

# Parsing Morphologically Rich Languages (PMRL)

## Class 1: Introduction

Reut Tsarfaty  
[reut.tsarfaty@weizmann.ac.il](mailto:reut.tsarfaty@weizmann.ac.il)

# Linguistics

Linguistics is the Scientific Study of Language

- Phonetics/Phonology
- Morphology
- Syntax
- Semantics
- Pragmatics

# Computational Linguistics

Computational Linguistics is the  
Scientific Study of Language from a  
computational perspective

- Phonetics/Phonology: Sound  $\longrightarrow$  Speech
- Morphology: Words  $\longrightarrow$  Morphemes
- Syntax: Sentence  $\longrightarrow$  Structure
- Semantics: Structure  $\longrightarrow$  Meaning
- Pragmatics: Discourse  $\longrightarrow$  Intention, referents

# Natural Language Processing

NLP is the science and engineering of designing systems that automatically map utterances to analyses

- Phonetics/Phonology: **f: Sound**  $\longrightarrow$  **Speech**
- Morphology: **f: Words**  $\longrightarrow$  **Morphemes**
- Syntax: **f: Sentence**  $\longrightarrow$  **Structure**
- Semantics: **f: Structure**  $\longrightarrow$  **Meaning**
- Pragmatics: **f: Text**  $\longrightarrow$  **Intention**

# Natural Language Processing

NLP is the science and engineering of designing systems that automatically map utterances to analyses

- Phonetics/Phonology: **f: Sound**  $\longrightarrow$  **Speech**
- Morphology: **f: Words**  $\longrightarrow$  **Morphemes**
- Syntax: **f: Sentence**  $\longrightarrow$  **Structure**
- Semantics: **f: Structure**  $\longrightarrow$  **Meaning**
- Pragmatics: **f: Text**  $\longrightarrow$  **Intention**

# NLP as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

# NLP as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



Input

# NLP as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



Input

Sounds  
Strings  
Sequences

# NLP as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



Input

The diagram illustrates the NLP structure prediction framework. At the top, the title 'NLP as Structure Prediction' is displayed. Below it, a function  $f : \mathcal{X} \rightarrow \mathcal{Y}$  is shown. A blue oval labeled 'Input' is connected by a line to the  $\mathcal{X}$  in the function. Another blue oval labeled 'Output' is connected by a line to the  $\mathcal{Y}$ . Below the 'Input' oval, the text 'Sounds', 'Strings', and 'Sequences' is listed, representing different types of input data.

Sounds  
Strings  
Sequences

Output

# NLP as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Input

Sounds  
Strings  
Sequences

Output

Boolean (T/F)  
Class Label  
Structure

# Success Stories

# Success Stories



# Success Stories



# Success Stories



# Success Stories



# Success Stories



# Success Stories



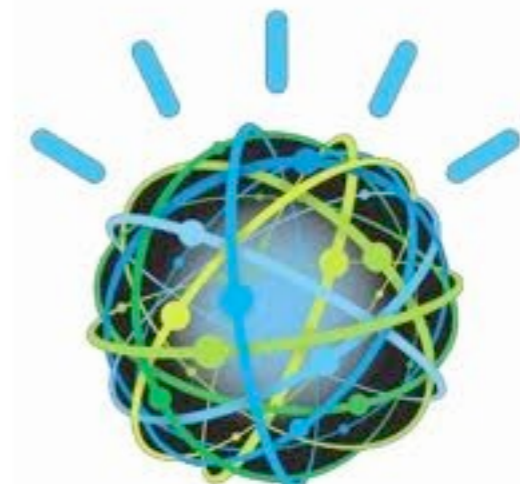
# Success Stories



# Success Stories

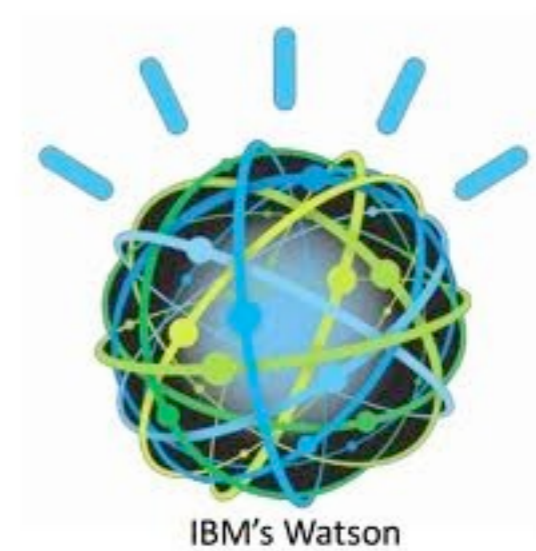


# Success Stories



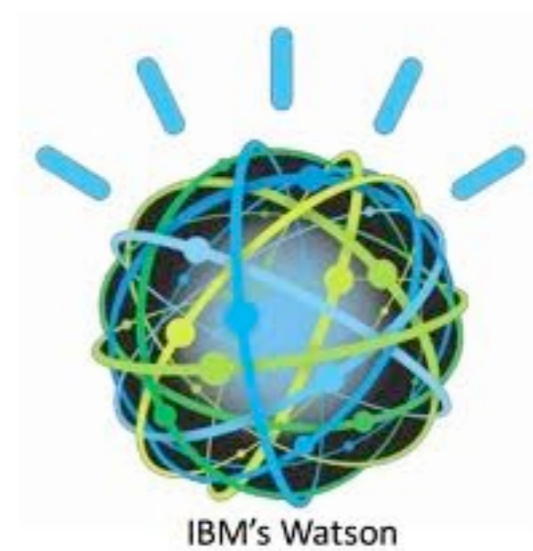
IBM's Watson

# Success Stories?



# Success Stories?

How does Siri respond to your own native language?

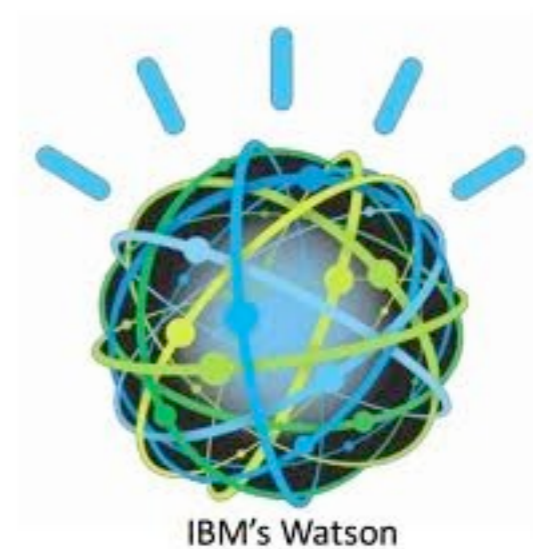


# Success Stories?

How does Siri respond to your own native language?



How does GT translate into your own native language?



# Success Stories?

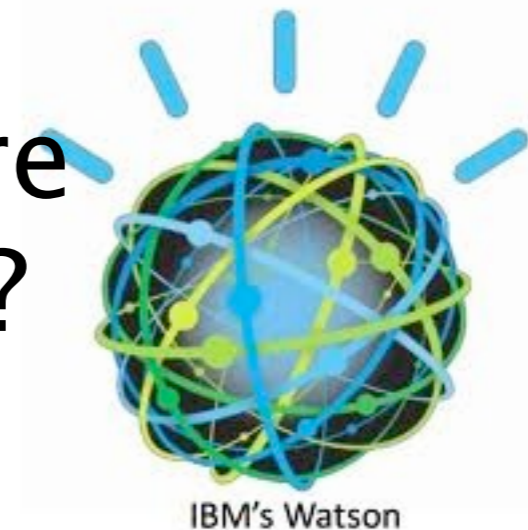
How does Siri respond to your own native language?



How does GT translate into your own native language?



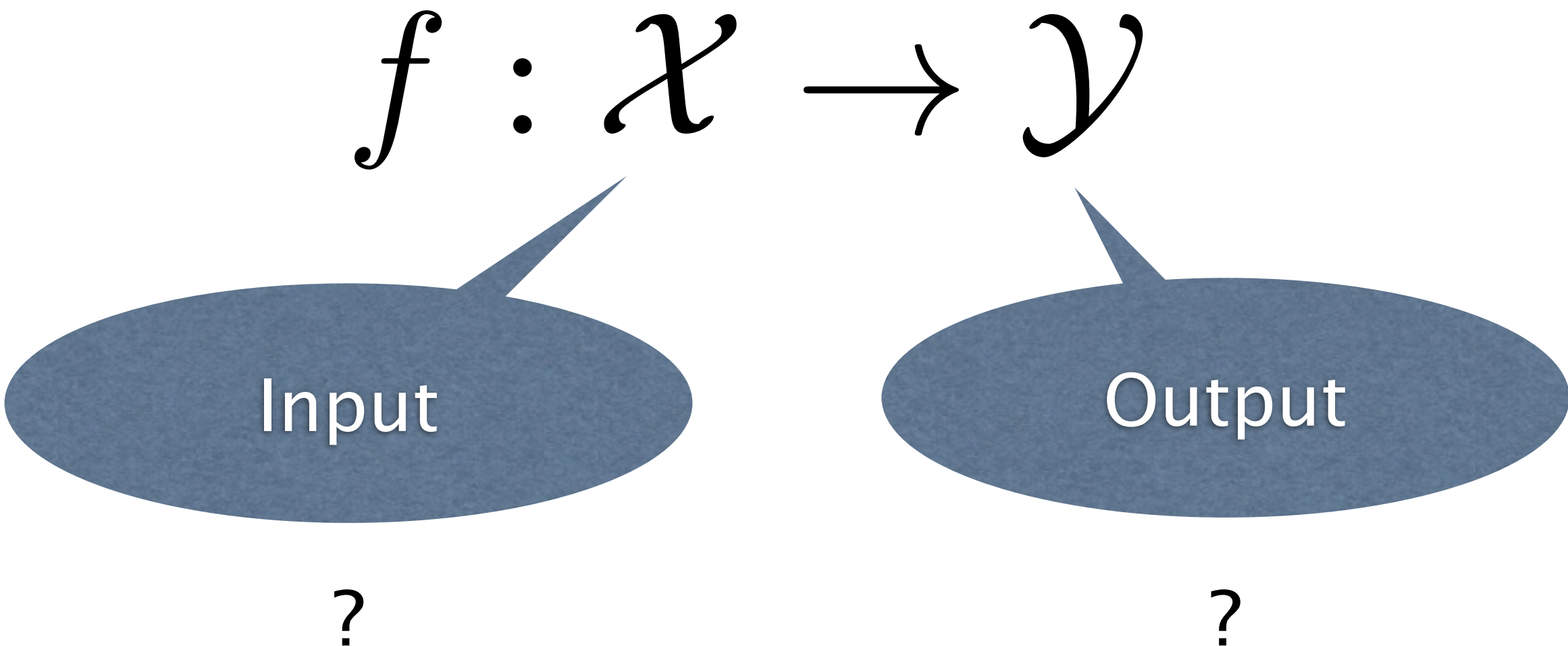
How much does Watson score in your own native language?



# Why is it Hard to Parse MRLs?

# PMRL as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



Input

The diagram illustrates the PMRL structure prediction framework. At the top, the function  $f : \mathcal{X} \rightarrow \mathcal{Y}$  is shown. Below this, two blue ovals represent the input and output spaces. The left oval is labeled 'Input' and has a callout line pointing to the  $\mathcal{X}$  in the function notation above. Below the 'Input' oval is a question mark. The right oval is labeled 'Output' and has a callout line pointing to the  $\mathcal{Y}$  in the function notation above. Below the 'Output' oval is a question mark.

Output

?

?

# PMRL as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



Complex Input

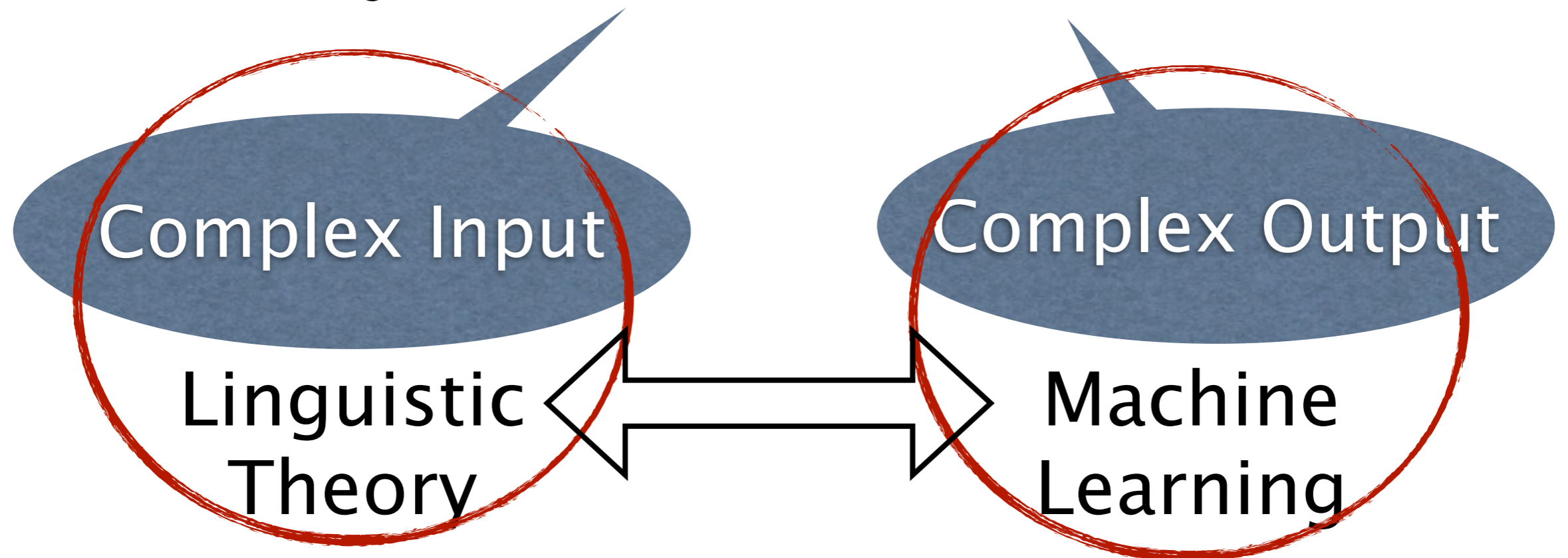
Linguistic  
Theory

Complex Output

Machine  
Learning

# PMRL as Structure Prediction

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$



How do we approach  
the PMRL challenge?

# This Course: Day 1

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Complex Input

Complex Output

# This Course: Day 1

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Complex Input

Linguistic  
Theory

Complex Output

# This Course: Day 2

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Complex Input

Complex Output

Phrase  
Structures

# This Course: Day 3

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Complex Input

Complex Output

Dependency  
Structures

# This Course: Day 4

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Complex Input

Complex Output

Relational  
Networks

# This Course: Day 5

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

Complex Input


Complex Output

# This Course

# This Course

- Day 1: Introduction
- Day 2: Phrase-structures
- Day 3: Dependency-structures
- Day 4: Relational-Realizational
- Day 5: Evaluation and Multilinguality

# This Course

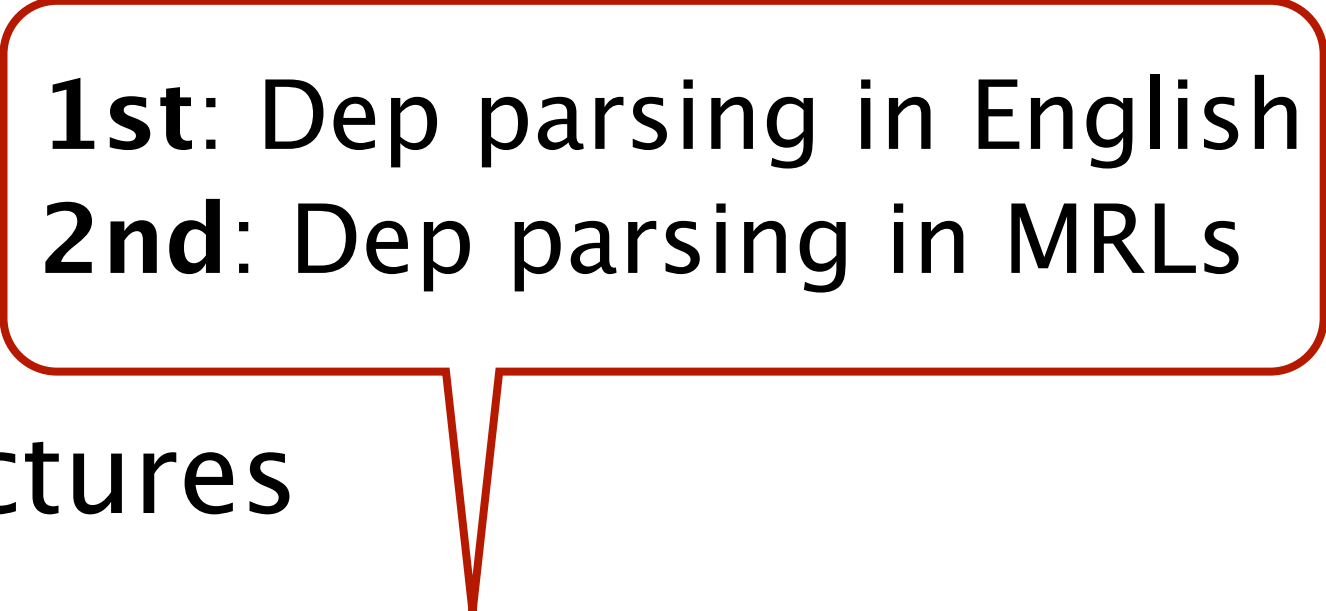
- Day 1: **Introduction** 
  - 1st: Basics in Parsing
  - 2nd: Basics in Morphology
- Day 2: Phrase-structures
- Day 3: Dependency-structures
- Day 4: Relational-Realizational
- Day 5: Evaluation and Multilinguality

# This Course

- Day 1: Introduction
- Day 2: **Phrase-structures**
- Day 3: Dependency-structures
- Day 4: Relational-Realizational
- Day 5: Evaluation and Multilinguality

1st: PS parsing in English  
2nd: PS parsing in MRLs

# This Course

- Day 1: Introduction
  - Day 2: Phrase-structures
  - Day 3: **Dependency-structures**
  - Day 4: Relational-Realizational
  - Day 5: Evaluation and Multilinguality
- 
- 1st: Dep parsing in English  
2nd: Dep parsing in MRLs

# This Course

- Day 1: Introduction
- Day 2: Phrase-structures
- Day 3: Dependency-structures
- Day 4: Relational-Realizational
- Day 5: Evaluation and Multilinguality

1st: Relational Networks  
2nd: RR parsing in MRLs

# This Course

- Day 1: Introduction
- Day 2: Phrase-structure
- Day 3: Dependency-structure
- Day 4: Relational-Realizational
- Day 5: **Evaluation and Multilinguality**

**1st: Across-Frameworks**  
**2nd: Across-Languages**

# This Course

- Day 1: Introduction
- Day 2: Phrase-structure
- Day 3: Dependency-structure
- Day 4: Relational-Realizational
- Day 5: Evaluation and Multilinguality

**Towards Universal Parsing**

# Today

- Day 1: Introduction
  - Introduction to Parsing
  - Introduction to Morphology
- Day 2: Phrase-structure
- Day 3: Dependency-structure
- Day 4: Relational-Realizational
- Day 5: Evaluation and Multilinguality

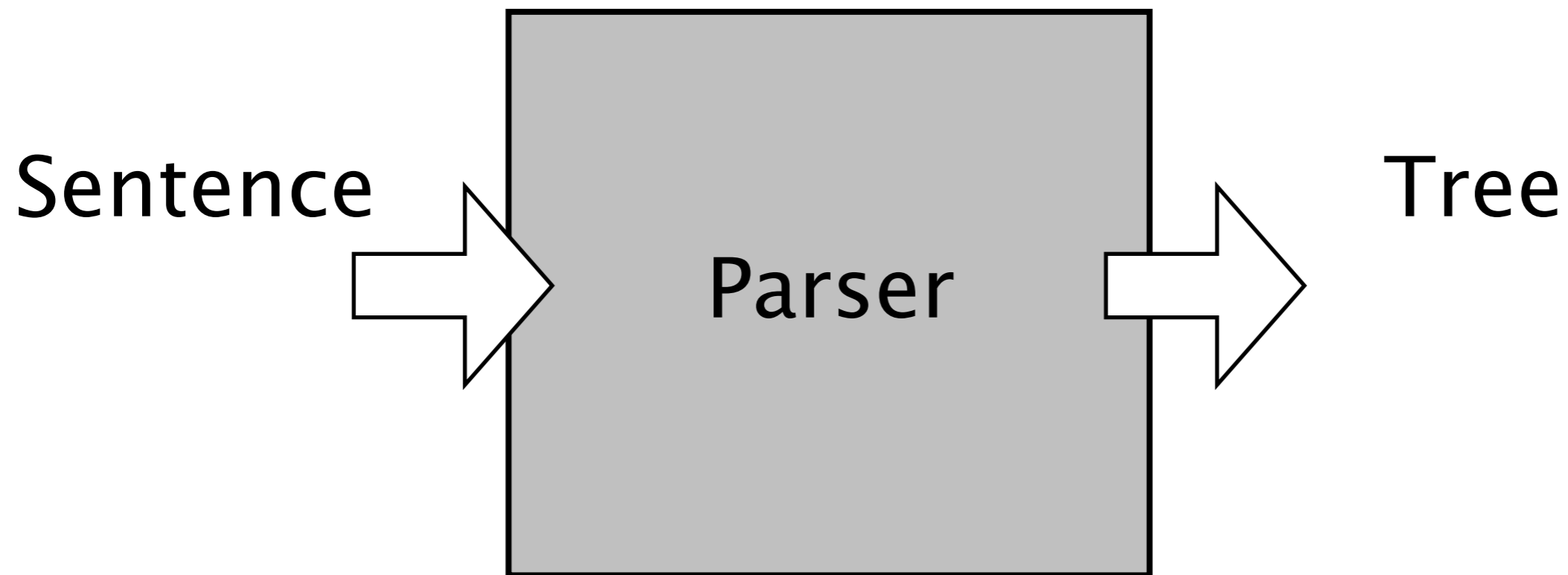
# Introduction#1

## Parsing

What's in  
a Parser?

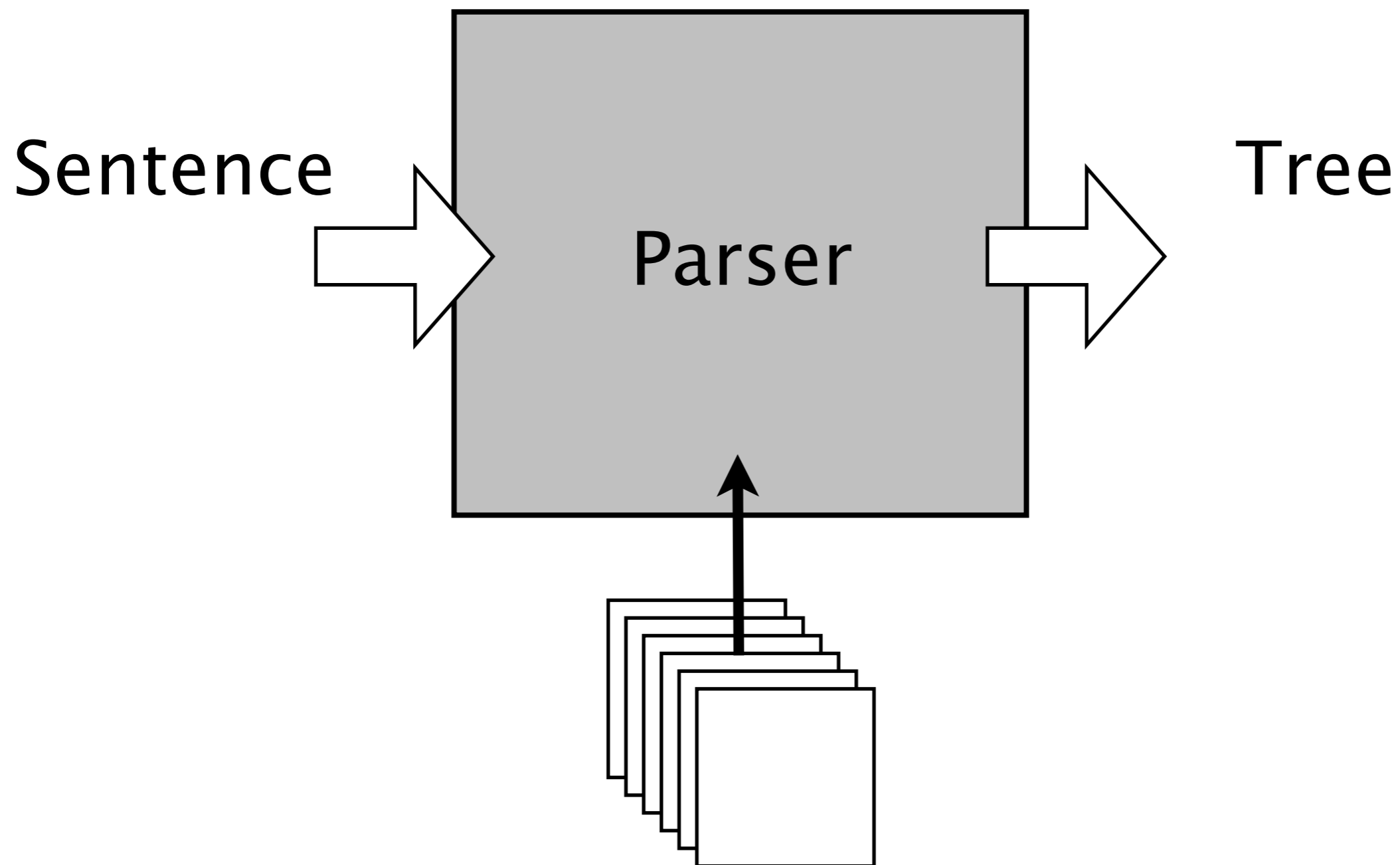
# Parsing

## An Architectural View



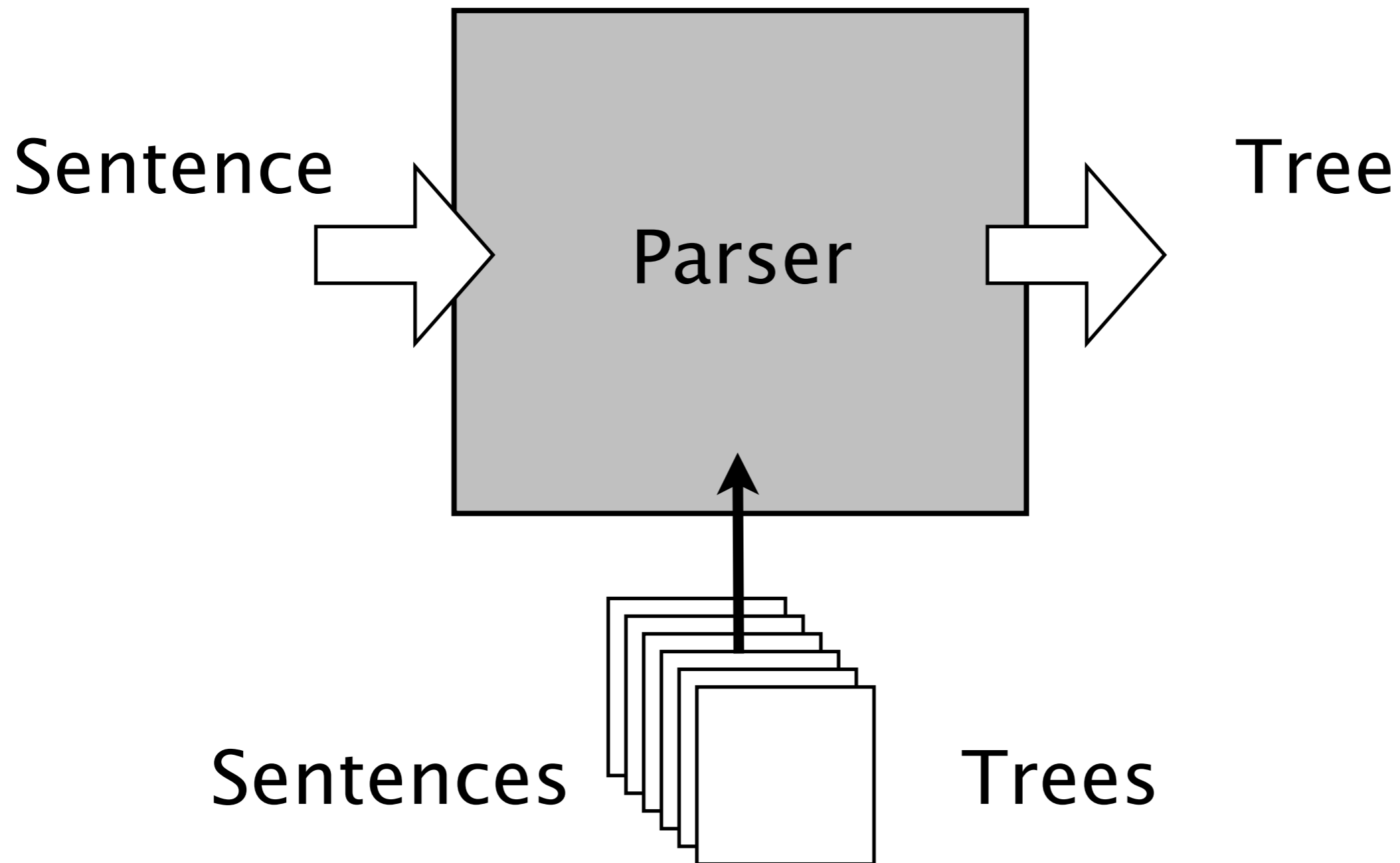
# Statistical Parsing

## An Architectural View



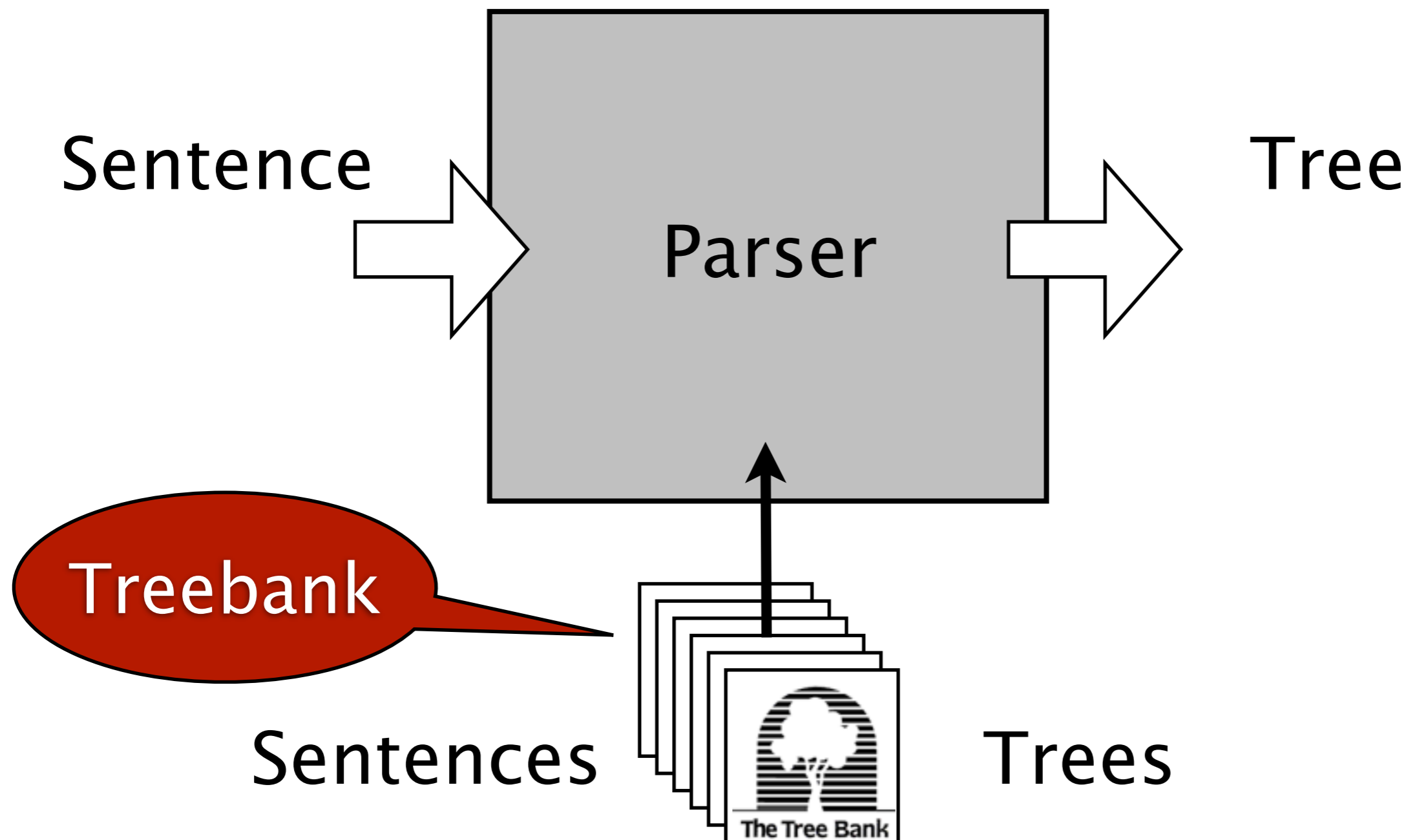
# Supervised Statistical Parsing

## An Architectural View



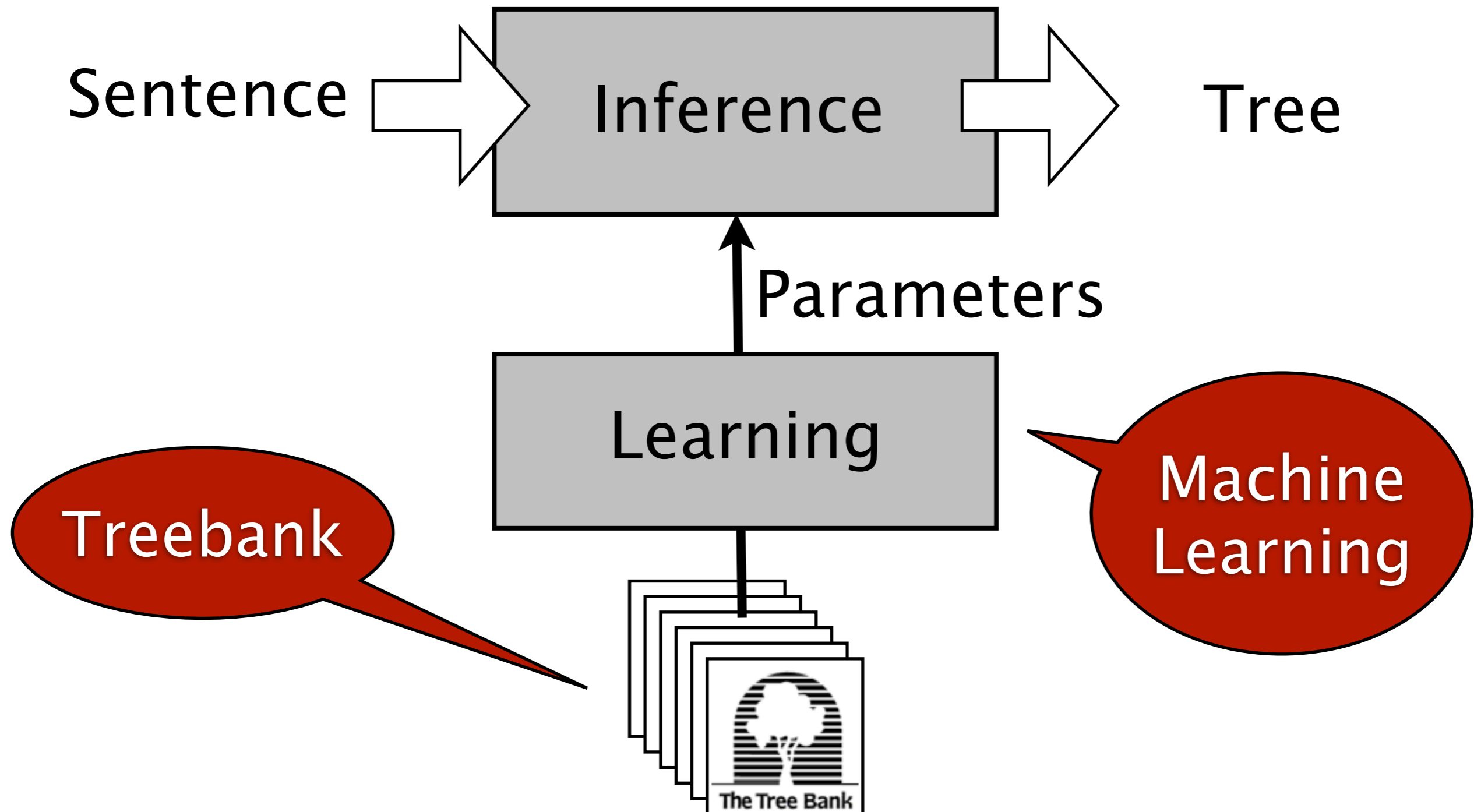
# Statistical Parsing

## An Architectural View



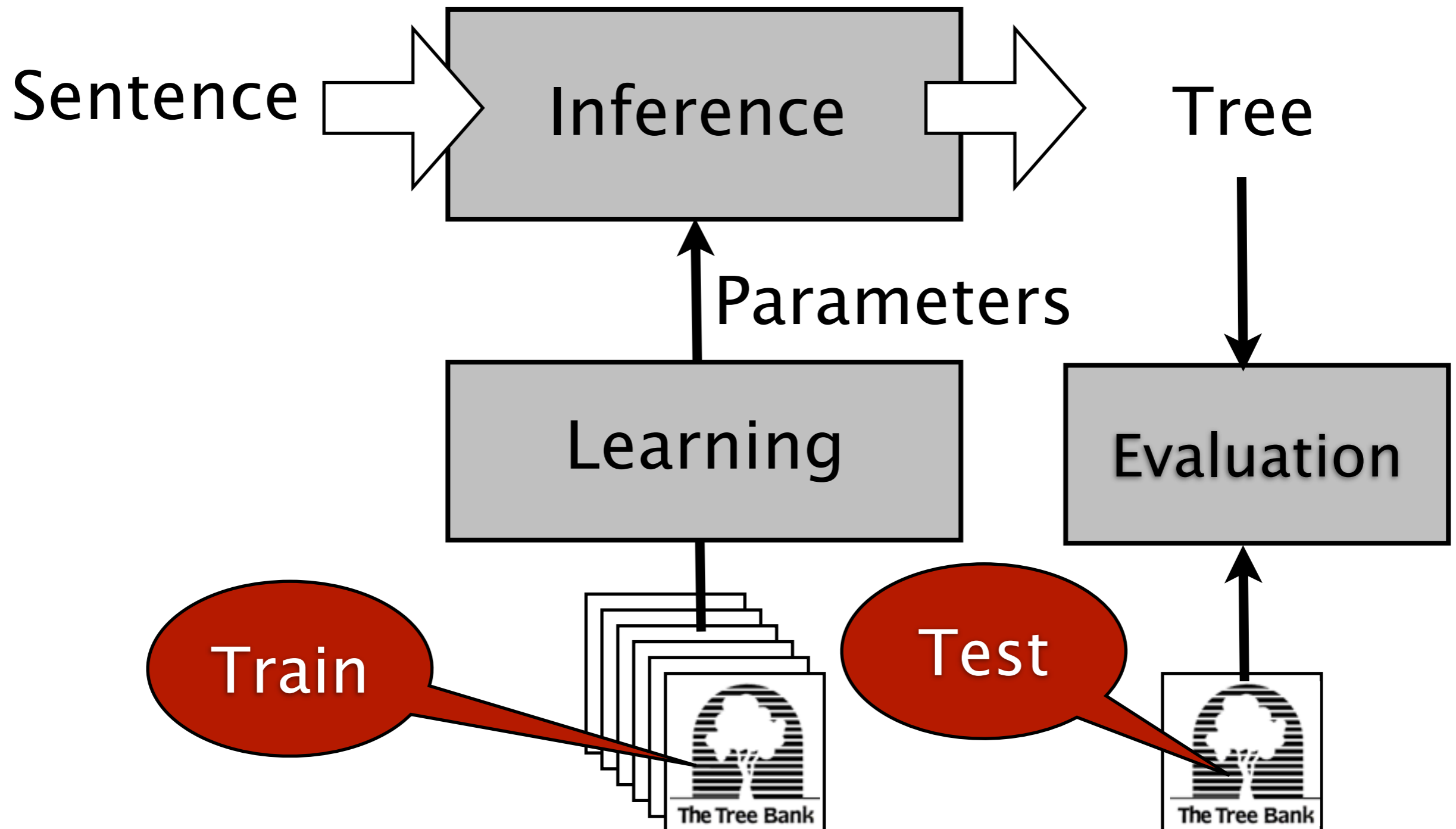
# Statistical Parsing

## An Architectural View



# Statistical Parsing

## An Architectural View



# Statistical Parsing

## An Architectural View

- Important Architectural decisions:
  - Representation (or: Input, Output)
  - Model (or: Parameters)
  - Learning (or: Training)
  - Inference (or: Decoding)
  - Evaluation

# Representation

גנן גידל דגן בגן

# Representation

גנן גידל דגן בגן



PARSER

# Representation

גנן גידל דגן בגן



PARSER



# Representation

How can  
**Structure**  
represent  
**Meaning?**



# Representation

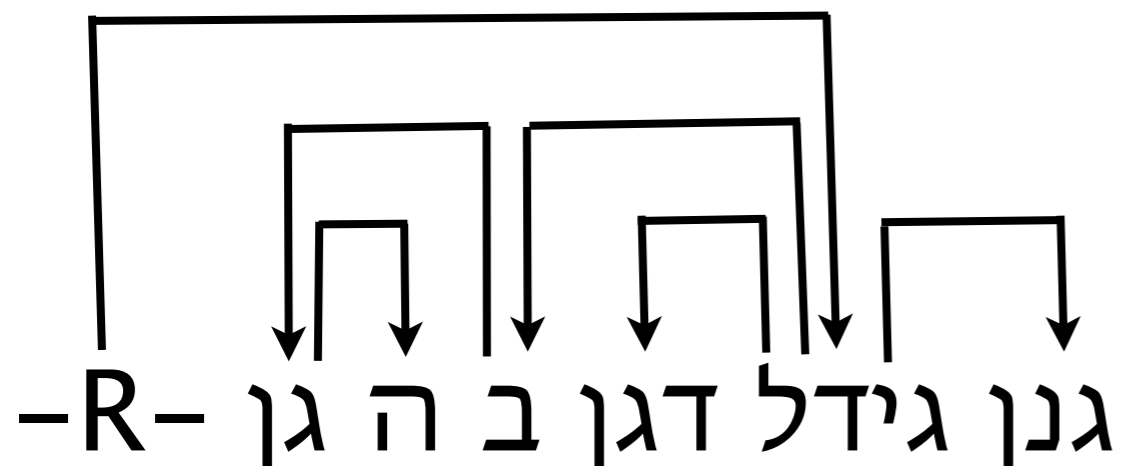
How can  
**Structure**  
represent  
**Meaning?**



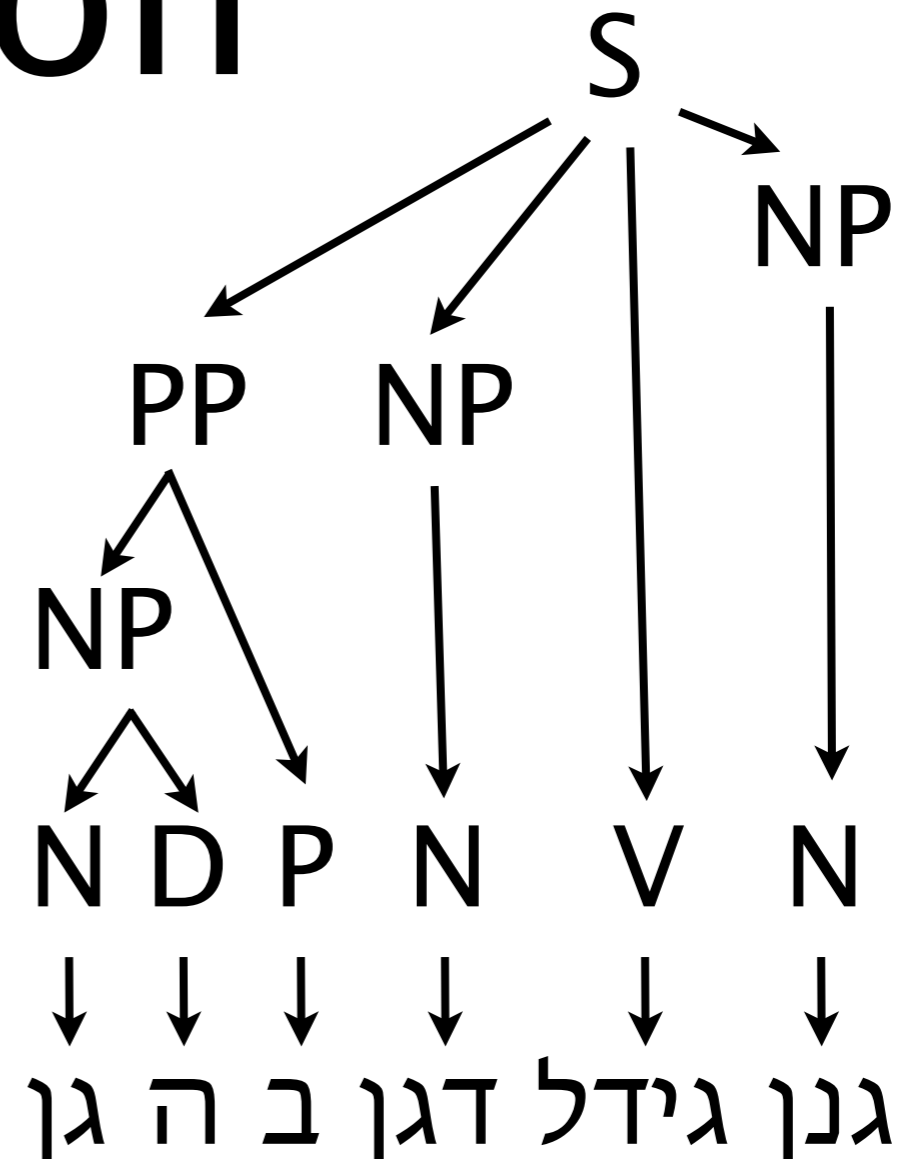
Dependency  
Trees

Constituency  
Trees

# Representation

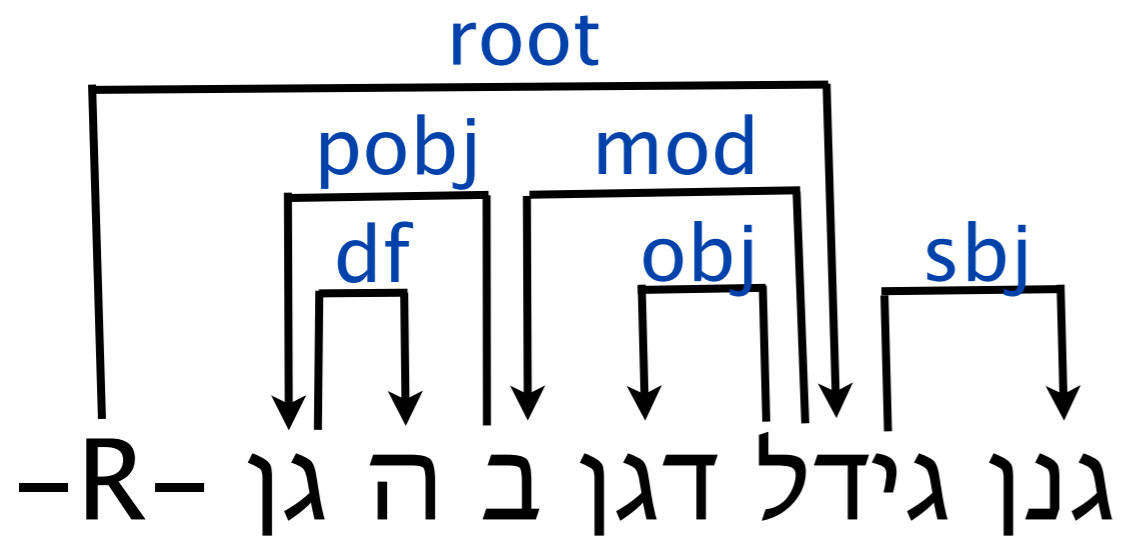


Dependency  
Trees

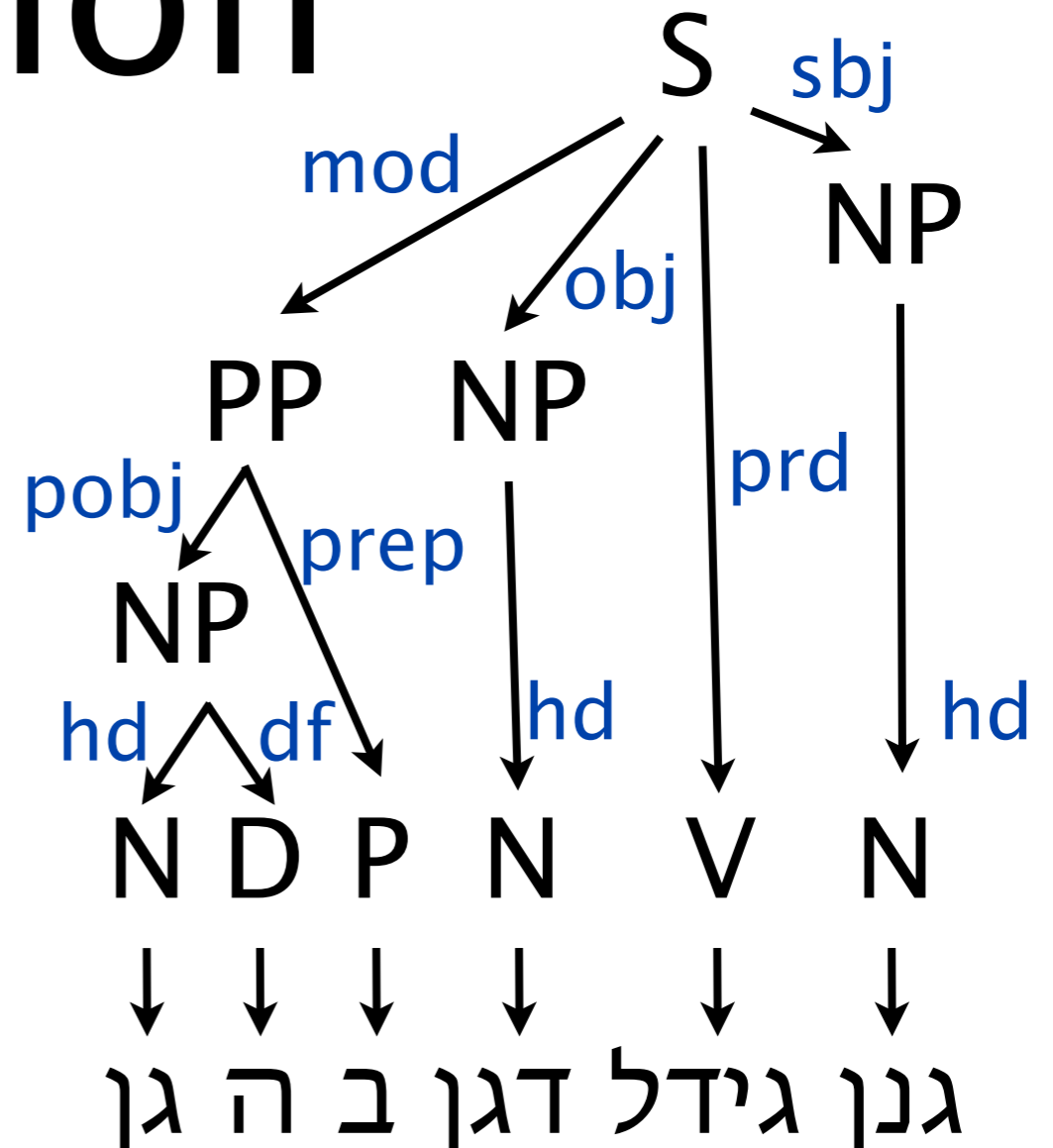


Constituency  
Trees

# Representation

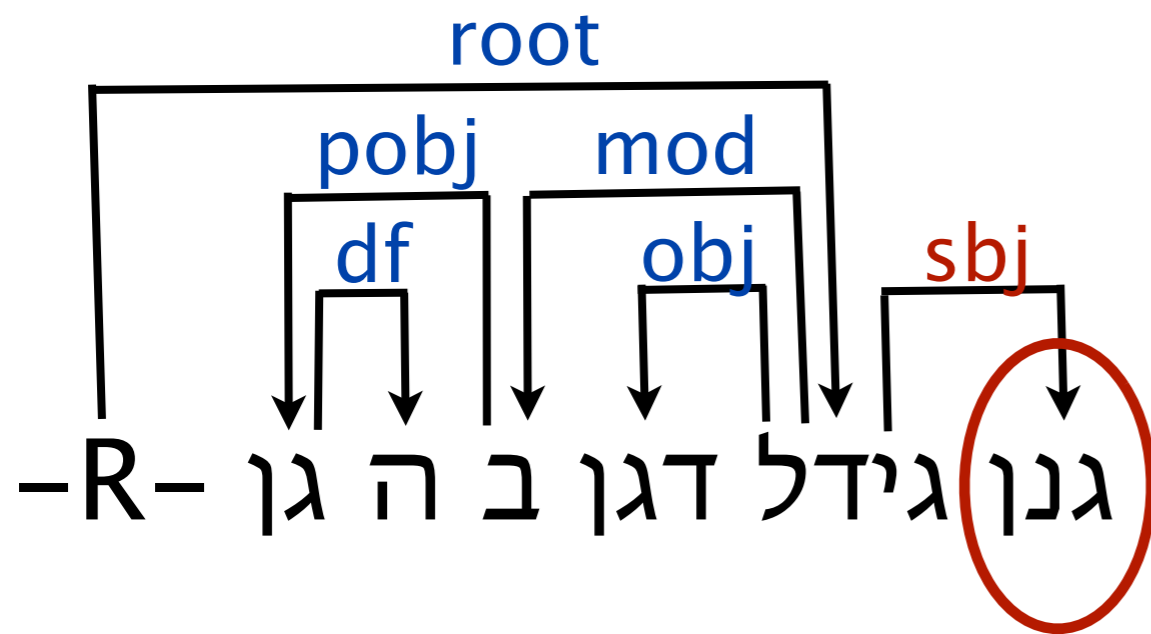


Dependency  
Trees

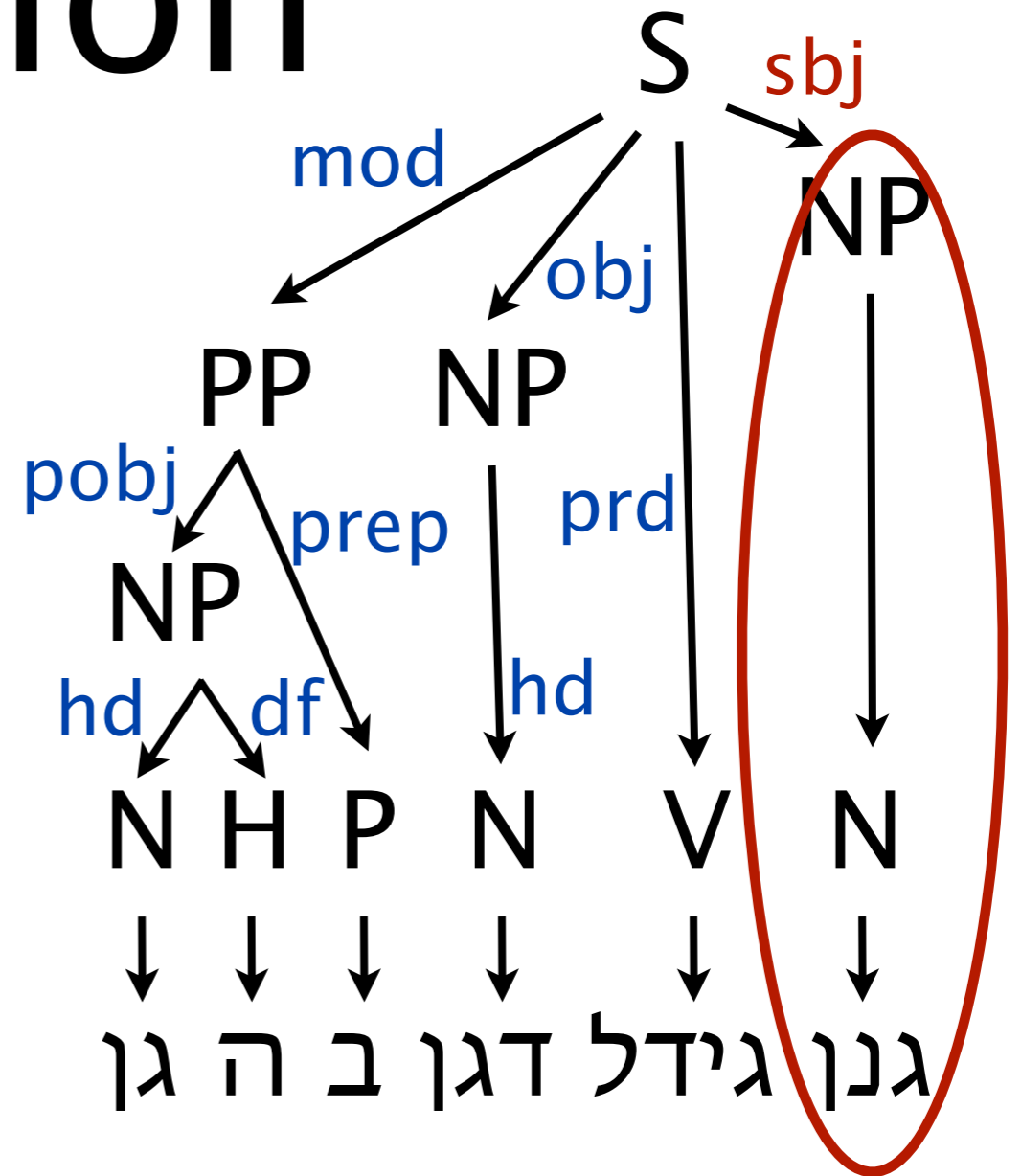


Constituency  
Trees

# Representation

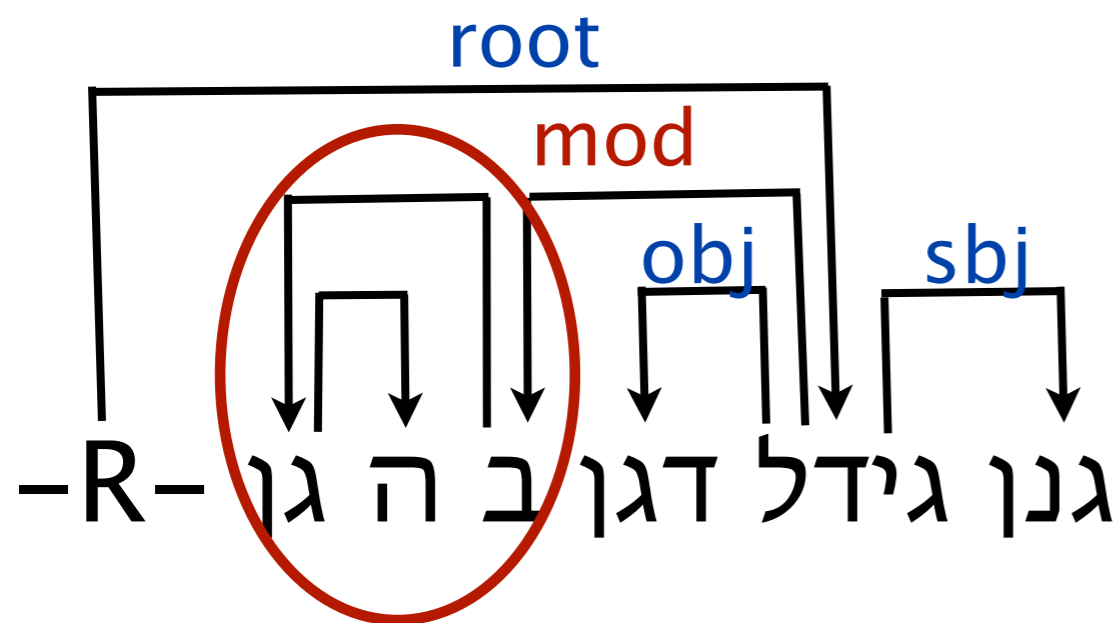


Dependency  
Trees

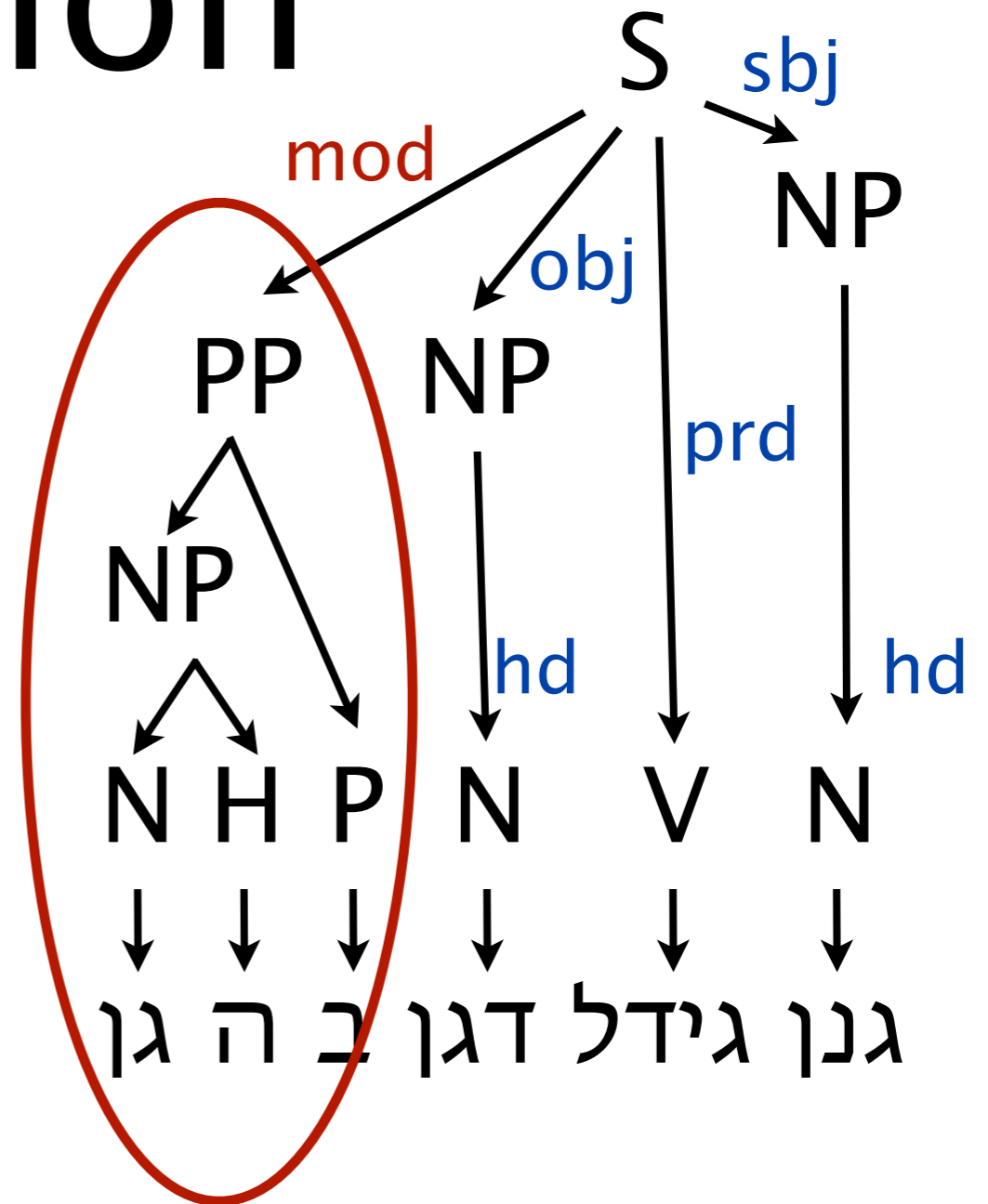


Constituency  
Trees

# Representation

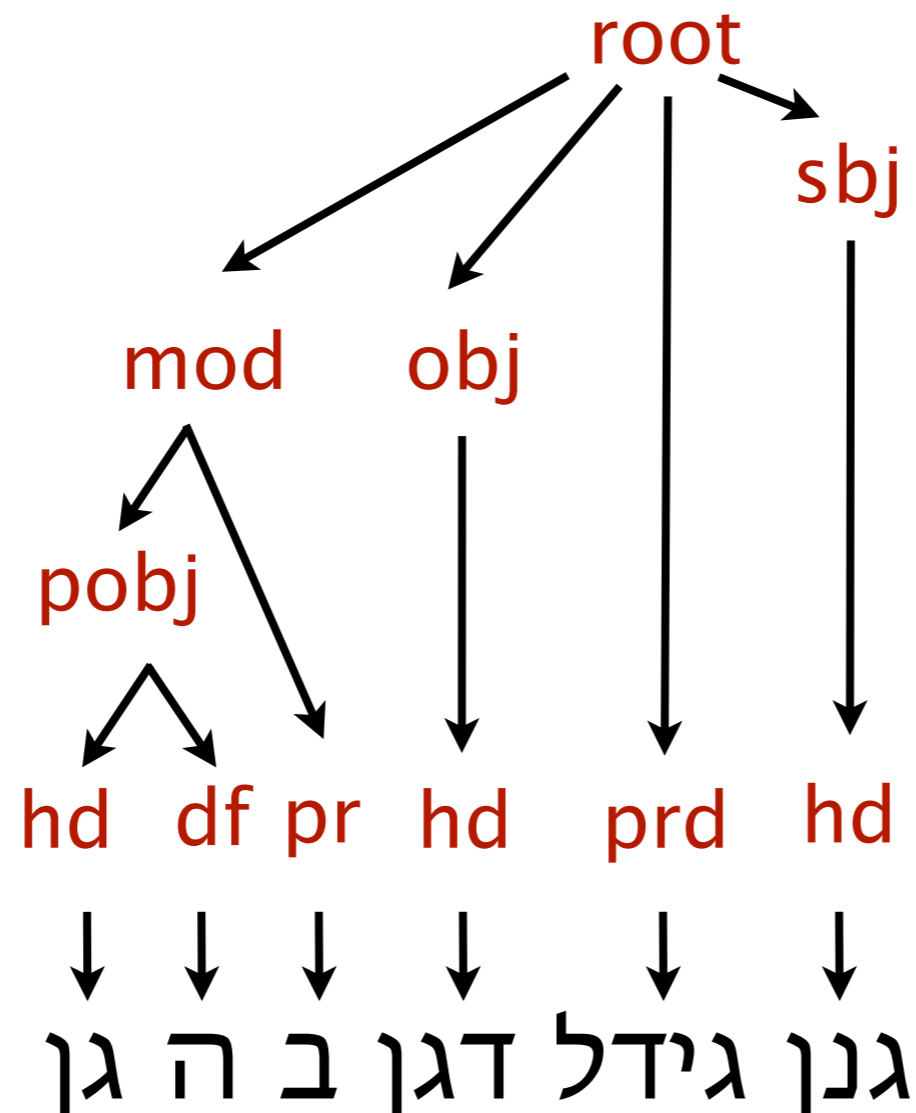


Dependency  
Trees



Constituency  
Trees

# Representation



Dependency  
Trees

Relational  
Networks

Constituency  
Trees

# Model

גנן גידל דגן בגן



a function



# Model

גנן גידל דגן בגן



$$f(x) = \arg \max_{y \in \mathcal{Y}} \text{Score}(x; y)$$



# Model

גנן גידל דגן בגן



$$f(x) = \arg \max_{y \in \mathcal{Y}} \text{Score}(x; y)$$



# Model

גנן גידל דגן בגן



$$f(x) = \arg \max_{y \in \mathcal{Y}} \text{Score}(x; y)$$



$$f(x) = \arg \max_y P(x, y)$$

**generative**

$$f(x) = \arg \max_y P(x|y)$$

**discriminative**

# Learning

גנן גידל דגן בגן



$$f(x) = \arg \max_{y \in \mathcal{Y}} \text{Score}(x; y)$$



# Inference

גנן גידל דגן בגן



$$f(x) = \arg \max_{y \in \mathcal{Y}} \text{Score}(x; y)$$



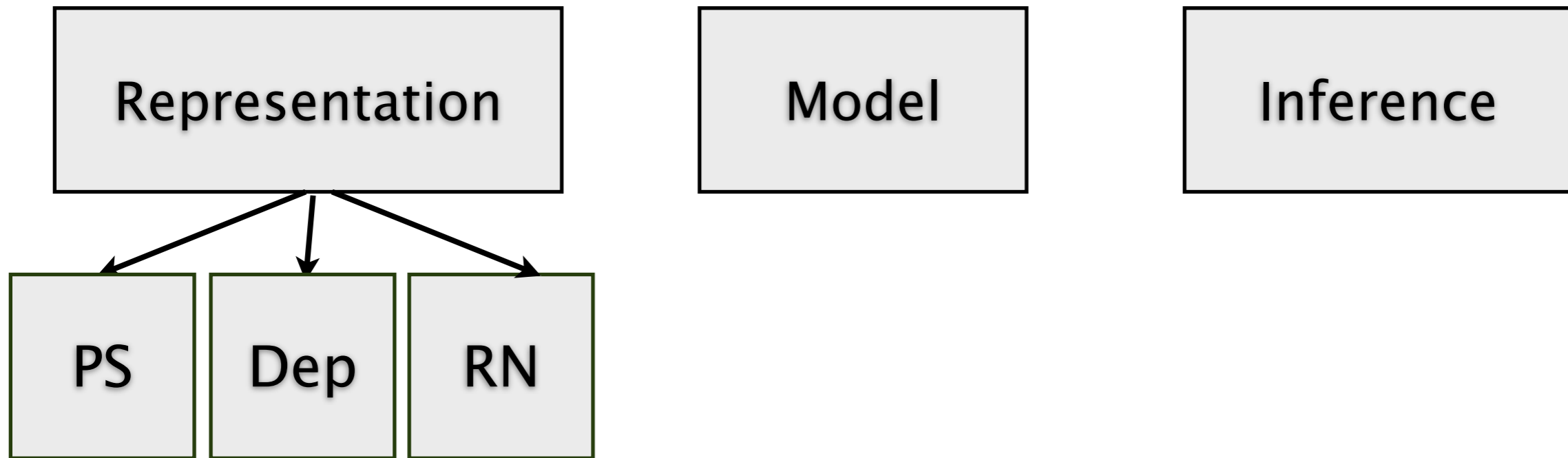
# -- The Parsing Universe --

Representation

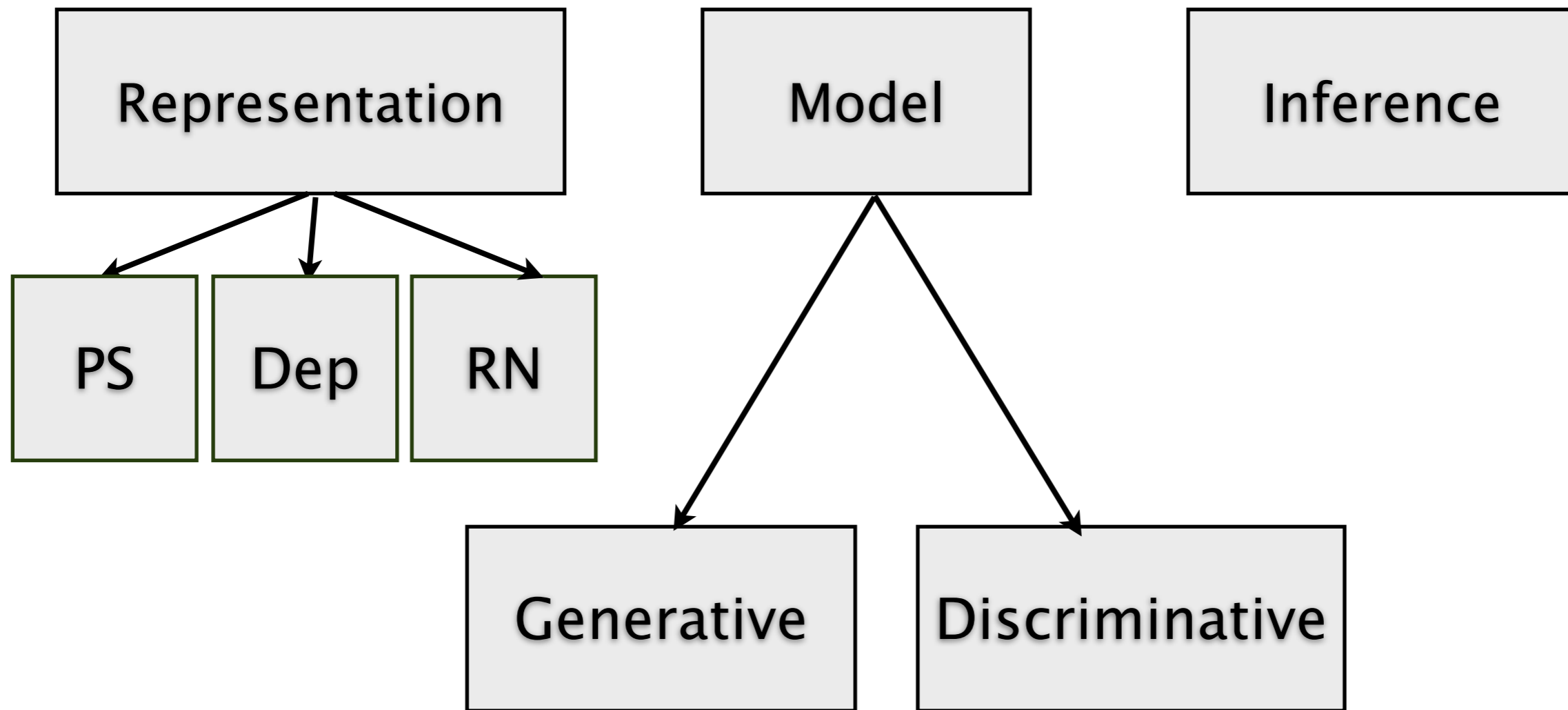
Model

Inference

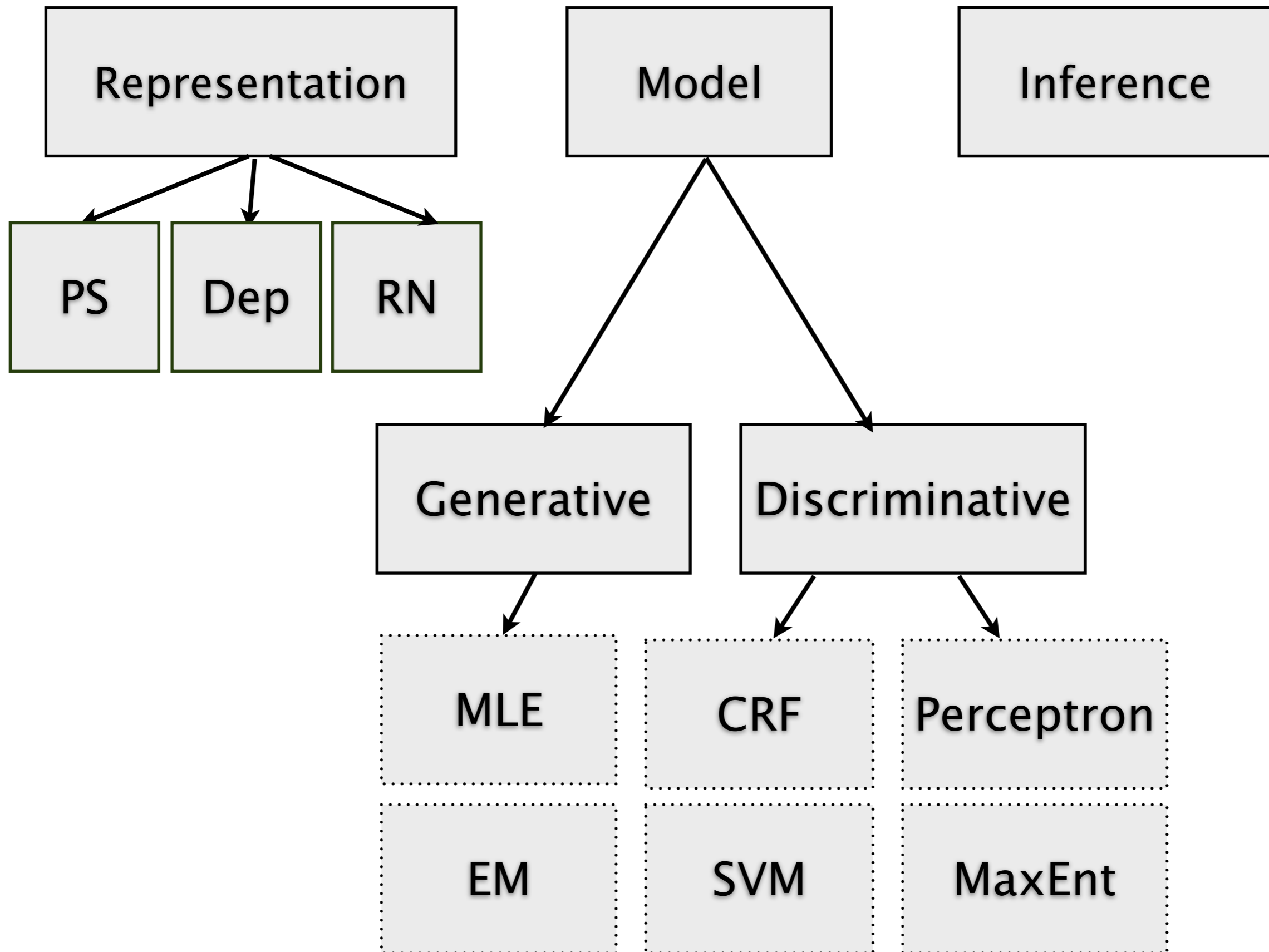
# -- The Parsing Universe --



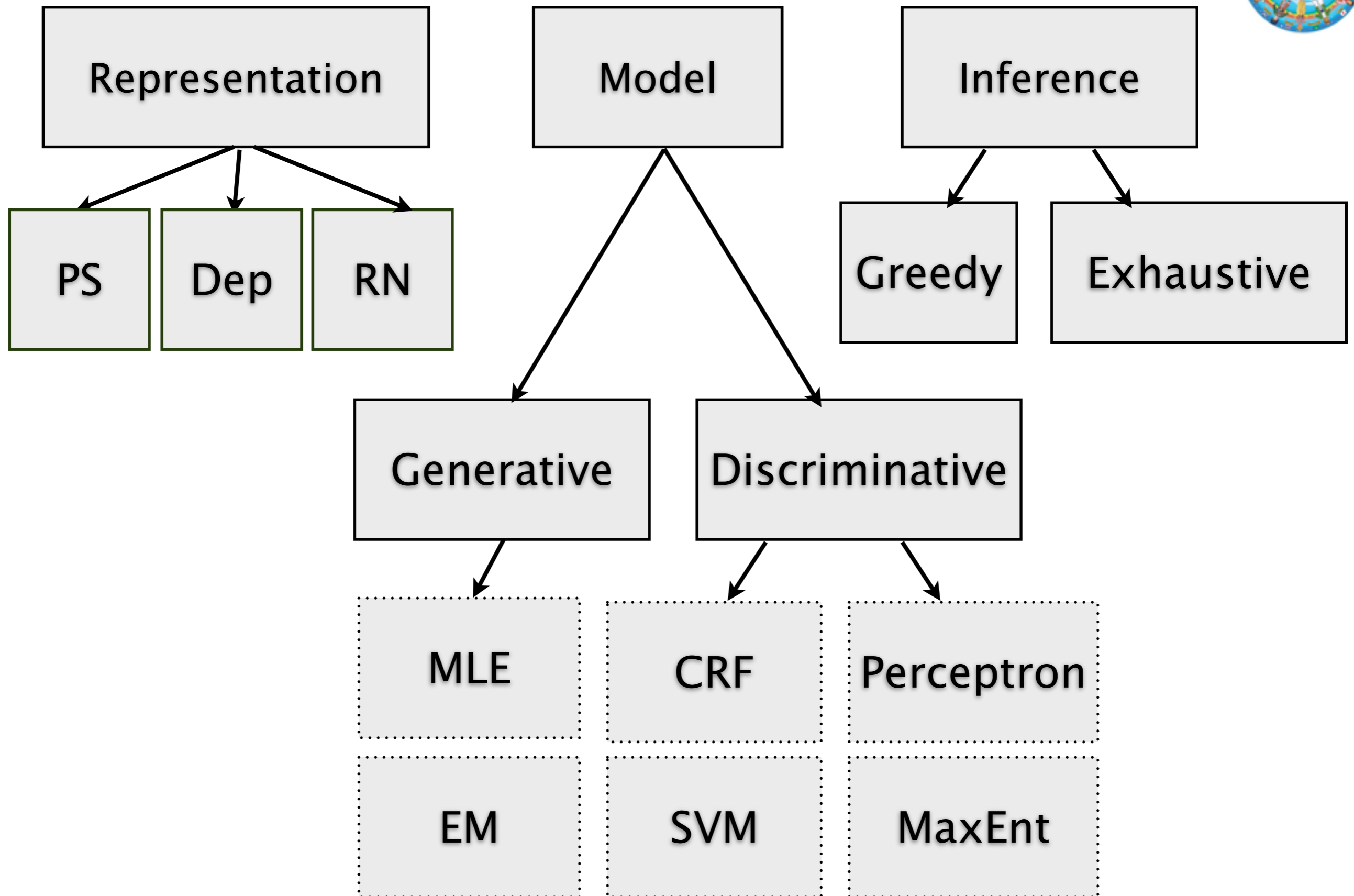
# -- The Parsing Universe --



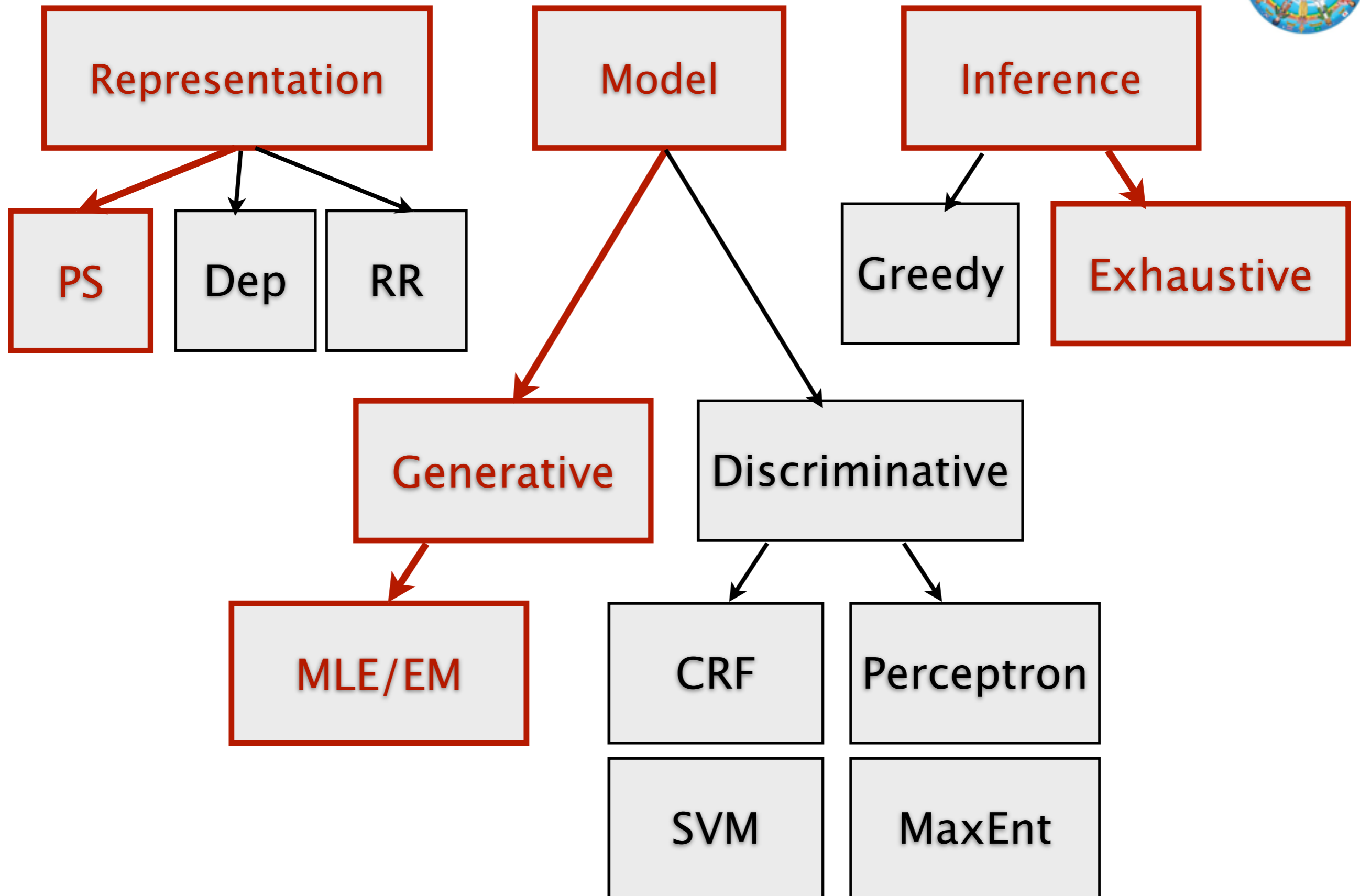
# -- The Parsing Universe --



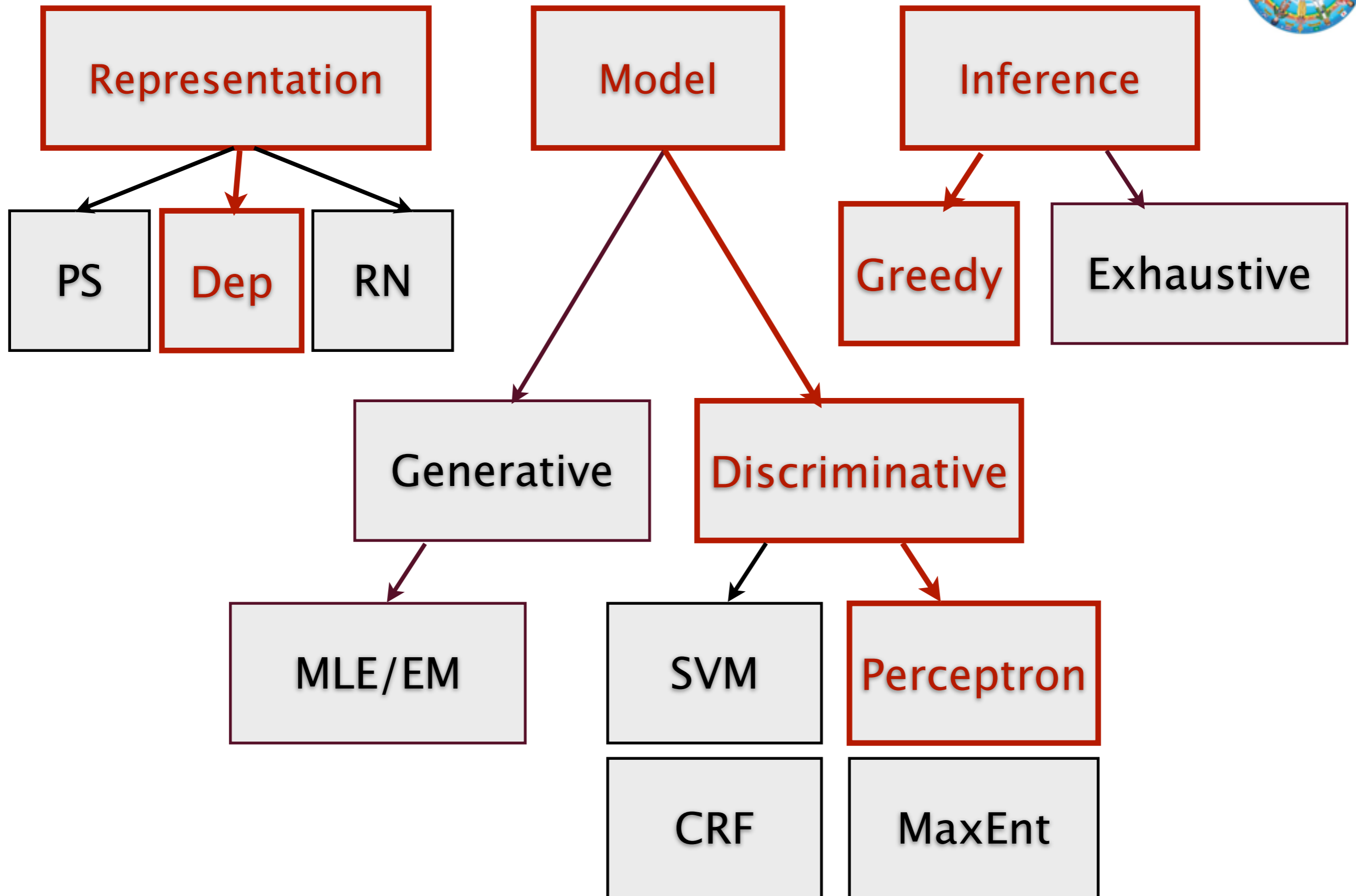
# -- The Parsing Universe --



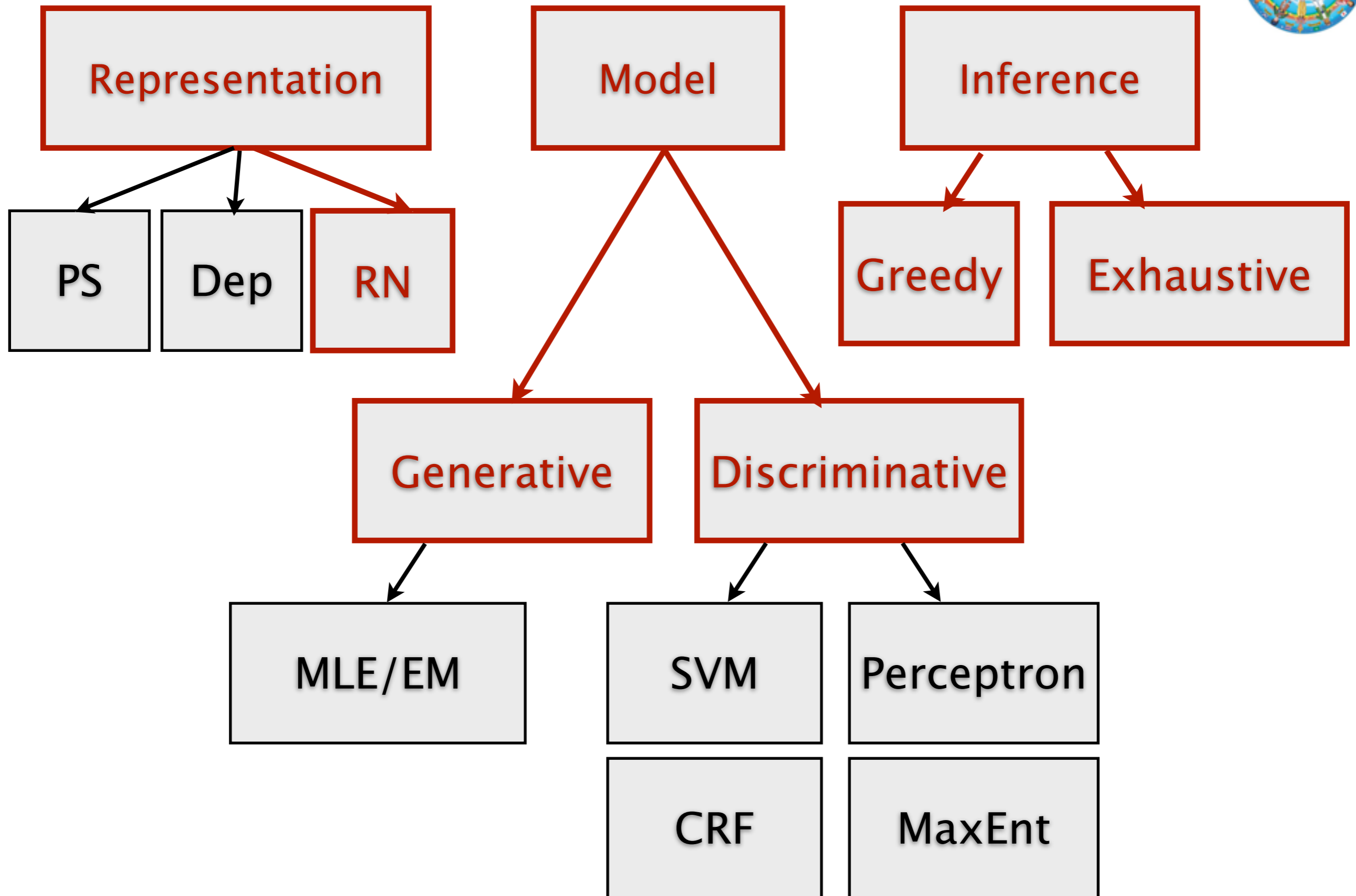
# -- Day 2 --



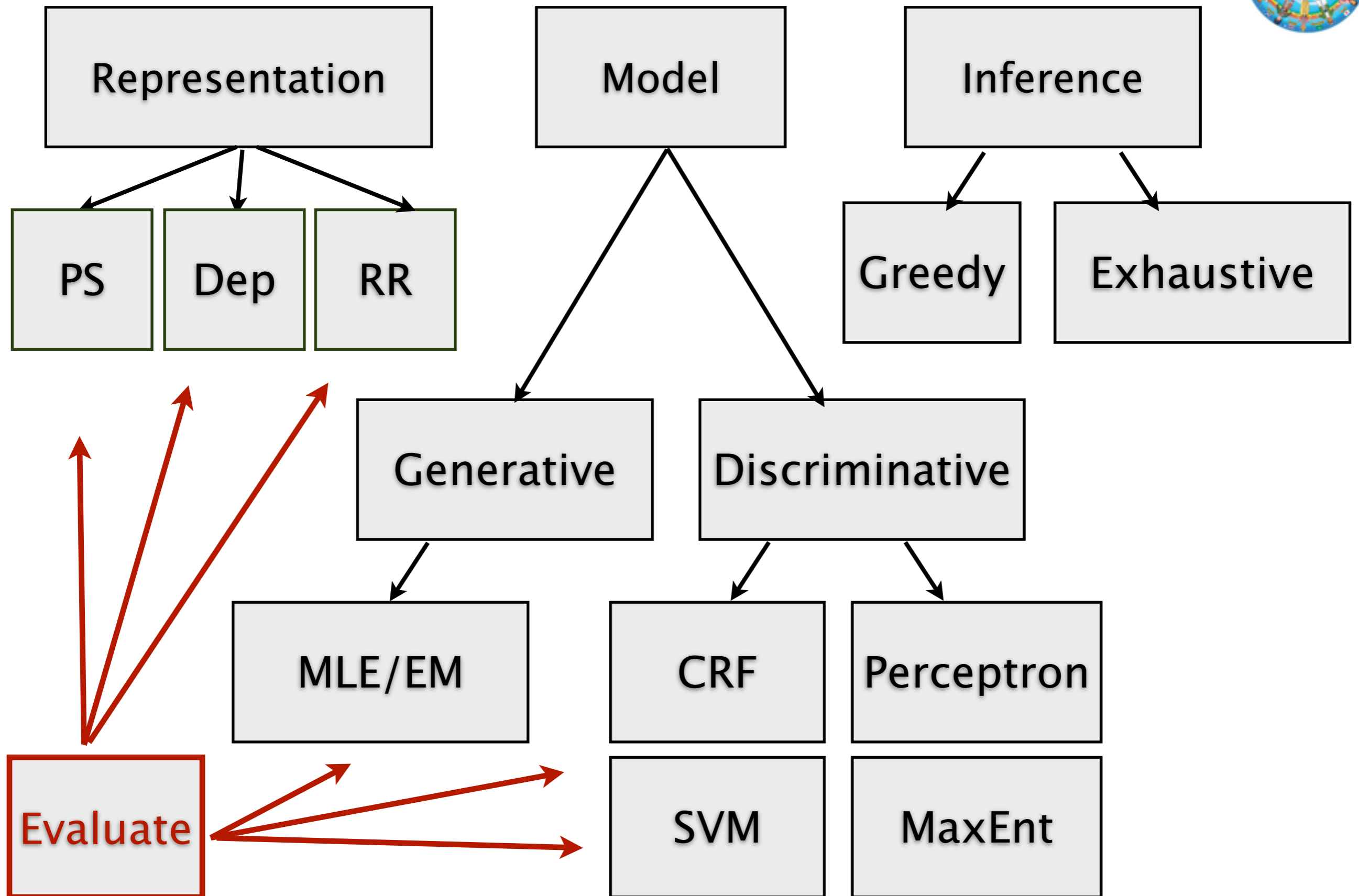
# -- Day 3 --



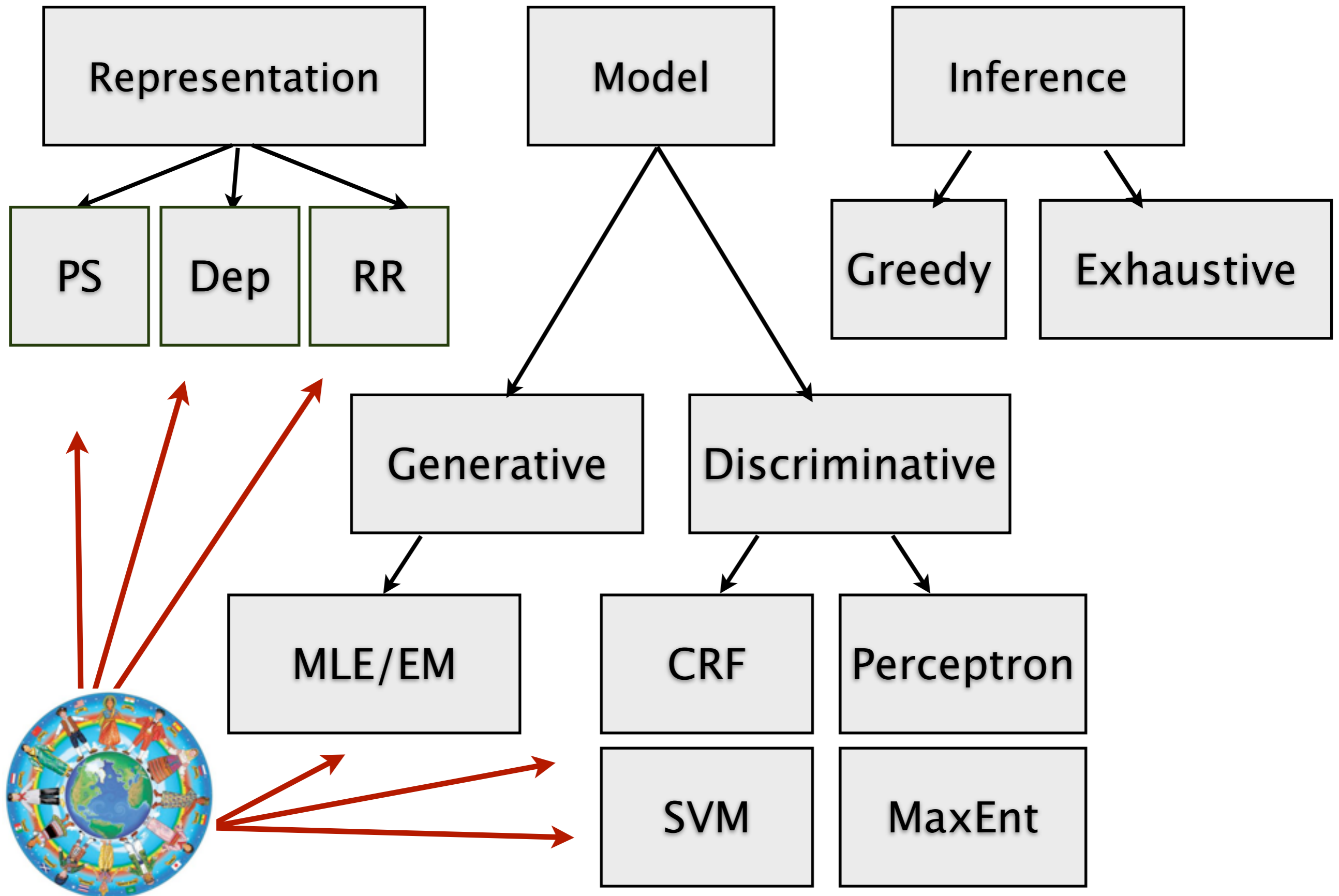
# -- Day 4 --



# -- Day 5 --



# -- Universal Parsing --



# Course Goal

- Day 1:
- Day 2:
- Day 3:
- Day 4:
- Day 5:

**Towards  
Universal  
Parsing?**

# Introduction#2

# Morphology

What's in a Word?

# What's a word? (1)

- The Unix command “wc -w” counts the words in a file.
- The Unix Command “wc -w” used spaces

# What's a word? (1)

- The Unix command “wc -w” counts the words in a file.

```
> cat example.txt
```

```
This company isn't New York-based anymore  
We moved to Chicago
```

```
> wc -w example.txt
```

```
10 example.txt
```

- The Unix Command “wc -w” used spaces

# What's a word? (1)

- The Unix command “wc -w” counts the words in a file.

```
> cat example.txt
```

```
This company isn't New York-based anymore  
We moved to Chicago
```

```
> wc -w example.txt
```

```
10 example.txt
```

- The Unix Command “wc -w” used spaces

```
This1 company2 isn't3 New4 York-based5 anymore6  
We7 moved8 to9 Chicago10
```

# What's a word? (1)

- A Problem: **Writing Conventions**

English: New York-based

Hebrew: הבית בבית ומהבית

- A Problem: **Contraction**

English: They're, doesn't, John's

Hebrew: ביתו שלחתינו שלנו

- A Problem: **Compounding**

English: Ice cream cone , New York

Hebrew: בית ספר , תל אביב

German: sommerschule

# What's a Word? (2)

“Of course he wants to take the  
advanced course too, he took  
two beginners' courses already”

# What's a Word? (2)

“Of course he wants to take the  
advanced course too, he took  
two beginners' courses already”

# What's a Word? (2)

“Of course he wants to take the  
advanced course too, he took  
two beginners' courses already”

# What's a word? (2)

“Of course he wants to take the  
advanced course too, he took  
two beginners' courses already”

# What's a word? (2)

- A **word-form** can have different **senses**
  - Sense1: of course
  - Sense2: the NLP course
- A **sense** can be in different **word-forms**
  - /TAKE/
  - take, takes, took, taking

# What's a word? (2)

- A **word-form** can have different **senses**
  - Sense1: of course
  - Sense2: the NLP course
- A **sense** can be in different **word-forms**
  - **/TAKE/** **Lemma**
  - take, takes, took, taking

# What's a word? (2)

- A **word-form** can have different **senses**
  - Sense1: of course
  - Sense2: the NLP course
- A **sense** can be in different **word-forms**
  - **/TAKE/** **Lemma**
  - take, takes, took, **taking** **Inflected-form**

# What's a word? (2)

- A **word-form** can have different **senses**
  - Sense1: of course
  - Sense2: the NLP course
- A **sense** can be in different **word-forms**
  - **/TAKE/** **Lemma**
  - take, takes, took, tak**ing** **morpheme**

# What's a morpheme?

- A minimal unit of form:function pairing
  - cat : NN
  - s : plural
  - do : VB
  - un : negation

# What's a morpheme?

- A minimal unit of form:function pairing

- cat : NN

- s : plural

Stems

- do : VB

- un : negation

# What's a morpheme?

- A minimal unit of form:function pairing

- cat : NN

- s : plural

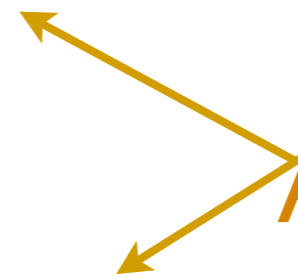
- do : VB

- un : negation

Suffixes

Affixes

Prefixes



# What's a morph?

- A morphological form without a function
  - s : plural
  - es : plural
  - null : plural
  - un : NEG
  - dis : NEG

# What's a morph?

- A morpheme may be realized by different morphs (and vice versa)
  - s : plural
  - es : plural
  - null : plural
  - eat : Verb
  - s : 3rd- singular

# What's a morph?

- A morpheme may be realized by different morphs (and vice versa)

- s : plural

- es : plural

- null : plural

- eat : Verb

- s : 3rd- singular

Allomorphs

# What's a property?

- A morpheme may be realized by different morphs (and vice versa)

- s : plural

Number=Plural

- es : plural

- null : plural

- eat : Verb

- s : 3rd- singular

Person=3rd

Number = Singular

# What's in a word? (3)

- Derivational Morphology



- Inflectional Morphology



- Concatenation/Compounding



# What's in a word? (3)

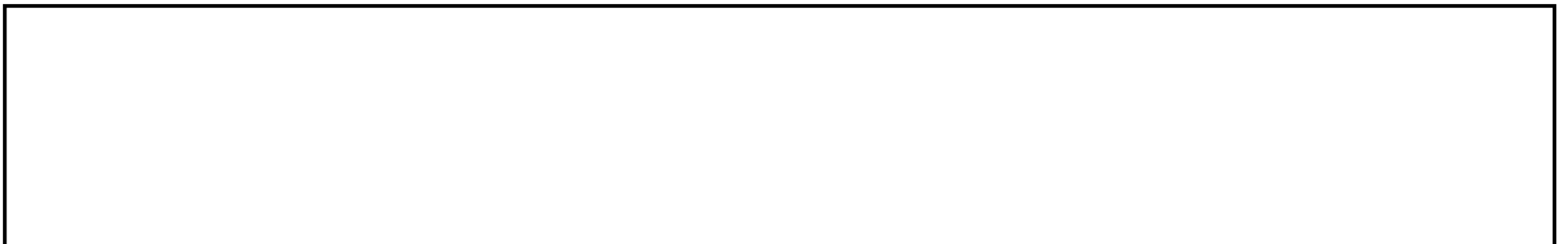
- Derivational Morphology

- Create lemmas (core meaning)
- available vs. availability

- Inflectional Morphology



- Concatenation/Compounding



# What's in a word? (3)

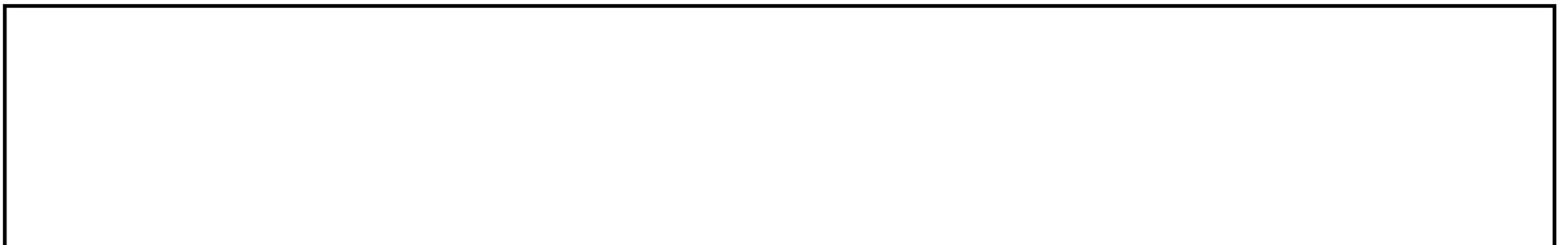
- Derivational Morphology

- Create lemmas (core meaning)
- available vs. availability

- Inflectional Morphology

- Create inflected word-forms
- /SEE/: see, seen, saw, seeing,

- Concatenation/Compounding



# What's in a word? (3)

- **Derivational Morphology**

- Create lemmas (core meaning)
- available vs. availability

- **Inflectional Morphology**

- Create inflected word-forms
- /SEE/: see, seen, saw, seeing,

- **Concatenation/Compounding**

- Join forms to a space-delimited token
- John's, aren't, she's, won't

# What's in a word? (3)

- Derivational Morphology



- Inflectional Morphology



- Concatenation/Compounding



# What's in a word? (3)

## ● Derivational Morphology

- לקח =  $\_\_\_ + \text{ל.ק.ח}$
- התלקח =  $\_\_\_ \text{הת} + \text{ל.ק.ח}$

## ● Inflectional Morphology

## ● Concatenation/Compounding

# What's in a word? (3)

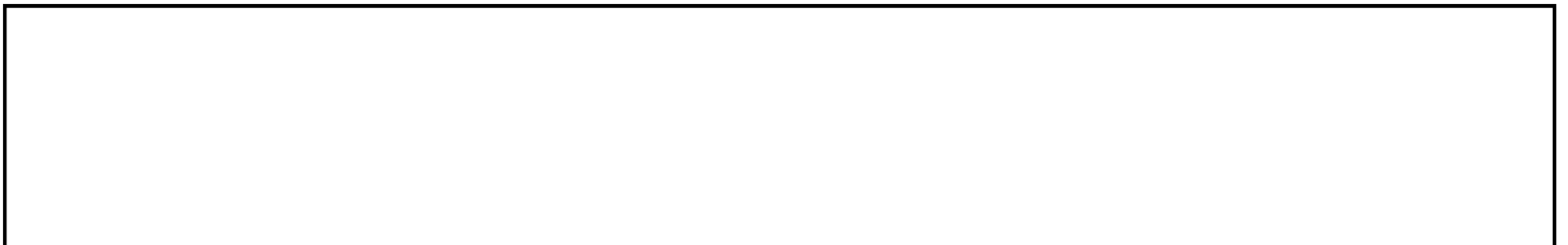
## ● Derivational Morphology

- לקח = \_ \_ \_ + ל.ק.ח
- התלקח = \_ \_ \_ + ה.ת.ל.ק.ח

## ● Inflectional Morphology

- /TAKE/ = לקחתי, יקח, תיקח
- /BURN/ = התלקחתי, יתלקח, תתלקח

## ● Concatenation/Compounding



# What's in a word? (3)

## ● Derivational Morphology

- ל.ק.ח + \_ \_ \_ = לקח
- ל.ק.ח + הת \_ \_ \_ = התלקח

## ● Inflectional Morphology

- /TAKE/ = לקחתי, יקח, תיקח
- /BURN/ = התלקחתי, יתלקח, תתלקח

## ● Concatenation/Compounding

- וכשמהבית = and when from the house
- Donaudampfschiffahrtsgesellschaftskapitän

# More Examples:

- Derivational Morphology

- unbelievably =

- un + believe + able + y

- Inflectional Morphology

- charming examples =

- charm + ing example + s

- Clitics/Compunds

- The king's unbelievably charming daughter

- [The king] + 's + [unbelievably..daughter]



Unbelievably  
charming  
examples

# More Examples:

- Turkish

- tan-ı<sup>2</sup>-tır-ıl-a-ma-dık-lar-ın-dan-dır.

- ‘It is because they cannot be introduced to each other.’

- Yu’pik (Central Alaska)

- qaya:liy’u:l’u:n’i

- ‘He was excellent at making kayaks’

- German

- lebensversicherungsgesellschaftsangestellter

- ‘Life insurance company employee’

# Morphological Paradigms

# Morphological Paradigms

/Lemma/VB

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past						
present						
gerund						
participial						

# Morphological Paradigms

/TAKE/VB

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past	took	took	took	took	took	took
present	take	take	takes	take	take	take
gerund	taking	taking	taking	taking	taking	taking
participial	taken	taken	taken	taken	taken	taken

# Morphological Paradigms

/TAKE/VB

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past	took	took	took	took	took	took
present	take	take	takes	take	take	take
gerund	taking	taking	taking	taking	taking	taking
participial	taken	taken	taken	taken	taken	taken

Syncretism

# Morphological Paradigms

/TAKE/ לקח

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past-feminine	לקחתי	לקחת	לקח	לקחנו	לקחתם	לקחו
Past-masculine	לקחתי	לקחת	לקחה	לקחנו	לקחתו	לקחו
future-feminine	אקח	תיקחי	תיקח	ניקח	תיקחו	יקחו
future-masculine	אקח	תיקח	ייקח	ניקח	תיקחנה	תיקחנה

Syncretism

# Morphological Paradigms

/TAKE/ לקח

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past-feminine	לקחתי	לקחת	לקח	לקחנו	לקחתם	לקחו
Past-masculine	לקחתי	לקחת	לקחה	לקחנו	לקחתו	לקחו
future-feminine	אקח	תיקחי	תיקח	ניקח	תיקחו	יקחו
future-masculine	אקח	תיקח	ייקח	ניקח	תיקחנה	תיקחנה

Syncretism

# Morphological Paradigms

/BURN/ הת+לקח

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past- feminine	התלקחתי	התלקחת	התלקחה	התלקחנו	התלקחתם	התלקחו
Past- masculine	התלקחתי	התלקחת	התלקח	התלקחנו	התלקחתו	התלקחו
future- feminine	אתלקח	תיתלקחי	תיתלקח	ניתלקח	תיתלקחו	תיתלקחנה
future- masculine	אתלקח	תיתלקח	ייתלקח	ניתלקח	תיתלקחנה	ייתלקחו

# Morphological Paradigms

/BURN/ התלקח

	1st Person	2nd Person	3rd Person	1st Person	2nd Person	3rd Person
	Singular	Singular	Singular	Plural	Plural	Plural
past-feminine	התלקחתי	התלקחת	התלקחה	התלקחנו	התלקחתם	התלקחו
Past-masculine	התלקחתי	התלקחת	התלקח	התלקחנו	התלקחתו	התלקחו
future-feminine	אתלקח	תיתלקחי	תיתלקח	ניתלקח	תיתלקחו	תיתלקחנה
future-masculine	אתלקח	תיתלקח	ייתלקח	ניתלקח	תיתלקחנה	ייתלקחו

Inflectional class

# Morphosyntactic Representation (MSR)

- In MRLs, instead of **POS** tags we'll refer to Morphosyntactic Representations (MSRs)
- An **MSR** contains:
  - A **Lemma** (Semantic information)
  - A **POS tag** (Syntactic information)
  - Set of **Features** (Attribute:Value pairs)
- We will organize MSRs in the lexicon of MRLs into morphological **paradigms**.

# Productivity

- Morphological processes are productive:
- They apply to new words all the time

# Productivity

- Morphological processes are productive:
- They apply to new words all the time

google, googling, googler, googlers,  
to google, to ungoogle, to misgoogle,  
googlification, googlified, ...

# Productivity

- Morphological processes are productive:
- They apply to new words all the time

google, googling, googler, googlers,  
to google, to ungoogle, to misgoogle,  
googlification, googlified, ...



# What's in a Word:

## Recap

- **Derivational Morphemes**
  - derive **lemmas** – express meaning
- **Inflectional Morphemes**
  - create **forms** – express properties
- **Clitics and Compounds**
  - contract **units** – express relations

# Introduction#3

Perspective on

**POS Tags**

MRL viewpoint

# Part-of-Speech Tags

- What are part of speech tags?
  - Exist since the Greek Grammarians (2–3 BC)
  - Express abstract grammatical categories
  - Signal grammatical features
  - Signaled in word morphology

# Part-of-Speech Tags

- What are part of speech tags?
  - Exist since the Greek Grammarians (2–3 BC)
  - Express abstract grammatical categories
  - Signal grammatical features
  - Signaled in word morphology

***The yinkish dripner blorked  
quastofically into the nindin  
with the pidibs.***

# Part-of-Speech Tags

Open Class (lexical, semantic)

Closed Class (Functional)

# Part-of-Speech Tags

Open Class (lexical, semantic)



Nouns

Closed Class (Functional)

# Part-of-Speech Tags

Open Class (lexical, semantic)

Nouns

Verbs

Closed Class (Functional)

# Part-of-Speech Tags

Open Class (lexical, semantic)

Nouns

Verbs

Adjectives

Closed Class (Functional)

# Part-of-Speech Tags

Open Class (lexical, semantic)

Nouns

Verbs

Adverbs

Adjectives

Closed Class (Functional)

# Part-of-Speech Tags

Open Class (lexical, semantic)

Nouns

Verbs

Adverbs

Adjectives

Numbers

Closed Class (Functional)

# Noun

A Noun is a person, a place or a thing

- A person
  - mother, boy, uncle, David
- A place
  - London, school, field, JFK
- A thing
  - Trees, point, rice, wikipedia

# Noun

A Noun is a person, a place or a thing

- A person
  - mother, boy, uncle, David
- A place Common vs. Proper
  - London, school, field, JFK
- A thing
  - Trees, point, rice, wikipedia

# Noun

A Noun is a person, a place or a thing

- A person

- mother, boy, uncle, David

- A place

- London, school, field, JFK

- A thing    Count vs. Mass

- Trees, point, rice, wikipedia

# Noun

A Noun is a person, a place or a thing

- A person **Feminine vs. Masculine**
  - mother, boy, uncle, David
- A place
  - London, school, field, JFK
- A thing
  - Trees, point, rice, wikipedia

# Noun

A Noun is a person, a place or a thing

- A person

- mother, boy, uncle, David

- A place

- London, school, field, JFK

- A thing Plural vs. Singular

- Trees, point, rice, wikipedia

# Verb

Verb is where it's all happening

- An Event

- When Harry met Sally.

- A Process

- The tree grows.

- An Activity

- It will be playing in Chicago.

- A State

- She knows everything.

# Verb

Verb is where it's all happening

- An Event

- When Harry met Sally.

- A Process

Singular+Past

- The tree grows.

- An Activity

- It will be playing in Chicago.

- A State

- She knows everything.

# Verb

Verb is where it's all happening

- An Event

- When Harry met Sally.

- A Process

- The tree grows.

- An Activity

Simple vs. Progressive

- It will be playing in Chicago.

- A State

- She knows everything.

# Adjectives

Adjectives describe people places and things

Who's afraid of  
the hairy scary biggish grayish bad wolf?

<http://www.youtube.com/watch?v=6Nuxeh4V1ng>

# Adjectives

Adjectives describe people places and things

- Adjectives

- Young, old

- Red, green, blue

- good, bad, awesome

- Comparatives

- Tall, taller, tal**est**

Superlatives

- Better, best, worse, wor**st**

<http://www.youtube.com/watch?v=6Nuxeh4V1ng>

# Adverbs

Adverbs modify verbs, adverbs, adjectives

Unfortunately, John walked home  
extremely slowly yesterday

<http://www.youtube.com/watch?v=x6dPHWemygY>

# Adverbs

Adverbs modify verbs, adverbs, adjectives

- Locatives

- Here, there, home, everywhere

- Temporal

- Yesterday, tomorrow, nowadays

- Manner

manner

- Quickly, accurately, gracefully

- Slowly, surely, definitely,

<http://www.youtube.com/watch?v=x6dPHWemygY>

# Numbers

Count and order people places and things

- Numerical expressions
  - 2, 1344, 2.14, 40,000, 500000
  - 1, eleven, twenty-seven, thousand
  - First, second, third, thirty-seven

# Part-of-Speech Tags

Open Class (lexical, semantic)

Nouns

Verbs

Adverbs

Adjective

Numbers

Closed Class (Functional)

# Pronouns

Pronouns take the place of a noun

- Personal Pronouns (nominative)
  - I, you, he, she, it, they
- Personal Pronouns (accusative)
  - me, you, him, her, it, them
- Possessive Pronouns (genitive)
  - my, your, his, her, its, their

<http://www.youtube.com/watch?v=GiPJs096bVE>

# Conjunctions

Hooking up words and phrases and clauses

- Nouns:
  - Milk or honey
- Verbs:
  - Erase and rewind
- Adjectives:
  - Sad but true
- Adverbs:

<http://www.youtube.com/watch?v=RPoBE-E8VOc>

# Interjection

Interjections show excitement or emotion

- Hey, Wow, Ahem, Oh, Oops, Darn, Duh, Yay, Ouch, Well, Phew!
- **Hey**, you're kinda cute.
- **Ouch**! that hurts.
- **Darn**, I'll never make it.
- I want a ticket, **uh**, just for one.

[http://www.youtube.com/watch?v=\\_e24kdjdbtw](http://www.youtube.com/watch?v=_e24kdjdbtw)

# Special Verbs

reflect state, modality, tense and aspect

- Copula

- am, is, are, was, were

- Auxiliaries

- will, have, been

- Modals

- should, would, could, ought

gender  
number  
tense  
aspect  
modality  
agency  
...

[http://www.youtube.com/watch?v=\\_e24kdjdbtw](http://www.youtube.com/watch?v=_e24kdjdbtw)

# Part-of-Speech Tags

## Open Class (lexical)

Nouns

Verbs

Adverbs

Adjective

Numbers

Pronouns

Interjections

Conjunctions

Auxiliaries  
Modals  
Copula

Punctuation

Determiners

Prepositions

Particles

## Closed Class (Functional)

# Tag Sets and Tagging Schemes

- The Google Corpus: 12 Tags
- The Brown Corpus: 87 Tags
- The Wall Street Journal Corpus: 45 Tags
- The British National Corpus: 146 Tags
- The Modern Hebrew Treebank: ~250 Tags

1.	AGR-gn	Agreement particle
2.	AT	Accusative marker
3.	AUX	Auxiliary verb
4.	CC	Coordinating conjunction
5.	CD-gn-(H)	Numeral (definite)
6.	CDT-gn-(H)	Numeral determiner (definite)
7.	COM	Complementizer
8.	DT	Determiner
9.	IN	Preposition
10.	JJ-gn-(H)	Adjective (definite)
11.	JJT-gn	Construct state adjective
12.	H	Definiteness marker
13.	HAM	
14.	MD-gnpt	
15.	MOD	
16.	NN-gn-(H)	
17.	NNG-gn-(H)	
18.	NNGT-gn	
19.	NNP-gn	
20.	NNT-gn	
21.	POS	
22.	PRP-gnp	
23.	QW	
24.	RB	
25.	RBR	
26.	REL	
27.	VB-gnpt	Verb, finite
28.	VB-M	Verb, infinite
29.	WDT-gn	Determiner question word
30.	ZVL	Garbage
31.	yy*	various symbols, see appendix A

## 27. VB-gnpt

g: gender  
n: number  
p: person  
t: tense

# Morphological Analysis/Tagging

- A **Part-of-Speech** tagger assigns each word possible POS tags.
- A **Morphological Analyzer** assigns to each word possible MSRs.
- POS/MSRs pass/share information between morphology and syntax
- Assignment is highly ambiguous

# Intro to PMRL (1)

- In MRL words may contain multiple morphemes:
  - **Derivational** – create paradigms
  - **Inflectional** – fill-in paradigms
  - **Clitics/Compounds** – conjoin MSRs
- Every word/paradigm-cell has an MSR:
  - at word-level : provide **abstraction**
  - at sentence-level : provide **specification**

# Intro to PMRL (2)

- MRLs are languages with rich, high-dimensional, morphological paradigms.
- In MRLs, every word-form is a paradigm cell, represented as a rich MSR.
- In MRLs, every word-token may consist of multiple MSRs.
- In MRLs, MSR assignment is ambiguous.

# Intro to PMRL (3)

- How to resolve challenges caused by rich morphology, for better parsing.



- How to exploit information provided by rich morphology, for better parsing.



UPPSALA  
UNIVERSITET



מכון  
ויצמן  
למדע

# Questions?

[reut.tsarfaty@weizmann.ac.il](mailto:reut.tsarfaty@weizmann.ac.il)

# Homework

- Define each application on slide 10 as a structure prediction task. What is the function? what is the input? what is the output?
- Define each of the following terms: Morpheme, Morph, Allomorph, Stem, Affix, Lemma, Feature, Gender, Paradigm, Syncretism, MSR.
- Write rich MSRs for the following words: Jack, snowing, apples, sings, a, the, could, his, yesterday, John's, won't. Hint: in some cases you may need more than one MSR per word.
- Find three examples of syncretism in three languages. Why might it challenge parsing?





UPPSALA  
UNIVERSITET



מכון  
ויצמן  
למדע

*Thanks*

[reut.tsarfaty@weizmann.ac.il](mailto:reut.tsarfaty@weizmann.ac.il)