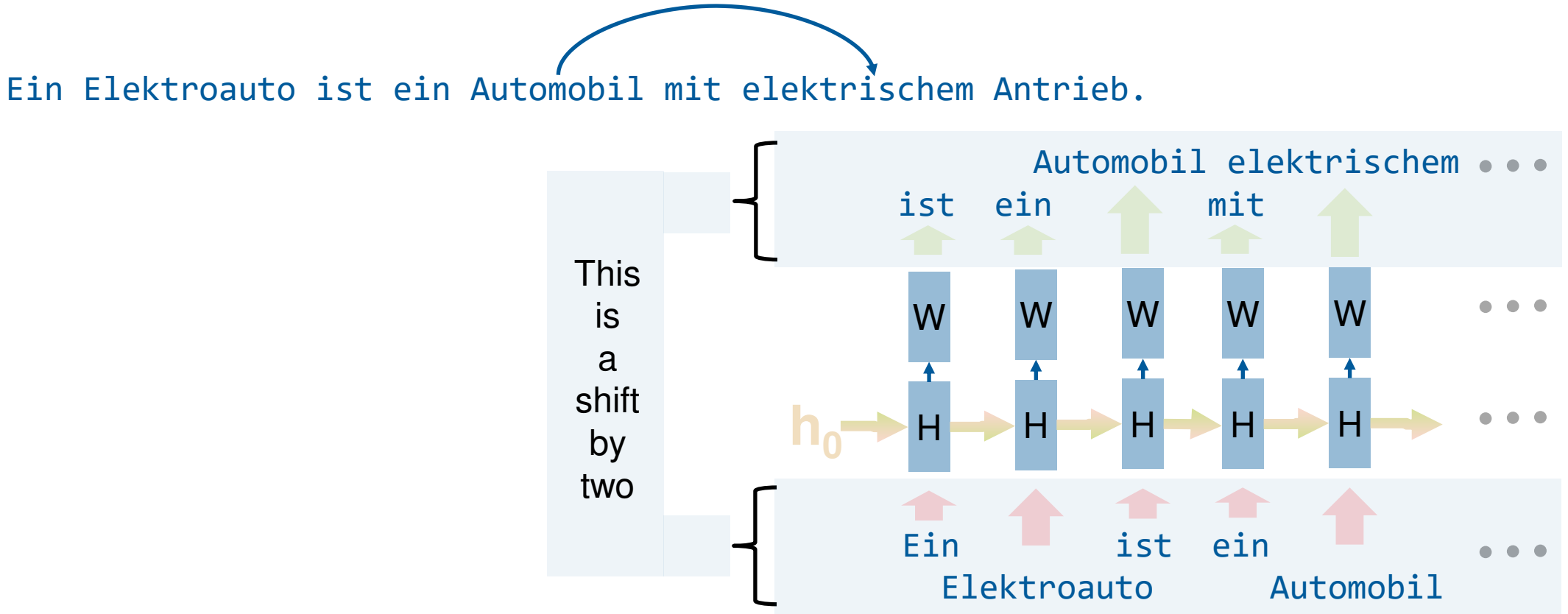
Ein Elektroauto ist ein Automobil mit elektrischem Antrieb.

? What about a shift by two ?

This is a shift by one



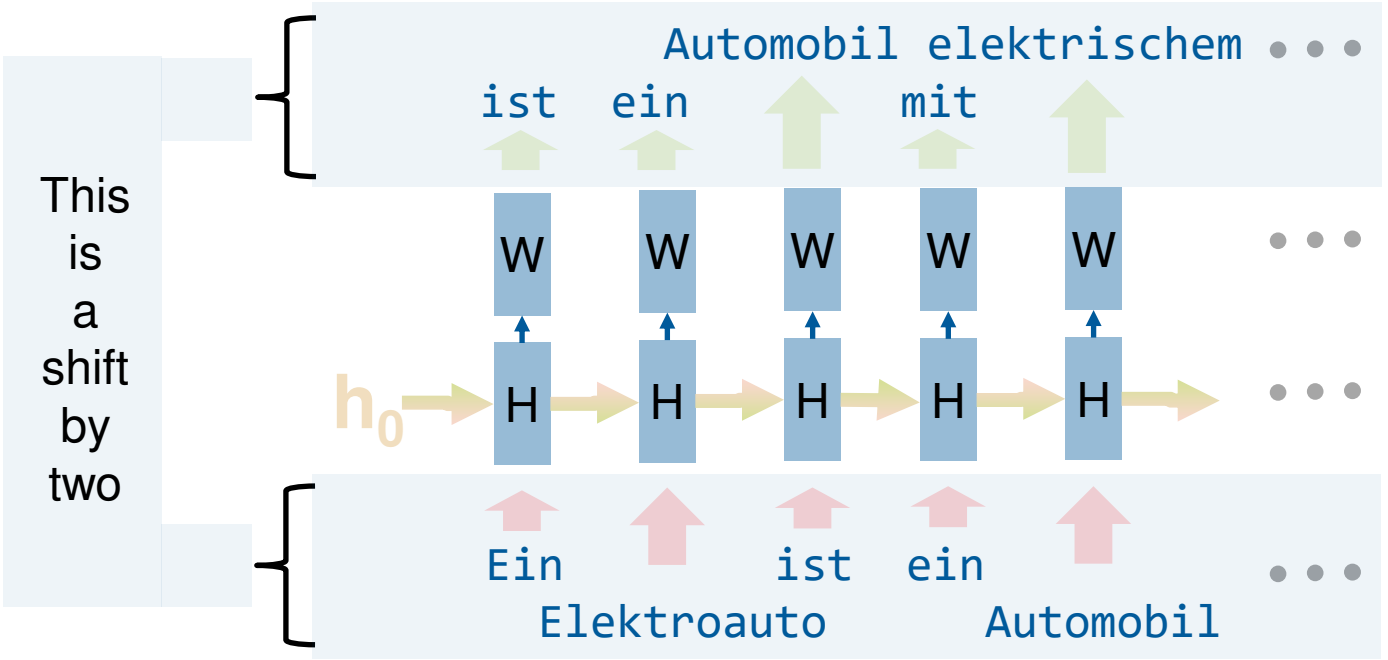
Statistical Language Model: How likely is the next-next word?



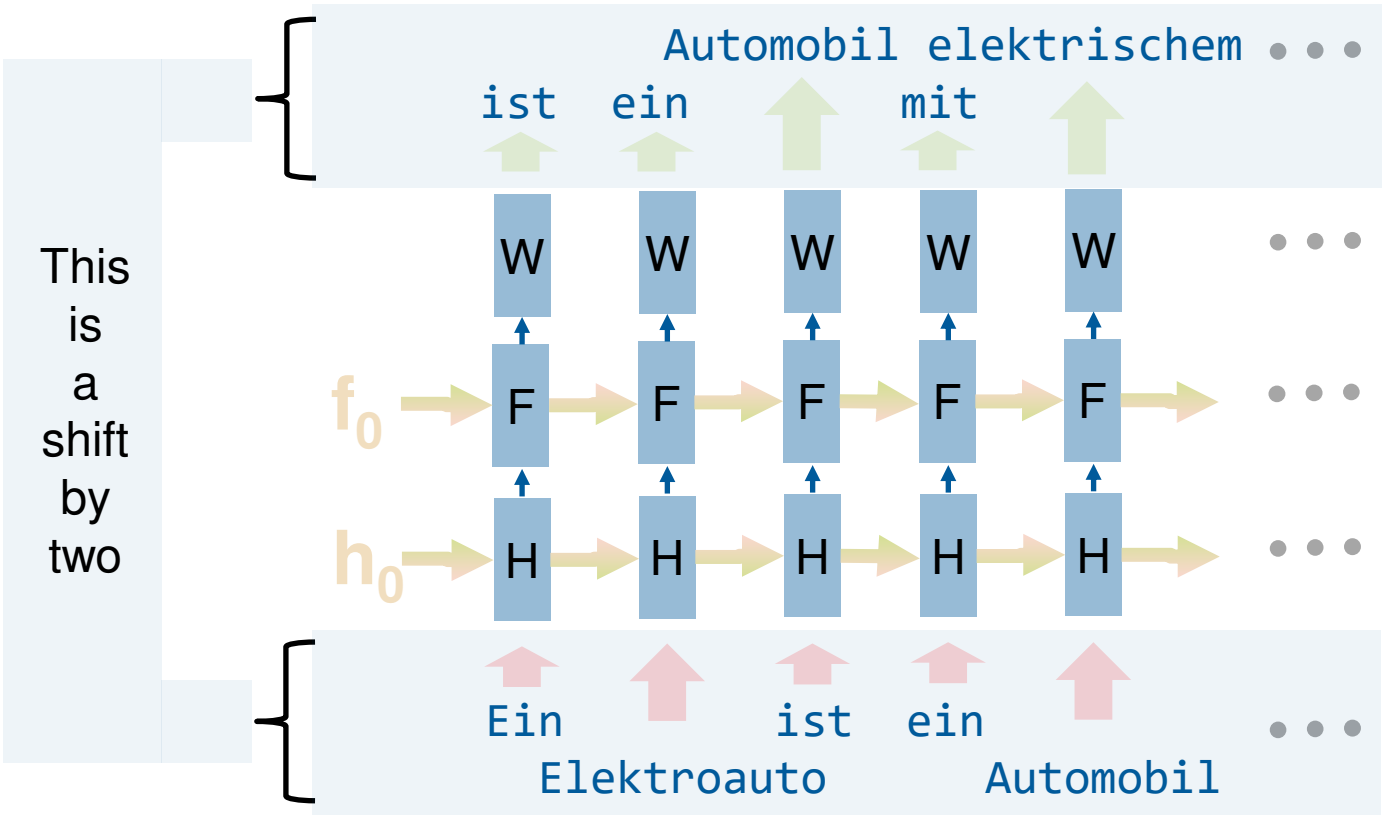
Statistical Language Model: How likely is the next-next word?

Remember the goal:

$P(w_i|w_0 \ldots w_{i-2})$



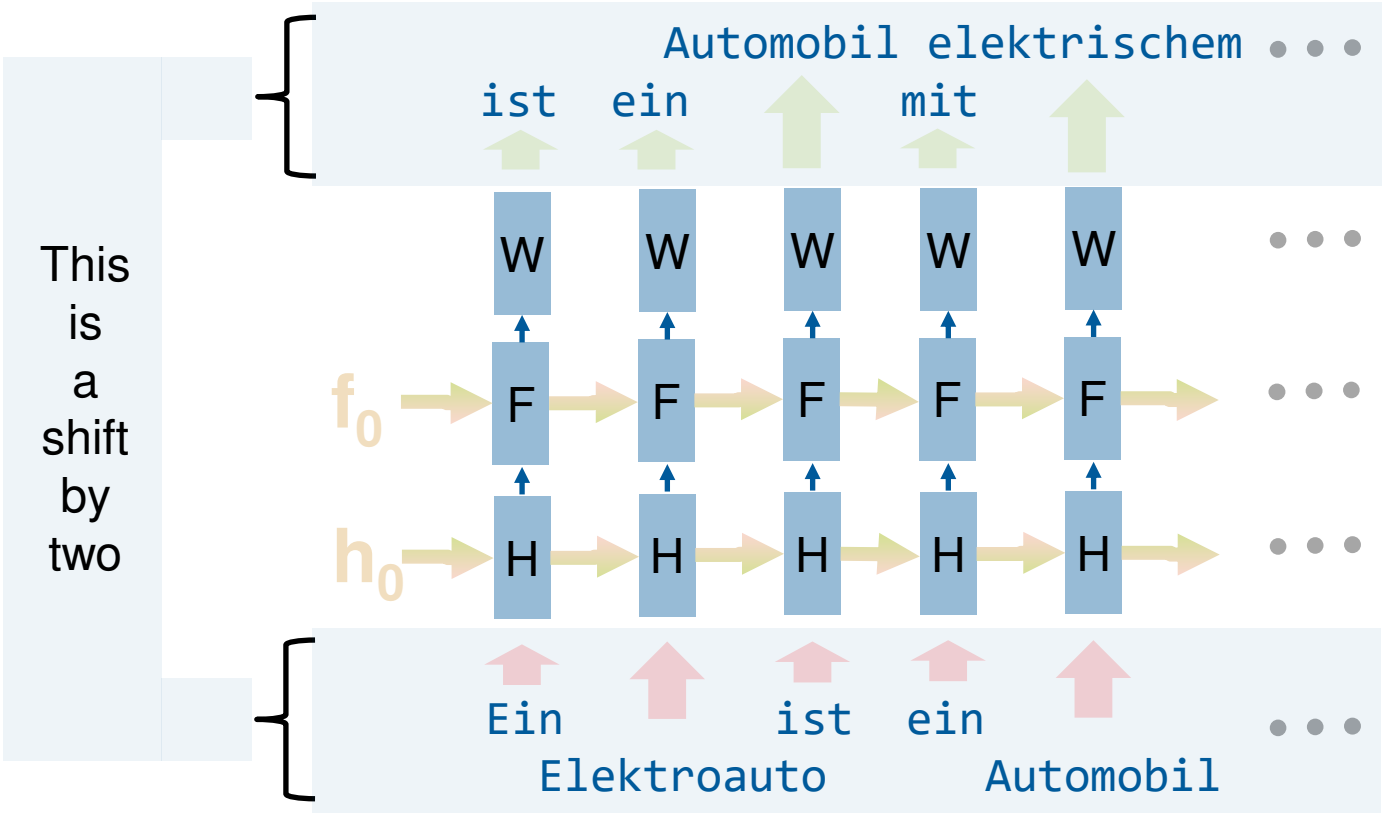
Stack of Recurrent Neural Networks



Stack of Recurrent Neural Networks

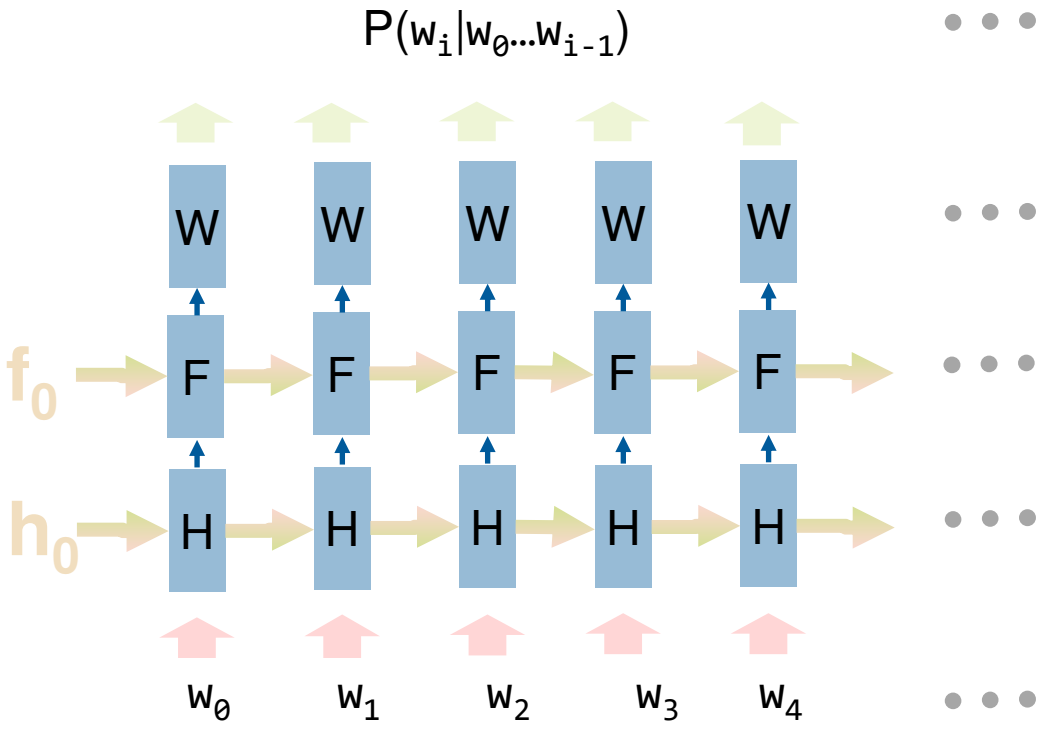
Stack == Delay?

We will discuss this later!



WikiText Long Term Dependency Language Modeling

Publication	Model	Parameters	Validation	Testing
Merity et al. 2016	Zoneout + Variational LSTM	20M	108.7	100.9
Grave et al. 2016	LSTM	-	-	99.3
Merity et al. 2016	Variational LSTM (code from Gal 2015)	20M	101.7	96.3
Grave et al. 2016	Neural cache model (size = 100)	-	-	81.6
Merity et al. 2016	Pointer LSTM (window = 100)	21M	84.8	80.8
Grave et al. 2016	Neural cache model (size = 2000)	-	-	68.9

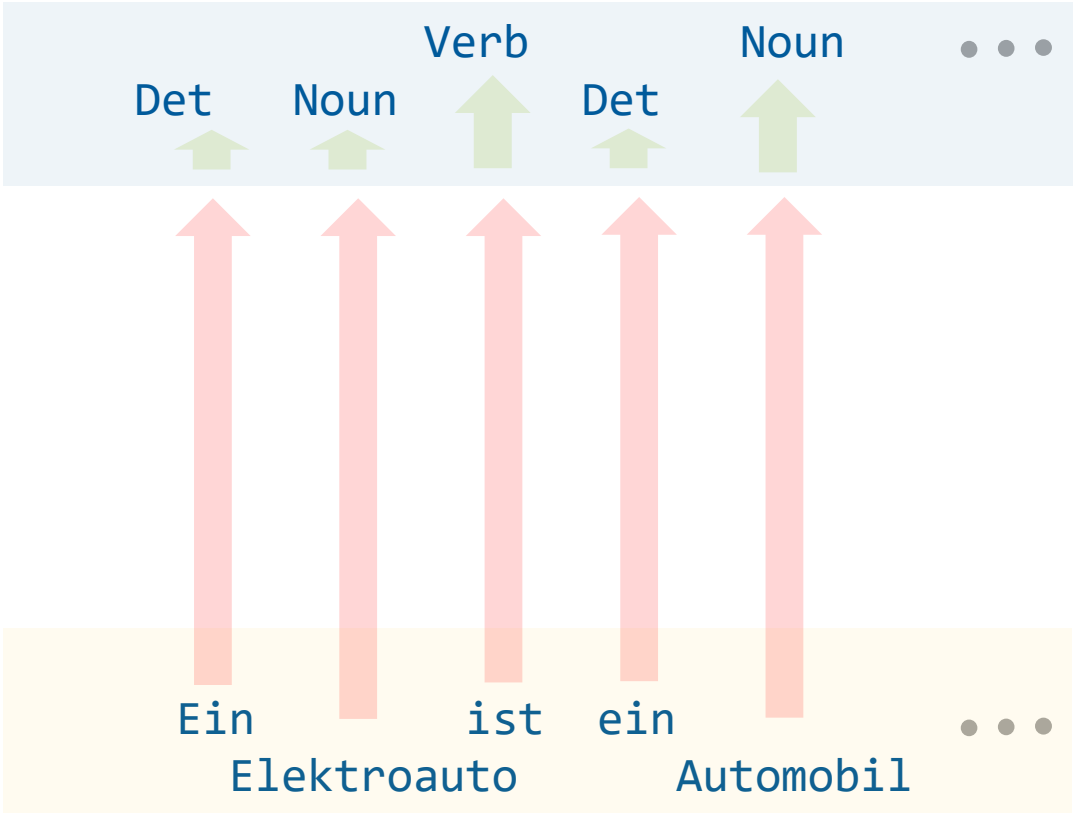


https://blog.einstein.ai/the-wikitext-long-term-dependency-language-modeling-dataset/

Constituent Analysis

Part-Of-Speech Tagging with Recurrent Neural Network

Assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.



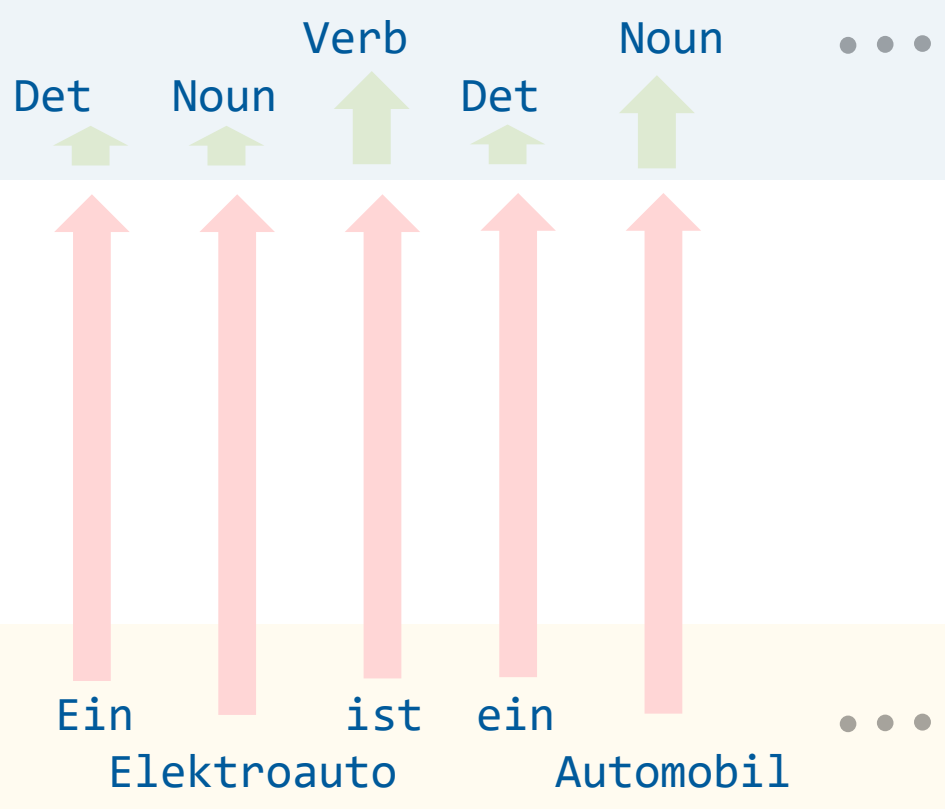
Constituent Analysis

Part-Of-Speech Tagging with Bi-Directional Recurrent Neural Network

Assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

Typically done with a Hidden Markov Model

***Det**erminer*	=>	*the|a|an|this|these|that|…*
Noun	=>	*flights|breeze|tri p|morning|…*
Verb	=>	*is|prefer|like|need|want|fly|…*
Proper Noun	=>	*Munich|Ingolstadt|…*
***Pro**noun*	=>	*me|I|you|it|…*



Constituent Analysis

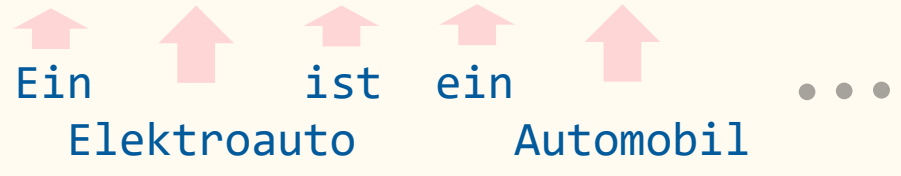
Part-Of-Speech Tagging with Bi-Directional Recurrent Neural Network

Assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

Typically done with a Hidden Markov Model

***Det**erminer*	=>	*the|a|an|this|these|that|…*
Noun	=>	*flights|breeze|tri p|morning|…*
Verb	=>	*is|prefer|like|need|want|fly|…*
Proper Noun	=>	*Munich|Ingolstadt|…*
***Pro**noun*	=>	*me|I|you|it|…*

The idea of basing a grammar on constituent structure date back to the psychologist Wilhelm Wundt (1900) but was not formalized until **Chomsky** (1956) and, independently, Backus (1959).



Hadumod Bußmann (Hrsg.): Lexikon der Sprachwissenschaft. 3., aktualisierte und erweiterte Auflage. Kröner, Stuttgart 2002, ISBN 3-520-45203-0 (Konstituentenanalyse).

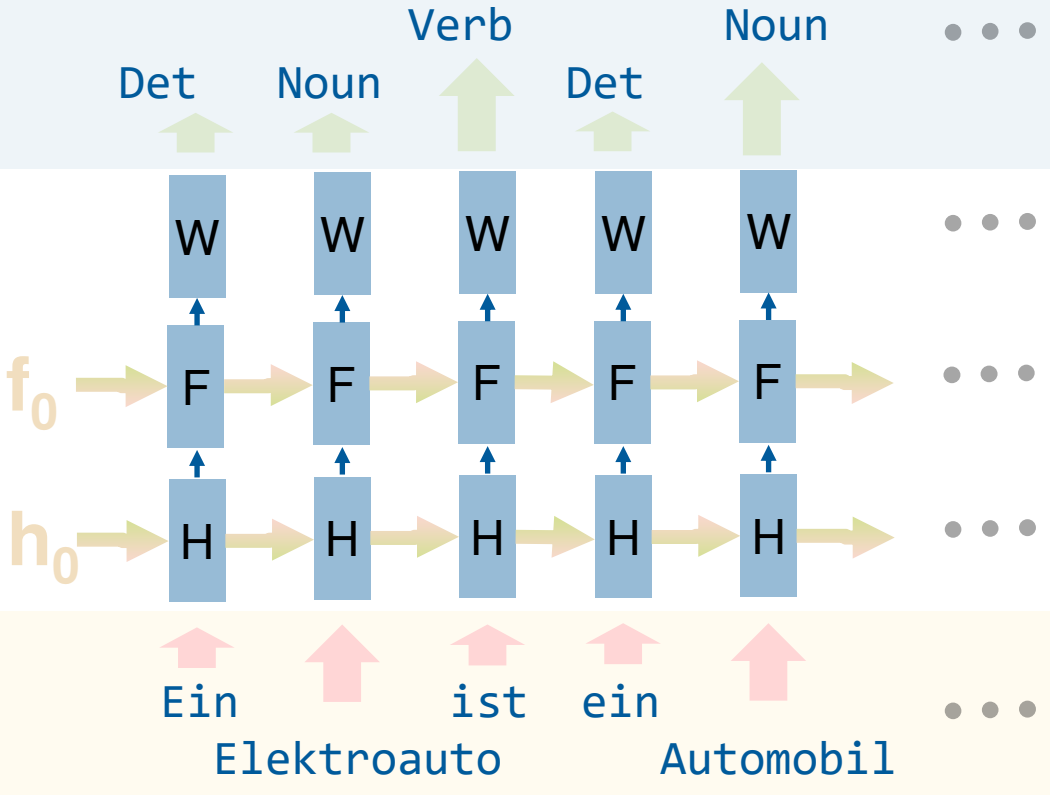
Constituent Analysis

Part-Of-Speech Tagging with Bi-Directional Recurrent Neural Network

Assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

Typically done with a Hidden Markov Model

or with Deep Neural Networks



***Det**erminer* => *the|a|an|this|these|that|…*

Noun => *flights|breeze|tri p|morning|…*

Verb => *is|prefer|like|need|want|fly|…*

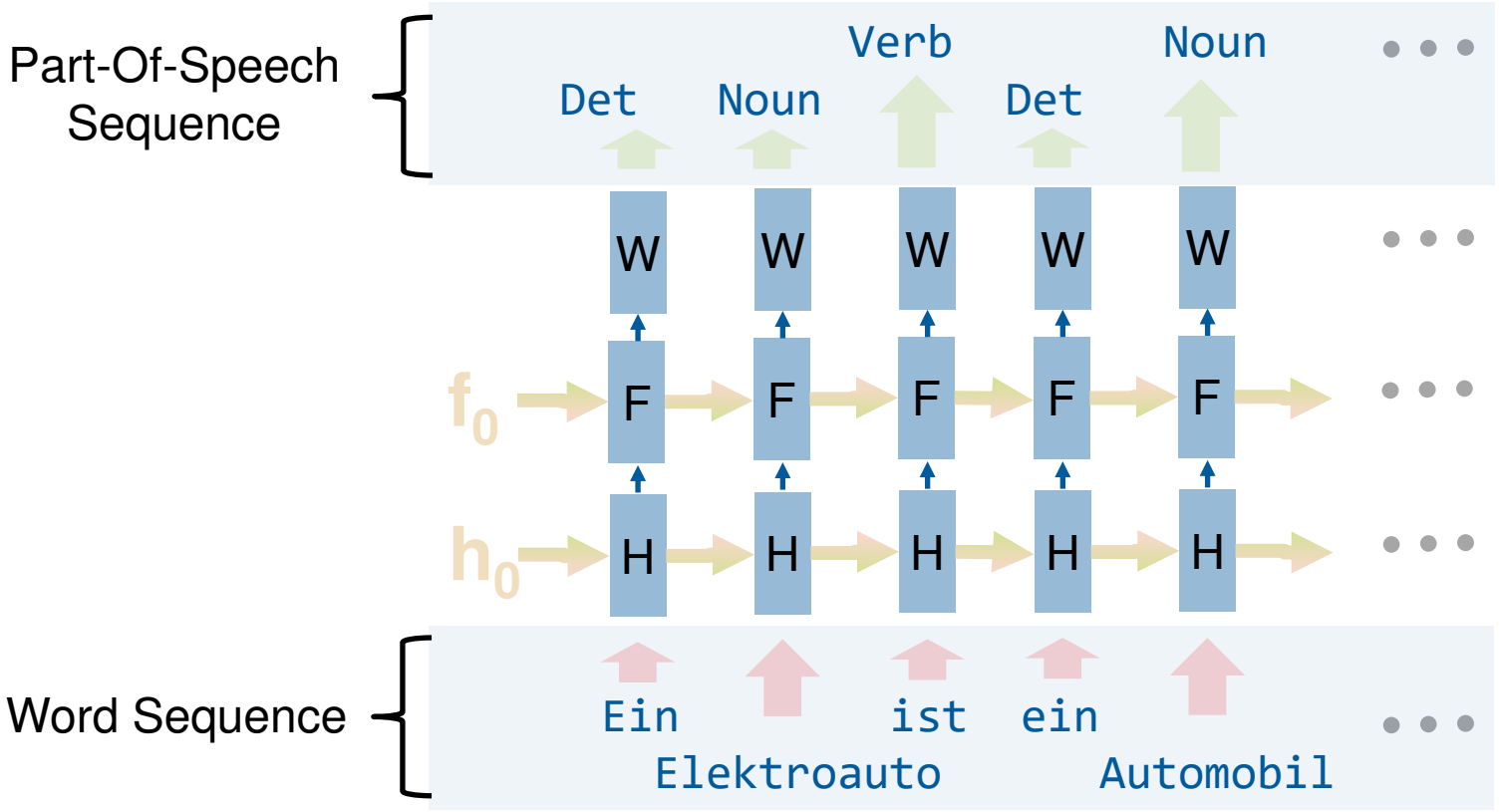
Proper Noun => *Munich|Ingolstadt|…*

***Pro**noun* => *me|I|you|it|…*

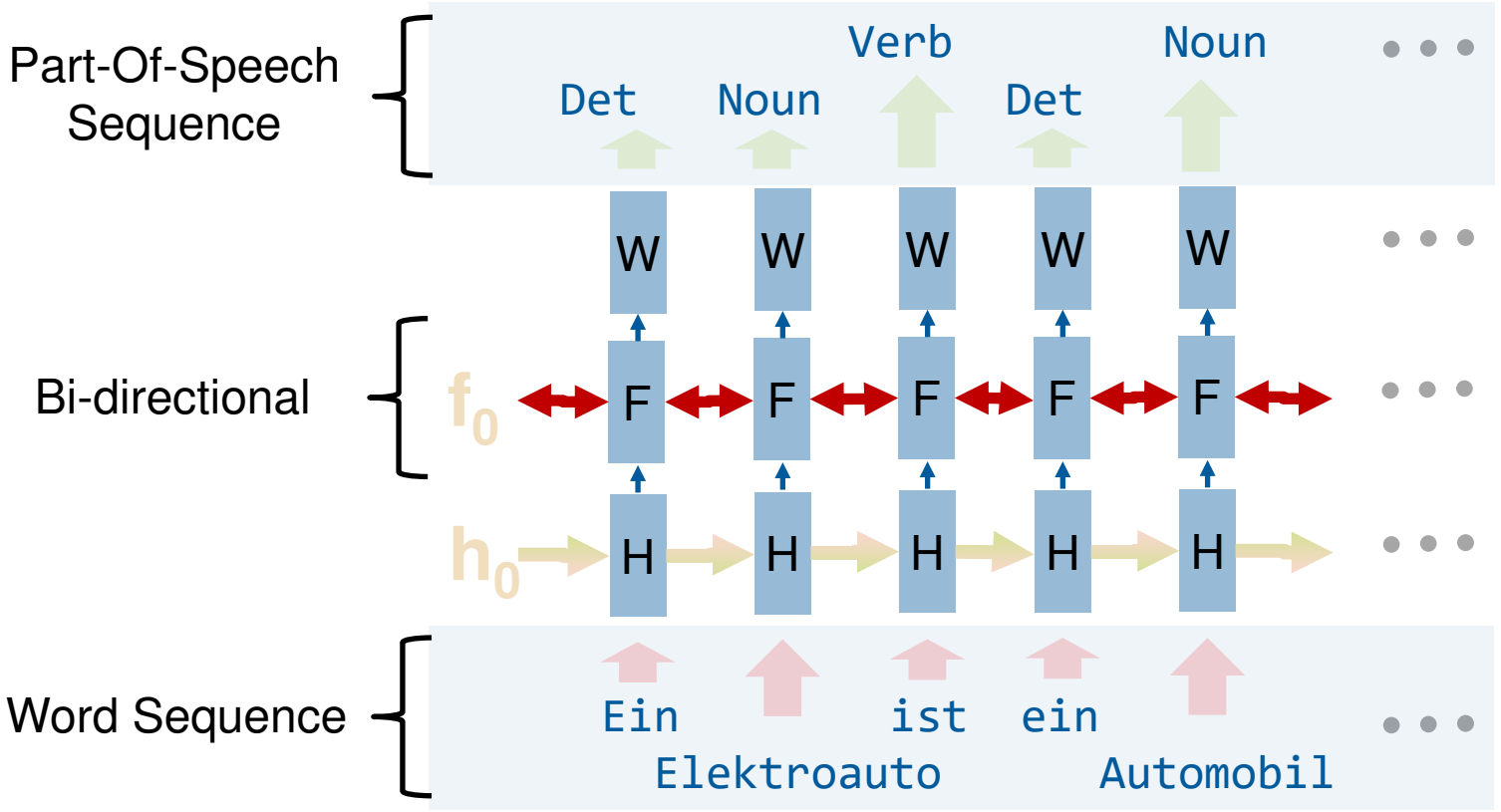
Constituent Analysis

Part-Of-Speech Tagging with Recurrent Neural Network

Assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.



Assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.



Sentence Structure

What is a correct sentence?

Syntax is the study of sentence structure and the rules of grammar. While people can do what they want with language (and many often do), syntax helps common users of a language understand how to organize words so that they make the most sense.

"Ein Kaninchen jagte den Hund über die Weide.“

Syntactically correct, but semantically not.

"Der Hund jagte ein Kaninchen durch die Weide"

Syntactically and semantically correct.

Syntax is the study of sentence structure and the rules of grammar. While people can do what they want with language (and many often do), syntax helps common users of a language understand how to organize words so that they make the most sense.

"The professor said on Monday he would give an exam."

This sentence means either that

"it was on Monday that the professor told the class about the exam"

or that

"the exam would be given on Monday".

Syntax is the study of sentence structure and the rules of grammar. While people can do what they want with language (and many often do), syntax helps common users of a language understand how to organize words so that they make the most sense.

"The chicken is ready to eat."

This sentence either means the chicken

"is cooked and can be eaten now"

or the chicken

"is ready to be fed".

Syntax is the study of sentence structure and the rules of grammar. While people can do what they want with language (and many often do), syntax helps common users of a language understand how to organize words so that they make the most sense.

"It's a small world, but I wouldn't want to paint it."

—American comedian Steven Wright

The ambiguity here is lies within the phrase "small world." While the adage, "It's a small world" is generally accepted to have one of several accepted figurative meanings (what a coincidence; we're not so different from one another, etc.), Wright has chosen to take the phrase literally. Comparatively speaking, the world—as in the Earth—may not be as large as other planets, but it would still be a Herculean chore to paint it.

Syntax is the study of sentence structure and the rules of grammar. While people can do what they want with language (and many often do), syntax helps common users of a language understand how to organize words so that they make the most sense.

"I shot an elephant in my pajamas".

It has two possible interpretations:

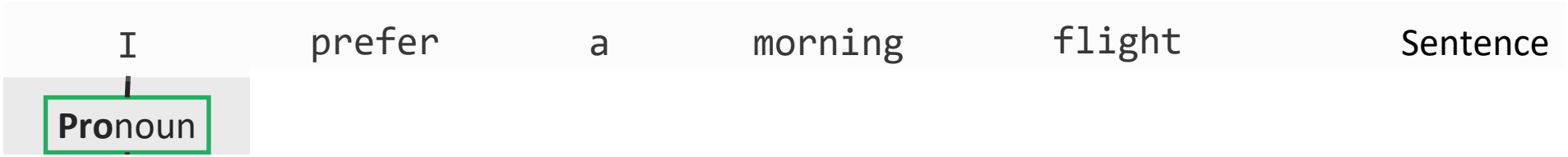
one where the man is wearing his pajamas while shooting the elephant, and

the other where the elephant is inside the man's pajamas.

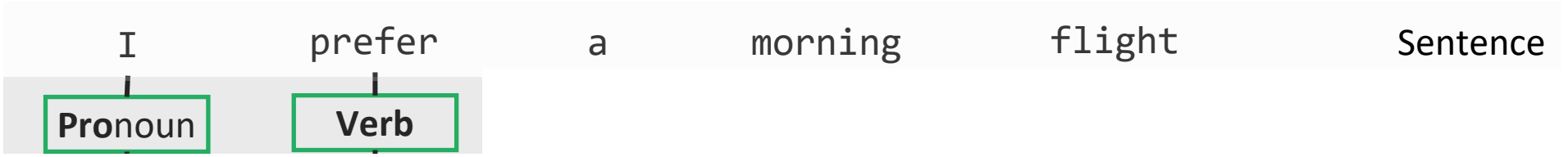
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.

```
I        prefer       a       morning       flight          Sentence
```

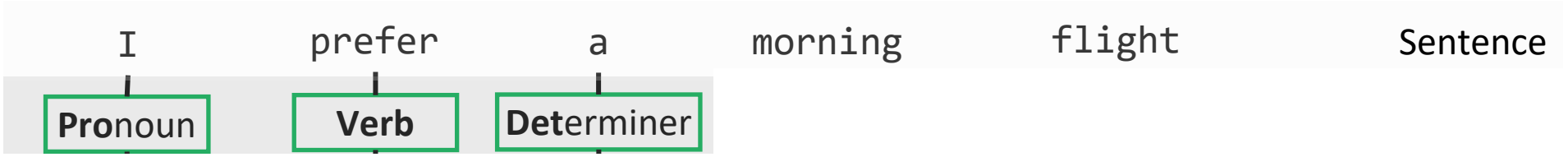
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



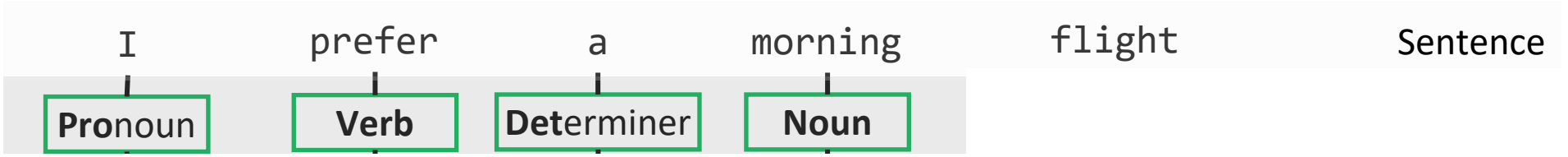
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



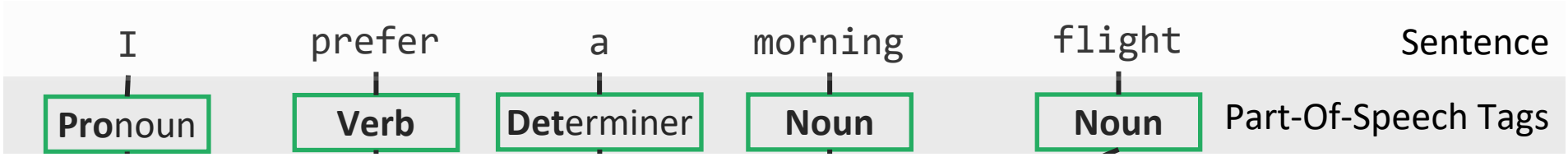
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



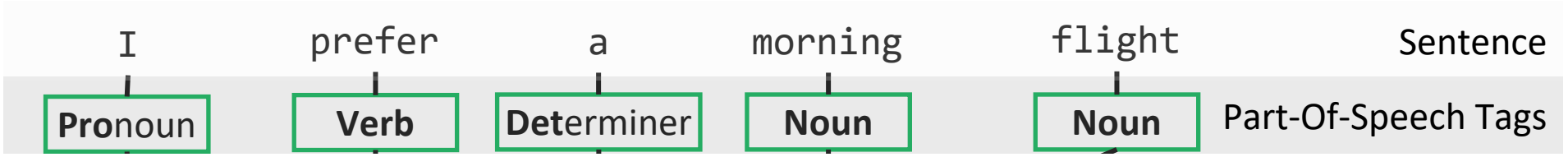
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



Constituent Analysis

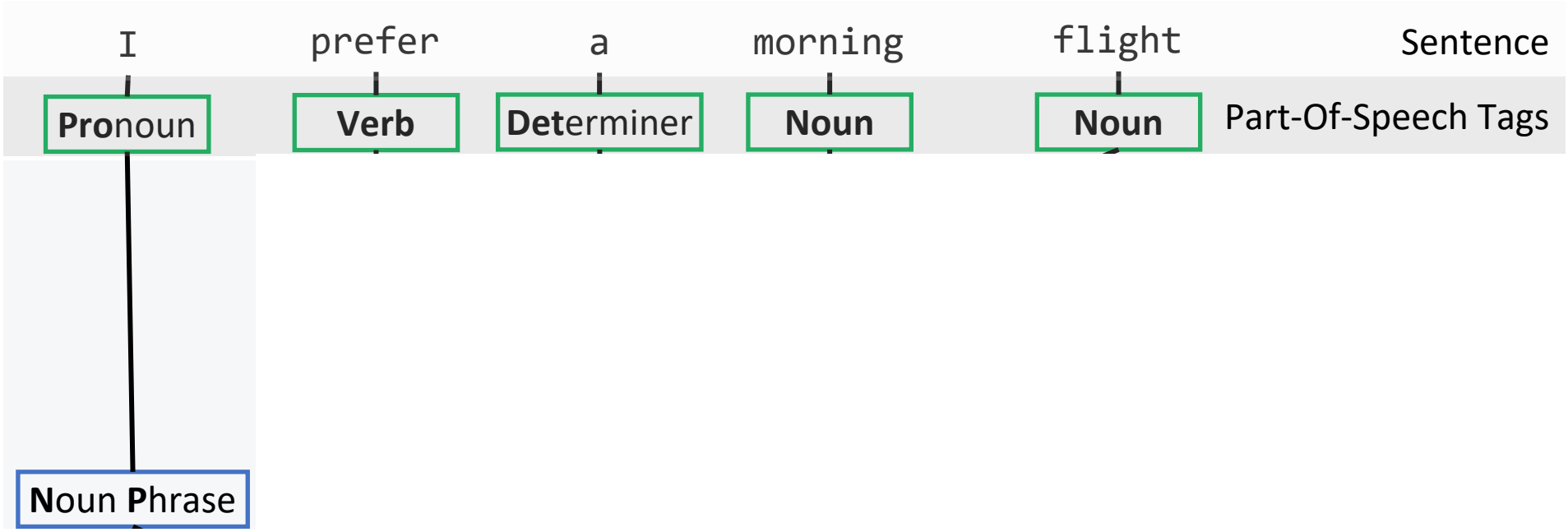
Determining Sentence Constituents

The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.

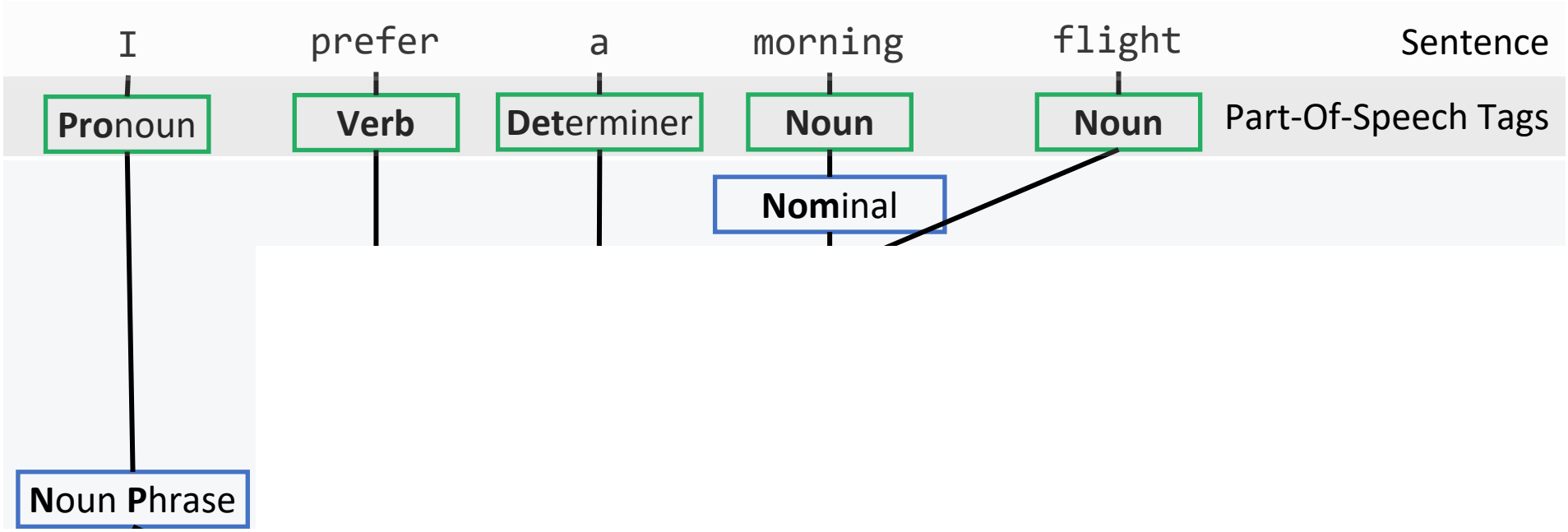


Remember: This can be done with a DNN

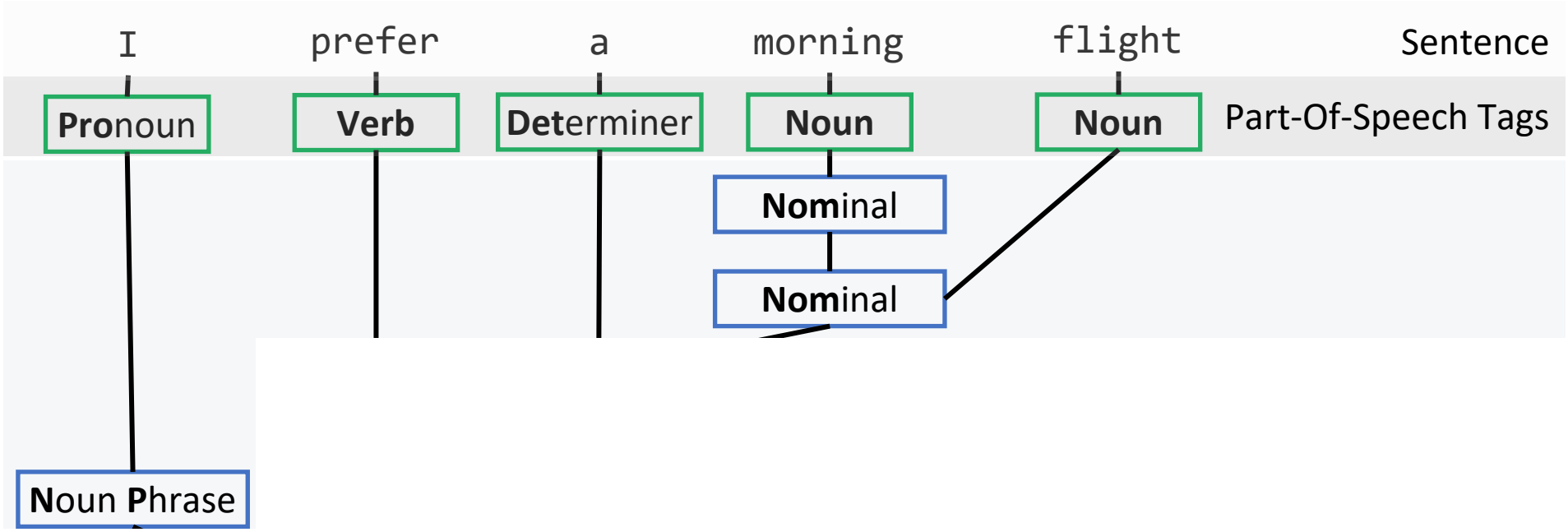
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



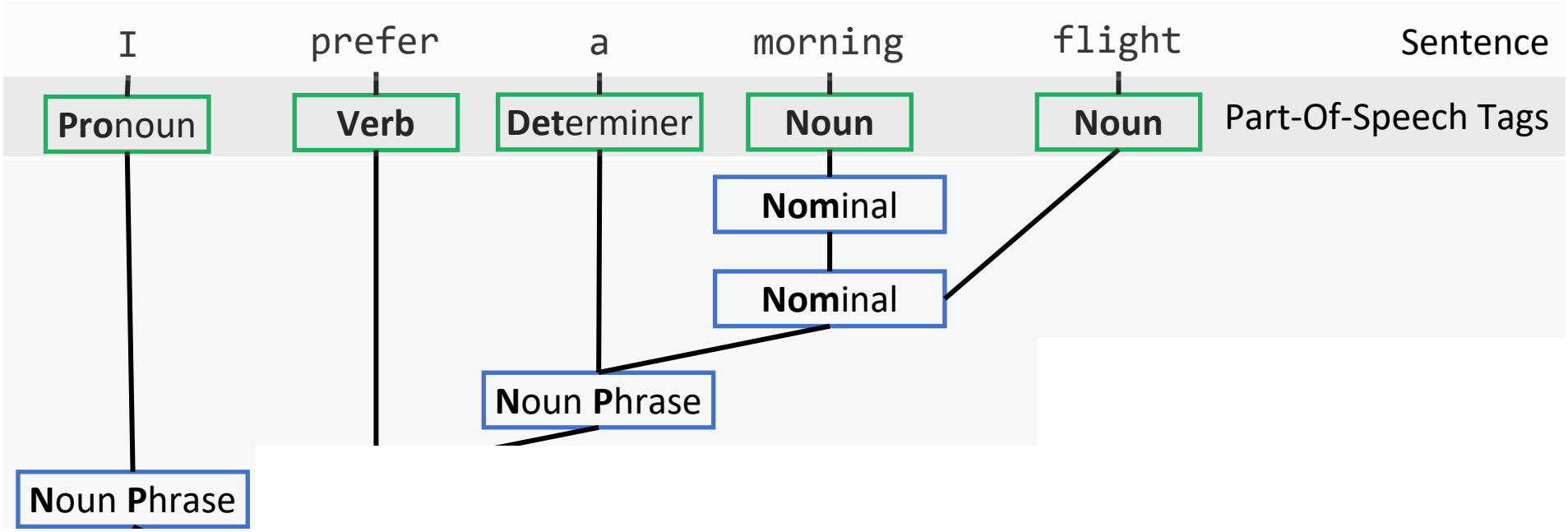
The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



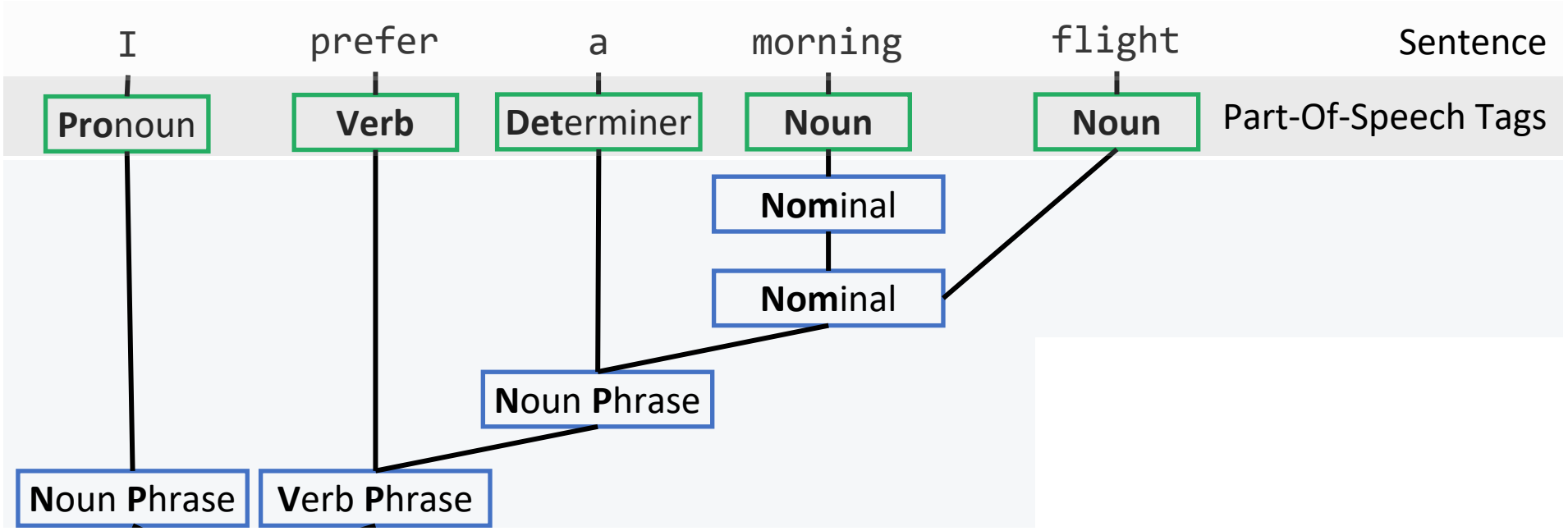
Ziel und Ergebnis der Konstituentenanalyse ist die Zerlegung eines sprachlichen Ausdrucks in eine hierarchisch definierte Abfolge von Konstituenten.



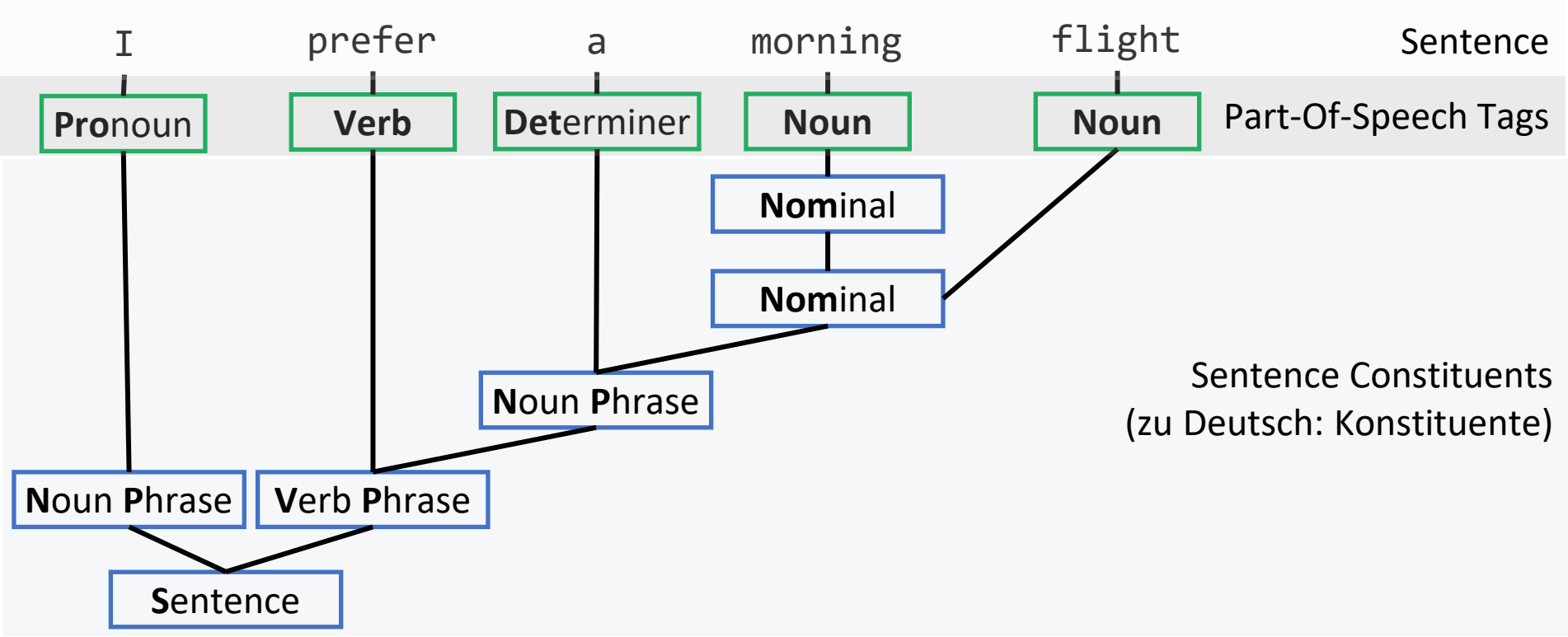
Ziel und Ergebnis der Konstituentenanalyse ist die Zerlegung eines sprachlichen Ausdrucks in eine hierarchisch definierte Abfolge von Konstituenten.



The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.



This is done with a grammar; but DNNs achieve promising results, too.

Divides the sentences into single phrasal constituents such as phrases, words, or morphemes. It's a usually a context-free grammar and highly hierarchical. 4.5k different rules for expanding Verb Phrase in English. It discover its deep meaning and explore alternative ways of expressing the meaning

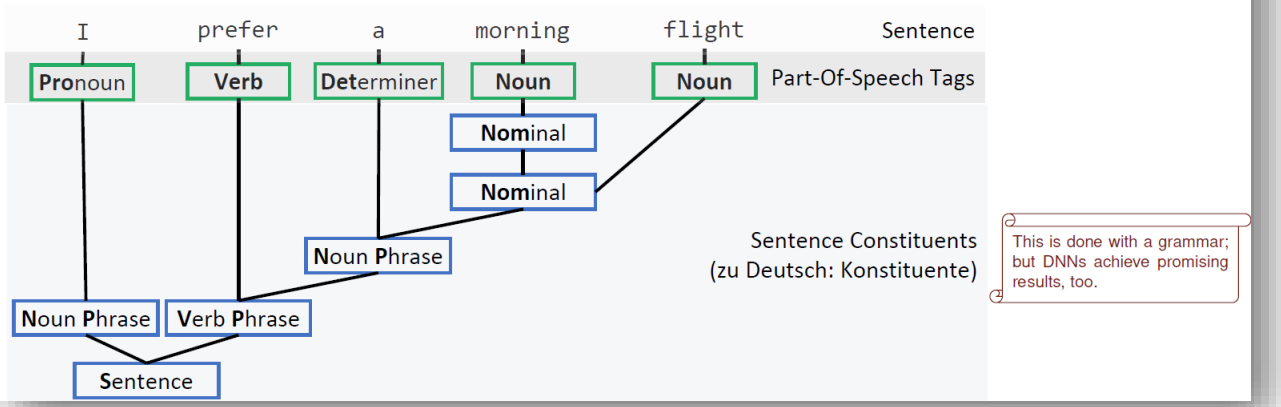
```
Sentence      => Noun Phrase + Verb Phrase
Verb Phrase   => Verb + Noun Phrase
Noun Phrase   => Pronoun | Determiner + Nominal
Nominal       => Noun | Nominal + Noun
```

Sentence Constituents
(zu Deutsch: Konstituente)

Recap

Linguistic Structure

The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.

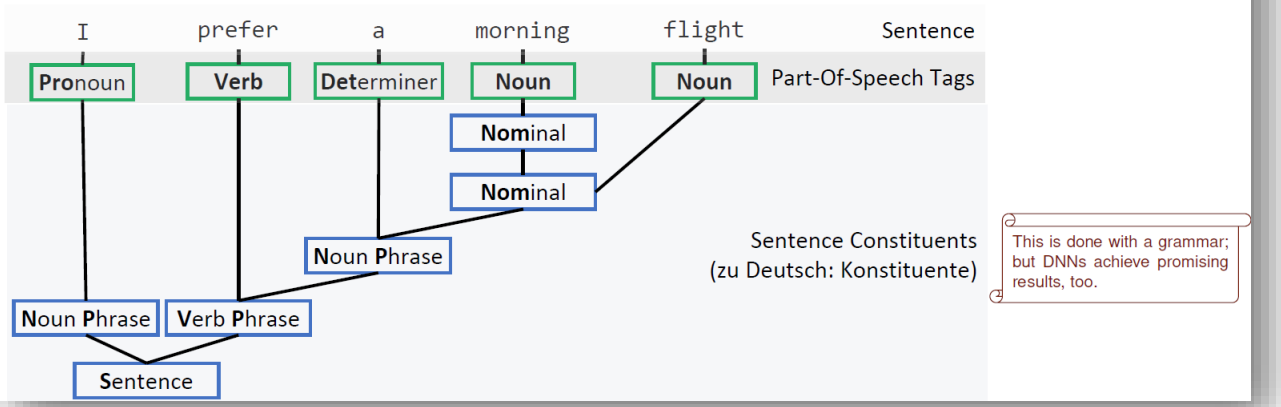


https://en.wikipedia.org/wiki/Syntactic_parsing_(computational_linguistics)#Constituency_parsing

Recap

Linguistic Structure

The aim and result of the constituent analysis is the decomposition of a linguistic expression into a hierarchically defined sequence of constituents.




```
Sentence      => Noun Phrase + Verb Phrase
Verb Phrase   => Verb + Noun Phrase
Noun Phrase   => Pronoun | Determiner + Nominal
Nominal       => Noun | Nominal + Noun
```

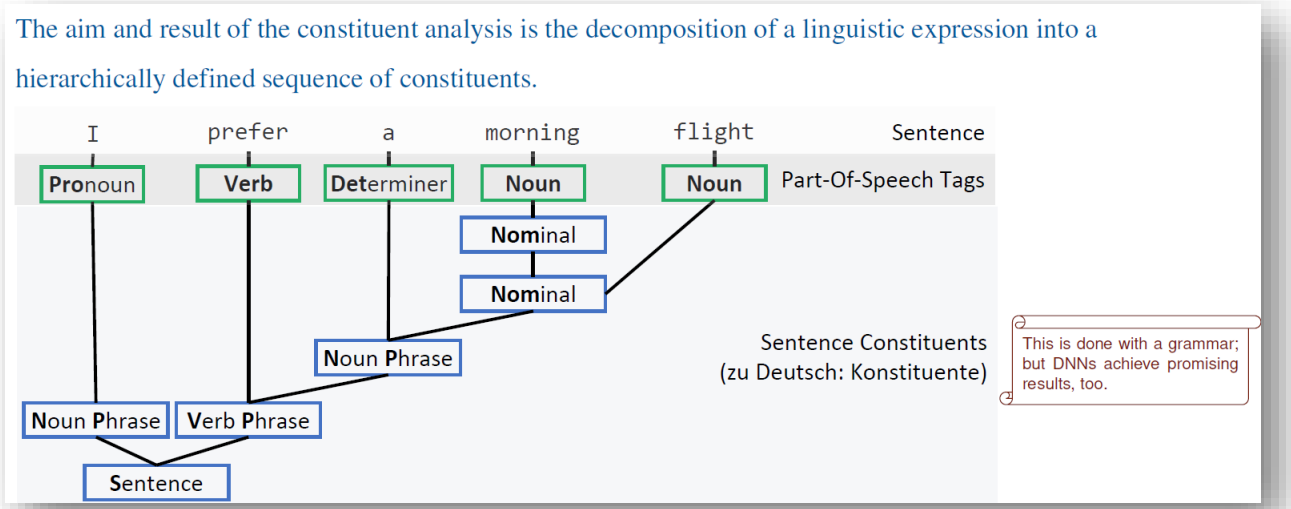
https://en.wikipedia.org/wiki/Syntactic_parsing_(computational_linguistics)#Constituency_parsing

Recap

Linguistic Structure

Last week

- constituency analysis,
- POS

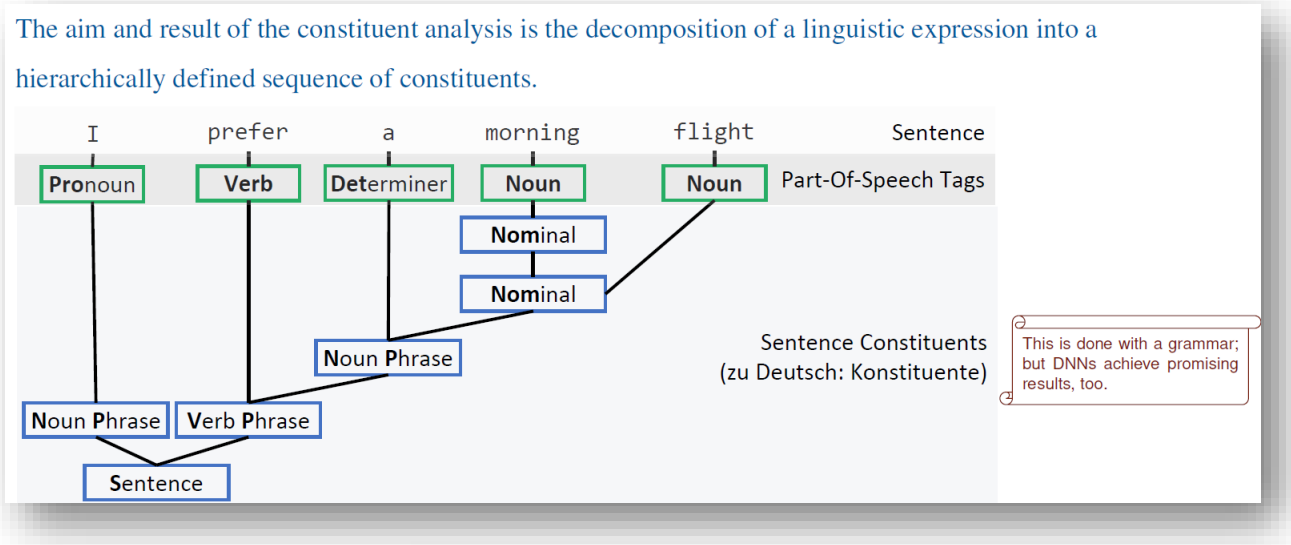



```
Sentence      => Noun Phrase + Verb Phrase
Verb Phrase   => Verb + Noun Phrase
Noun Phrase   => Pronoun | Determiner + Nominal
Nominal       => Noun | Nominal + Noun
```

https://en.wikipedia.org/wiki/Syntactic_parsing_(computational_linguistics)#Constituency_parsing

Recap

Linguistic Structure




```
Sentence     => Noun Phrase + Verb Phrase
Verb Phrase  => Verb + Noun Phrase
Noun Phrase  => Pronoun | Determiner + Nominal
Nominal      => Noun | Nominal + Noun
```

Last week

- constituency analysis,
- POS

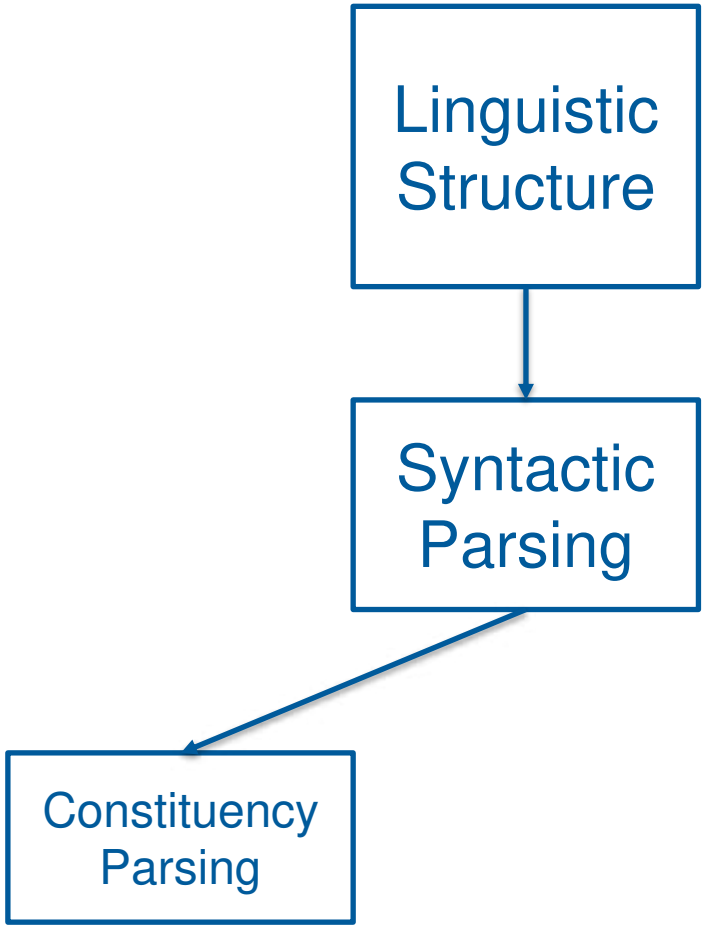
Remember

"Telling which spans are constituents (e.g. [The man] is here.) and what kind of constituent it is (e.g. [The man] is a noun phrase) on the basis of a context-free grammar (CFG) which encodes rules for constituent formation and merging"

- https://en.wikipedia.org/wiki/Syntactic_parsing_(computational_linguistics)#Constituency_parsing
- Chapter 17, https://web.stanford.edu/~jurafsky/slp3/

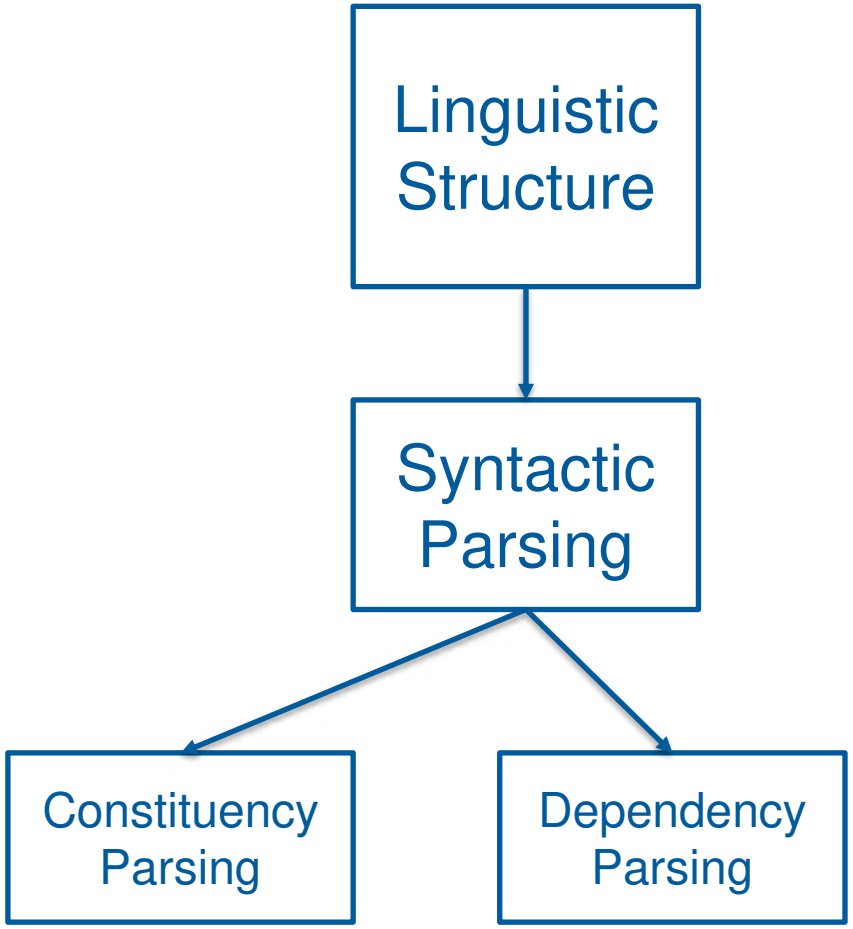
Detecting Linguistic Structure

Overview



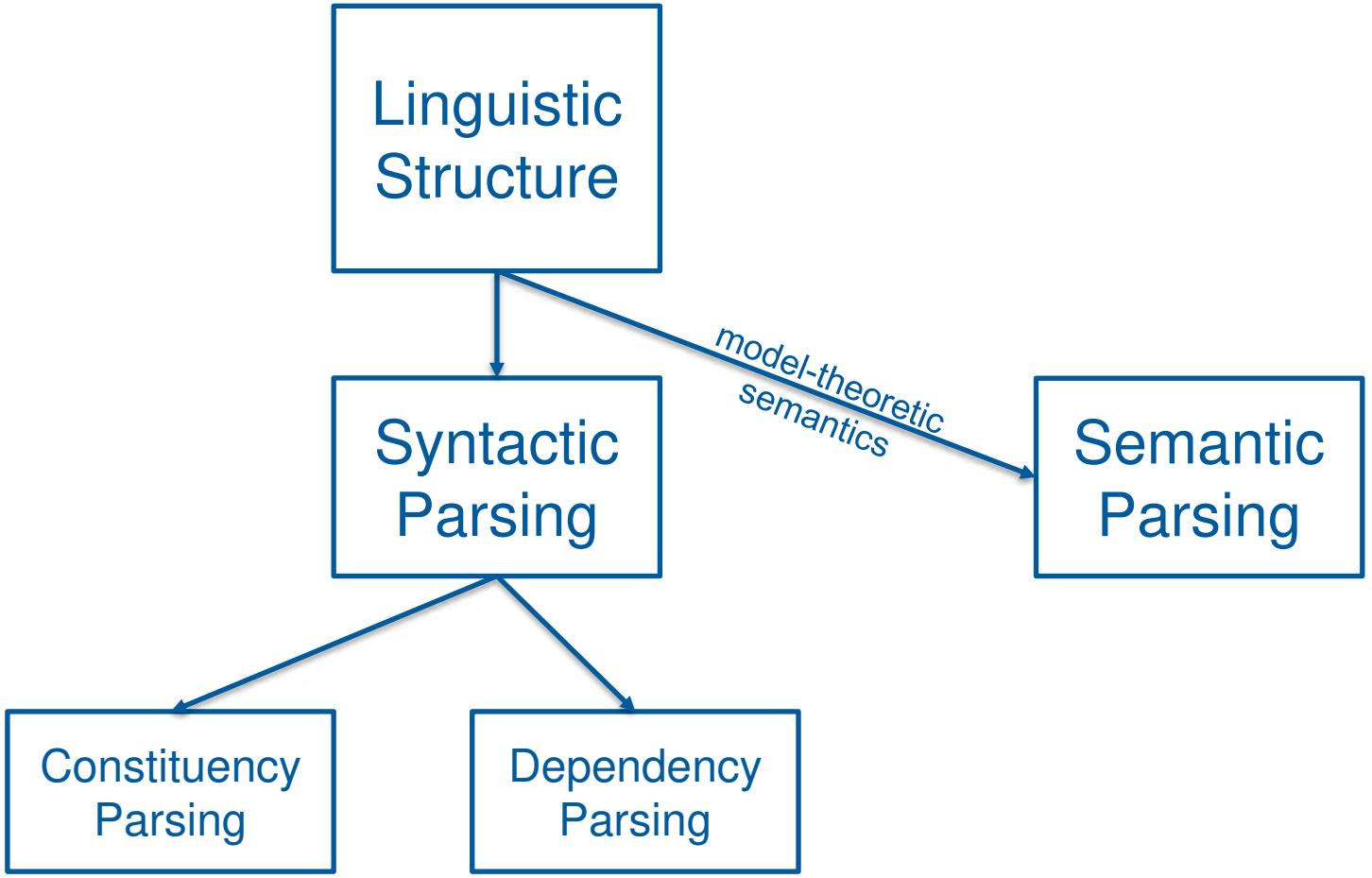
Detecting Linguistic Structure

Overview



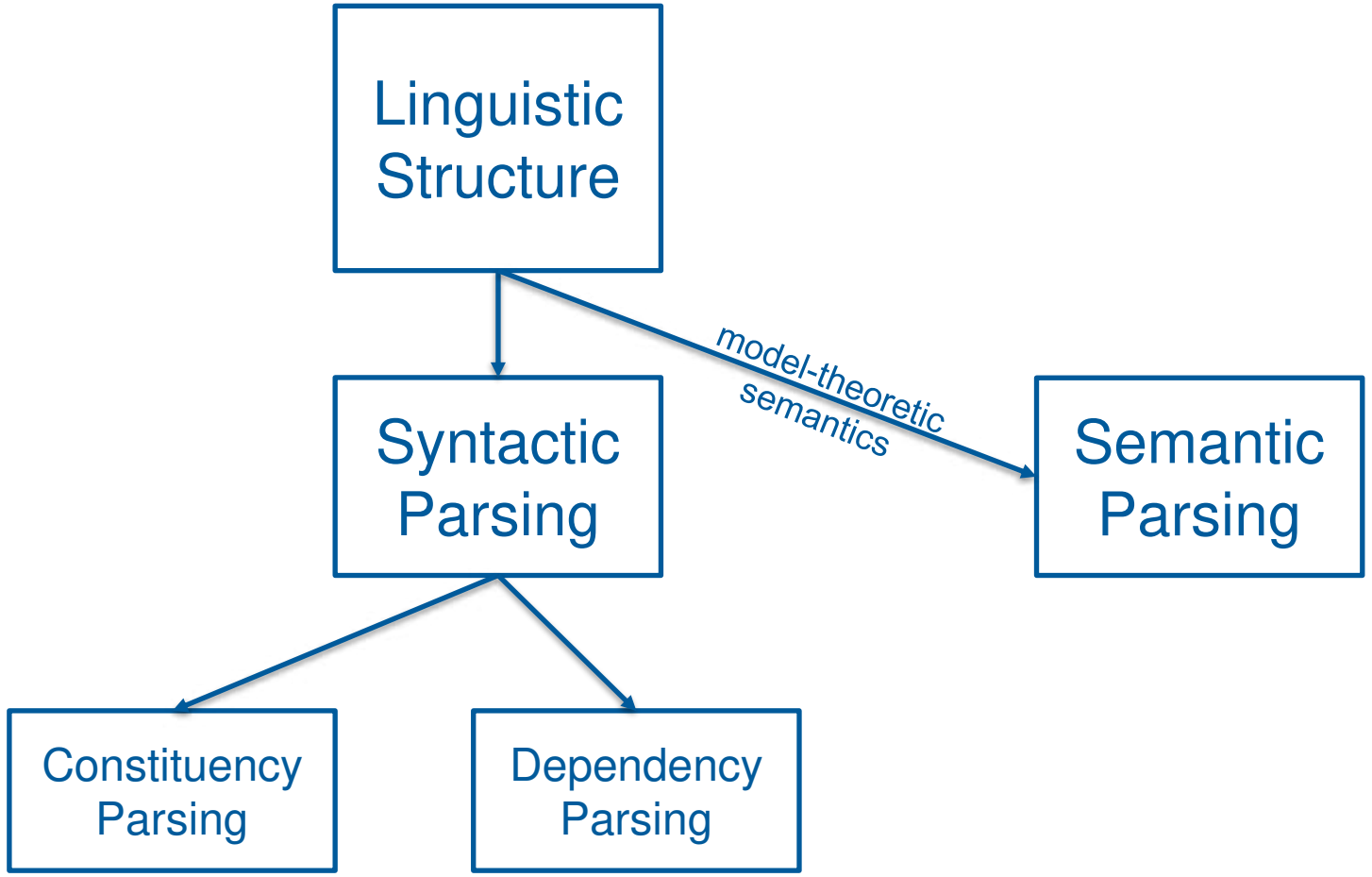
Detecting Linguistic Structure

Overview



Detecting Linguistic Structure

Overview

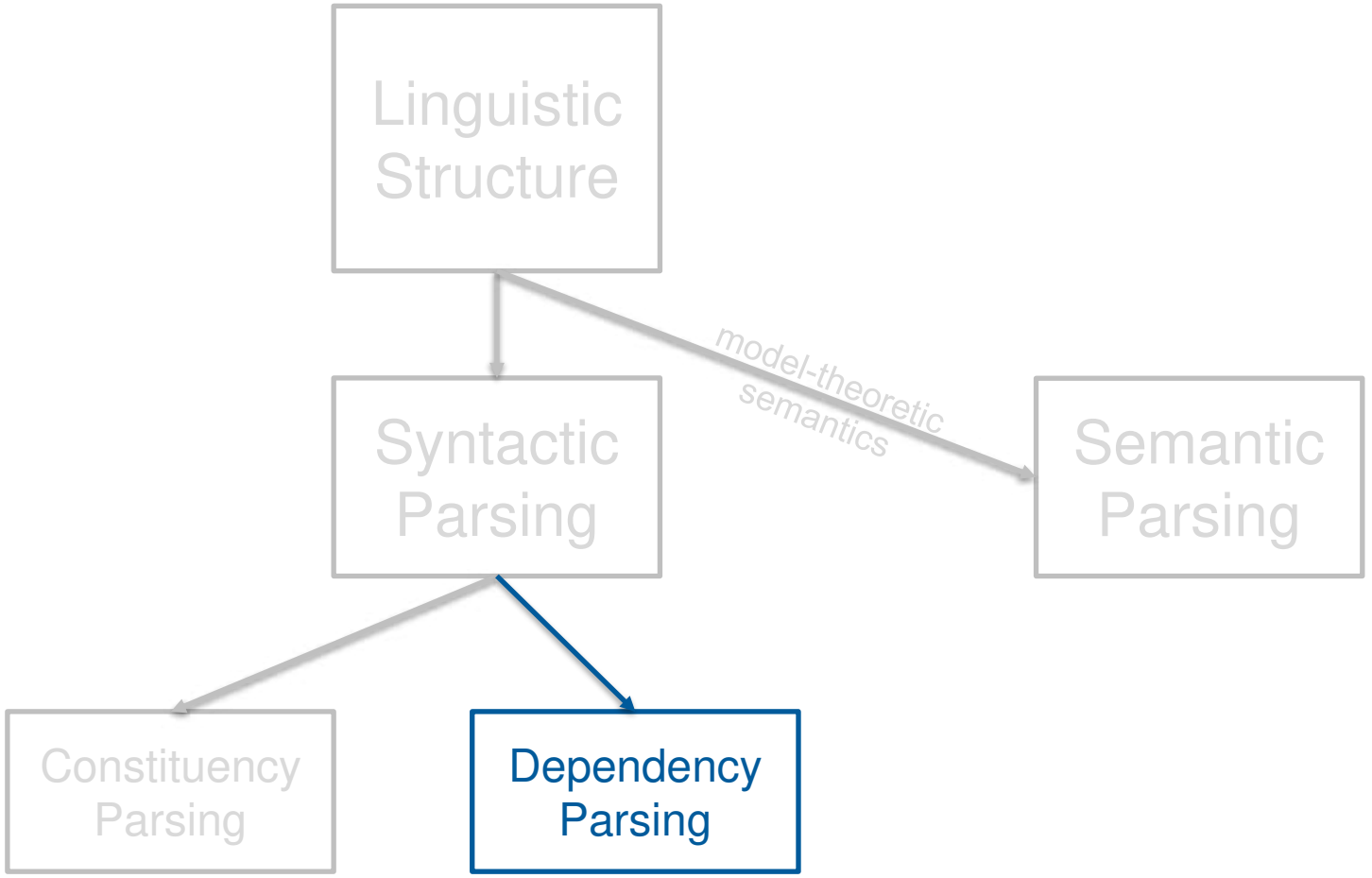


There are further structures related to meaning:

- semantic roles,
- word senses,
- entity relations,
- events…

Detecting Linguistic Structure

Overview



Scientists count whales from space

Scientists count whales from space

Image from: https://youtu.be/PSGIodTN3KE?t=1685

Scientists count whales from space

Image from: https://youtu.be/PSGIodTN3KE?t=1685

Dependency Parsing

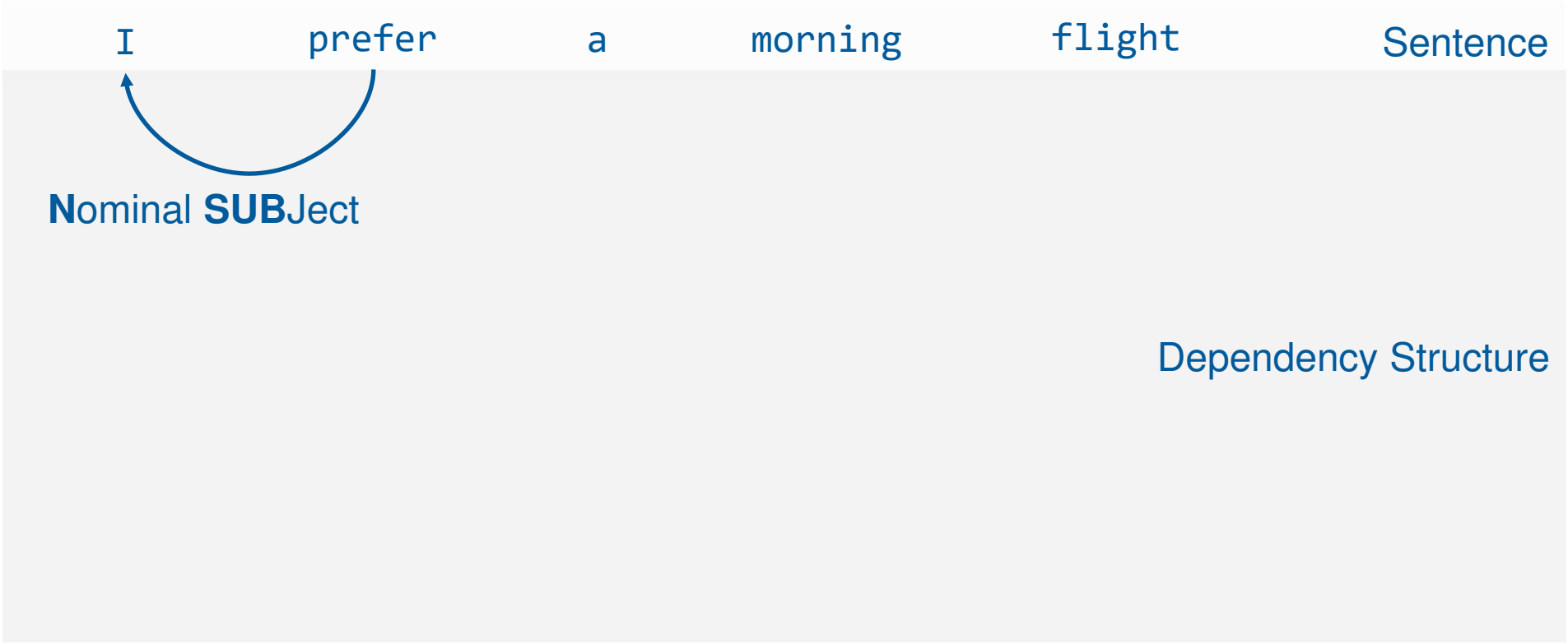
What is it?

Dependency Parsing

What is it?

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.

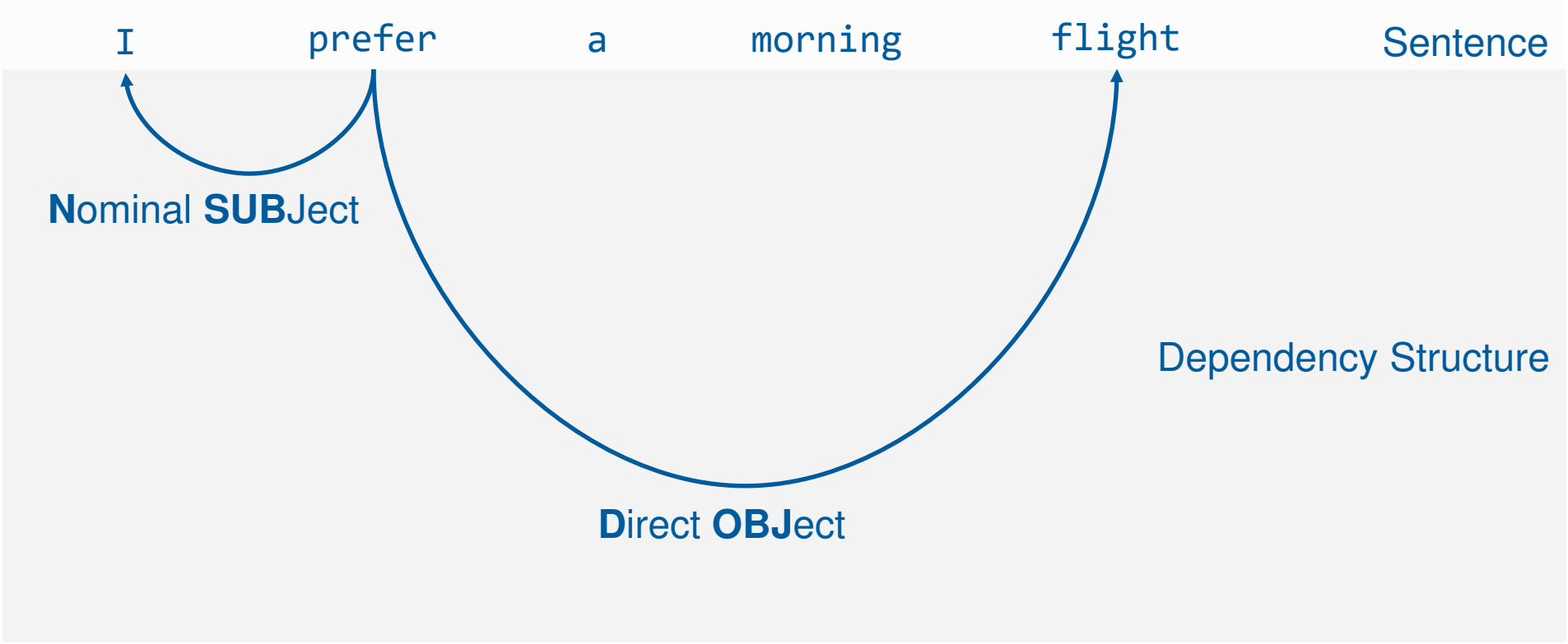
The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



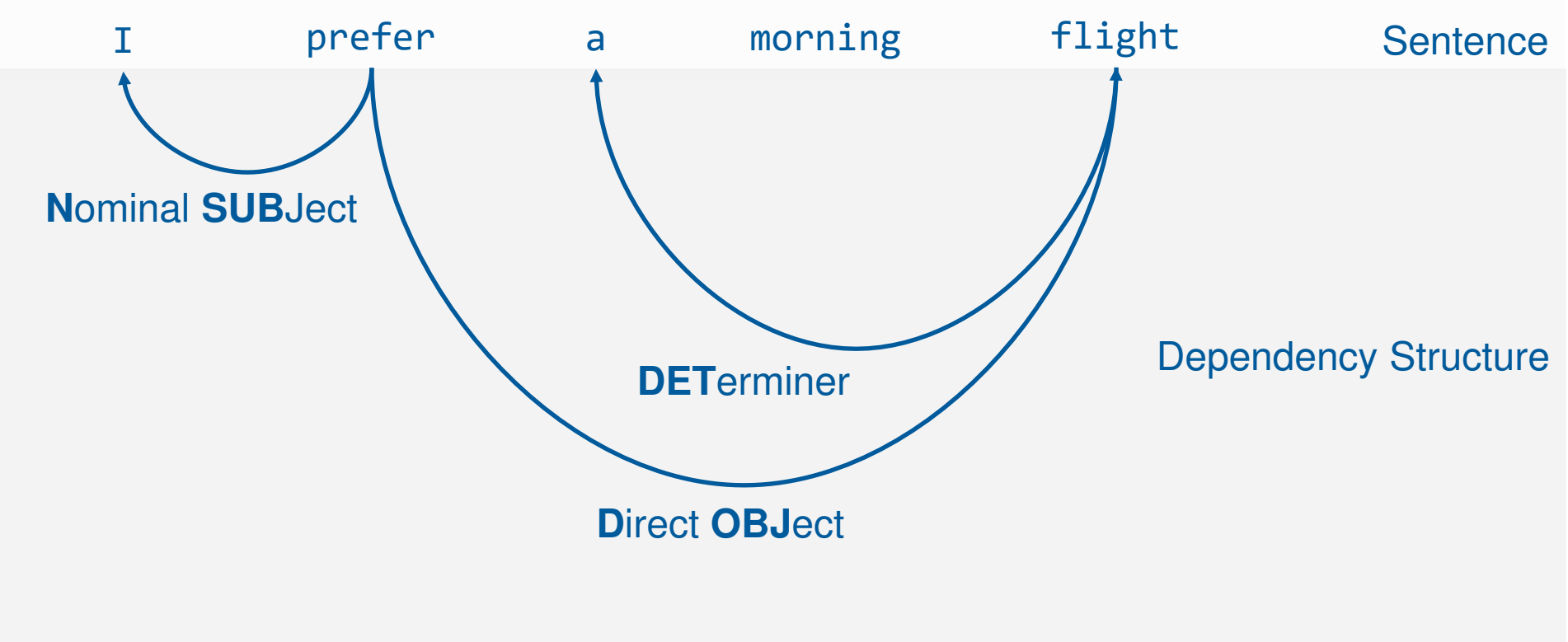
Dependency Parsing

Example

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



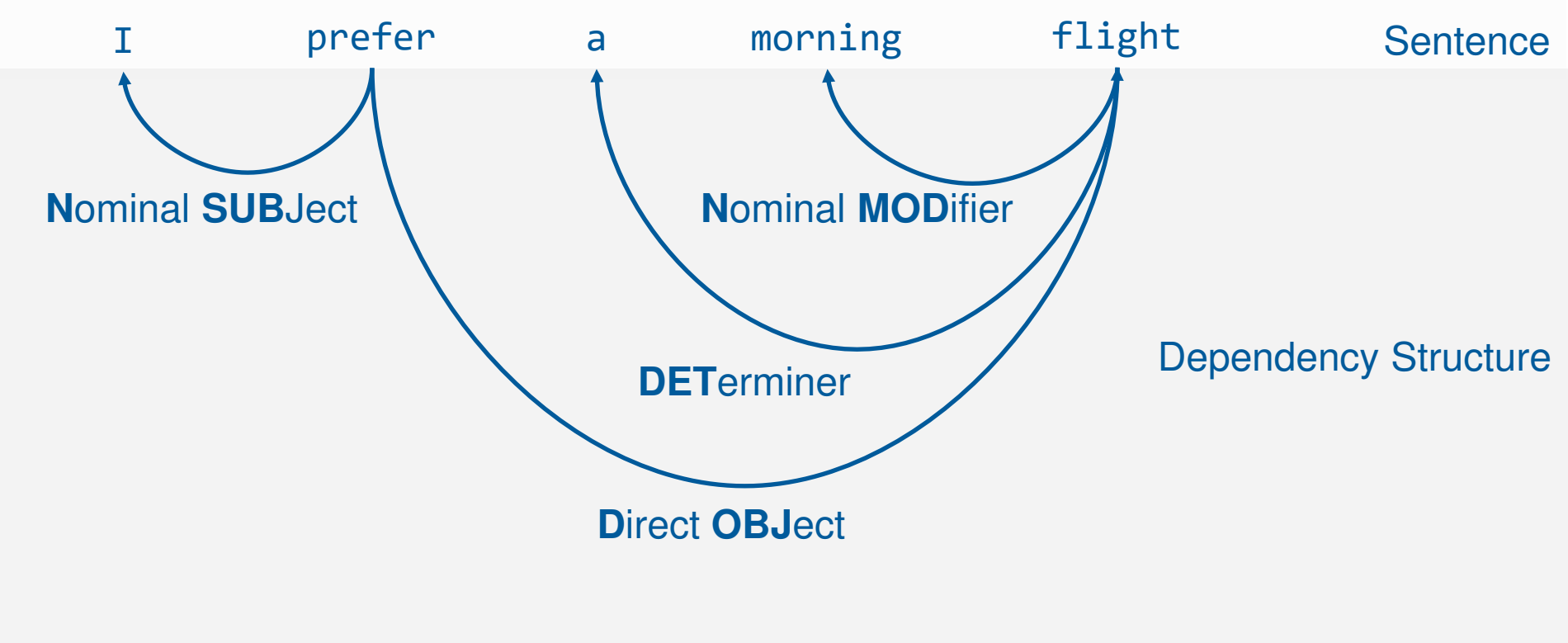
The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



Dependency Parsing

Example

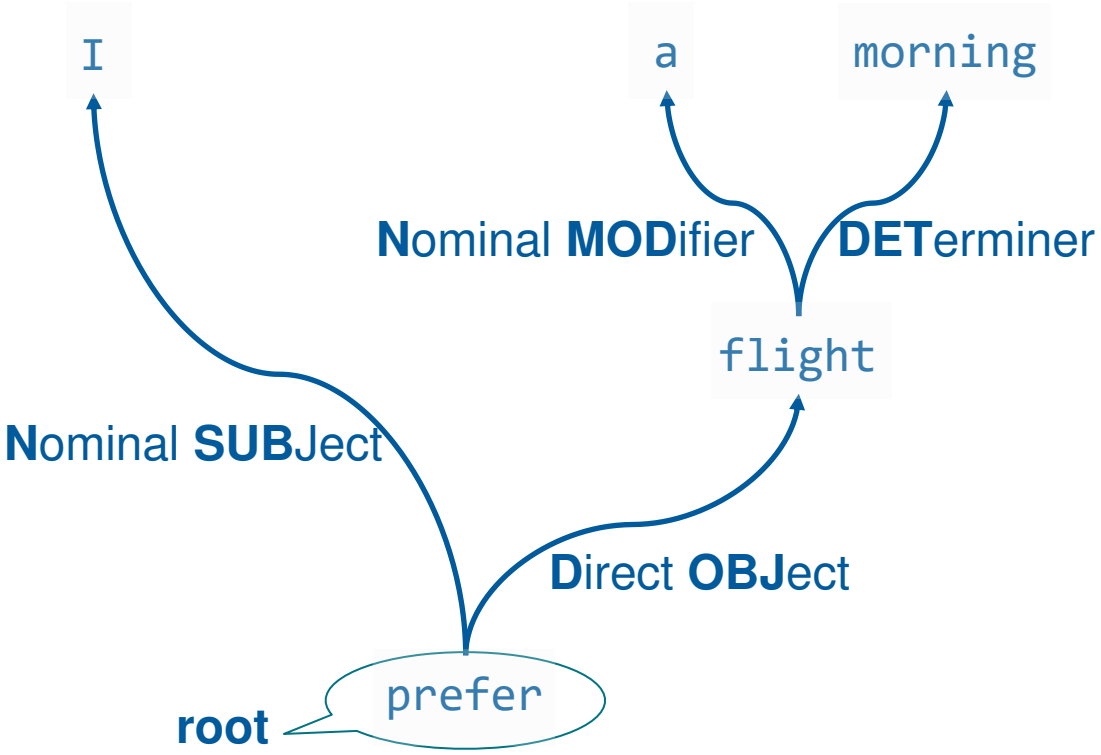
The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



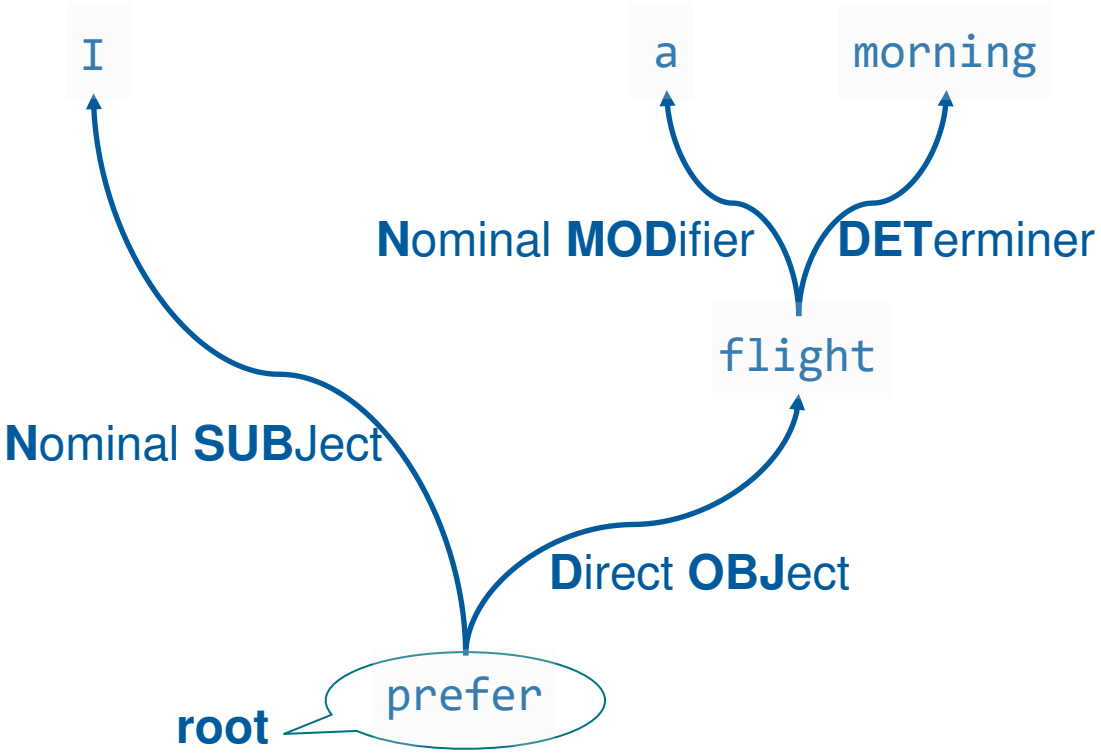
Dependency Parsing

Example, restructured

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.

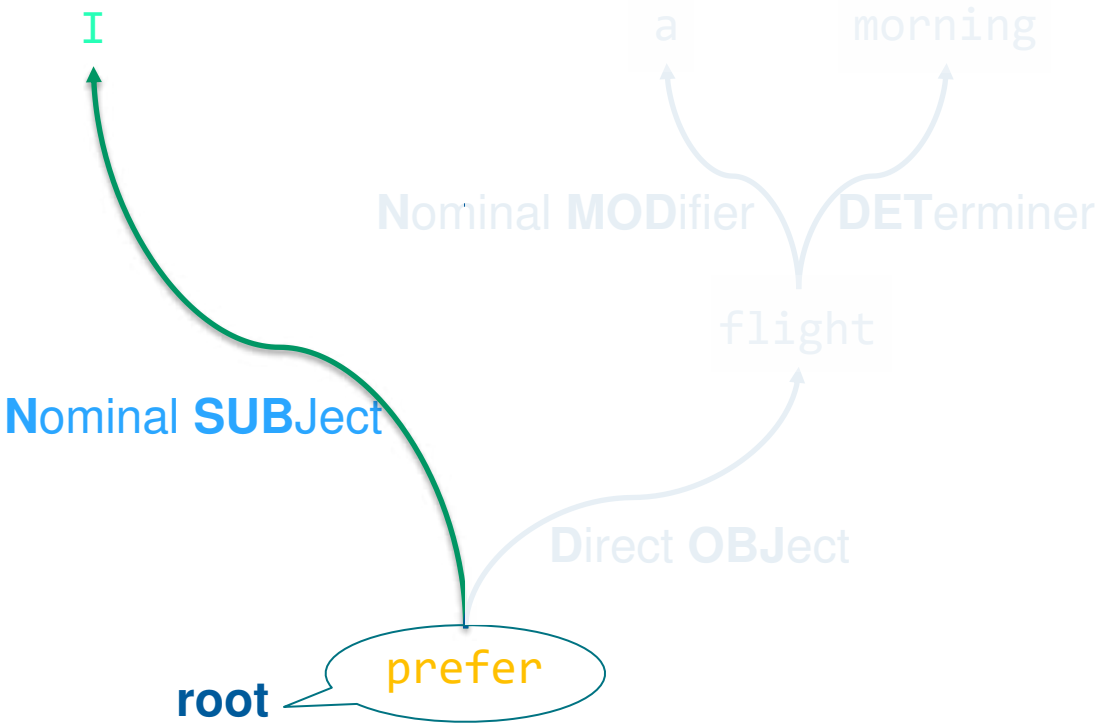


Syntax of the sentence is expressed in terms of *dependencies* between words

Dependency Parsing

And directed graphs

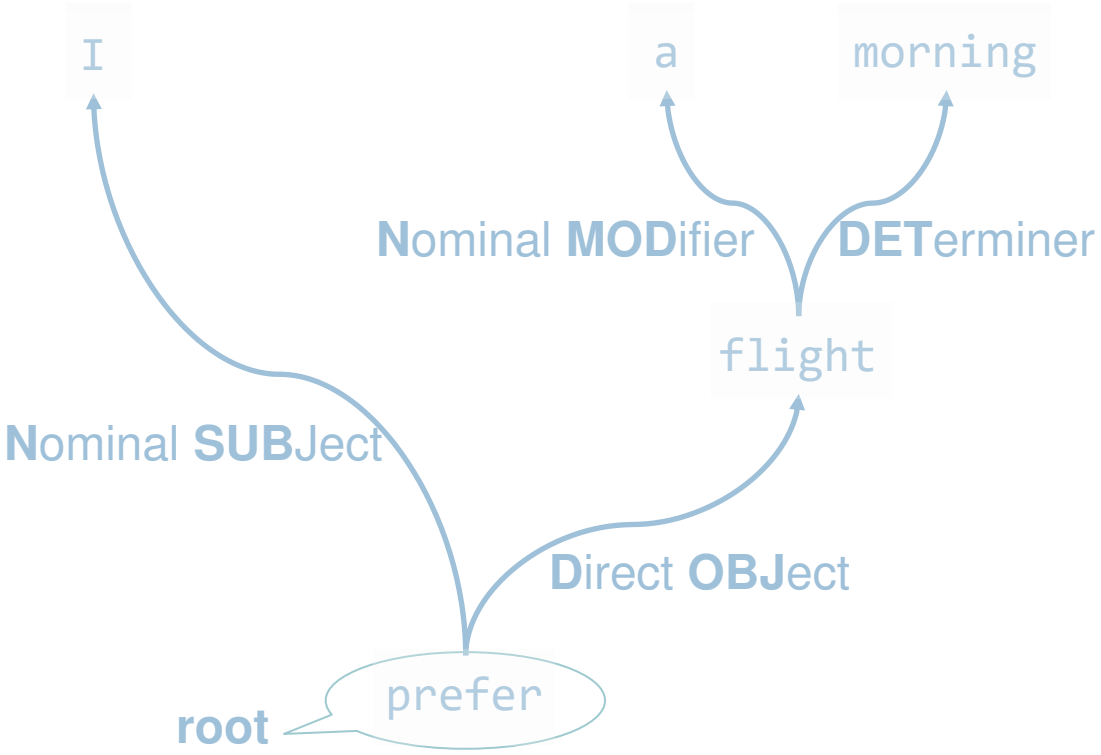
The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



Syntax of the sentence is expressed in terms of *dependencies* between words

directed, labeled
typed arcs from
heads to dependents

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



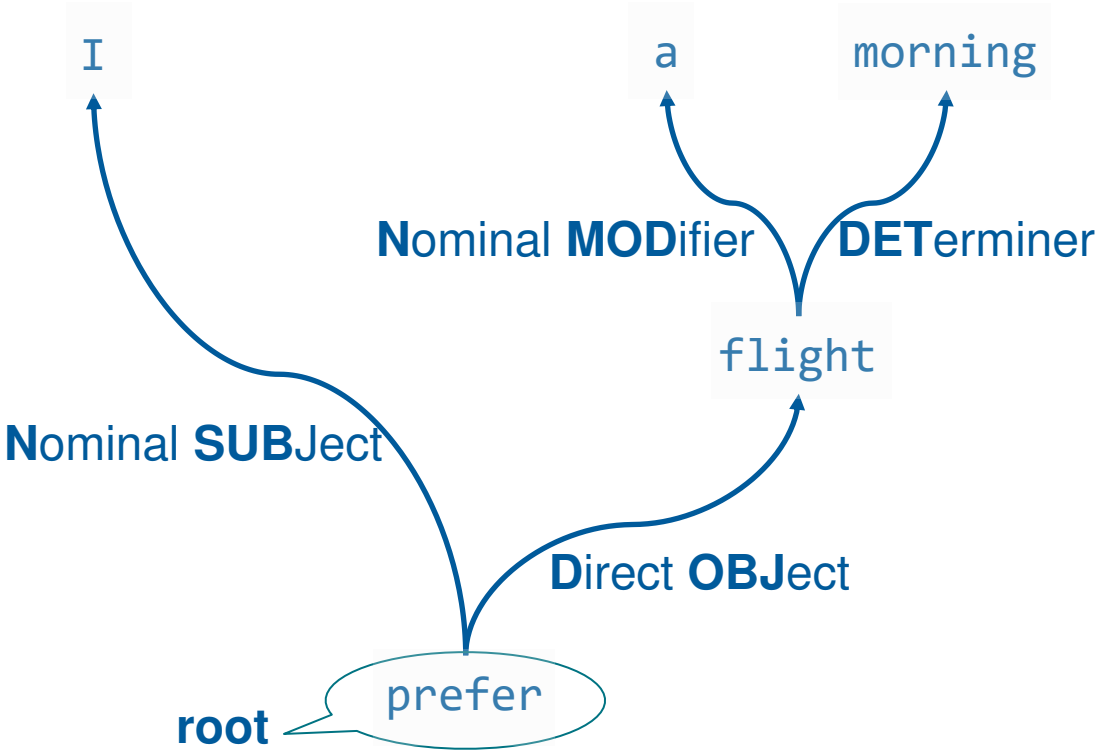
Dependence tree
= directed graph that satisfies:

1. There is a single designated root node that has no incoming arcs.
2. With exception of the root node, each vertex has exactly one incoming arc.
3. There is a unique path from the root node to each vertex in V.

Dependency Parsing

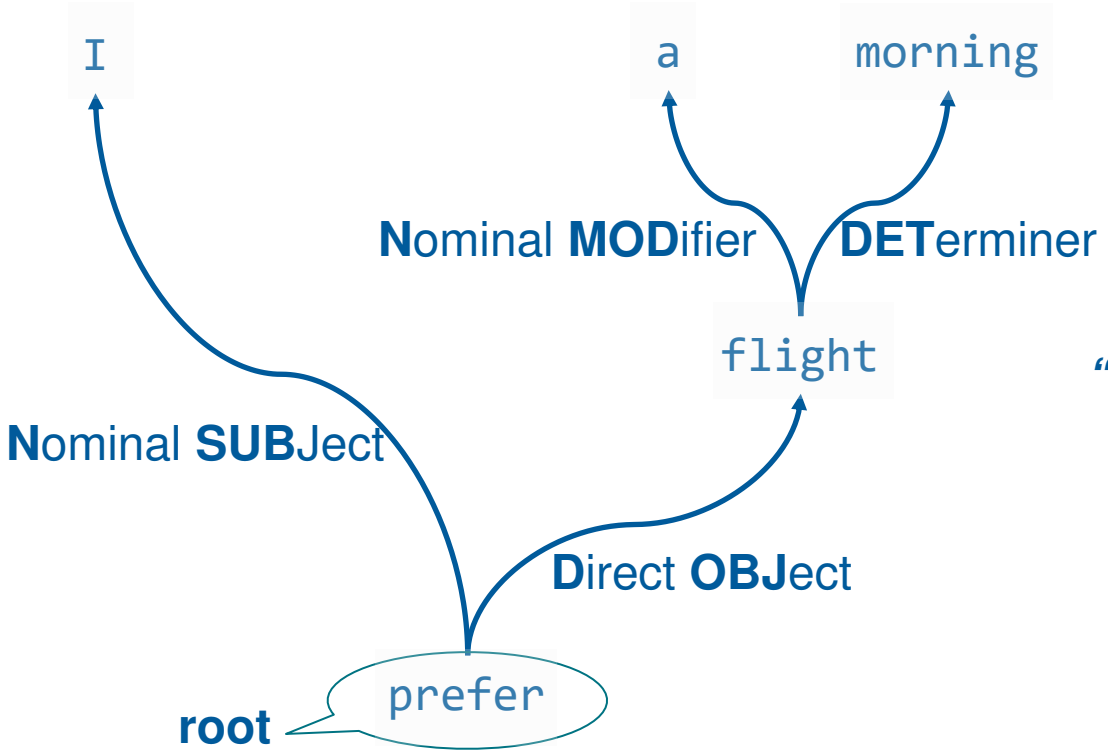
Example

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



This is done with a grammar; but DNNs achieve promising results, too.

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.

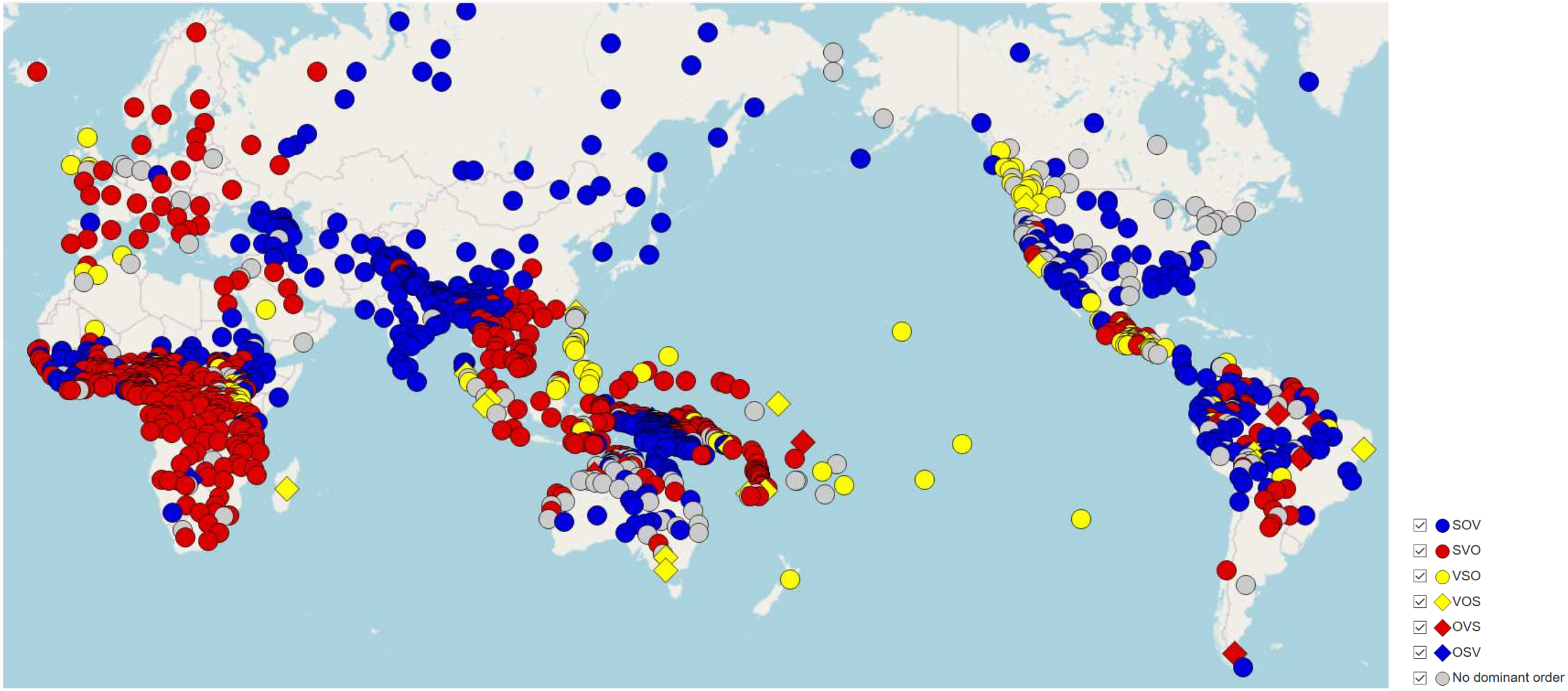


"free word order languages" such as Polish: *These languages don't impose a specific order to the words in a sentence.*

https://en.wikipedia.org/wiki/Word_order#Pragmatic_word_order
https://linguistics.stackexchange.com/questions/1476/what-is-word-order-used-for-in-free-word-order-languages
https://www.baeldung.com/cs/constituency-vs-dependency-parsing

Subject, Verb, Object?

Word order around the world

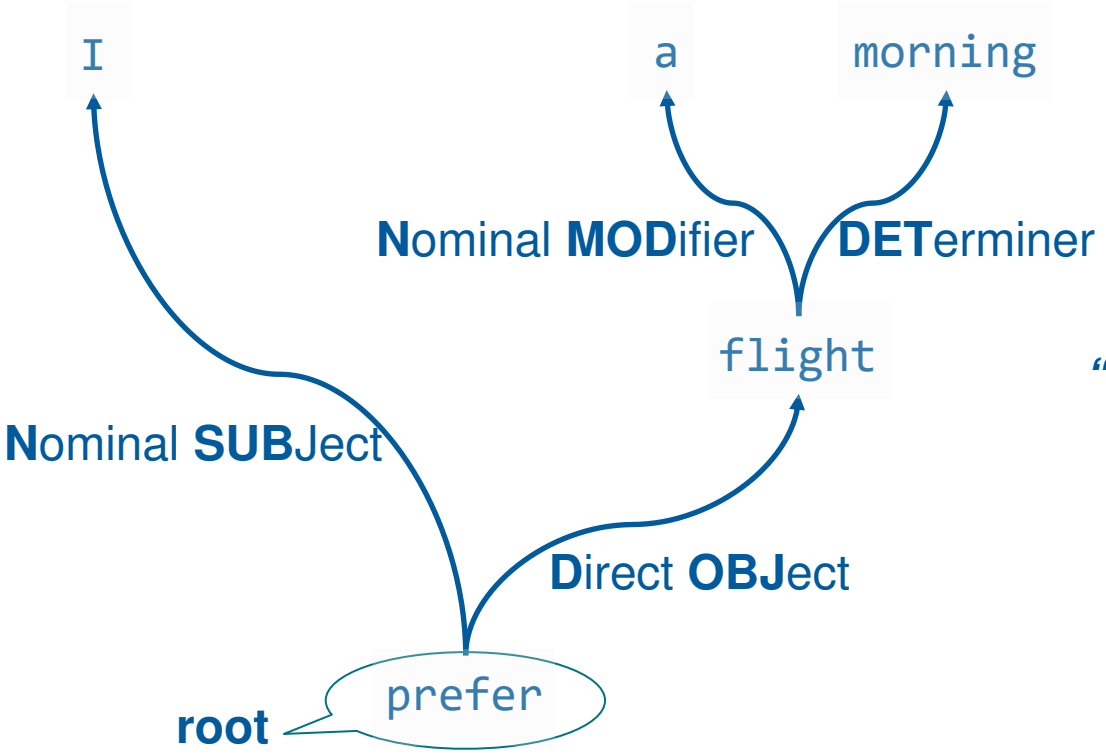


https://wals.info/feature/81A#2/16.0/165.2

Dependency Parsing

Example

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.



"free word order languages" such as Polish: *These languages don't impose a specific order to the words in a sentence.*

"Due to the nature of the underlying grammars, dependency parsing performs better than constituency in this kind of scenario."

https://en.wikipedia.org/wiki/Word_order#Pragmatic_word_order
https://linguistics.stackexchange.com/questions/1476/what-is-word-order-used-for-in-free-word-order-languages
https://www.baeldung.com/cs/constituency-vs-dependency-parsing

The task of analyzing the grammatical structure of a sentence and establishing the relationships between "head" words and the words which modify those heads.

It makes it easy to extract

subject-verb-object triples *(Who did what to whom)*

that are often

indicative of semantic relations between predicates.

Some Dependency Relations

Further Relation Examples

Relation	Examples with *head* and **dependent**
NSUBJ	**United** *canceled* the flight.
OBJ	United *diverted* the **flight** to Reno.
	We *booked* her the first **flight** to Miami.
IOBJ	We *booked* **her** the flight to Miami.
NMOD	We took the **morning** *flight*.
AMOD	Book the **cheapest** *flight*.
NUMMOD	Before the storm JetBlue canceled **1000** *flights*.
APPOS	*United*, a **unit** of UAL, matched the fares.
DET	**The** *flight* was canceled.
	Which *flight* was delayed?
CONJ	We *flew* to Denver and **drove** to Steamboat.
CC	We flew to Denver **and** *drove* to Steamboat.
CASE	Book the flight **through** *Houston*.

Clausal Argument Relations	**Description**
NSUBJ	Nominal subject
OBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
Nominal Modifier Relations	**Description**
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	**Description**
CONJ	Conjunct
CC	Coordinating conjunction

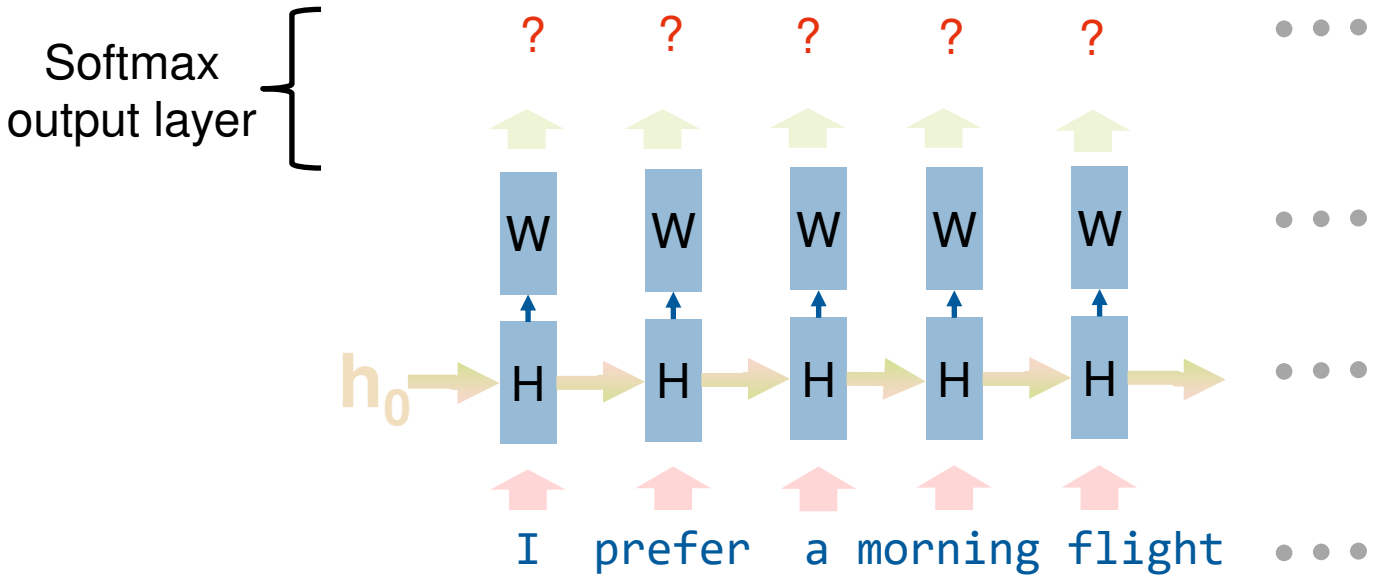
Tables from Section 18.1: https://web.stanford.edu/~jurafsky/slp3/

- **Transition-based methods**
- **Graph-based methods**

More about those techniques -> chapter 17 in Jurafsky

A tree can be represented as a sequence of nodes.

This sequence of nodes can be learned by an RNN.



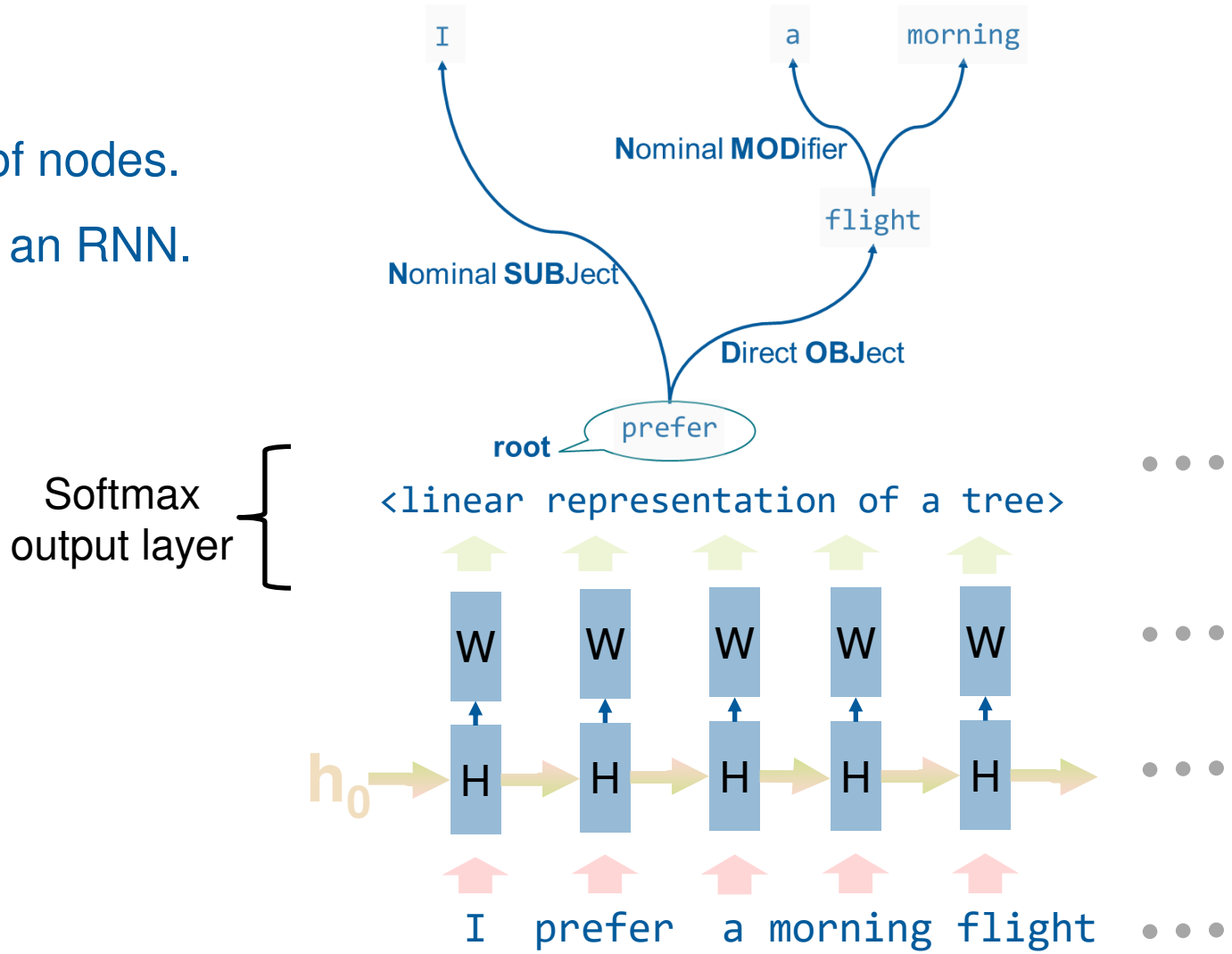
https://arxiv.org/pdf/2003.04807.pdf

Dependency Parsing or Constituent Analysis

Recurrent Neural Networks

A tree can be represented as a sequence of nodes.

This sequence of nodes can be learned by an RNN.

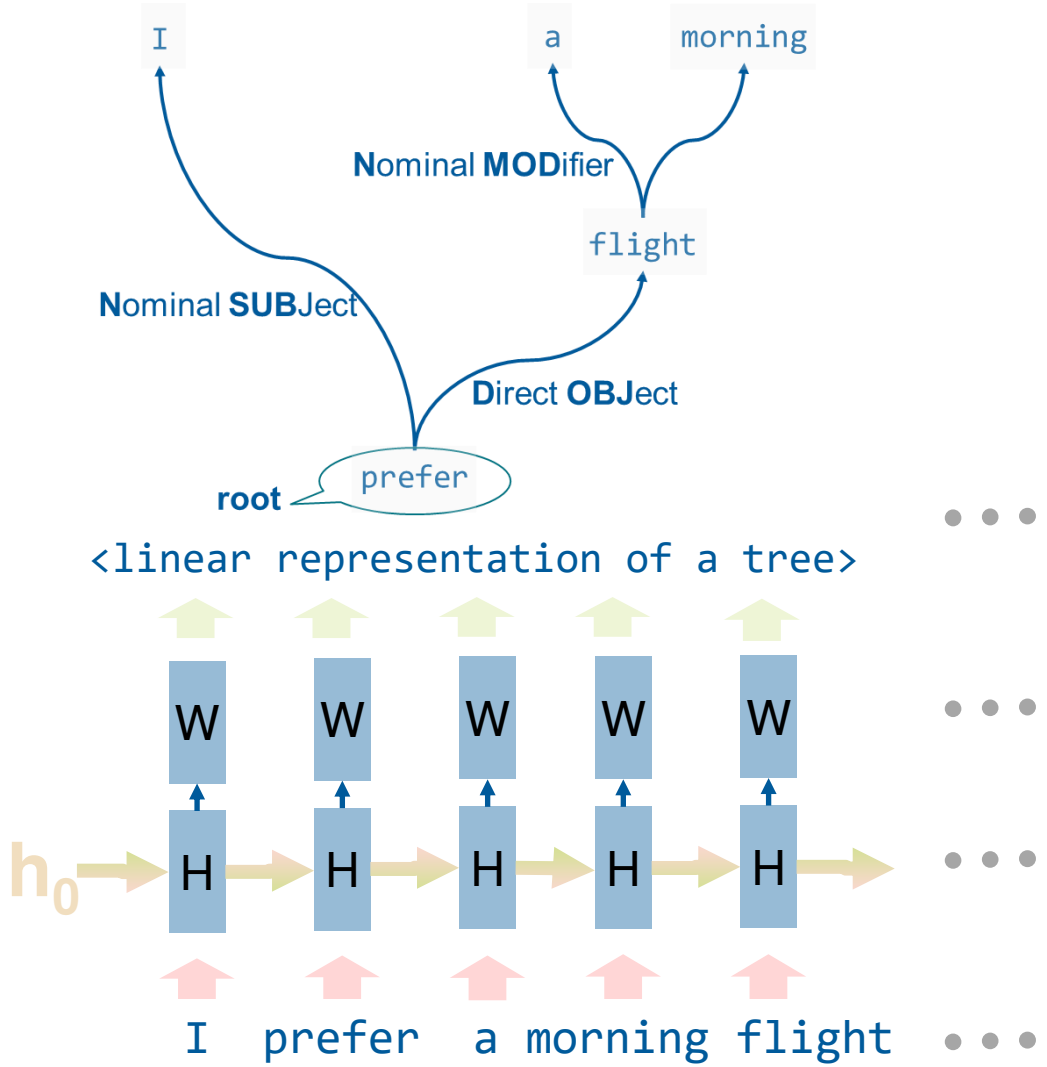


https://arxiv.org/pdf/2003.04807.pdf

A tree can be represented as a sequence of nodes.

This sequence of nodes can be learned by an RNN.

Interested in learning tree structures with DNNs?



https://arxiv.org/pdf/2003.04807.pdf

Dependency Parsing

What is it good for?

- **Extract Relations (open information extraction)**

  ⇔ build Knowledge bases from text

- **for building Q&A systems**

Dependency Parsing

Further Resources

- **Dependency Parser are often evaluated against Penn Treebank (PTB) data set**
- **https://github.com/sebastianruder/NLP-progress/blob/master/english/dependency_parsing.md**
- **UD:** project that is developing cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective