

Introduction to Natural Language Processing

DATE15/EDA132 – Lecture 2 Phrase-Structure Grammars

Pierre Nugues

Pierre.Nugues@cs.lth.se

www.cs.lth.se/~pierre



LUNDS TEKNISKA
HÖGSKOLA
Lunds universitet

Grammar

Grammar is the focus of natural language processing in the textbook (Russell and Norvig 2003, Chapters 22 and 23).

Two main (modern) traditions: constituent grammars (Chomsky, main advocate) and dependency grammars (Tesnière).

Constituent grammars are still dominant for English, although declining.

But they do not work well for Swedish, as well as many other languages.

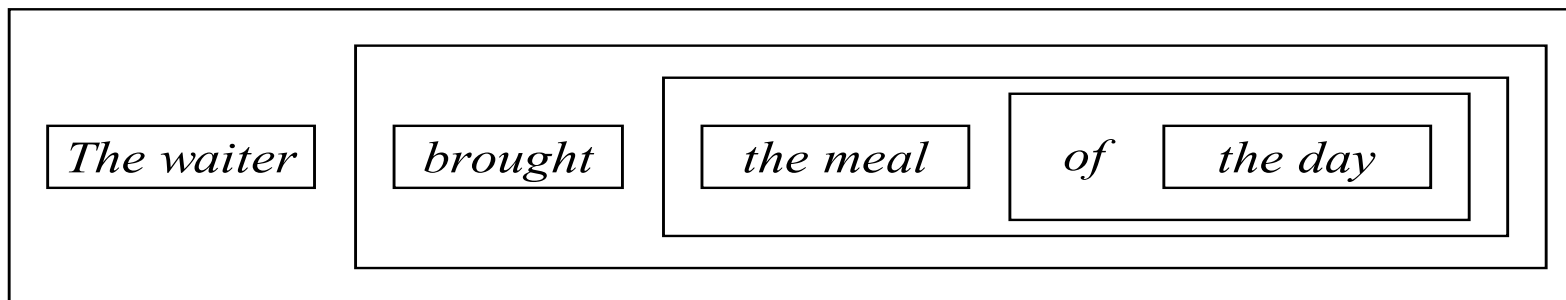
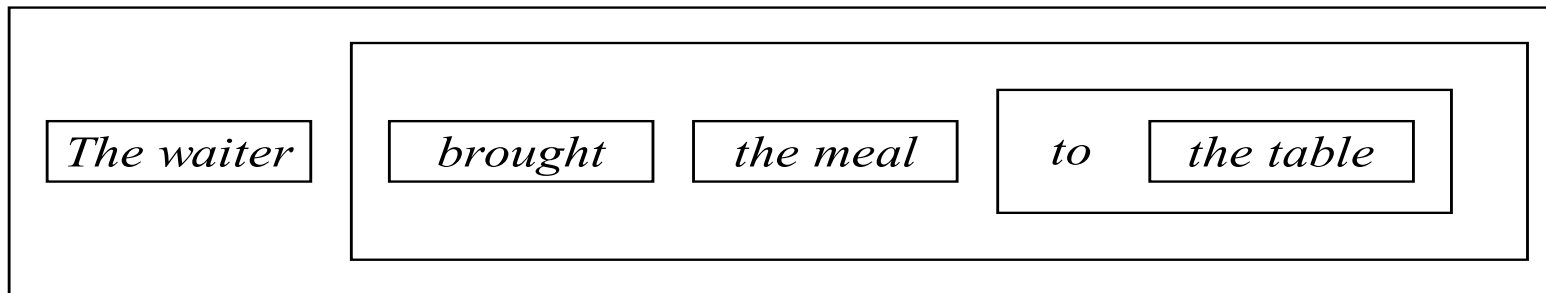
Dependency grammars are more or less universal

We will review both formalisms and you will implement a dependency parser in Swedish, English, or a language for which an annotated corpus is available.

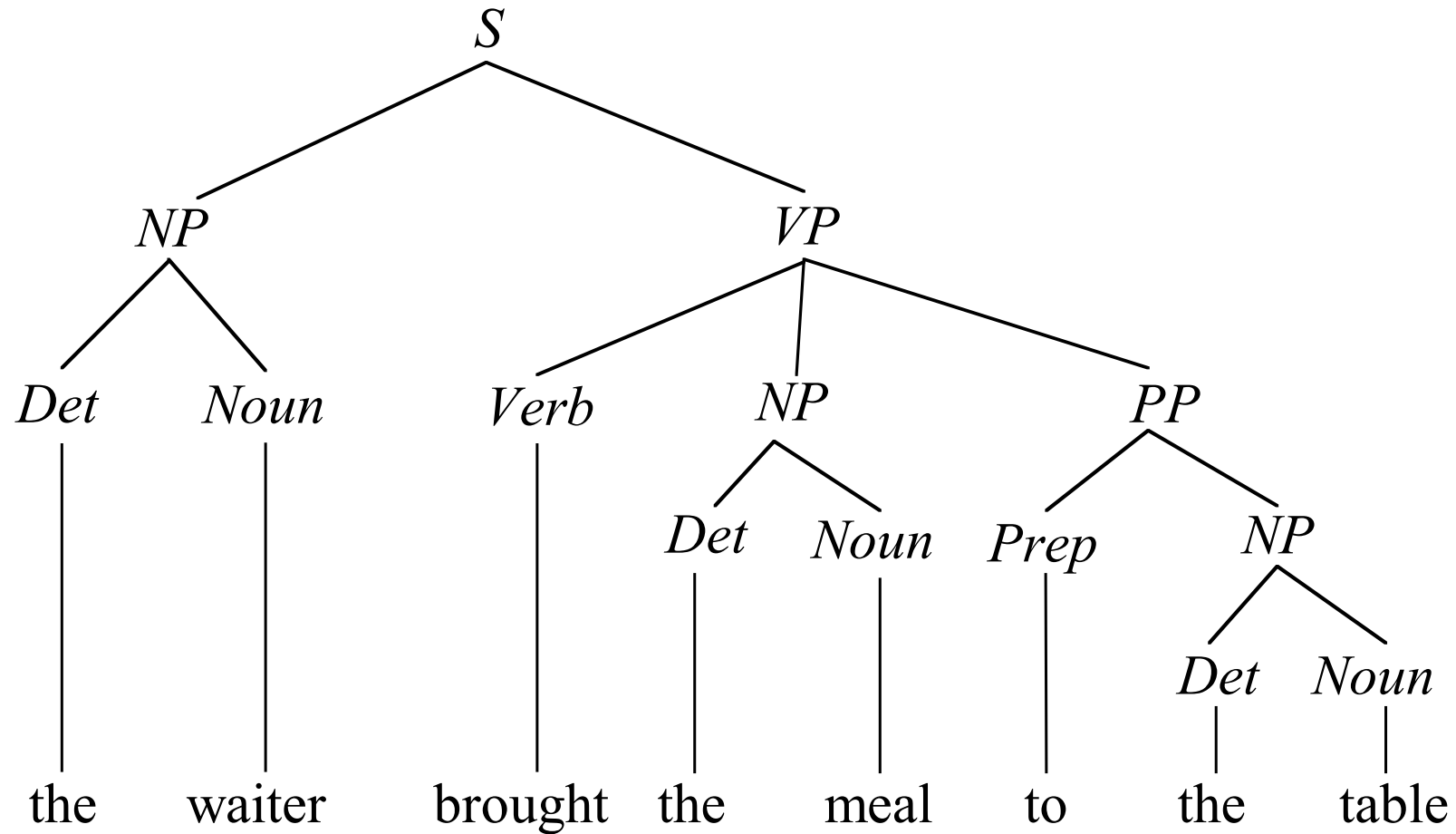
Constituents

The waiter brought the meal to the table

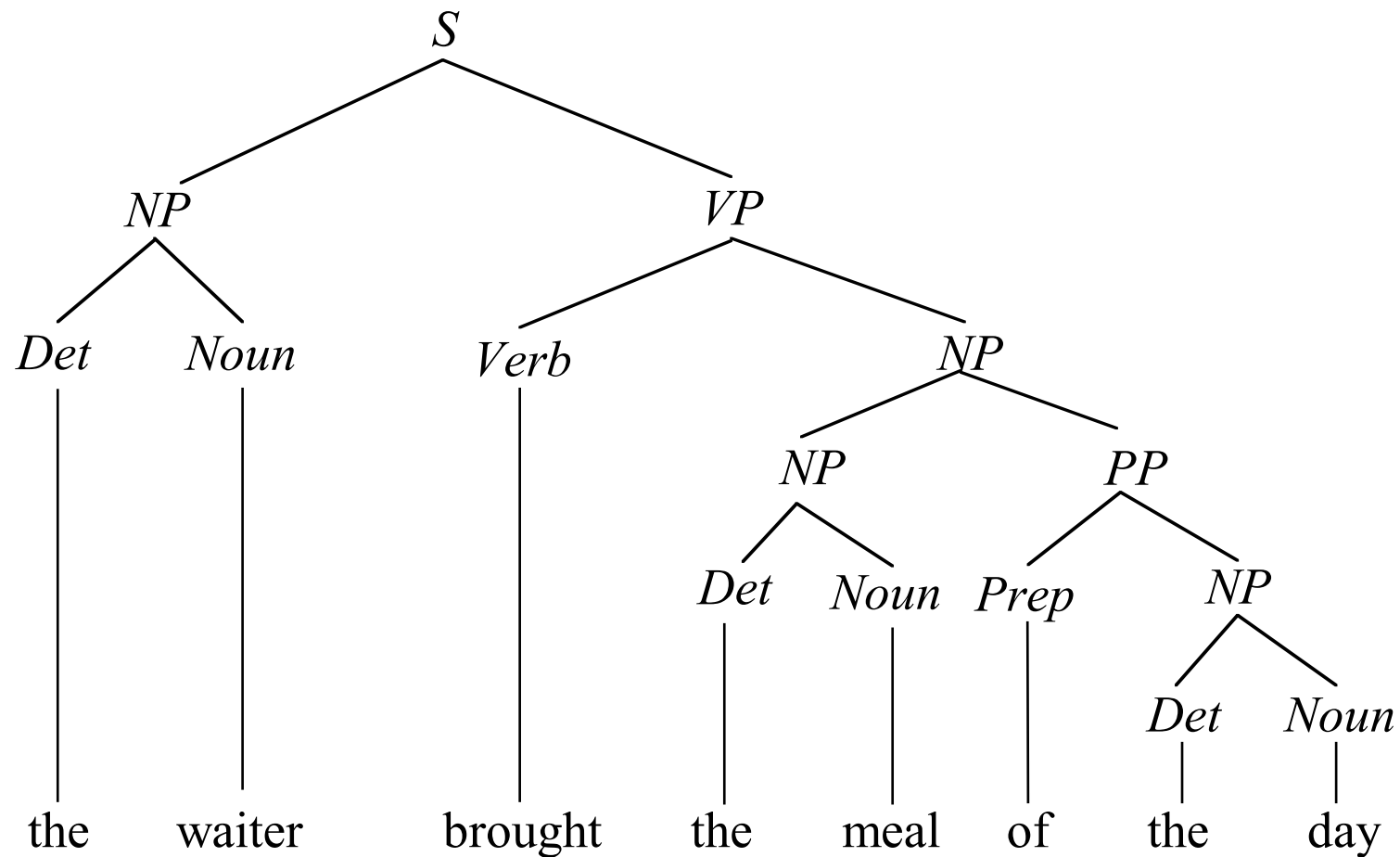
The waiter brought the meal of the day



Syntactic Trees



Syntactic Trees



Lexicon (DCG)

noun --> [stench] ; [breeze] ; [glitter] ; [nothing] ; [wumpus]
; [pit] ; [pits] ; [gold] ; [east].

verb --> [is] ; [see] ; [smell] ; [shoot] ; [feel] ; [stinks] ;
[go] ; [grab] ; [carry] ; [kill] ; [turn].

adjective --> [right] ; [left] ; [east] ; [south] ; [dead] ;
[back] ; [smelly].

adverb --> [here] ; [there] ; [nearby] ; [ahead] ; [right] ;
[left] ; [east] ; [south] ; [back].

pronoun --> [me] ; [you] ; ['I'] ; [it] ; [she] ; [he].

pnoun --> ['John'] ; ['Mary'] ; ['Boston'] ; ['UCB'] ; ['PAJC'].

article --> [the] ; [a] ; [an].

preposition --> [to] ; [in] ; [on] ; [near].

conjunction --> [and] ; [or] ; [but].

digit --> [0] ; [1] ; [2] ; [3] ; [4] ; [5] ; [6] ; [7] ; [8] ;
[9].

Grammar Rules (DCG)

s --> np, vp. % I + feel a breeze
s --> s, conjunction, s.
np --> pronoun. %I
npx --> pnoun.
np --> noun. %pits
np --> article, noun. %the + wumpus
np --> digit, digit. %3 4
np --> np, pp. %the wumpus + to the east
np --> np, rel_clause. %the wumpus + that is smelly
vp --> verb. %stinks
vp --> vp, np. %feel + a breeze
vp --> vp, adjective. %is + smelly
vp --> vp, pp. %turn + to the east
vp --> vp, adverb. %go + ahead
pp --> preposition, np. %to + the east
rel_clause --> [that], vp. %that + is smelly

Parsing and Generation

Parsing tells if a sentence is correct according to the grammar

```
?-s([the, wumpus, is, dead], []).
```

yes.

```
?- s([the, wumpus, that, stinks, is, in, 2, 2], []).
```

yes.

The parser can generate all the solutions

```
?- s(L, []).
```

```
L = [me, is] ;
```

```
L = [me, see] ;
```

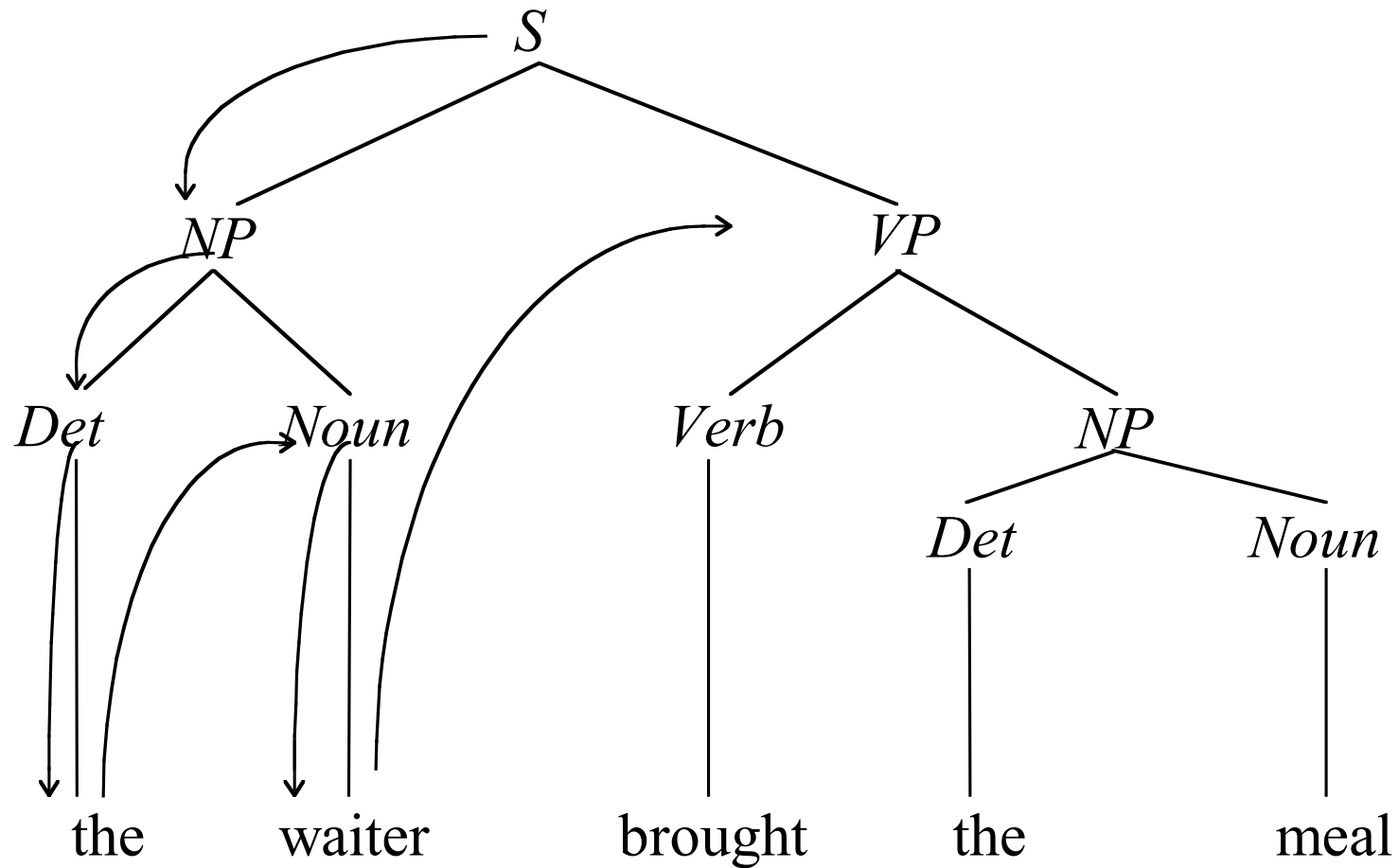
```
L = [me, smell] ;
```

```
L = [me, shoot] ;
```

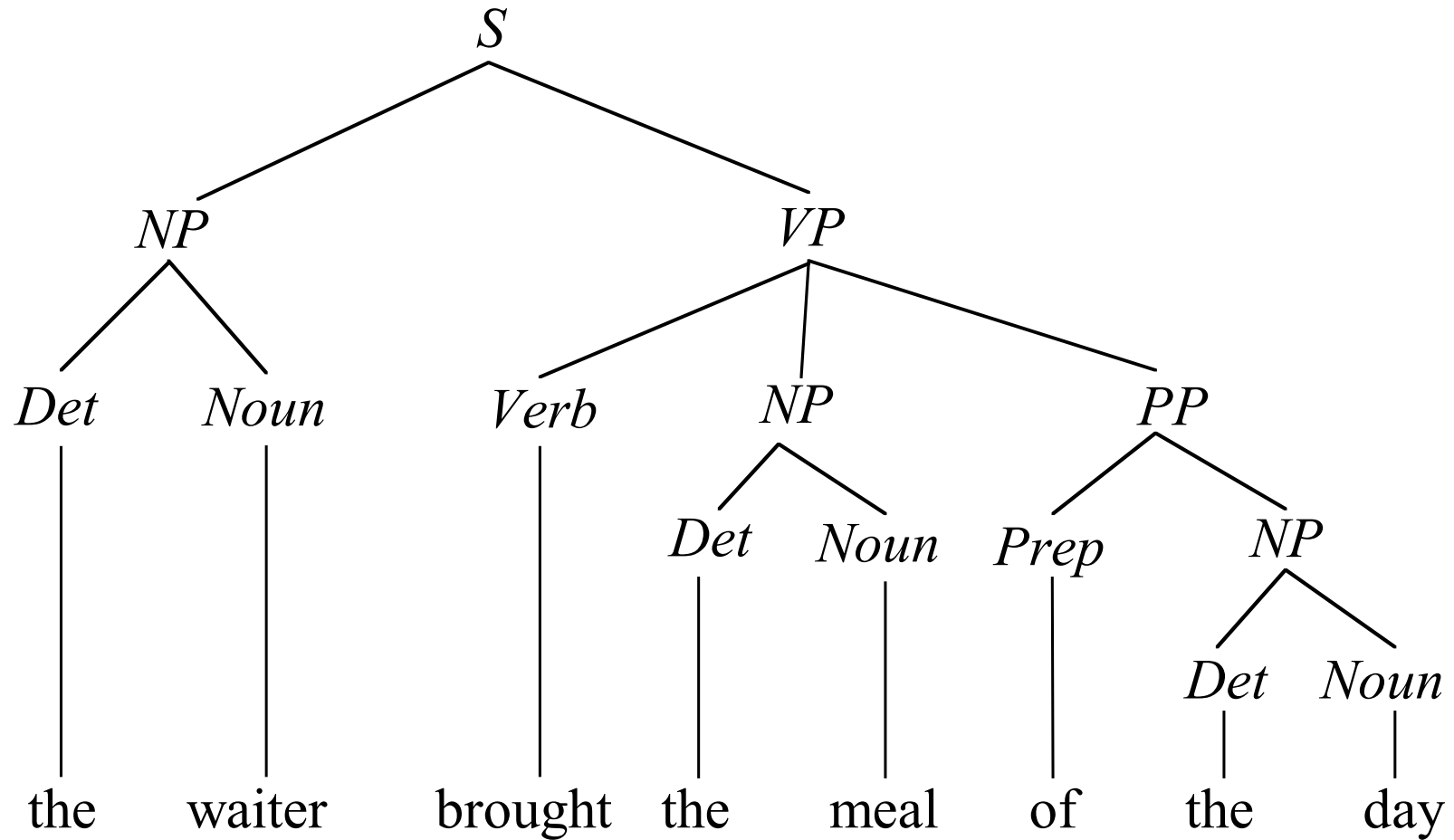
```
L = [me, feel] ;
```

```
L = [me, stinks] ;
```

Top-Down Search



Ambiguity



Left-Recursive Rules

$np \rightarrow np, pp.$

The sentence:

The wumpus in the pit is dead

traps the parser in an infinite recursion.

We can use auxiliary symbols to remove left recursion:

$np_x \rightarrow det, noun.$

$np \rightarrow np_x.$

$np \rightarrow np_x, pp.$

Variables

Overgeneration:

```
?- s(X, []).
```

```
X = [me, is] ;
```

```
X = [me, see] ;
```

```
X = [me, smell] ;
```

Solution: Add variables to differentiate between subject and object pronouns.

```
s --> np(s), vp.
```

```
np(Case) --> pronoun(Case).
```

```
pronoun(s) --> [you] ; ['I'] ; [it]; [she]; [he].
```

```
pronoun(o) --> [me] ; [you] ; [it].
```

Subcategorization Frames

Valence is a model of verb construction. It can be extended to more specific patterns as in the *Oxford Advanced Learner's Dictionary*

Verb	Complement structure	Example
<i>slept</i>	None (Intransitive)	<i>I slept</i>
<i>bring</i>	NP	<i>The waiter brought the meal</i>
<i>bring</i>	NP + to + NP	<i>The waiter brought the meal to the patron</i>
<i>depend</i>	on + NP	<i>It depends on the waiter</i>
<i>wait</i>	for + NP + to + VP	<i>I am waiting for the waiter to bring the meal</i>
<i>keep</i>	VP(ing)	<i>He kept working</i>
<i>know</i>	that + S	<i>The waiter knows that the patron loves fish</i>

Semantic Parsing

Converts sentences to first-order logic or predicate-argument structures

Example:

Mr. Schmidt called Bill

to

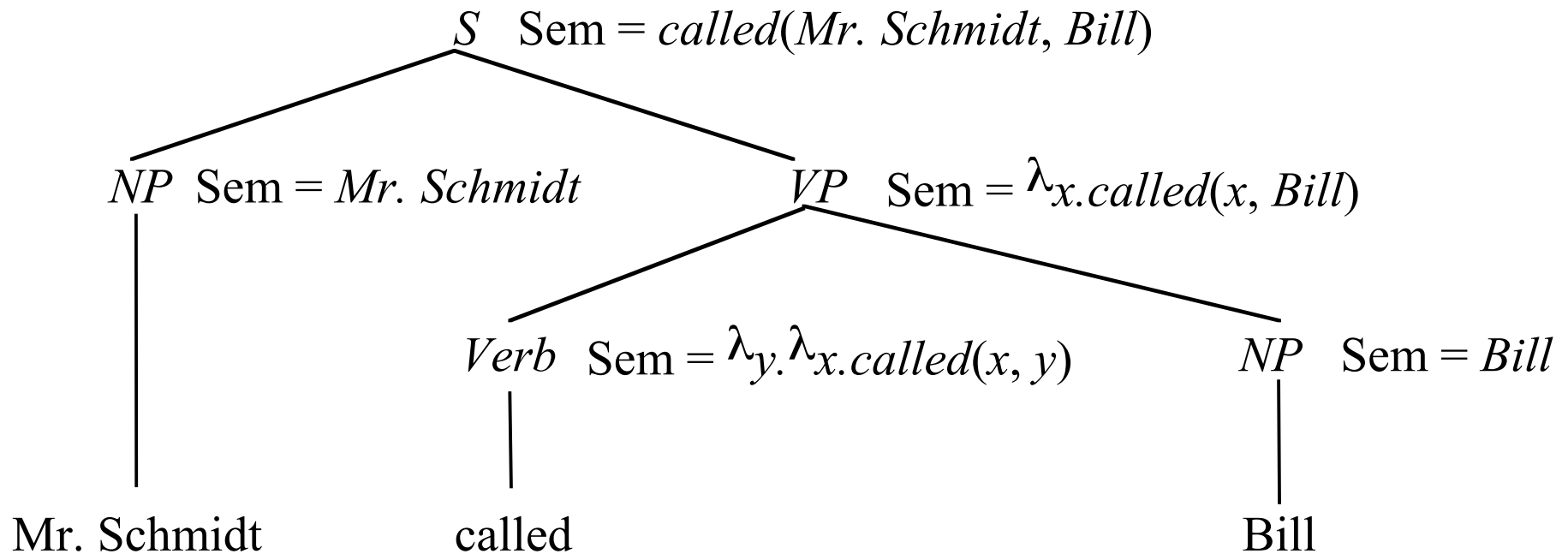
called('Mr. Schmidt', 'Bill').

Assumption: We can compose sentence fragments (phrases) into logical forms while parsing

This corresponds to the compositionality principle

Semantic Composition

Semantic composition can be viewed as a parse tree annotation



Encoding the Semantic Structure

Bill rushed `rushed('Bill').`

$\lambda x.rushed(x)$ $\lambda x.rushed(x)(Bill) = rushed(Bill)$

In Prolog \rightarrow `X^rushed(X).`

The patron ordered a meal `ordered(patron, meal)`

ordered a meal `X^ordered(X, meal)`

ordered `Y^X^ordered(X, Y)`

Getting the Semantic Structure

`s(Semantics) --> np(Subject), vp(Subject^Semantics).`

`np(X) --> det, noun(X).`

`vp(Subject^Predicate) --> verb(Subject^Predicate).`

`vp(Subject^Predicate) -->`

`verb(Object^Subject^Predicate), np(Object).`

`noun(waiter) --> [waiter].`

`det --> [a].`

`noun(patron) --> [patron].`

`det --> [the].`

`noun(meal) --> [meal].`

`verb(X^rushed(X)) --> [rushed].`

`verb(Y^X^ordered(X, Y)) --> [ordered].`

`verb(Y^X^brought(X, Y)) --> [brought].`

`?- s(Semantics, [the, patron, ordered, a, meal], []).`

`Semantics = ordered(patron, meal)`

Time and Tense

Possible treatment: reify events, quantify them existentially, and connect them to other objects using predicates.

John saw Mary in London on Tuesday has the logical form:

$\exists e[Saw(e, John, Mary) \wedge Place(e, London) \wedge Time(e, Tuesday)]$

Other temporal relations include:

before (a, b)

after (b, a)

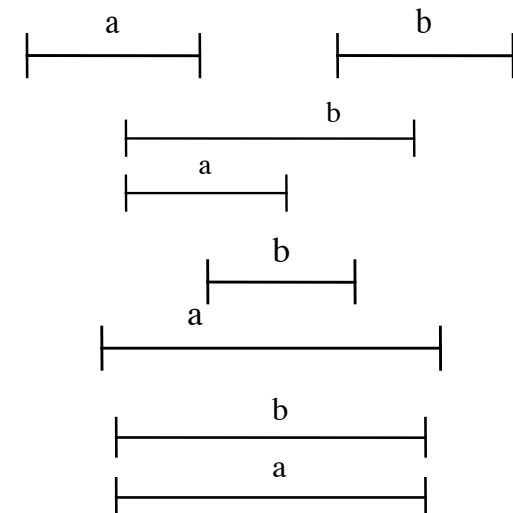
starts (a, b)

started_by (b, a)

during (b, a)

contains (a, b)

equals (a, b)



Quantifiers

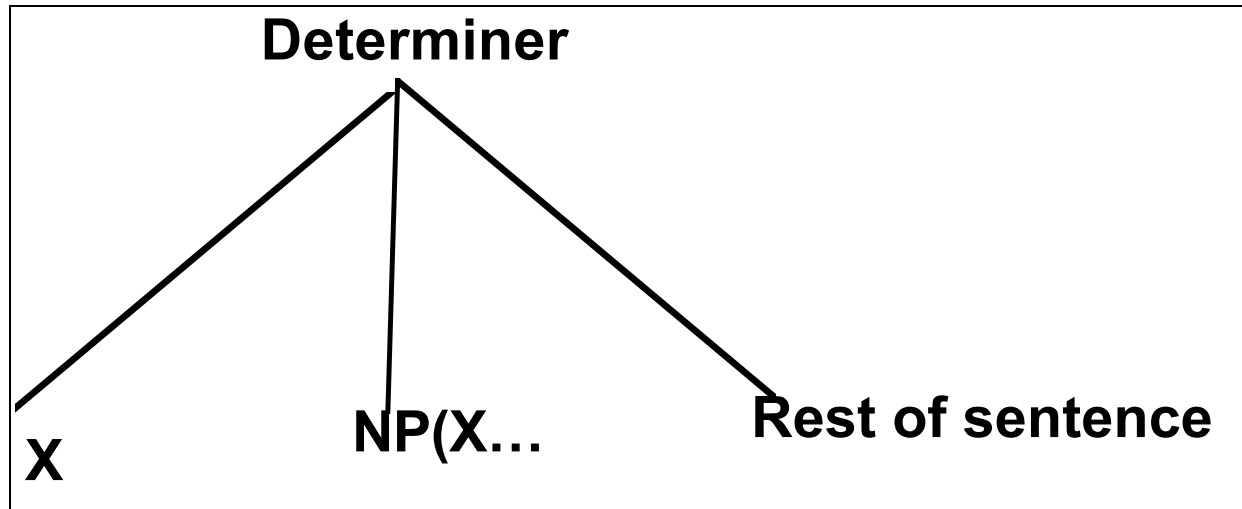
A caterpillar is eating $\exists x, \text{caterpillar}(x) \wedge \text{eating}(x)$, or
exists (X, caterpillar(X), eating(X))

Every caterpillar is eating $\forall x, \text{caterpillar}(x) \Rightarrow \text{eating}(x)$, or
all (X, caterpillar(X), eating(X))

A caterpillar is eating a hedgehog $\exists x, \text{caterpillar}(x) \wedge (\exists y, \text{hedgehog}(y) \wedge \text{eating}(x, y))$, or
exists (X, caterpillar(X), exists (Y, hedgehog(Y), eating(X, Y))

Every caterpillar is eating a hedgehog $\forall x, \text{caterpillar}(x) \Rightarrow (\exists y, \text{hedgehog}(y) \wedge \text{eating}(x, y))$, or
all (X, caterpillar(X), exists (Y, hedgehog(Y), eating(X, Y))

Determiners



Compositionality

Some rules to generate the logical form:

`noun (X^dog (X)) --> [dog] .`

`noun (X^wumpus (X)) --> [wumpus] .`

`determiner ((X^NP) ^exists (X, NP)) --> [a] .`

`np (SemDet) -->`

`determiner ((X^NP) ^SemDet) ,`

`noun (X^NP) .`

?- `np (Semantics, [a, wumpus], []).`

`Semantics = exists (_4, wumpus (_4))`

A Definition of Discourse

A discourse is a sequence of sentences: a text or a conversation

A discourse is made of words or phrases that refer to things: the **discourse entities**

A discourse normally links the entities together to address topics

Within a single sentence, grammatical structures provide with a model of relations between entities.

Discourse models extend relations to more sentences

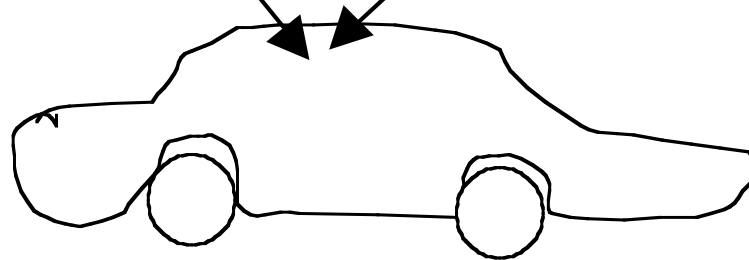
References and Coreference

Discourse entities are the objects introduced by the discourse.

Referring expressions are mentions of the discourse entities

[_{entity1} Garcia Alvarado], 56, was killed when [_{entity2} a bomb] placed by [_{entity3} urban guerrillas] on [_{entity4} his vehicle] exploded as [_{entity5} it] came to [_{entity6} a halt] at [_{entity7} an intersection] in [_{entity8} downtown] [_{entity9} San Salvador].

on his vehicle exploded as it came to a halt



Solving Coreferences: A Simplistic Method

Coreferences define a class of equivalent references

Backward search with a compatible gender and number

*Garcia Alvarado, 56, was killed when a **bomb** placed by urban*

2 ←

*Guerrillas on **his vehicle** exploded as **it** came to a halt at an*

← 1

intersection in downtown San Salvador

Rhetorical Structure Theory

The rhetorical structure theory (RST) can be used to represent a discourse

It is a text grammar that analyzes argumentation.

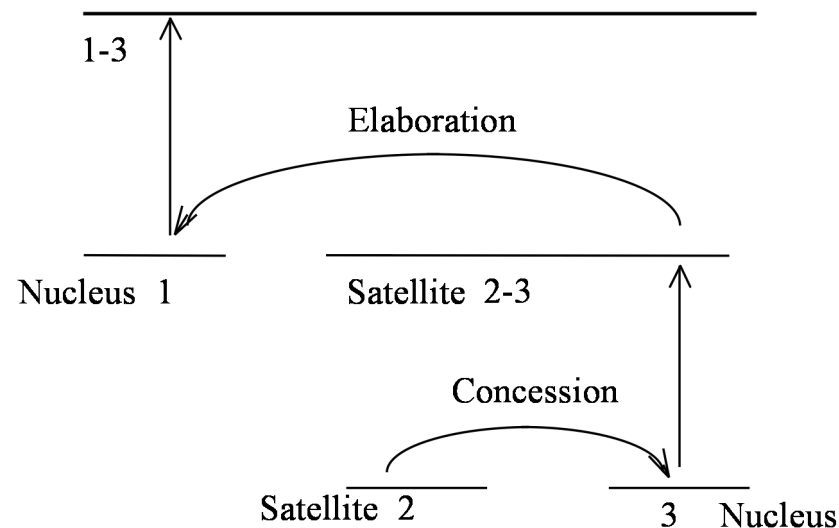
A text consists of:

- **Text spans** that can be sentences or clauses
- **Rhetorical relations** that link the text spans
- Relation can be asymmetrical and involve a nucleus and a satellite or symmetrical between two nuclei

Graphical Representation

Example cited by Mann and Thompson (1987):

- 1. Concern that this material is harmful to health or the environment may be misplaced.*
- 2. Although it is toxic to certain animals,*
- 3. evidence is lacking that it has any serious long-term effect on human beings.*



RST Relations

Nucleus- satellite relations

Circumstance	Evidence	Otherwise
Solutionhood	Justify	Interpretation
Elaboration	Cause	Evaluation
Background	Antithesis	Restatement
Enablement	Concession	Summary
Motivation	Condition	

Multi-nucleus relations

Sequence Contrast Joint