Applying Logic to the Study of Human Language Syntax

Geoffrey K. Pullum

Brown University and the University of Edinburgh

August 2012
East Asian School in Logic, Language and Computation

Lecture 1: Preliminaries

Introduction: human language

Humans can not only **entertain propositional thoughts** and **store, manipulate, and reflect** on them, but also

- **transmit** propositions overtly and intentionally
- with the **intent of altering epistemic states** of conspecifics
- **independently of stimulus control**

- **receive** them from conspecifics
- **acquire** new propositional information in that way

- using a **huge and apparently open** variety of signs with
- **arbitrary internal complexity**, and
- **learn** this system of signs rapidly and early

Introduction: human language

Other animals communicate, over various channels:

- **vision** (bodily shape, position, or color)
- **sound** (growling, squealing, calling)
- **odour**
- **touch**

and so on.

What other animals do not have is **grammar**:

- **phonology** (sound patterns)
- **morphology** (word structure)
- **semantics** (literal meaning)
- **syntax** (expression structure)

Introduction: human language

Other animals communicate, over various channels:

- **vision** (bodily shape, position, or color)
- **sound** (growling, squealing, calling)
- **odour**
- **touch**

and so on.

What other animals do not have is **grammar**:

- **phonology** (sound patterns)
- **morphology** (word structure)
- **semantics** (literal meaning)
- **syntax** (expression structure) ⟵ MY FOCUS HERE

Introduction: human language

What is syntax?

Studying the syntax of a language means studying the **structure** of the expressions in that language:

- the elements they contain
- the structure of those elements
- the forms the elements take in particular contexts
- the order in which they can be put together
- the structural relations that hold between them.

Introduction: human language

The internal structure of words is called **morphology**; the structure of the units that are composed of words is called **syntax**.

Languages differ radically in both their morphological and their syntactic properties.

- In languages like Chinese, a large proportion of the words have no morphological subparts — there is very little internal structure to words.
- In languages like English, there is some morphological structure but not very much.
- In languages like Swahili, there is a prefix system of considerable complexity in the internal structure of nouns, verbs, and adjectives.
- In languages like Turkish, words can contain a dozen word-forming suffix elements, or more.
- In Eskimoan languages words can be of any length and can contain whole incorporated noun phrases inside verbs.

Introduction: human language

We can define ‘word’ fairly precisely. In Turkish, it happens that vowels have to match in backness throughout a word, so we know that this is one word:

Slovakyalılaştıramadıklarımızdanmıymışsınız?

‘Aren’t you one of the people we were said to be unable to Slovakianize?’

Replacing *Slovakya* (‘Slovakia’) by *Çin* (‘China’):

Çinlileştirmediklerimizdenmiymišsiniz?

‘Aren’t you one of the people we were said to be unable to make Chinese?’

The change of vowels runs right through a word, but not beyond a word.

Introduction: human language

Some examples of syntactic facts in human languages:

- Closed interrogative clauses in English begin with an auxiliary verb.
- Object noun phrases precede the verb in a Japanese clause.
- The form of a verb in Basque depends on the properties of the subject, the direct object (if there is one), and the indirect object (if there is one).
- A relative clause in an English noun phrase follows the head noun.
- A subject noun phrase in Latin is in the nominative case.

(Underlined words are technical terms in grammar.)

Introduction: human language

Such facts are not **universal** (common to all languages); they are **parochial**: they must be specified differently in different languages.

Consider just the order of Subject (S), Direct Object (O), and verb (V) in simple main clauses:

- In Turkish, Korean, and Mongolian the usual order is SOV.
- In English, Swahili, and Finnish the usual order is SVO.
- In Irish, Maasai, and Tahitian the usual order is VSO.
- In Malagasy (Madagascar) the usual order is VOS.
- In Hixkaryana (Brazil) the usual order is OVS.
- In Xavante (Brazil) the usual order is OSV.

Introduction: human language

And there are other possibilities:

- In Mandarin Chinese the order is often SVO but sometimes SOV.
- In languages like Welsh and colloquial Arabic some declarative clauses are VSO and some are SVO.
- In English, closed interrogative clauses (*Can they see us?*) have a VSX order, where V is an auxiliary verb and X is a predicate constituent such as a Verb Phrase (VP).
- In German the order is SOV in subordinate clauses, but in main clauses it is any order in which the verb comes second.
- In Dyirbal (an aboriginal language of Queensland, Australia) the order is extremely free, and within a clause almost any order may occur.

Introduction: human language

A **grammar** for a language L is a statement of the structure expressions in L must have in order to be properly formed.

It should state (or entail) all the grammatically relevant properties of all expressions.

However, the traditional grammars of English published between 1600 and 2000 are, to a massive extent,

- **incomplete** (they don't cover every kind of expression),
- **inaccurate** (what they say about some expressions is wrong), and
- **imperfect** (their generalizations are often not the right ones).

Grammaticality and ungrammaticality

Traditional grammars may tell you that adjectives modify nouns, but they will never note that repetition for emphasis of this sort is possible:

There are some ugly, ugly people out there.

I want you to be very, very, very careful.

We have many, many topics to discuss.

Grammaticality and ungrammaticality

Traditional grammars may tell you that adjectives modify nouns, but they will never note that repetition for emphasis of this sort is possible:

There are some ugly, ugly people out there.

I want you to be very, very, very careful.

We have many, many topics to discuss.

They may tell you that adverbs modify words other than nouns, but they never warn you about this difference:

I eagerly opened the box.

**I opened eagerly the box.*

‘ * ’ indicates ‘not grammatically correct in Standard English’

Grammaticality and ungrammaticality

They never point out that adverbs can be built up into phrases:

separately

[*separately* [*from* [*the rest*]]]

[*separately* [*from* [*the rest* [*of the company*]]]]

[*quite* [*separately* [*from* [*the rest* [*of the company*]]]]]

Tree diagrams of expression structure

To represent the syntactic structure of expressions, linguists typically use **labeled ordered trees**.

A **tree**, as understood here, is a **singly-rooted**, **directed**, **acyclic** graph.

The vertices are labeled from a vocabulary of **category** names, and are referred to by linguists as **nodes**.

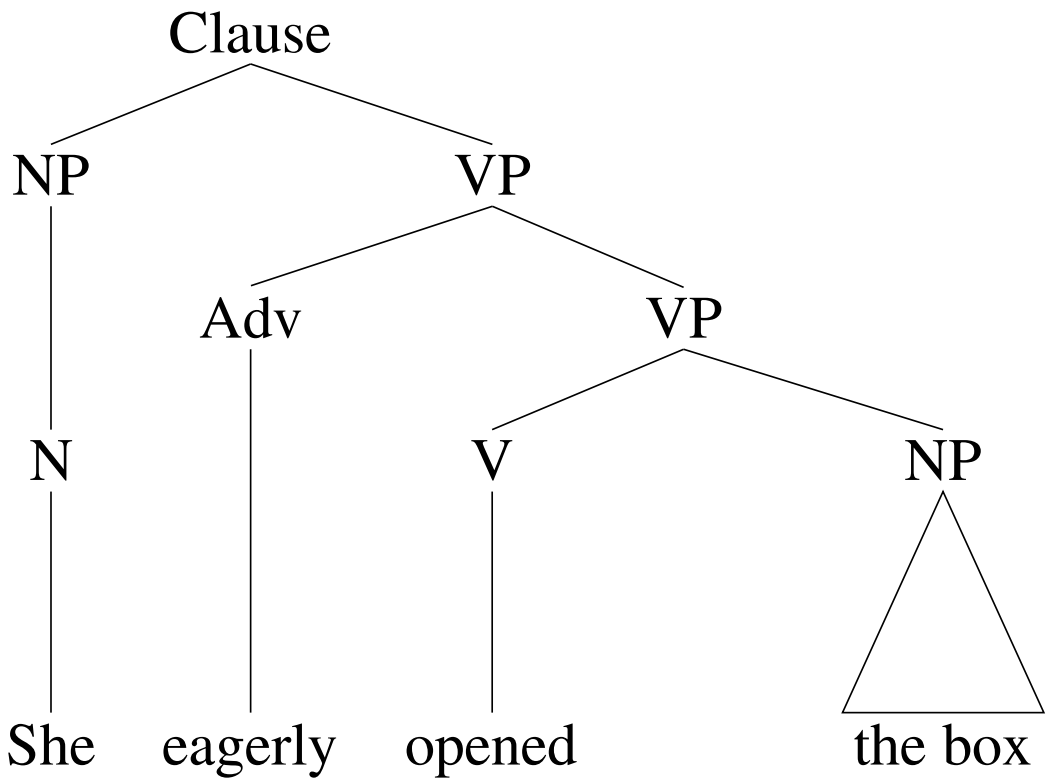
Terminal nodes (those not dominating any other node) are labeled with the words or symbols of the language.

Sometimes trees may be embellished in certain ways:

- the edges may be labeled, e.g. with **grammatical function** names like 'Subject' or 'Head'
- the category names may be internally complex **feature–value matrices**
- the nodes may be indexed with numbers, and nodes bearing the same number treated as related (e.g. **anaphorically**)

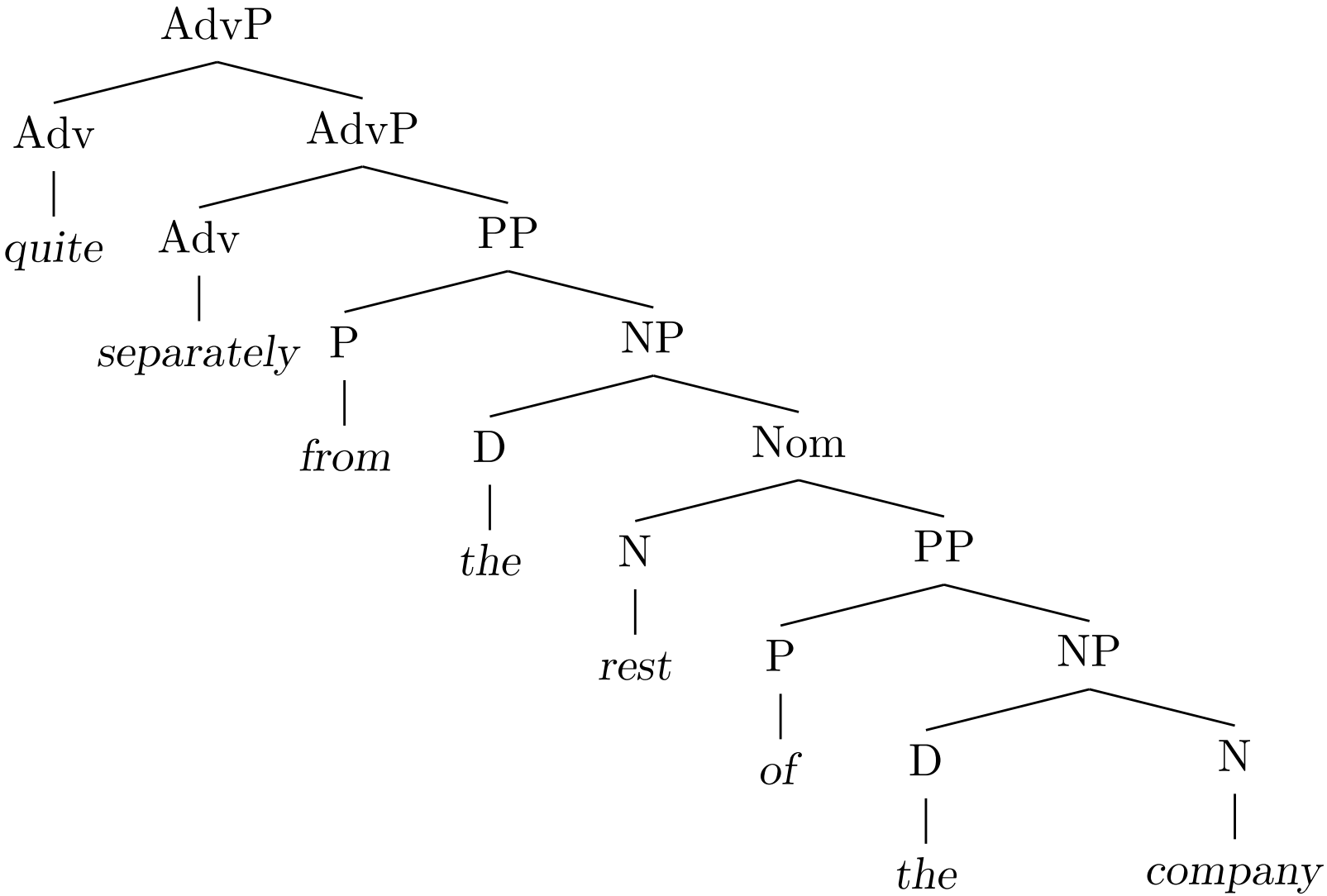
Tree diagrams of expression structure

An example of a labeled tree representing the declarative clause *She eagerly opened the box*:



Tree diagrams of expression structure

The structure of *quite separately from the rest of the company*:



Tree diagrams of expression structure

Some grammatical categories that will be assumed in these lectures:

Clause	clause	*that almost every word from his or her friends seemed quite obviously insincere*
Adj	adjective	*insincere*
AdjP	adjective phrase	*quite obviously insincere*
Adv	adverb	*obviously*
AdvP	adverb phrase	*quite obviously*
N	noun	*word*
Nom	nominal group	*word from his or her friends*
NP	noun phrase	*almost every word from his or her friends*

Tree diagrams of expression structure

Further grammatical categories that will be assumed in these lectures:

Crd	coordinator	*or*
D	determinative	*every*
DP	determinative phrase	*almost every*
P	preposition	*from*
PP	preposition phrase	*from his or her friends*
Sbr	subordinator	*that*
V	verb	*seemed*
VP	verb phrase	*seemed quite obviously insincere*

Tree diagrams of expression structure

Some important grammatical functions:

Subject: *they* in the clause *They depress me so much.*

Object: *that* in the clause *I suspected that.*

Head: *delightful* in the phrase *perfectly delightful*

Predicative Complement: *sorry for that* in the clause *I'm sorry for that.*

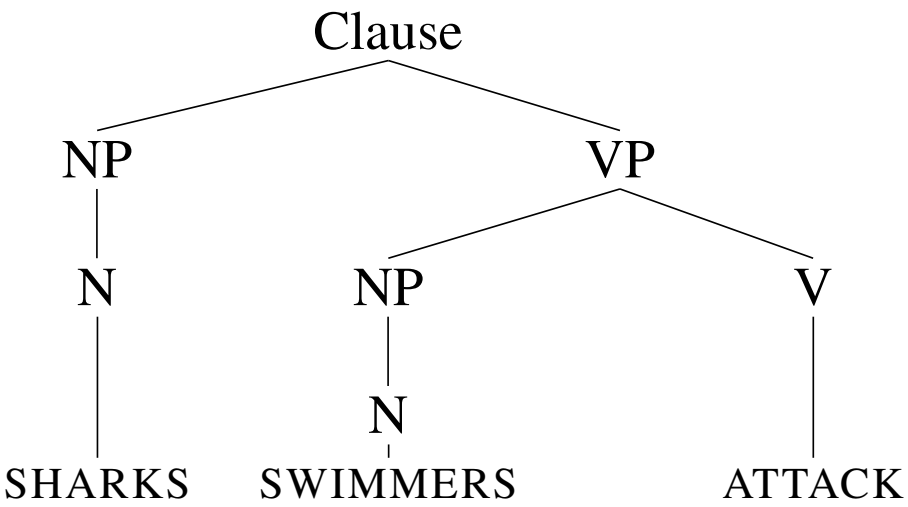
Complement: *what I was playing* in the clause
Did you hear what I was playing?

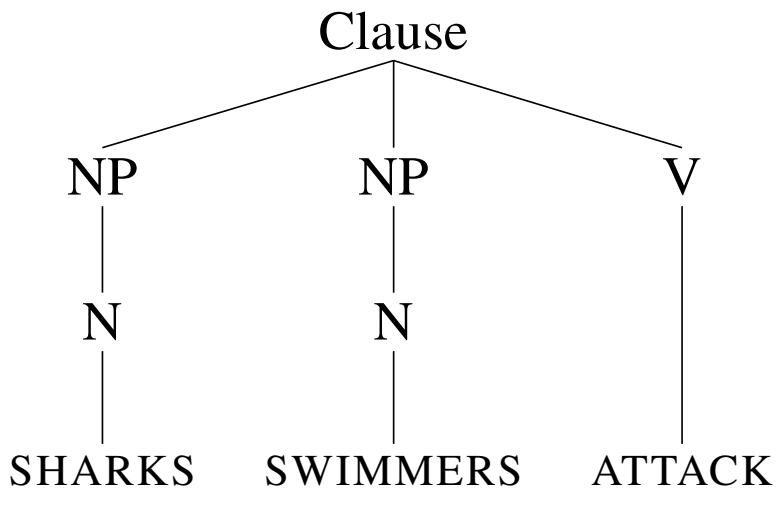
Modifier: *accurately* in the clause *Anyone can play accurately.*

Tree diagrams of expression structure

I will assume here that there are **empirical** questions about the **syntactic structures** of particular expressions.

For example, an SOV language like Japanese might or might not have a VP constituent:

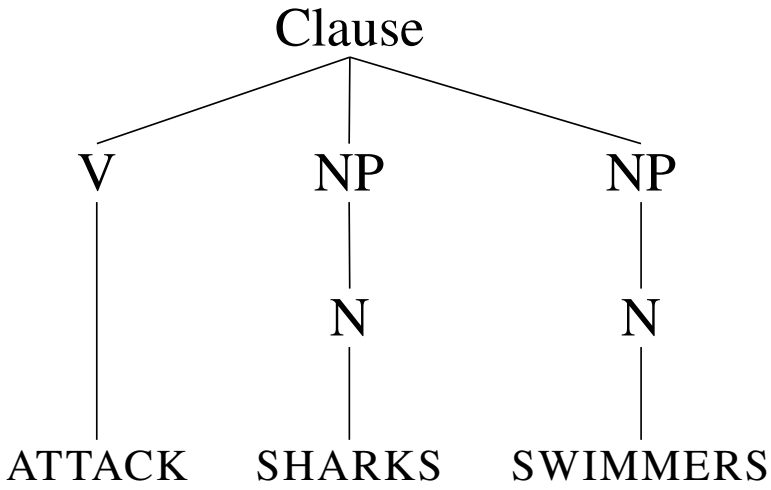




[English words in capitals are shown instead of Japanese words.]

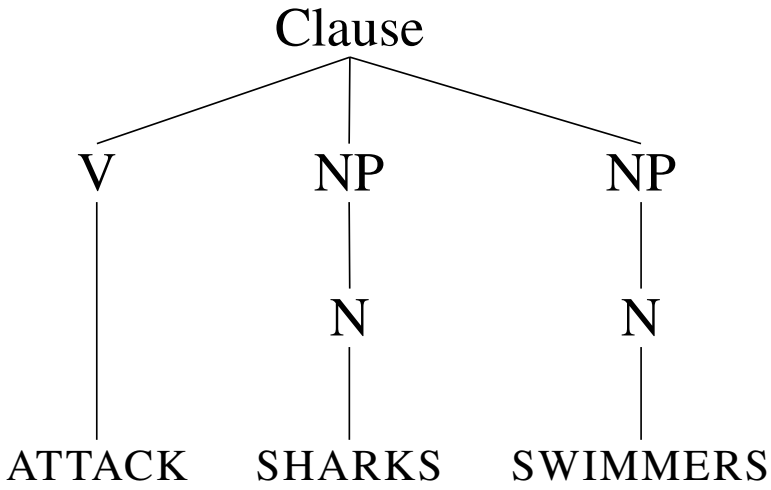
Tree diagrams of expression structure

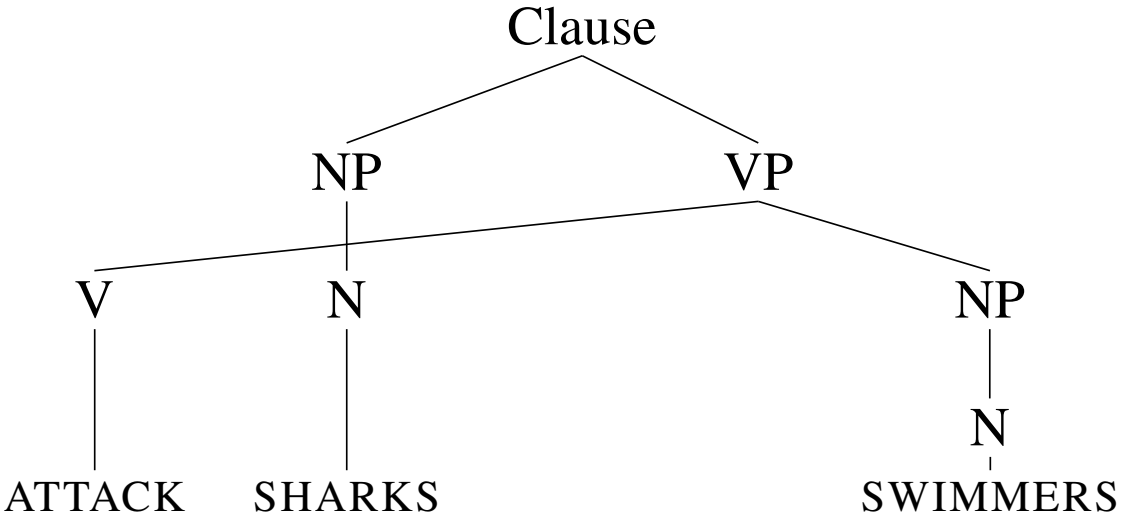
But a VSO language cannot have a VP:



Tree diagrams of expression structure

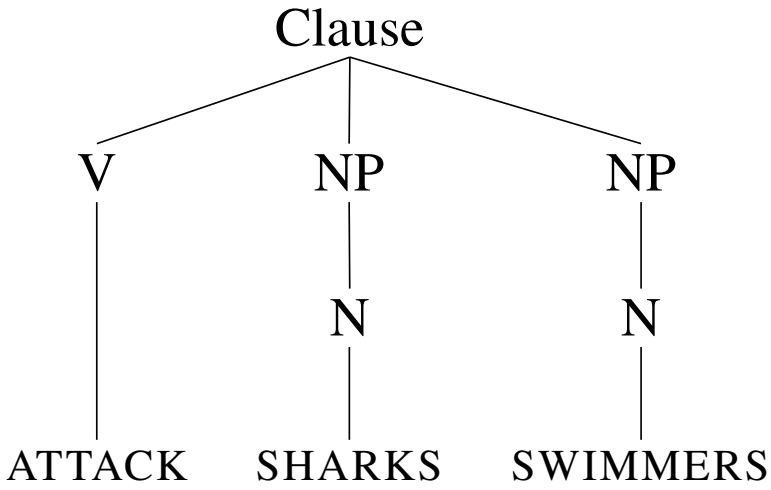
But a VSO language cannot have a VP:

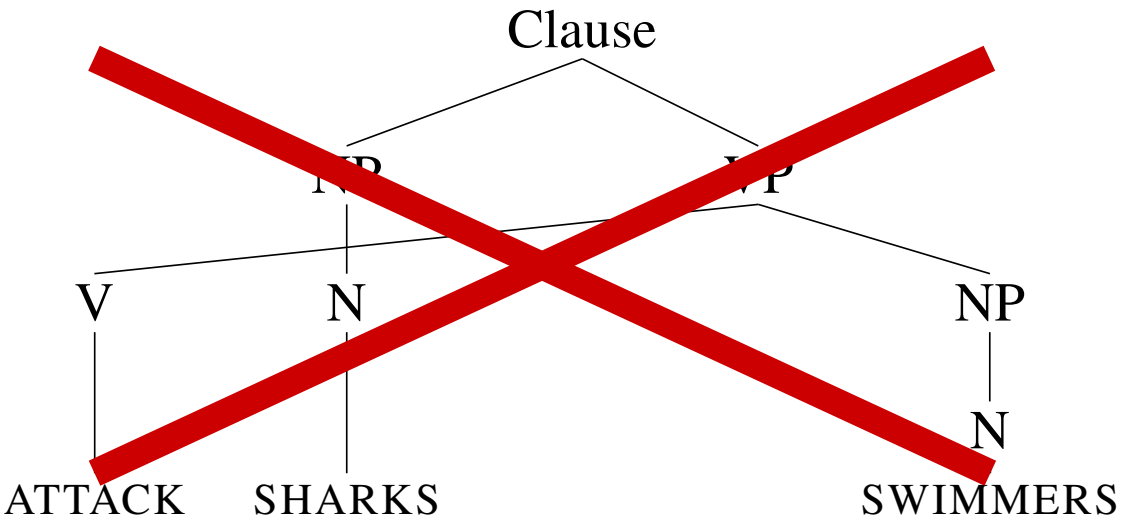




Tree diagrams of expression structure

But a VSO language cannot have a VP:





⟵ NOT A TREE!

Tree diagrams of expression structure

How can we tell which structure is right? Sometimes there are restrictions that offer potential evidence about structure. For example, in English:

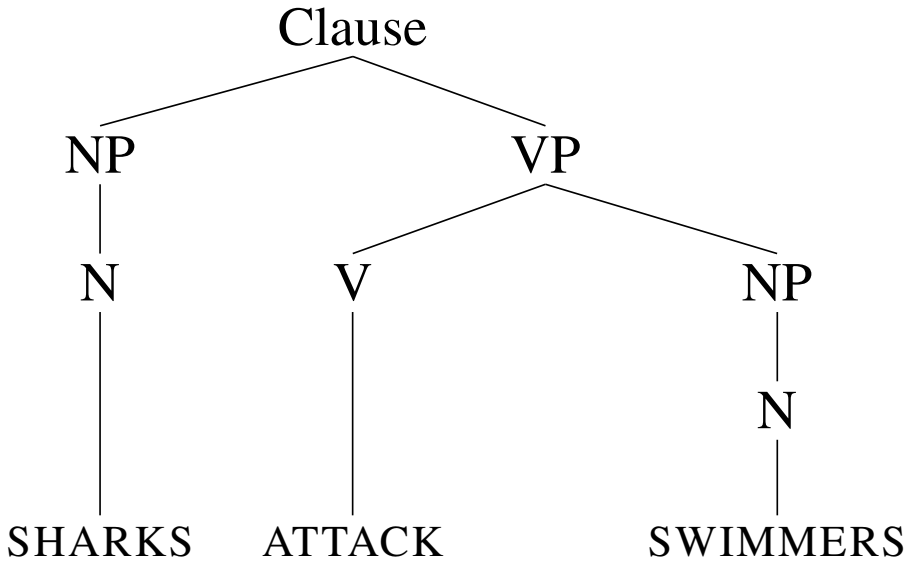
Sometimes sharks attack swimmers.	[INITIAL]
Sharks sometimes attack swimmers.	[BEFORE VP]
**Sharks attack sometimes swimmers.*	[INTERRUPTING VP]
Sharks attack swimmers sometimes.	[FINAL]

Tree diagrams of expression structure

How can we tell which structure is right? Sometimes there are restrictions that offer potential evidence about structure. For example, in English:

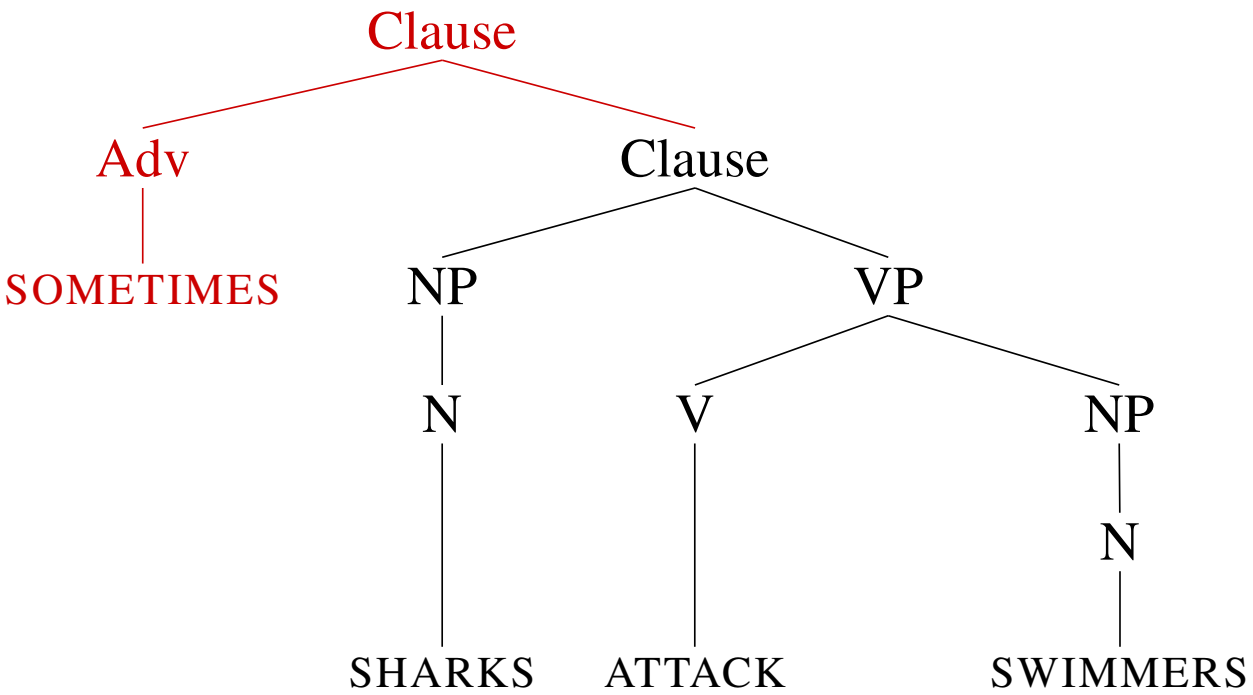
Sometimes sharks attack swimmers.	[INITIAL]
Sharks sometimes attack swimmers.	[BEFORE VP]
**Sharks attack sometimes swimmers.*	[INTERRUPTING VP]
Sharks attack swimmers sometimes.	[FINAL]

Suppose the clause without the modifier *sometimes* has roughly this structure:



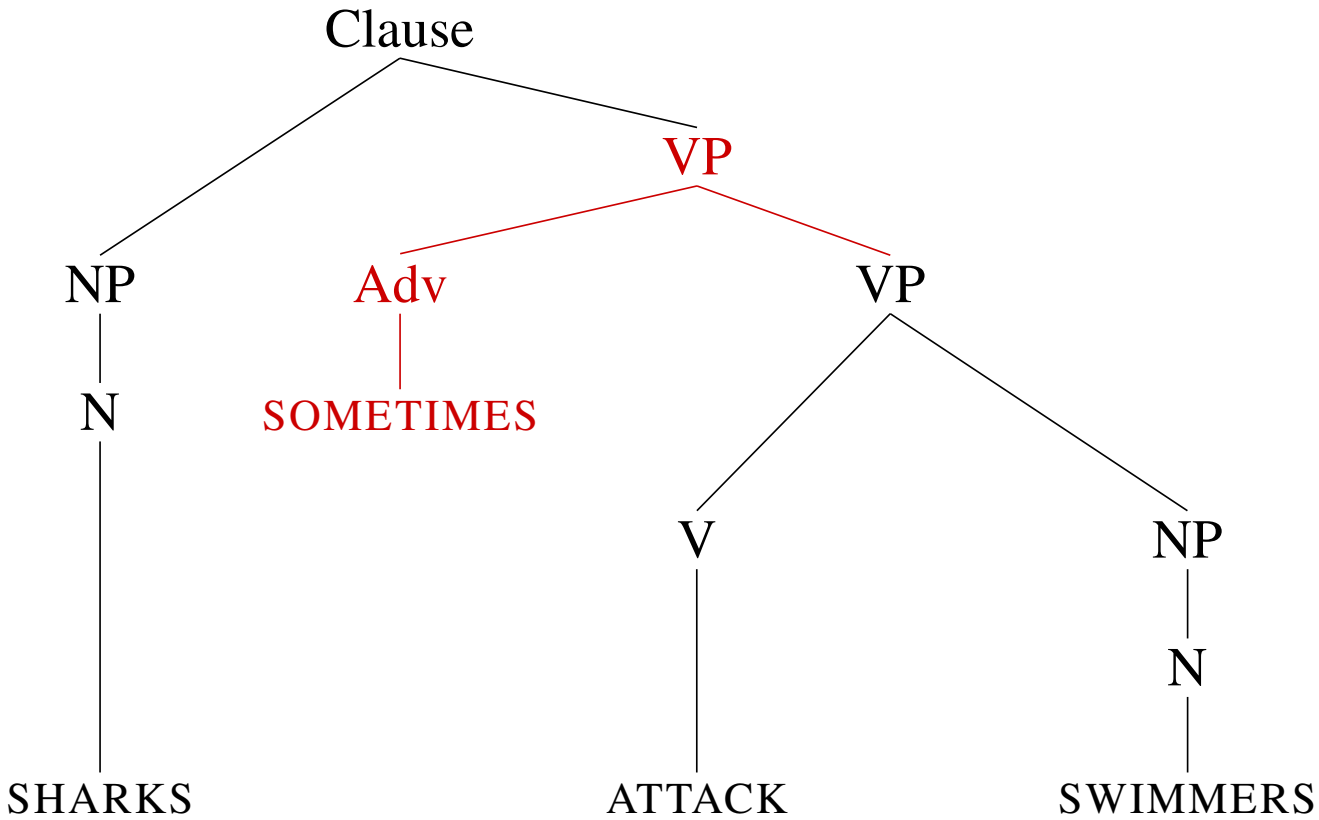
Tree diagrams of expression structure

Hypothesis: You can attach an Adverb to the beginning of the Clause. . .



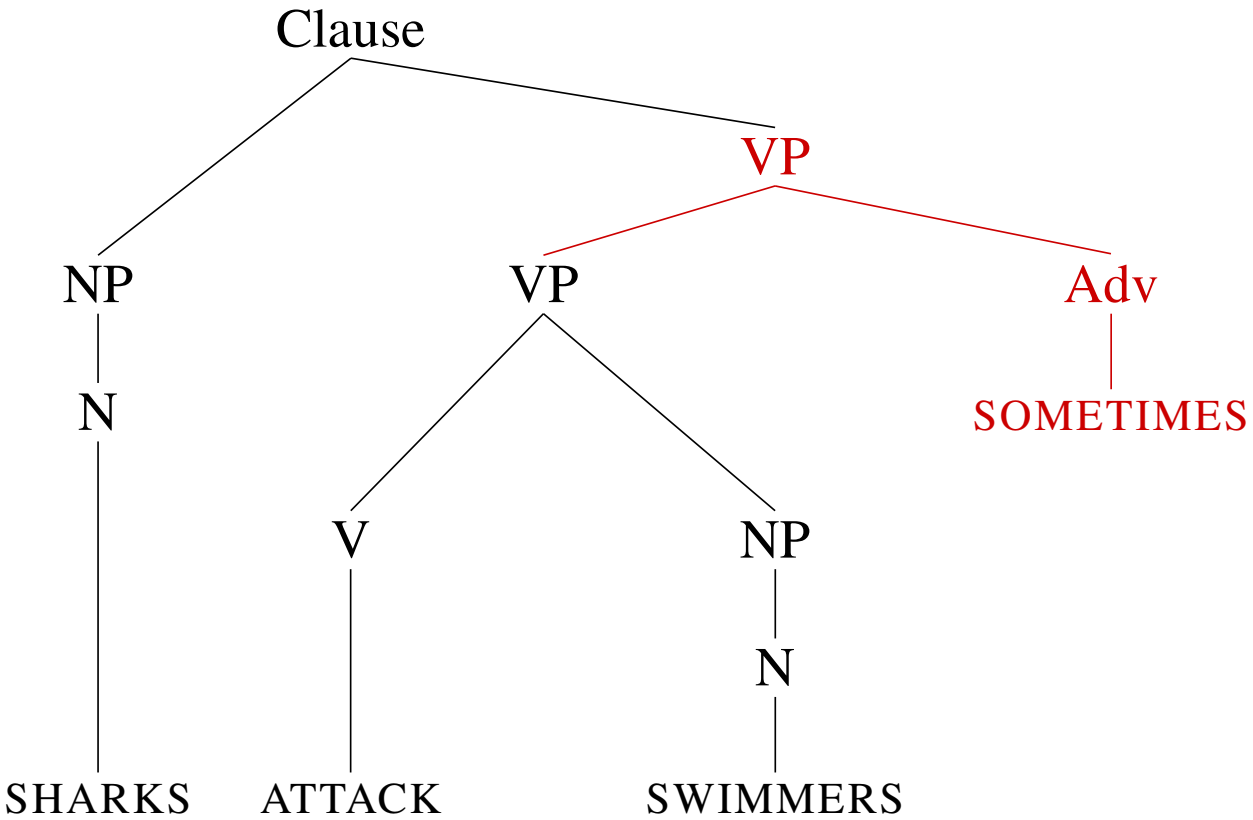
Tree diagrams of expression structure

Hypothesis: You can attach an Adverb to the beginning of the Clause, or to the beginning of the VP...



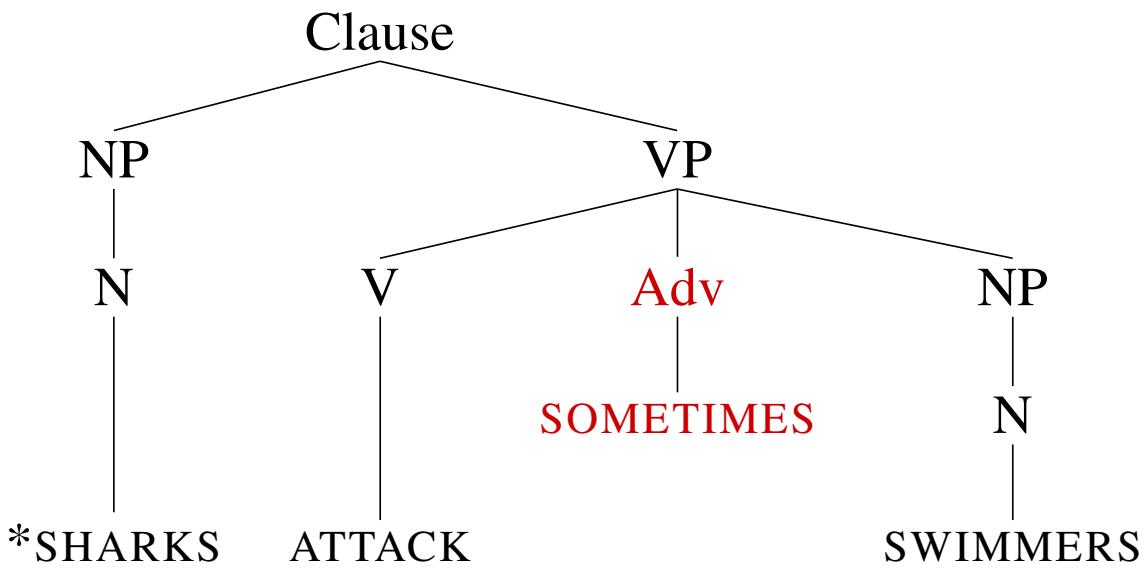
Tree diagrams of expression structure

Hypothesis: You can attach an Adverb to the beginning of the Clause, or to the beginning of the VP, or to the end of the Clause. . .



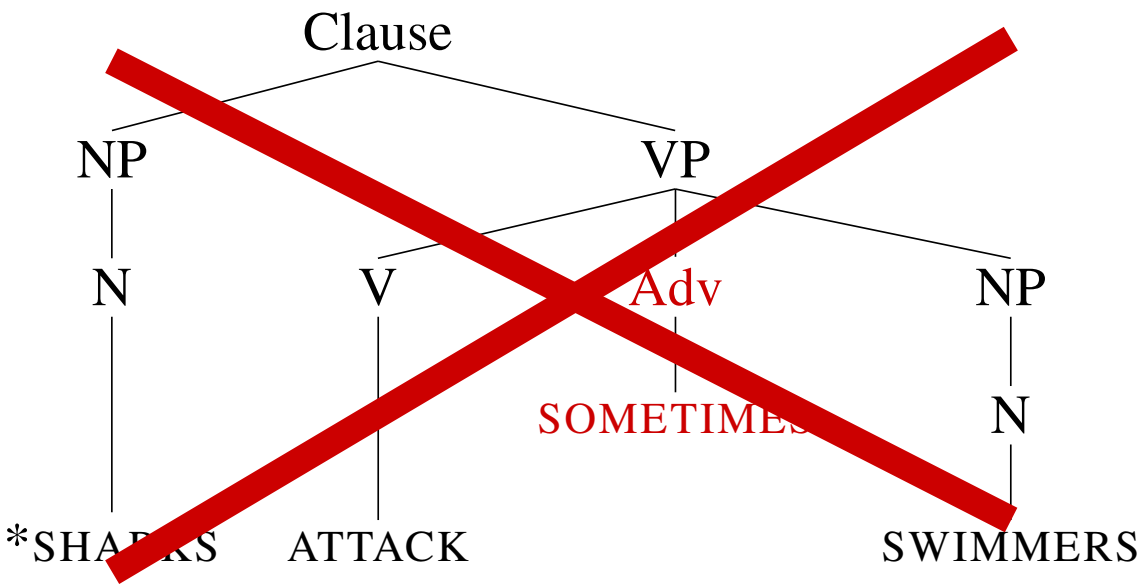
Tree diagrams of expression structure

Hypothesis: You can attach an Adverb to the beginning of the Clause, or to the beginning of the VP, or to the end of the Clause, but **the VP must not be interrupted**.



Tree diagrams of expression structure

Hypothesis: You can attach an Adverb to the beginning of the Clause, or to the beginning of the VP, or to the end of the Clause, but **the VP must not be interrupted**.



Tree diagrams of expression structure

The idea is that assuming a VP permits an elegant description of the restriction: **An adverb modifying a clause must not separate the words of the VP**.

In fact a more general statement can be given—

Tree diagrams of expression structure

The idea is that assuming a VP permits an elegant description of the restriction: **An adverb modifying a clause must not separate the words of the VP**.

In fact a more general statement can be given—**An adverb modifying a clause must not separate the words of any of its subconstituents**:

Tree diagrams of expression structure

The idea is that assuming a VP permits an elegant description of the restriction: **An adverb modifying a clause must not separate the words of the VP**.

In fact a more general statement can be given—**An adverb modifying a clause must not separate the words of any of its subconstituents**:

Sometimes sharks [attack [lone swimmers.]]

Sharks sometimes [attack [lone swimmers.]]

*Sharks [attack sometimes [lone swimmers.]]

*Sharks [attack [lone sometimes swimmers.]]

Sharks [attack [lone swimmers] sometimes.]

Tree diagrams of expression structure

A grammar should accurately describe the structures of **all expressions of the language**, and assign meanings in a compositional way (cf. the Westerstahl and Jacobson courses), and ideally should provide adequate explanations for general facts about them.

How can a grammar do all this?

Tree diagrams of expression structure

A grammar should accurately describe the structures of **all expressions of the language**, and assign meanings in a compositional way (cf. the Westerstahl and Jacobson courses), and ideally should provide adequate explanations for general facts about them.

How can a grammar do all this?

Answers began to emerge in the 20th century from mathematical logic.

A logic must provide for **each of unboundedly many formulas** a structure that permits inference rules to derive other formulas from it.

It must also provide **for each formula** a statement of the conditions under which it is true.

The methods for doing these things led to ways of constructing comprehensive and explicit grammars for human languages.