A very short introduction to Language Technologies and Natural Language Processing

Jose F Quesada & Jose Luis Pro

Language

Set of conventional spoken or written symbols used for commucation between entities.

So we can see a language as the linking between meaning (semantic side) and expression (syntantic side).

Types of languages

- **Natural languages:** Used for the communication between human beings.
- **Formal languages:** Used by computers and in mathematical areas.

Language

Set of conventional spoken or written symbols used for commucation between entities.
So we can see a language as the linking between meaning (semantic side) and expression (syntantic side).

Types of languages

- **Natural languages:** Used for the communication between human beings.
- **Formal languages:** Used by computers and in mathematical areas.

Language

Set of conventional spoken or written symbols used for commucation between entities.
So we can see a language as the linking between meaning (semantic side) and expression (syntantic side).

Types of languages

- **Natural languages:** Used for the communication between human beings.
- **Formal languages:** Used by computers and in mathematical areas.

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.
2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.
2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.
2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Differences between natural and formal languages:

1. Computers don't understand natural languages, (normal) humans don't understand computer languages.
2. Formal languages shouldn't have ambiguities, but natural languages do have.

Language Technologies (LT's) are the set of technologies that aim to create software that has some kind of knowledge about natural languages.

Natural language processing (NLP) is the scientific field concerned with the interactions between computers and human by means of natural languages.

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. **Dialogue Systems.**

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. **Dialogue Systems.**

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. **Dialogue Systems.**

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. **Dialogue Systems.**

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. **Dialogue Systems.**

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. Dialogue Systems.

Applications of Language Technologies:

1. Machine Translation.
2. Question - Answering.
3. Automatic Text Classification.
4. Automatic Text Summarization.
5. Social Analytics.
6. Sentiment Analysis.
7. **Dialogue Systems.**

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.
2. Transactional agents.
3. Educational and learning tutoring.

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.
2. Transactional agents.
3. Educational and learning tutoring.

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.
2. Transactional agents.
3. Educational and learning tutoring.

Dialogue Systems (written or spoken) are also known as:

- Conversational interfaces.
- Chatbots.

Applications of Dialogue Systems:

1. Information providers.
2. Transactional agents.
3. Educational and learning tutoring.

Dialogue systems main issue

The most difficult challenge in the design of conversational interfaces are related with the highly ambigous nature of spoken languages.

Example

```
Peter come yesterday.
Yesterday Peter come.
```

Two syntatic expressions $\Longleftrightarrow$ One semantic form

Dialogue systems main issue

The most difficult challenge in the design of conversational interfaces are related with the highly ambigous nature of spoken languages.

Example

```
Peter come yesterday.
Yesterday Peter come.
```

Two syntatic expressions $\Longleftrightarrow$ One semantic form

Dialogue systems main issue

The most difficult challenge in the design of conversational interfaces are related with the highly ambigous nature of spoken languages.

Example

`Peter come yesterday.`
`Yesterday Peter come.`

Two syntatic expressions $\Longleftrightarrow$ One semantic form

Example

`Peter said John came yesterday.`

- Was it yesterday when Peter said that?
- Was it yesterday when John came?

One syntatic expression $\Longleftrightarrow$ Two semantic forms

Example

`Peter said John came yesterday.`

- Was it yesterday when Peter said that?
- Was it yesterday when John came?

One syntatic expression $\Longleftrightarrow$ Two semantic forms

Example

`Peter said John came yesterday.`

- Was it yesterday when Peter said that?
- Was it yesterday when John came?

One syntatic expression $\Longleftrightarrow$ Two semantic forms

Humans can deal with these ambiguities applying what is called "psicolinguistic preferences" and, of course, logic and common-sense reasoning:

Example

```
Peter said John will come yesterday.
```

From the computer point of view this sentence is such ambigous like previous one but humans know that nobody *"will come yesterday"*.

Humans can deal with these ambiguities applying what is called "psicolinguistic preferences" and, of course, logic and common-sense reasoning:

Example

```
Peter said John will come yesterday.
```

From the computer point of view this sentence is such ambigous like previous one but humans know that nobody *"will come yesterday"*.

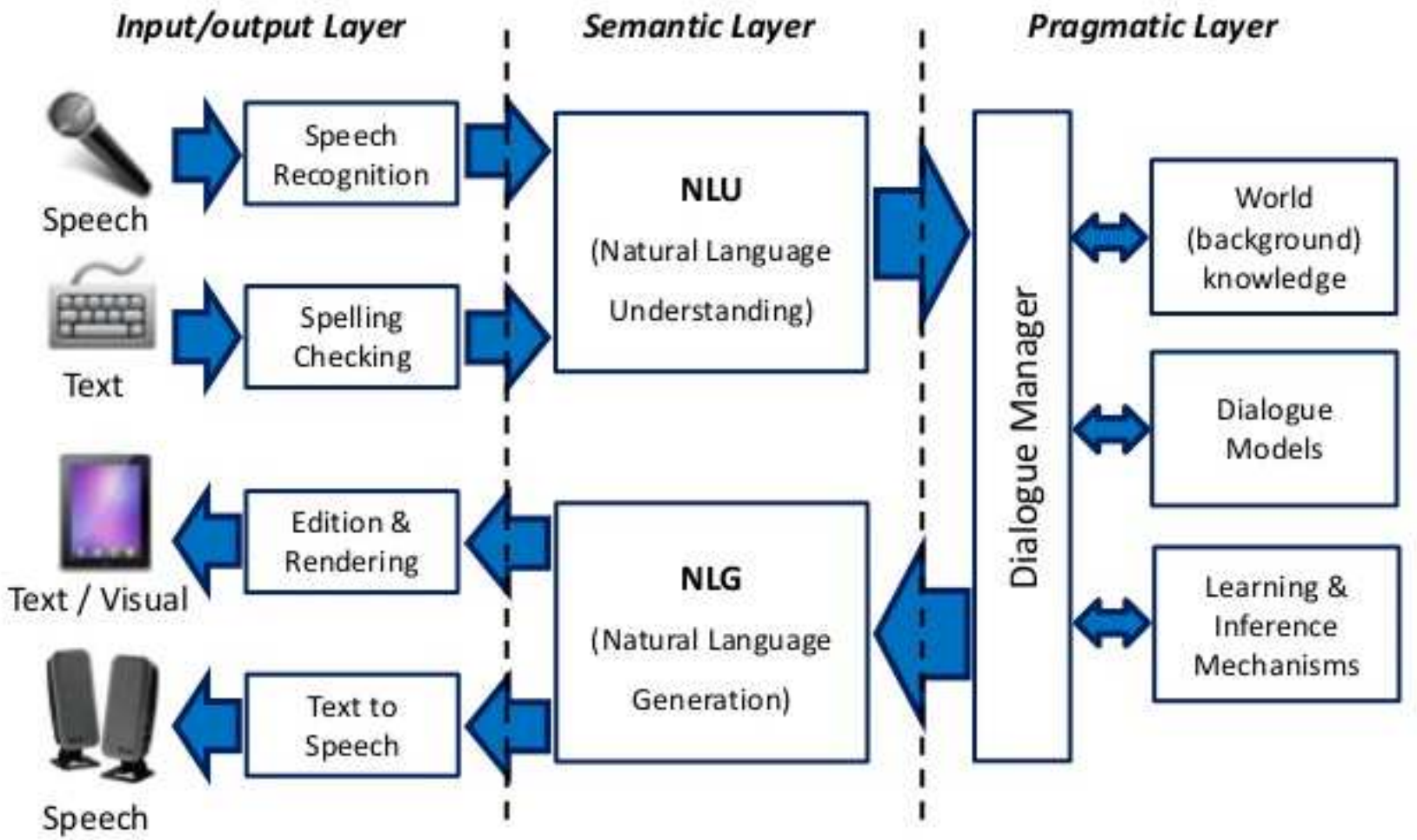
Humans can deal with these ambiguities applying what is called "psicolinguistic preferences" and, of course, logic and common-sense reasoning:

Example

`Peter said John will come yesterday.`

From the computer point of view this sentence is such ambigous like previous one but humans know that nobody *"will come yesterday"*.

Dialogue System architecture



Picture from Institute for Infocomm Research (Singapore)

Natural Language Understanding (NLU)

NLU main goal

The goal of NLU stage is to transform an input string (let's say *user proference*) in an abstract representation of its meaning easier for computer programs to manipulate it, in order to execute some kind of reasoning.

There is a wide variety of possible meaning representations.

- Topic maps.
- Concepts maps.
- Mind maps.
- Onthologies.
- **Feature structures.**

Natural Language Understanding (NLU)

NLU main goal

The goal of NLU stage is to transform an input string (let's say *user proference*) in an abstract representation of its meaning easier for computer programs to manipulate it, in order to execute some kind of reasoning.

There is a wide variety of possible meaning representations.

- Topic maps.
- Concepts maps.
- Mind maps.
- Onthologies.
- **Feature structures.**

Example

$$\texttt{John came yesterday} \longrightarrow \begin{bmatrix} \textsc{Subject:} & \texttt{John} \\ \textsc{Action:} & \texttt{come} \\ \textsc{Tense:} & \texttt{past} \\ \textsc{Offsetdate:} & \texttt{-1 day} \end{bmatrix}$$

Example

$$\texttt{John will talk in two days} \longrightarrow \begin{bmatrix} \textsc{Subject:} & \texttt{John} \\ \textsc{Action:} & \texttt{talk} \\ \textsc{Tense:} & \texttt{future} \\ \textsc{Offsetdate:} & \texttt{+2 day} \end{bmatrix}$$

Feature structures

- A feature structure is a set of features.
- With no particular order between them.
- Every feature may have (but it's not required) an associated value.
- The value associated to every feature can be **atomic** or **complex**.

$$\texttt{comes} \longrightarrow \begin{bmatrix} \textsc{Action}: & \texttt{come} \\ \textsc{Tense}: & \texttt{present} \\ \textsc{Agreement}: & \begin{bmatrix} \textsc{Number}: & \texttt{singular} \\ \textsc{Person}: & 3 \end{bmatrix} \end{bmatrix}$$

Feature structures

- A feature structure is a set of features.
- With no particular order between them.
- Every feature may have (but it's not required) an associated value.
- The value associated to every feature can be **atomic** or **complex**.

$$\texttt{comes} \longrightarrow \begin{bmatrix} \textsc{Action}: & \texttt{come} \\ \textsc{Tense}: & \texttt{present} \\ \textsc{Agreement}: & \begin{bmatrix} \textsc{Number}: & \texttt{singular} \\ \textsc{Person}: & 3 \end{bmatrix} \end{bmatrix}$$

NLU components

Trying to convert user proference to feature structures is not trivial. So we need to divide the process in some functional modules:

- Tokenization.
- Speller checker.
- Part Of Speech tagging (POS tagging).
- Parsing.
- Unifier.

NLU components

Trying to convert user proference to feature structures is not trivial. So we need to divide the process in some functional modules:

- Tokenization.
- Speller checker.
- Part Of Speech tagging (POS tagging).
- Parsing.
- Unifier.

Tokenization

Goal

Convert a sequence of characters into a sequence of tokens.

We must take into account:

- Separators: (␣-_)
- Punctuation marks: (, . ; : ! ?)
- Special symbols: ($€%?)
- Numbers and its own separators: (1234, .)
- Alphanumeric codes: (ES772024···)

Example

	tk1		tk2		tk3		tk4		tk5		tk6	
·	The	·	dog	·	is	·	in	·	the	·	park	·
0		1		2		3		4		5		6

Tokenization

Goal

Convert a sequence of characters into a sequence of tokens.

We must take into account:

- Separators: (␣-_)
- Punctuation marks: (,.;:!?)
- Special symbols: ($€%?)
- Numbers and its own separators: (1234,.)
- Alphanumeric codes: (ES772024···)

Example

	tk1		tk2		tk3		tk4		tk5		tk6	
·	The	·	dog	·	is	·	in	·	the	·	park	·
0		1		2		3		4		5		6

Tokenization

Goal

Convert a sequence of characters into a sequence of tokens.

We must take into account:

- Separators: (␣-_)
- Punctuation marks: (,.;:!?)
- Special symbols: ($€%?)
- Numbers and its own separators: (1234,.)
- Alphanumeric codes: (ES772024···)

Example

	tk1		tk2		tk3		tk4		tk5		tk6	
·	The	·	dog	·	is	·	in	·	the	·	park	·
0		1		2		3		4		5		6

Speller checker (only in written dialogue systems)

London

- Insertion: `Loondon`
- Deletion: `Lndon`
- Substitution: `Lpndon`
- Switching: `Lonodn`
- Bad separators: `Lon don`

Part Of Speech tagging (POS tagging)

Goal

To mark up lexical items with some lexical category depending on its definition and the context.

In natural language we can have some common lexical categories:

- Determiners: `a, the`
- Nouns: `London, dog`
- Pronouns: `you, me`
- Prepositions: `to, for`
- Adjectives: `blue, long`

Part Of Speech tagging (POS tagging)

Goal

To mark up lexical items with some lexical category depending on its definition and the context.

In natural language we can have some common lexical categories:

- Determiners: `a, the`
- Nouns: `London, dog`
- Pronouns: `you, me`
- Prepositions: `to, for`
- Adjectives: `blue, long`

POS tagging

So in the lexicon definition we can classify lexical items into categories:

- `("the", det)`
- `("dog", noun)`
- `("me", pronoun)`
- `("to", preposition)`

But in natural languages we can have several lexical categories corresponding to a single lexical item (especially in little inflectional ones, like in english):

- `("plans", noun)` $\longrightarrow$ plural of plan.
- `("plans", verb)` $\longrightarrow$ present of third person (singular) of verb to plan.

POS tagging

So in the lexicon definition we can classify lexical items into categories:

- `("the", det)`
- `("dog", noun)`
- `("me", pronoun)`
- `("to", preposition)`

But in natural languages we can have several lexical categories corresponding to a single lexical item (especially in little inflectional ones, like in english):

- `("plans", noun)` $\longrightarrow$ plural of plan.
- `("plans", verb)` $\longrightarrow$ present of third person (singular) of verb to plan.

POS tagging: Garden path problem

Example

The	government	plans	to	raise	taxes	...
det	noun	**verb**	prep	verb	noun	

Example

The	government	plans	to	raise	taxes	were	defeated
det	noun	**noun**	prep	verb	noun	verb	adj

POS tagging: Garden path problem

Example

The	government	plans	to	raise	taxes	...
det	noun	**verb**	prep	verb	noun	

Example

The	government	plans	to	raise	taxes	were	defeated
det	noun	**noun**	prep	verb	noun	verb	adj

Formal grammars

- Only study the purely syntactical aspects of language.
- **Alphabet:** Set of terminal symbols. $\{a, b\}$
- **Sentences:** Strings of symbols. $a, b, aa, ab, bb, ba, aaa, \cdots$
- **Language:** Set of sentences. $L = \{ab, aabb, aaabbb\}$
- **Formal grammar:** Set of formation rules used to define a language. $\{S \to aA, A \to bA, S \to \varepsilon \cdots\}$ Where S, A are non-terminal symbols.
- Depending on the syntax of production rules, grammars can be classified (Noam Chomsky, 1956).

Chomsky grammar hierarchy

Type	Grammar accepted	Rules	Observations
Type 0	Unrestricted grammar	$X \to Y$	$X, Y \in (N \cup T)^*$
Type 1	Context-sensitive grammar	$\alpha X \beta \to \alpha a \beta$	$X \in N$ $\alpha, a, \beta \in (N \cup T)^*$
Type 2	Context-free grammar	$X \to a$	$X \in N$ $a \in (N \cup T)^*$
Type 3	Regular grammar	$X \to a$ $X \to aY$	$X, Y \in N$ $X \in T$

Where:

- N is the set of non-terminal symbols.
- T is the set of terminal symbols.
- S^* is a string of elements in set S.

Easy context-free grammar example: $a^n b^m$

Consider:

- **Alphabet:** $\{a, b\}$
- **Language:** $L = \{a^n b^m\}\ n, m \geq 1$ (i.e. all strings with at least one "a" followed by at least one "b").
- **Grammar:**

$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

And let's see that a grammar defined with such production rules can be used to **generate** or either **recognize** the target language.

Generating $a^n b^m$ sentences

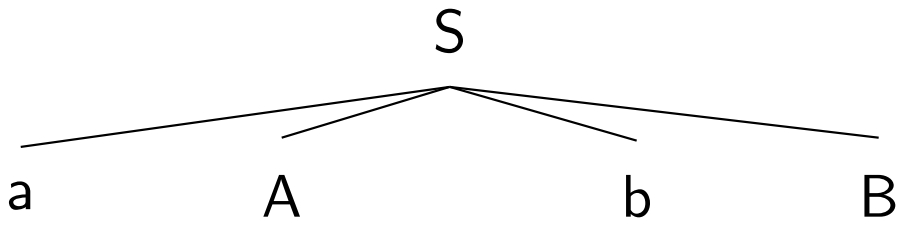
S

$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences

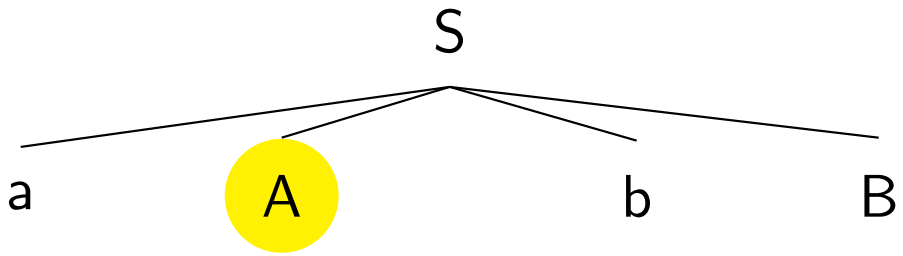
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



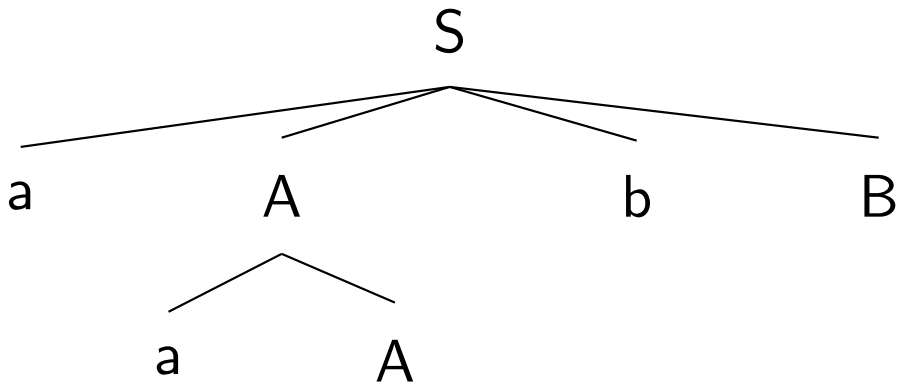
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



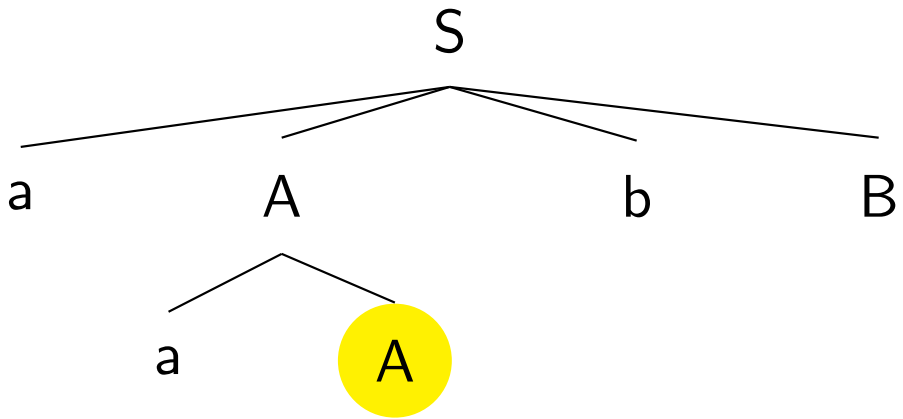
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



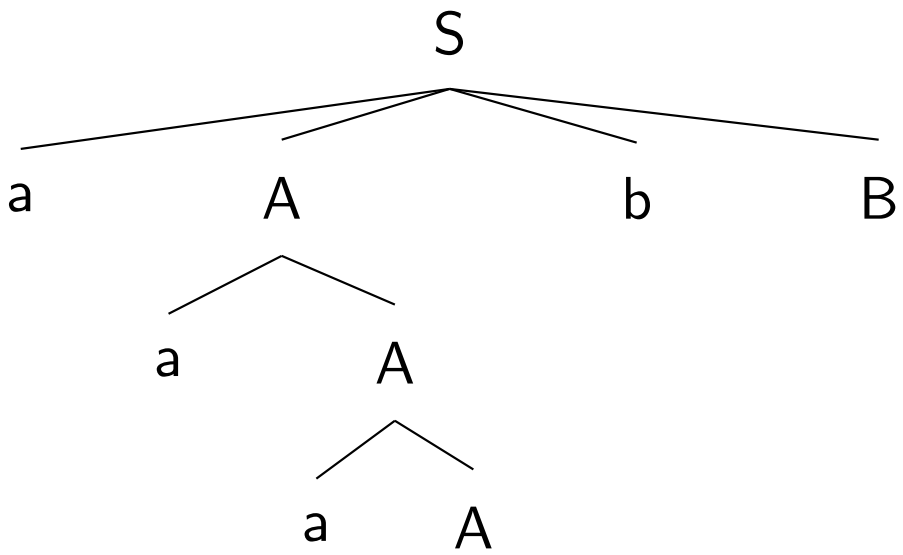
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



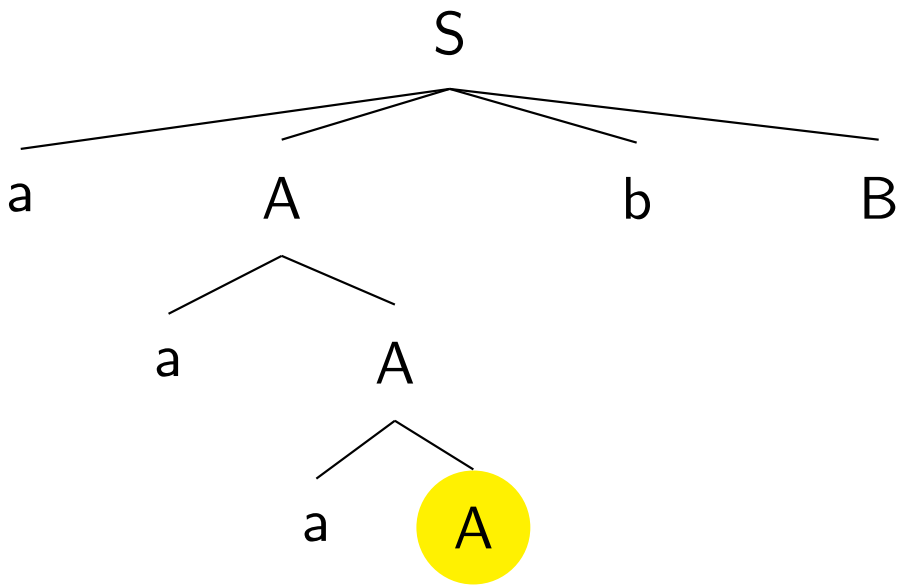
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



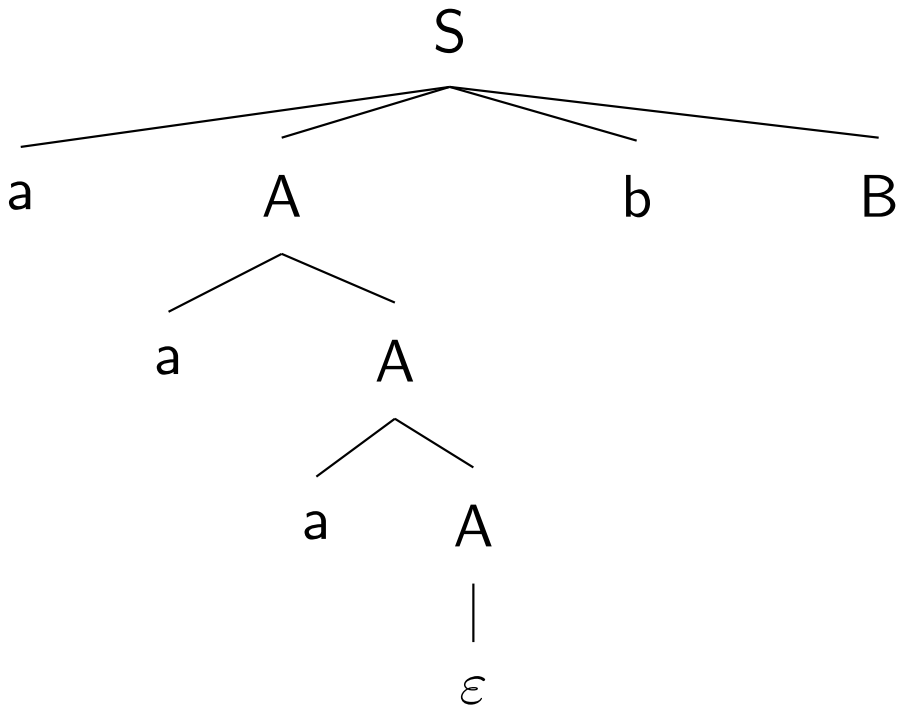
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



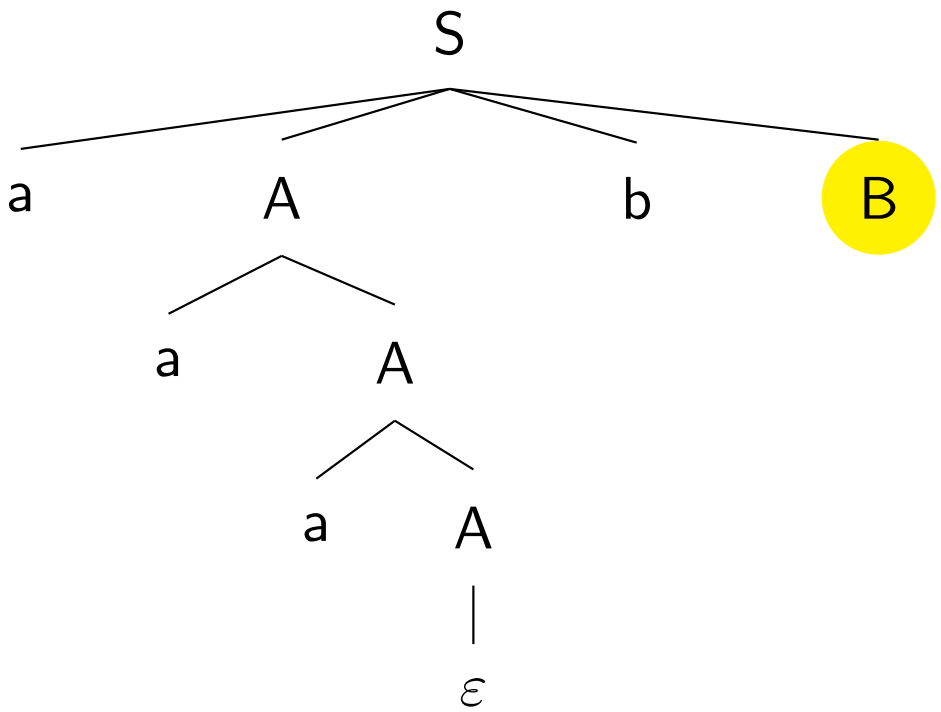
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



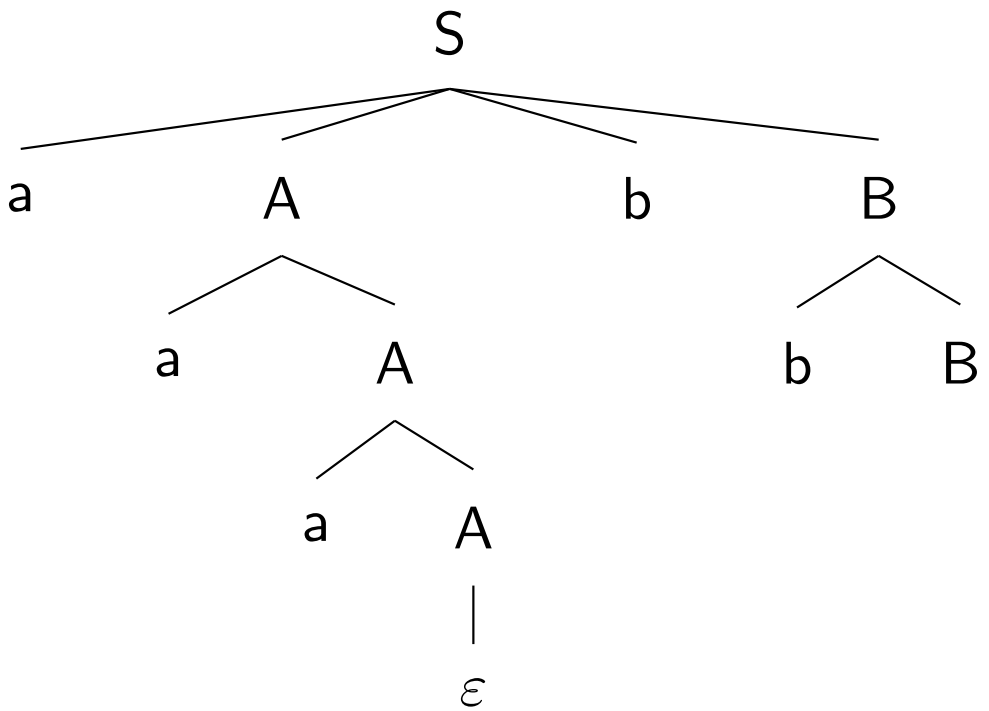
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



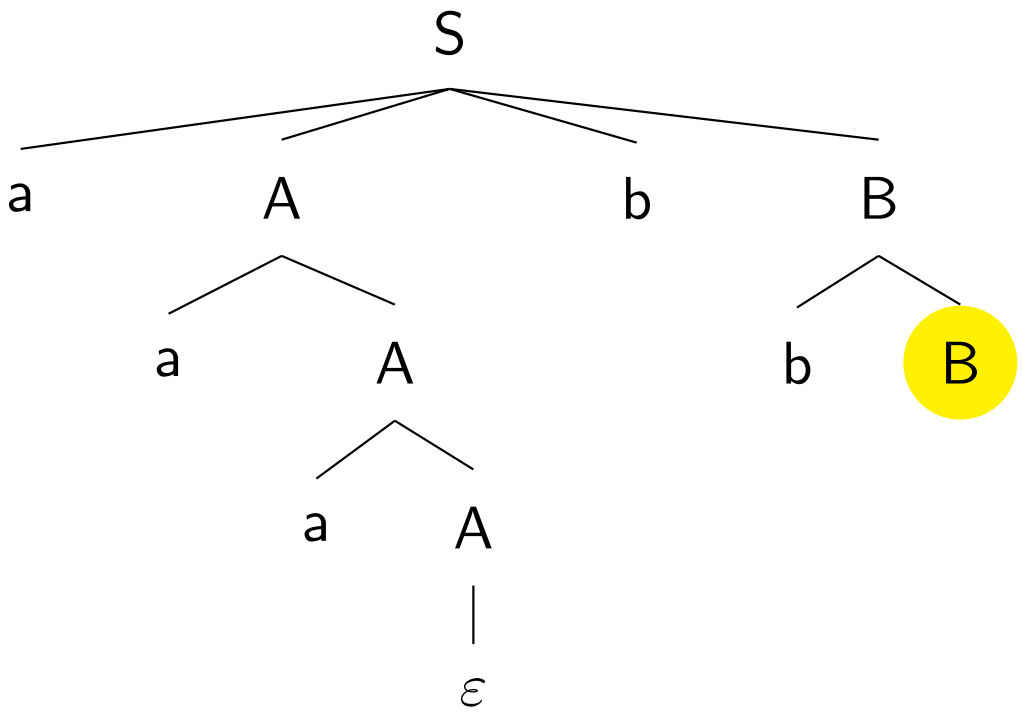
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



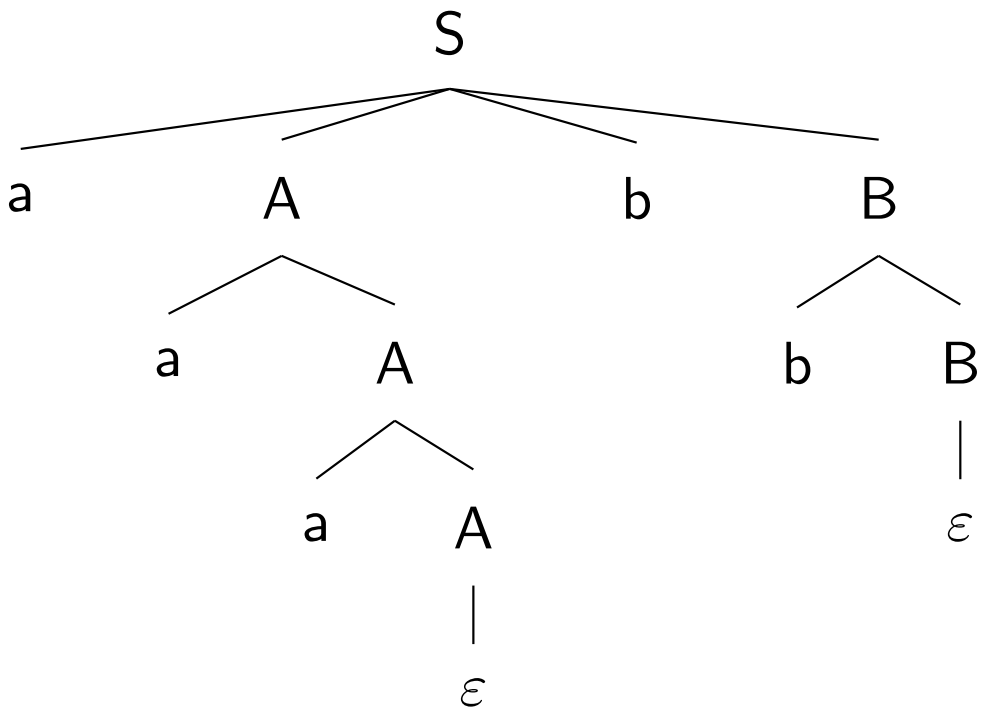
$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Generating $a^n b^m$ sentences



$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

So we have generated $aaabb$ sentence!



$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Recognizing $aaabb$ sentence

a

a

a

b

b

$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Recognizing $aaabb$ sentence

a

a

a

ε

b

b

ε

$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Recognizing $aaabb$ sentence

a

a

a

ε

b

b

ε

$S \rightarrow a\ A\ b\ B$

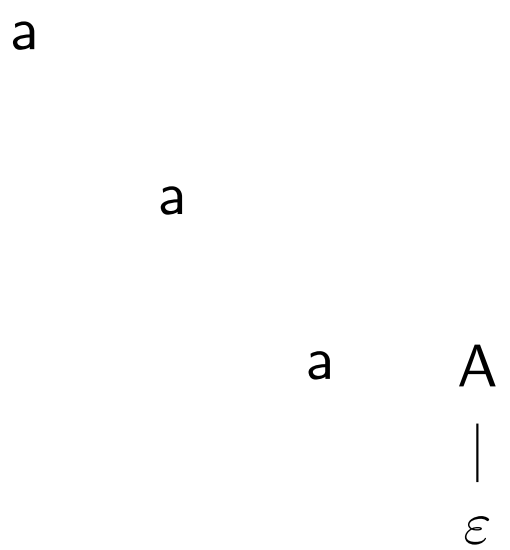
$A \rightarrow \varepsilon$

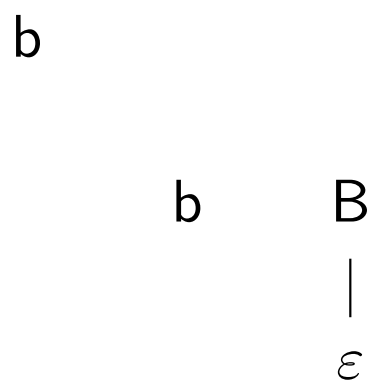
$A \rightarrow a\ A$

$B \rightarrow \varepsilon$

$B \rightarrow b\ B$

Recognizing $aaabb$ sentence



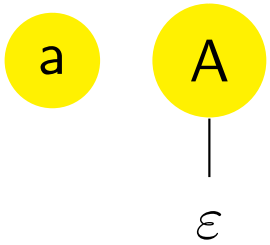


$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

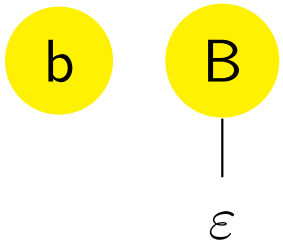
Recognizing $aaabb$ sentence

a

a

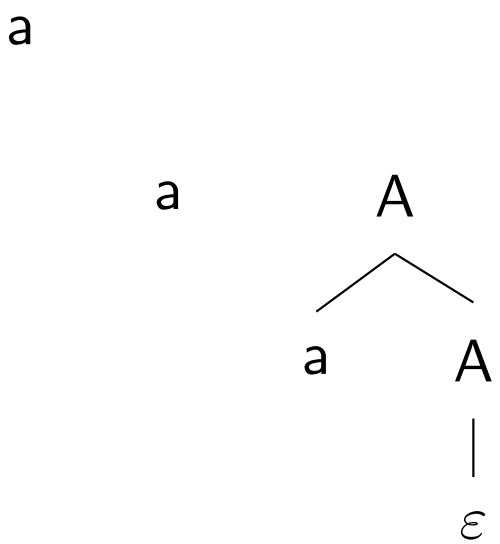


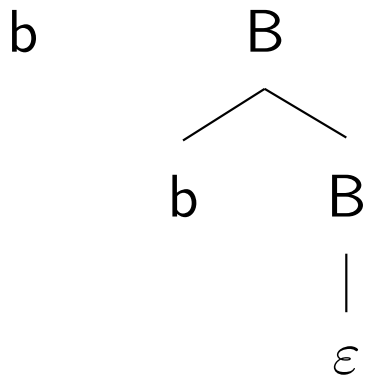
b



$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

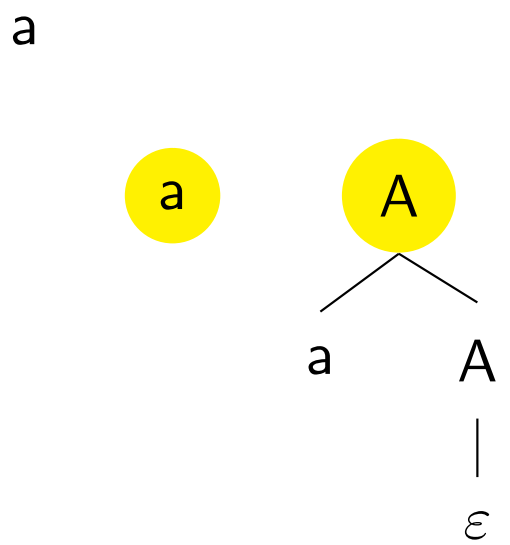
Recognizing $aaabb$ sentence

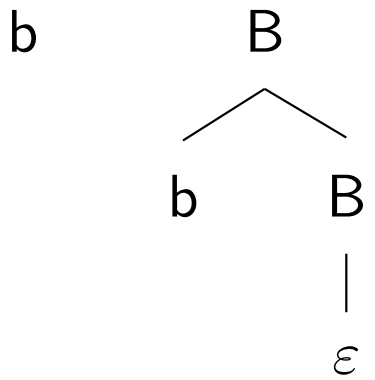




$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

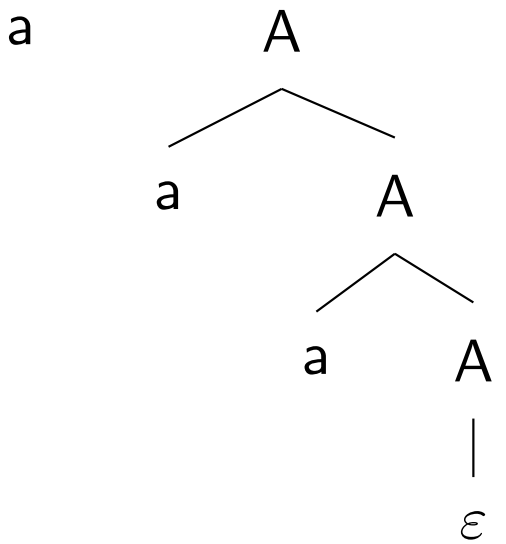
Recognizing $aaabb$ sentence

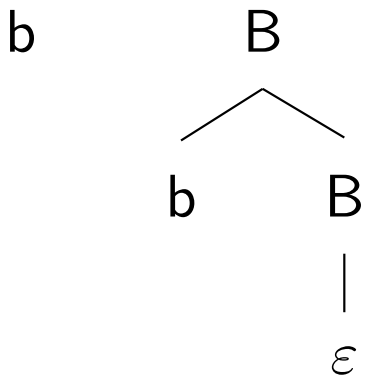




$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

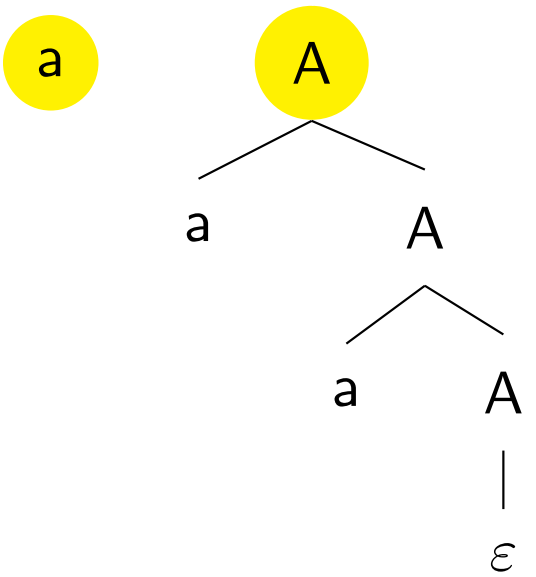
Recognizing $aaabb$ sentence

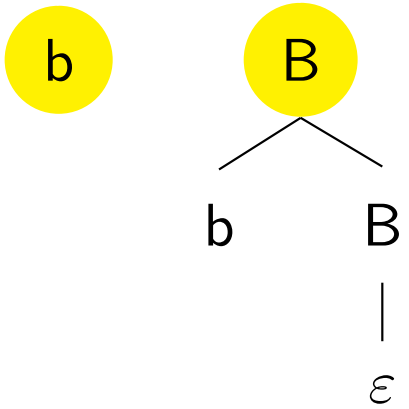




$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

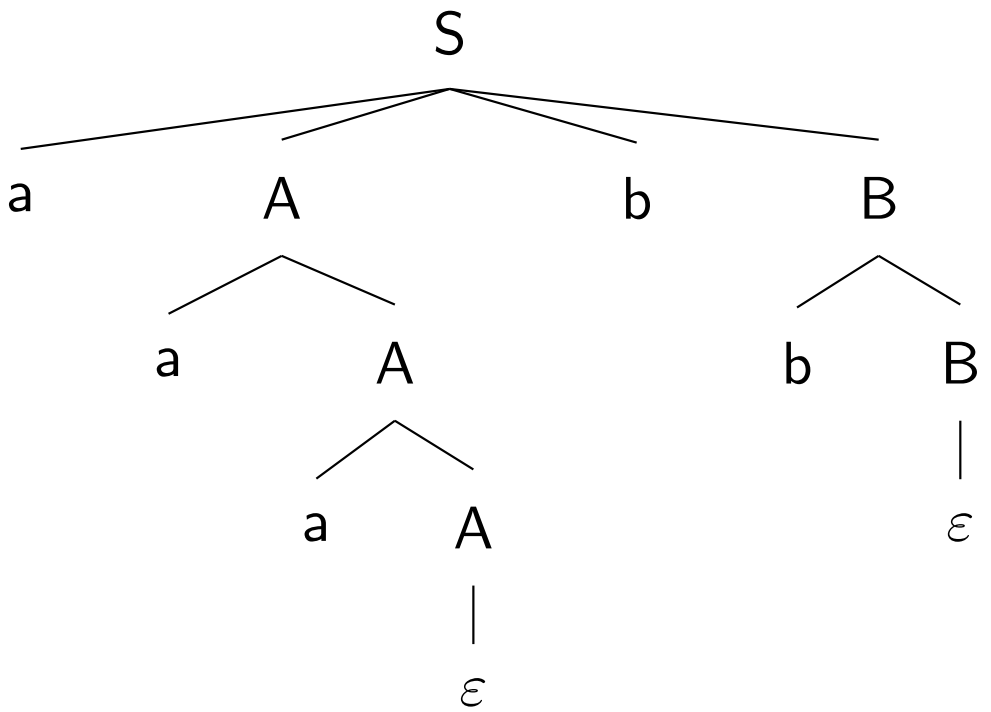
Recognizing $aaabb$ sentence





$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

We have recognized $aaabb$ sentence!



$$S \rightarrow a\ A\ b\ B$$
$$A \rightarrow \varepsilon$$
$$A \rightarrow a\ A$$
$$B \rightarrow \varepsilon$$
$$B \rightarrow b\ B$$

Parsing

- Is the process of analysing a sentence, according to the rules of a certain grammar, resulting in the creation of a "parse tree".
- That parse tree includes syntactic relations between components in the sentence but may also contains semantic information (e.g. a feature structure).

Parsing

- Is the process of analysing a sentence, according to the rules of a certain grammar, resulting in the creation of a "parse tree".
- That parse tree includes syntactic relations between components in the sentence but may also contains semantic information (e.g. a feature structure).

Parsing

- Is the process of analysing a sentence, according to the rules of a certain grammar, resulting in the creation of a "parse tree".
- That parse tree includes syntactic relations between components in the sentence but may also contains semantic information (e.g. a feature structure).

Parsing applied to natural languages (english)

Some typical lexical categories:

- `det`: Determiners.
- `noun`: Nouns.
- `verb`: Verbs.
- `prep`: Prepositions.

Some typical non-terminal symbols:

- `S`: Sentence.
- `NP`: Noun Phrase.
- `VP`: Verb Phrase.
- `PP`: Prepositional Phrase.

Some typical production rules

- `S` $\rightarrow$ `NP VP`
- `NP` $\rightarrow$ `det noun`
- `VP` $\rightarrow$ `verb`
- `VP` $\rightarrow$ `verb NP`
- `PP` $\rightarrow$ `prep NP`
- `NP` $\rightarrow$ `NP PP`
- `VP` $\rightarrow$ `VP PP`

Parsing applied to natural languages (english)

Some typical lexical categories:

- `det`: Determiners.
- `noun`: Nouns.
- `verb`: Verbs.
- `prep`: Prepositions.

Some typical non-terminal symbols:

- `S`: Sentence.
- `NP`: Noun Phrase.
- `VP`: Verb Phrase.
- `PP`: Prepositional Phrase.

Some typical production rules

- $\texttt{S} \rightarrow \texttt{NP VP}$
- $\texttt{NP} \rightarrow \texttt{det noun}$
- $\texttt{VP} \rightarrow \texttt{verb}$
- $\texttt{VP} \rightarrow \texttt{verb NP}$
- $\texttt{PP} \rightarrow \texttt{prep NP}$
- $\texttt{NP} \rightarrow \texttt{NP PP}$
- $\texttt{VP} \rightarrow \texttt{VP PP}$

Parsing applied to natural languages (english)

Some typical lexical categories:

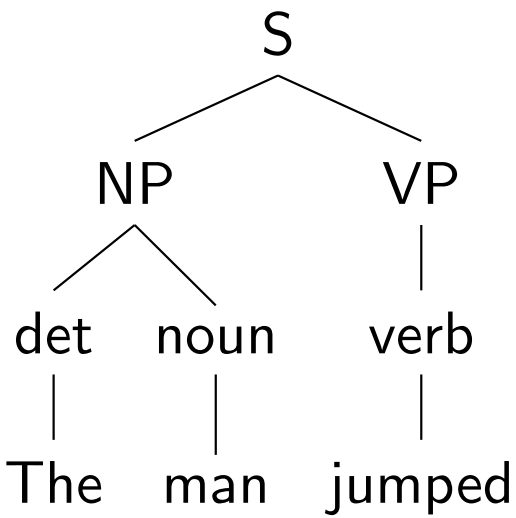
- `det`: Determiners.
- `noun`: Nouns.
- `verb`: Verbs.
- `prep`: Prepositions.

Some typical non-terminal symbols:

- `S`: Sentence.
- `NP`: Noun Phrase.
- `VP`: Verb Phrase.
- `PP`: Prepositional Phrase.

Some typical production rules

- $\texttt{S} \rightarrow \texttt{NP VP}$
- $\texttt{NP} \rightarrow \texttt{det noun}$
- $\texttt{VP} \rightarrow \texttt{verb}$
- $\texttt{VP} \rightarrow \texttt{verb NP}$
- $\texttt{PP} \rightarrow \texttt{prep NP}$
- $\texttt{NP} \rightarrow \texttt{NP PP}$
- $\texttt{VP} \rightarrow \texttt{VP PP}$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

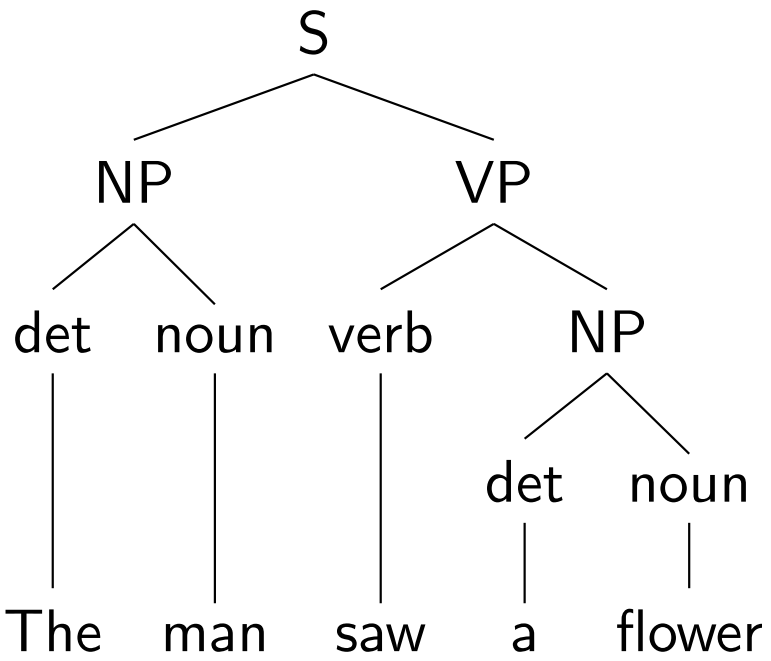
$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

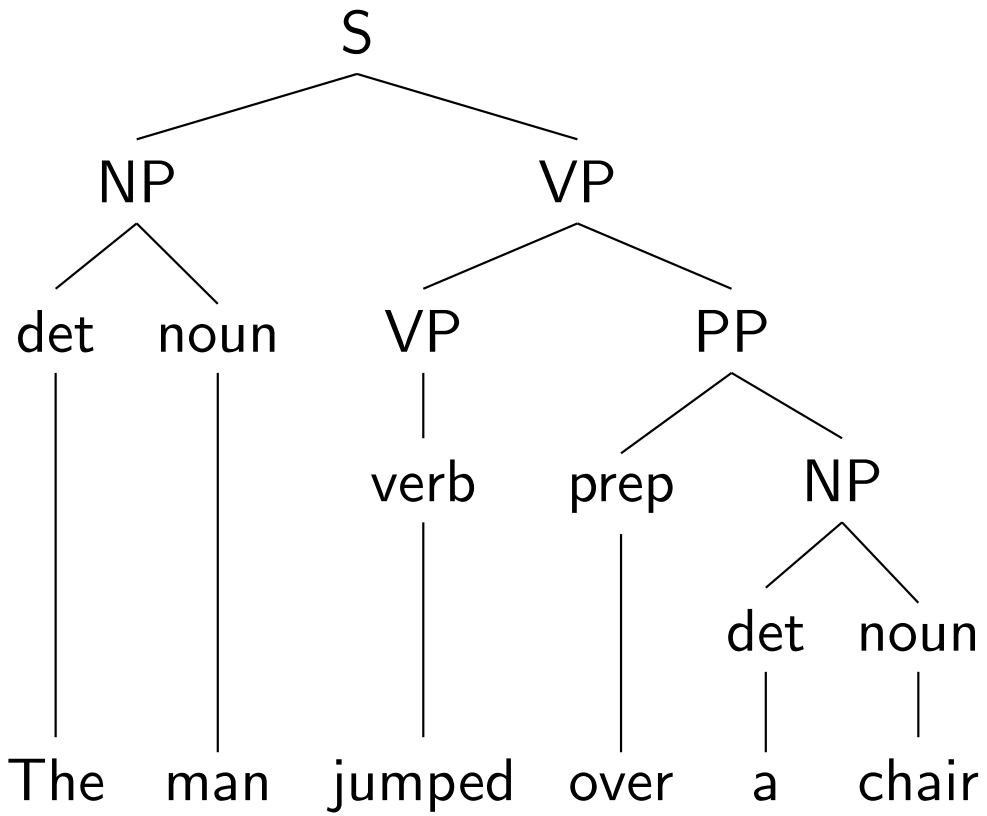
$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

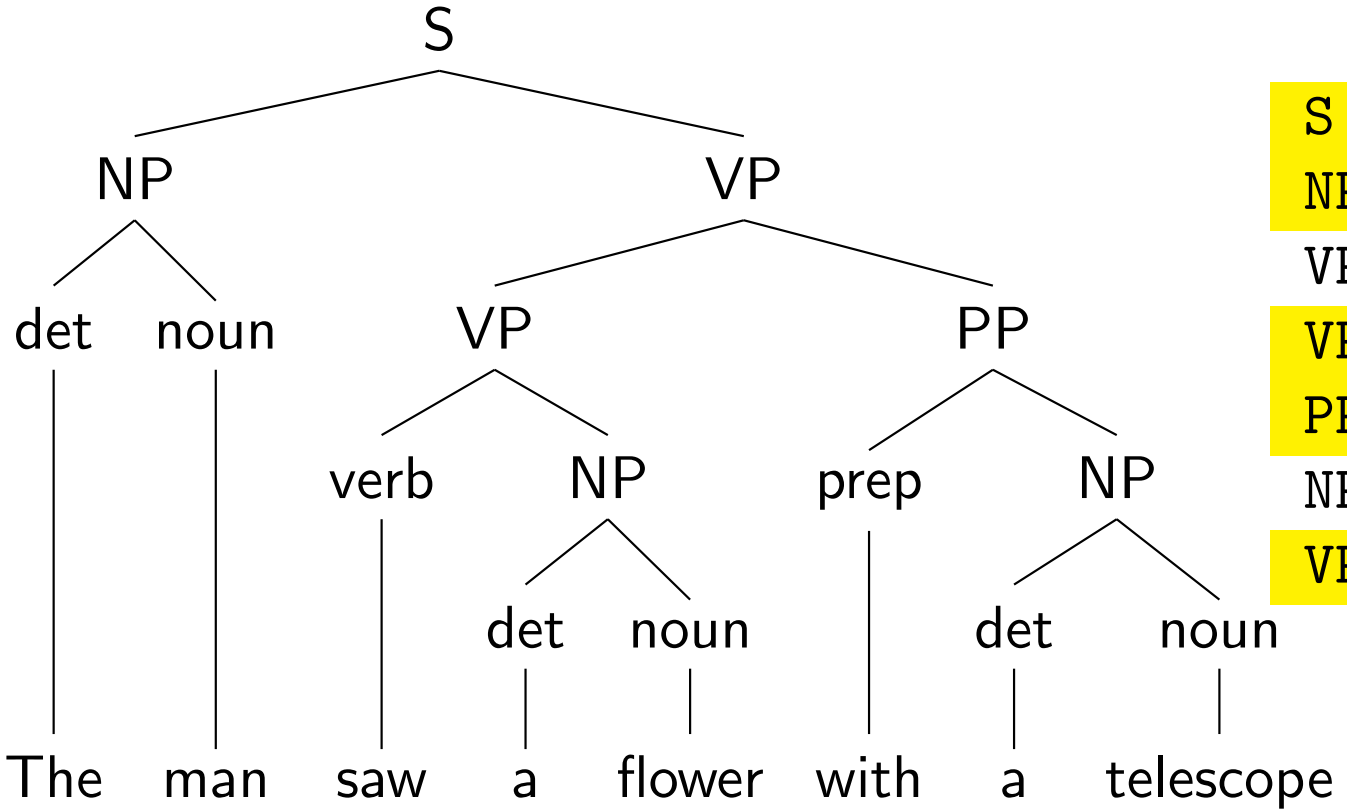
$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

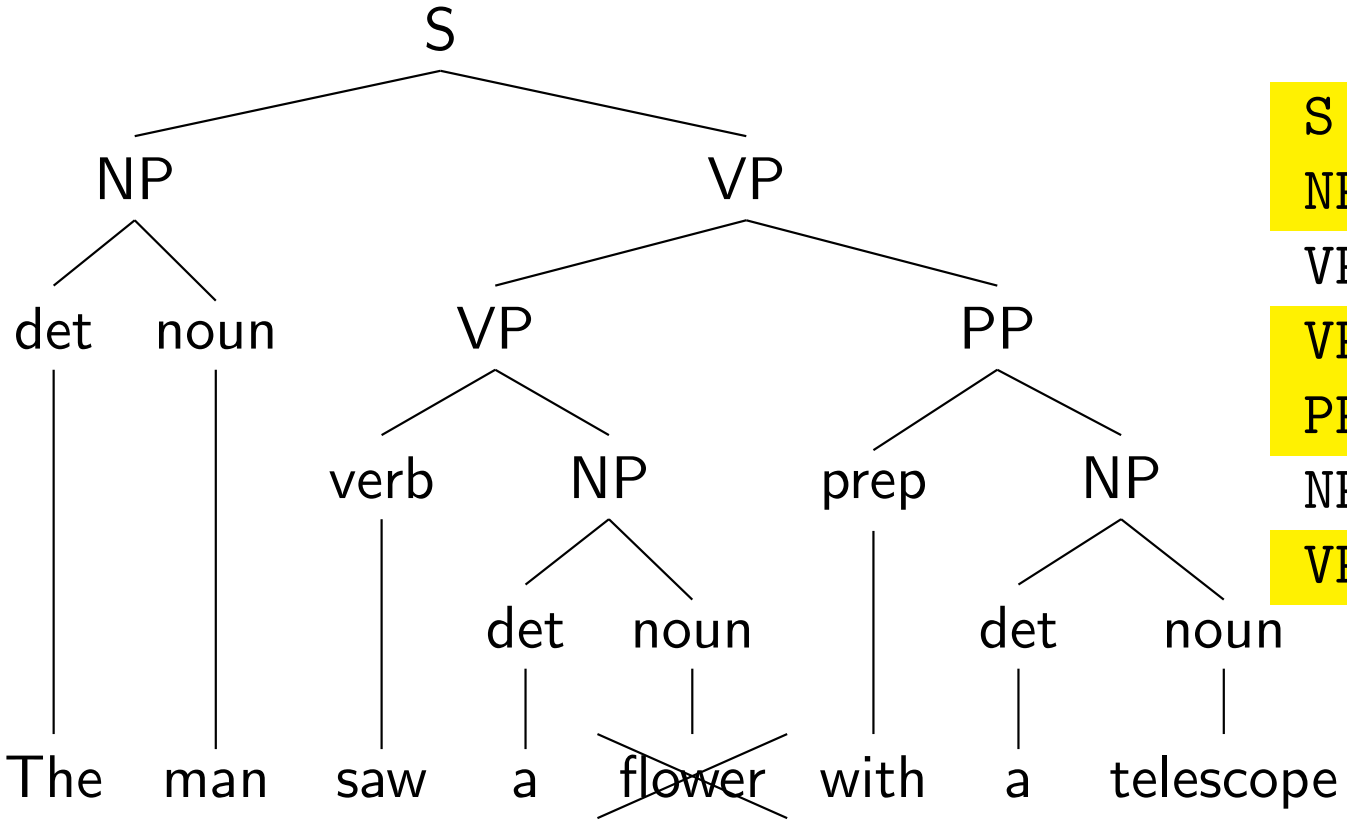
$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

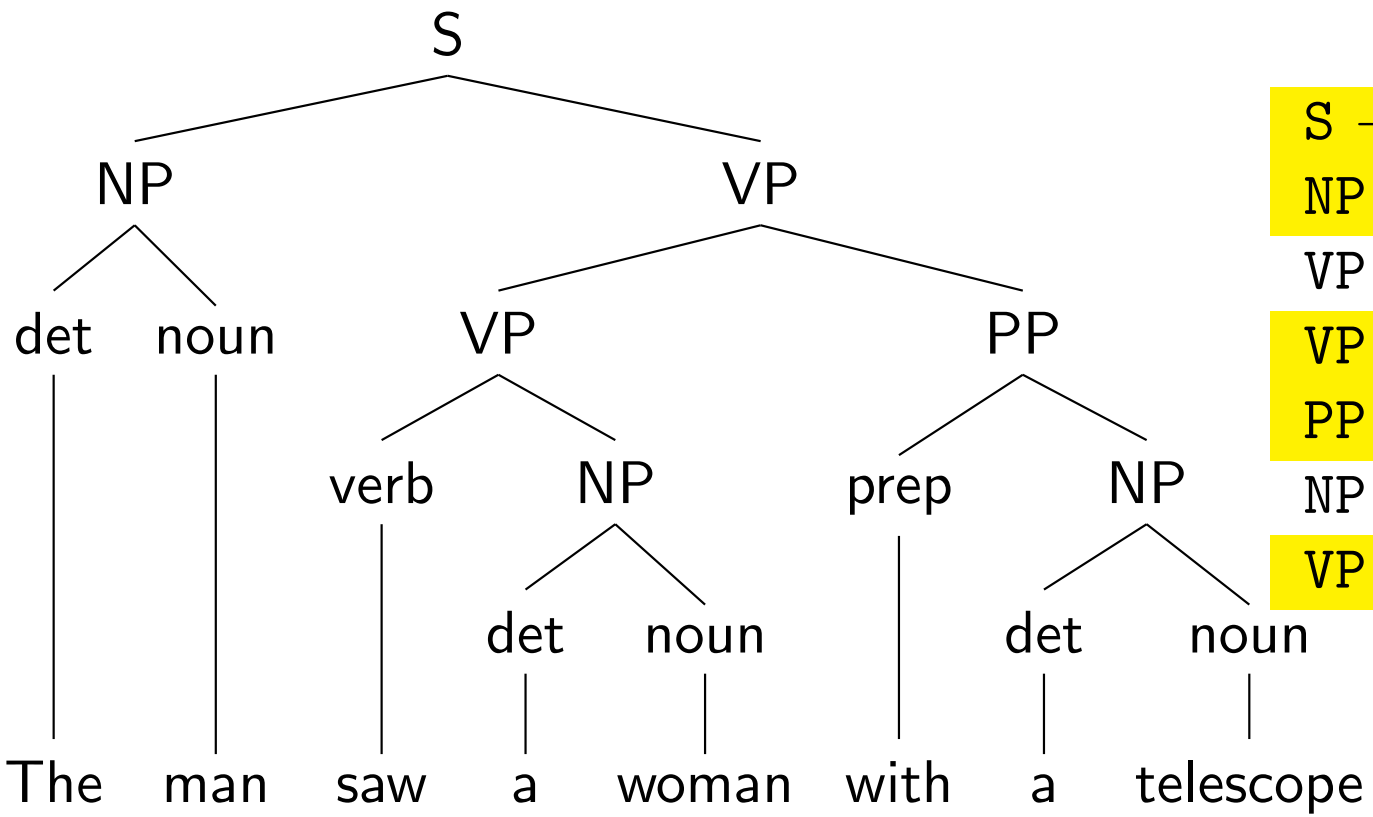
$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

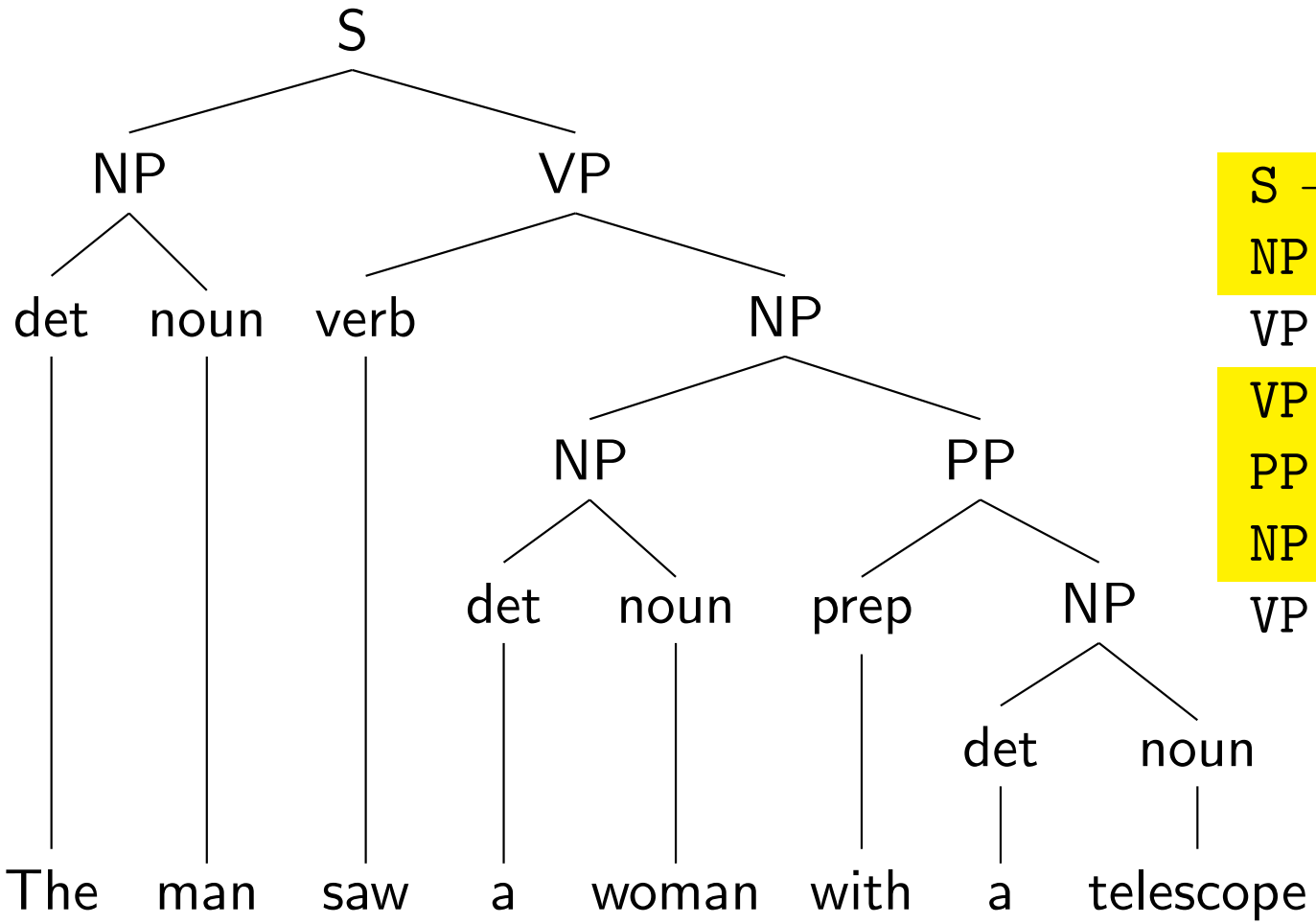
$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$



$S \rightarrow NP\ VP$

$NP \rightarrow det\ noun$

$VP \rightarrow verb$

$VP \rightarrow verb\ NP$

$PP \rightarrow prep\ NP$

$NP \rightarrow NP\ PP$

$VP \rightarrow VP\ PP$

Conclusions

- With so little production rules we have ambiguities.
- In practice, creating a general grammar results in too much ambiguities and poor performance.
- **Solution:** Create simple grammars for simple parameter extraction.

Conclusions

- With so little production rules we have ambiguities.
- In practice, creating a general grammar results in too much ambiguities and poor performance.
- **Solution:** Create simple grammars for simple parameter extraction.

Conclusions

- With so little production rules we have ambiguities.
- In practice, creating a general grammar results in too much ambiguities and poor performance.
- **Solution:** Create simple grammars for simple parameter extraction.

Conclusions

- With so little production rules we have ambiguities.
- In practice, creating a general grammar results in too much ambiguities and poor performance.
- **Solution:** Create simple grammars for simple parameter extraction.

Unification

- The main goal of unification is to find out that two different terms are in fact identical.
- Using unification **with** feature structures (Lexical Object Theory - Jose F Quesada):
  - Adequate for linguistic phenomena.
  - Computationally efficient.
  - Formal soundness.

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Number}: & \texttt{plural}] \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Person}: & 3] \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\&>$$

$$\begin{bmatrix} \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{plural} \\ \textsc{Person}: & 3 \end{bmatrix} \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

Unification

- The main goal of unification is to find out that two different terms are in fact identical.
- Using unification **with** feature structures (Lexical Object Theory - Jose F Quesada):
  - Adequate for linguistic phenomena.
  - Computationally efficient.
  - Formal soundness.

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Number}: & \texttt{plural}] \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Person}: & 3] \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\&>$$

$$\begin{bmatrix} \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{plural} \\ \textsc{Person}: & 3 \end{bmatrix} \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

Unification

- The main goal of unification is to find out that two different terms are in fact identical.
- Using unification **with** feature structures (Lexical Object Theory - Jose F Quesada):
  - Adequate for linguistic phenomena.
  - Computationally efficient.
  - Formal soundness.

$$[\textsc{Subject}: \quad [\textsc{Number}: \quad \texttt{plural}]]$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Person}: \quad 3] \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\&> \qquad \begin{bmatrix} \textsc{Subject}: & [\textsc{Number}: \quad \texttt{plural} \\ & \textsc{Person}: \quad 3] \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

Unification

- The main goal of unification is to find out that two different terms are in fact identical.
- Using unification **with** feature structures (Lexical Object Theory - Jose F Quesada):
  - Adequate for linguistic phenomena.
  - Computationally efficient.
  - Formal soundness.

$$\begin{bmatrix}\textsc{Subject}: & [\textsc{Number}: & \texttt{plural}]]\end{bmatrix}$$

$$\begin{bmatrix}\textsc{Subject}: & [\textsc{Person}: & 3]\\ \textsc{Tense}: & \texttt{present} & \end{bmatrix}$$

$$\&>$$

$$\begin{bmatrix}\textsc{Subject}: & \begin{bmatrix}\textsc{Number}: & \texttt{plural}\\ \textsc{Person}: & 3\end{bmatrix}\\ \textsc{Tense}: & \texttt{present}\end{bmatrix}$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Number}: & \texttt{plural}] \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & [\textsc{Number}: \quad \texttt{singular}] \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Person}: & 3] \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & [\textsc{Number}: \quad \texttt{singular}] \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{plural} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{singular} \end{bmatrix} \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & \begin{bmatrix} \textsc{Person}: & 3 \end{bmatrix} \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{singular} \end{bmatrix} \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Number}: & \texttt{plural}] \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & [\textsc{Number}: & \texttt{singular}] \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & [\textsc{Person}: & \texttt{3}] \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & [\textsc{Number}: & \texttt{singular}] \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{plural} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{singular} \end{bmatrix} \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

$$\begin{bmatrix} \textsc{Subject}: & \begin{bmatrix} \textsc{Person}: & 3 \end{bmatrix} \\ \textsc{Tense}: & \texttt{present} \end{bmatrix}$$

$$\begin{bmatrix} \textsc{Tense}: & \texttt{past} \\ \textsc{Subject}: & \begin{bmatrix} \textsc{Number}: & \texttt{singular} \end{bmatrix} \end{bmatrix} \quad \&> \quad \text{FAIL!!}$$

CC by Kat at the English language Wikipedia SA. 3.0

Dialogue Manager

Definition of Dialogue Manager (**DM**)

It's the part of a dialogue system responsible for the state and flow of the conversation.

- The input of **DM** is the human proference converted to a semantic representation by **NLU** component.
- **DM** maintains some state variables (dialogue history, latest unanswered question, etc.) depending on the system (ie. the **context**).
- The output of **DM** will be another semantic representation of what system want to communicate to human. This will be converted to natural language by the **NLG** component.

Dialogue Manager

Definition of Dialogue Manager (**DM**)

It's the part of a dialogue system responsible for the state and flow of the conversation.

- The input of **DM** is the human proference converted to a semantic representation by **NLU** component.
- **DM** maintains some state variables (dialogue history, latest unanswered question, etc.) depending on the system (ie. the **context**).
- The output of **DM** will be another semantic representation of what system want to communicate to human. This will be converted to natural language by the **NLG** component.

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: How old are you?
U: 24 $\rightarrow$ no need of confirmation

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: How old are you?
U: 24 $\rightarrow$ no need of confirmation

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: How old are you?
U: 24 $\rightarrow$ no need of confirmation

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: How old are you?
U: 24 $\rightarrow$ no need of confirmation

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: ~~How old are you?~~
U: 24

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: How many pizzas do you want?
U: 24 $\rightarrow$ we seriously need to confirm this!

Pragmatics

- Pragmatics is a subfield in linguistics that studies the way in which **context** contributes to meaning.
- So pragmatics studies how the transmission of meaning depends not only on grammar or lexicon, but also on the context.
- These techniques explain how human beings are able to overcome apparent ambiguities or misunderstandings.

Example

The man saw a woman with a telescope.

S: How many pizzas do you want?
U: 24 $\rightarrow$ we seriously need to confirm this!

Natural Language Generation (**NLG**)

- It's the last component of the Dialogue System.
- Receives as input some semantic representation (from **DM**) of what system must say.
- The output will be a string in natural language trying to communicate that semantic representation.
- Traditionally, the process is decomposed into two stages:
  - **Conceptualizer**: What to say.
  - **Formulator**: How to say.
- But frontiers between these modules are subtle: sequential, integrated, interactive, blackboard, revision-based,...

Natural Language Generation (**NLG**)

- It's the last component of the Dialogue System.
- Receives as input some semantic representation (from **DM**) of what system must say.
- The output will be a string in natural language trying to communicate that semantic representation.
- Traditionally, the process is decomposed into two stages:
  - **Conceptualizer**: What to say.
  - **Formulator**: How to say.
- But frontiers between these modules are subtle: sequential, integrated, interactive, blackboard, revision-based,...

Natural Language Generation (**NLG**)

- It's the last component of the Dialogue System.
- Receives as input some semantic representation (from **DM**) of what system must say.
- The output will be a string in natural language trying to communicate that semantic representation.
- Traditionally, the process is decomposed into two stages:
  - **Conceptualizer**: What to say.
  - **Formulator**: How to say.
- But frontiers between these modules are subtle: sequential, integrated, interactive, blackboard, revision-based,...

Natural Language Generation (**NLG**)

- It's the last component of the Dialogue System.
- Receives as input some semantic representation (from **DM**) of what system must say.
- The output will be a string in natural language trying to communicate that semantic representation.
- Traditionally, the process is decomposed into two stages:
  - **Conceptualizer**: What to say.
  - **Formulator**: How to say.
- But frontiers between these modules are subtle: sequential, integrated, interactive, blackboard, revision-based,...

Natural Language Generation (**NLG**)

- It's the last component of the Dialogue System.
- Receives as input some semantic representation (from **DM**) of what system must say.
- The output will be a string in natural language trying to communicate that semantic representation.
- Traditionally, the process is decomposed into two stages:
  - **Conceptualizer**: What to say.
  - **Formulator**: How to say.
- But frontiers between these modules are subtle: sequential, integrated, interactive, blackboard, revision-based,...

Template-based systems

- But, in practice, a simple and efficient model for language generation is templated-based systems.
- These systems map their input (semantic representation) directly to linguistic structure. This structure may contain "gaps" which are filled in during output.

Example

$$\textsc{Departure}: \begin{bmatrix} \textsc{Flight}: & \texttt{Delta 2047} \\ \textsc{Gate}: & \texttt{B37} \\ \textsc{Time}: & \texttt{15:00} \end{bmatrix}$$

```
{Flight} flight will depart from gate {Gate} at {Time}.
```

Template-based systems

- But, in practice, a simple and efficient model for language generation is templated-based systems.
- These systems map their input (semantic representation) directly to linguistic structure. This structure may contain "gaps" which are filled in during output.

Example

$$\textsc{Departure}: \begin{bmatrix} \textsc{Flight}: & \texttt{Delta 2047} \\ \textsc{Gate}: & \texttt{B37} \\ \textsc{Time}: & \texttt{15:00} \end{bmatrix}$$

```
{Flight} flight will depart from gate {Gate} at {Time}.
```

Template-based systems

- But, in practice, a simple and efficient model for language generation is templated-based systems.
- These systems map their input (semantic representation) directly to linguistic structure. This structure may contain "gaps" which are filled in during output.

Example

$$\textsc{Departure}: \begin{bmatrix} \textsc{Flight}: & \texttt{Delta 2047} \\ \textsc{Gate}: & \texttt{B37} \\ \textsc{Time}: & \texttt{15:00} \end{bmatrix}$$

`{Flight} flight will depart from gate {Gate} at {Time}.`

CC by Kat at the English language Wikipedia SA. 3.0