Trees, Recursion, and Natural Language

John DeNero

Source code: http://denero.org/content/misc/parse.zip

Ambiguity

Syntactic Ambiguity in English

Syntactic Ambiguity in English

Programs must be written for people to read[1]

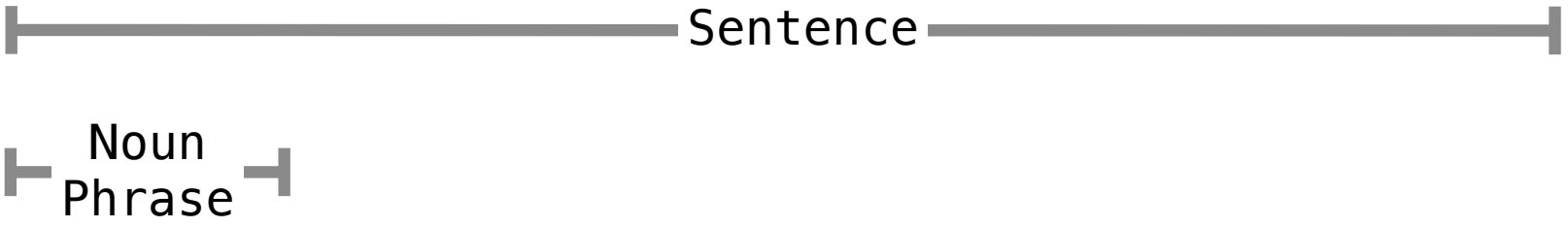
[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

Syntactic Ambiguity in English

Programs must be written for people to read[1]

[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

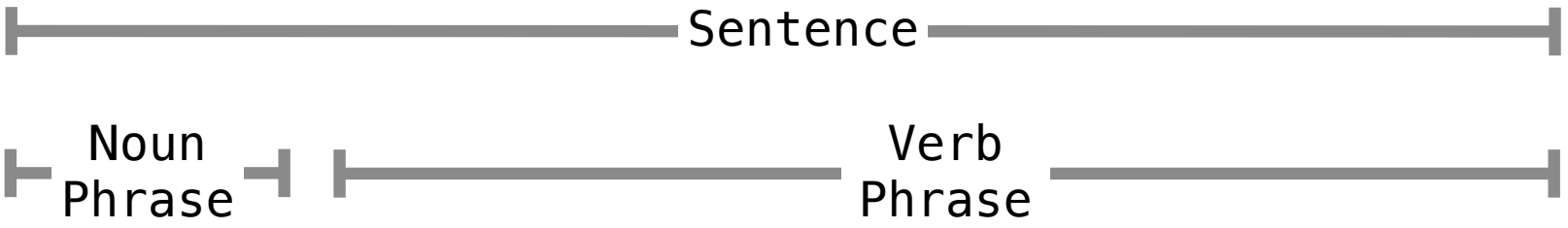
Syntactic Ambiguity in English



Programs must be written for people to read[1]

[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

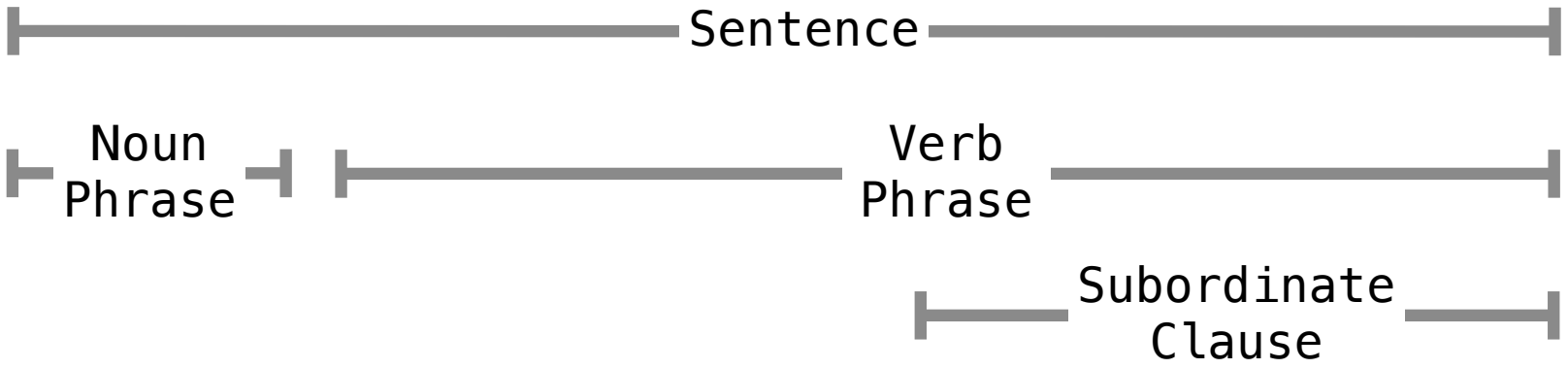
Syntactic Ambiguity in English



Programs must be written for people to read[1]

[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

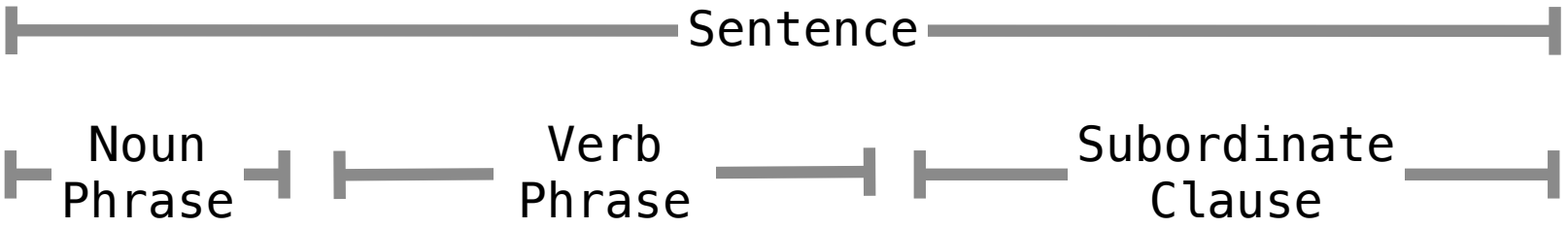
Syntactic Ambiguity in English



Programs must be written for people to read[1]

[1]Preface of Structure and Interpretation of Computer Programs by Harold Abelson and Gerald Sussman with Julie Sussman

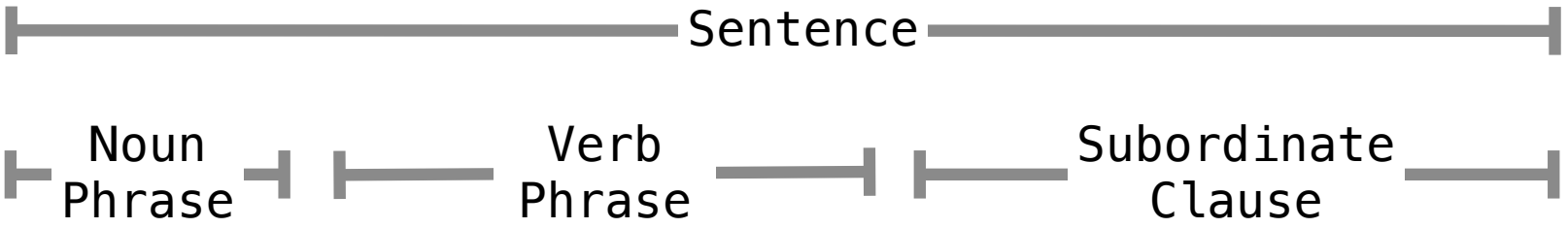
Syntactic Ambiguity in English



Programs must be written for people to read[1]

[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

Syntactic Ambiguity in English

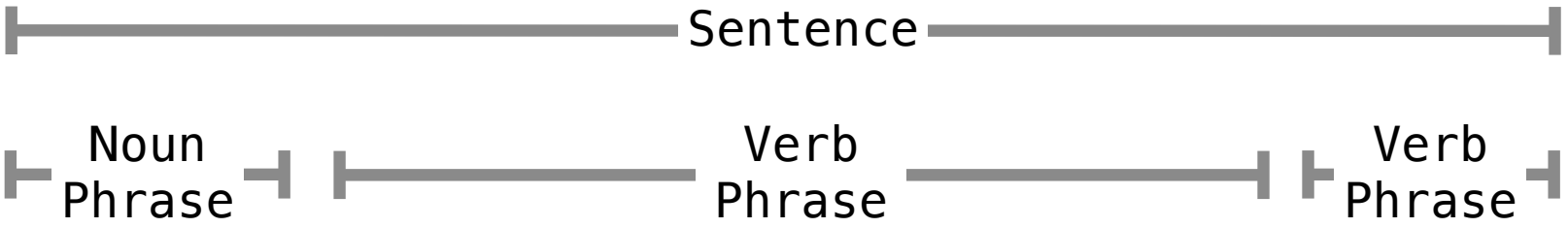


Programs must be written for people to read[1]

A program must first be written for it to crash

[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

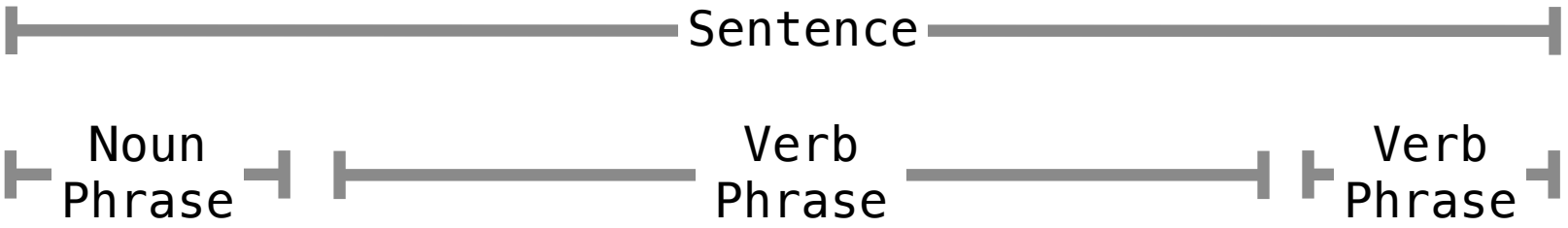
Syntactic Ambiguity in English



Programs must be written for people to read[1]

[1]Preface of Structure and Interpretation of Computer Programs by Harold Abelson and Gerald Sussman with Julie Sussman

Syntactic Ambiguity in English



Programs must be written for people to read[1]

Programs must be written for people to be useful

[1]Preface of Structure and Interpretation of Computer Programs
by Harold Abelson and Gerald Sussman with Julie Sussman

Syntactic Ambiguity in English

Programs must be written for people to read

Definitions from the New Oxford American Dictionary

Syntactic Ambiguity in English

pro•gram (*noun*)
  a series of coded software instructions.

pro•gram (*verb*)
  provide a computer with coded instructions.

Programs must be written for people to read

Definitions from the New Oxford American Dictionary

Syntactic Ambiguity in English

pro•gram (*noun*)
  a series of coded software instructions.

pro•gram (*verb*)
  provide a computer with coded instructions.

Programs must be written for people to read

must (*verb*)
  be obliged to.

must (*noun*)
  dampness or mold.

Definitions from the New Oxford American Dictionary

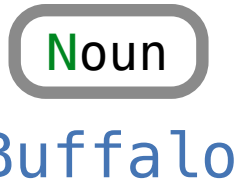
Syntax Trees

Representing Syntactic Structure

Representing Syntactic Structure

Buffalo

Representing Syntactic Structure



Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/

Noun

Buffalo

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/

Noun

Buffalo buffalo

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/

Noun Verb

Buffalo buffalo

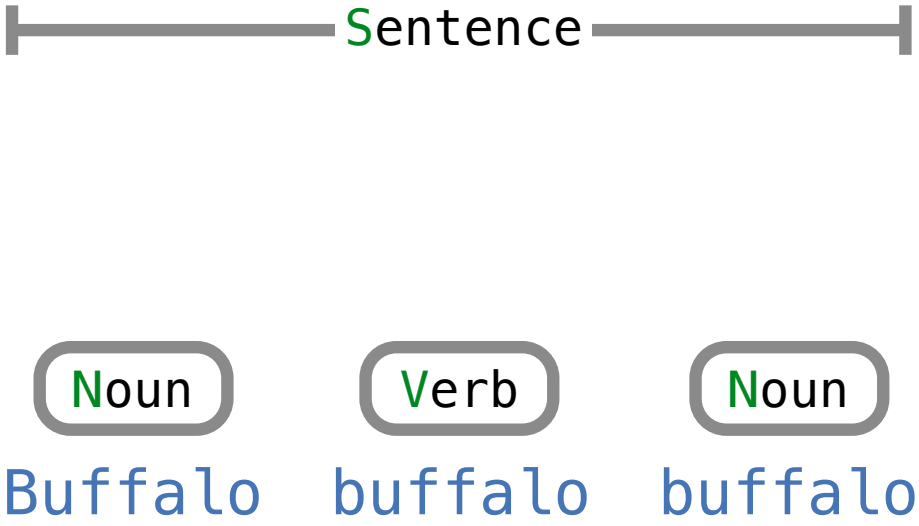
Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/

Noun	Verb	Noun
Buffalo	buffalo	buffalo

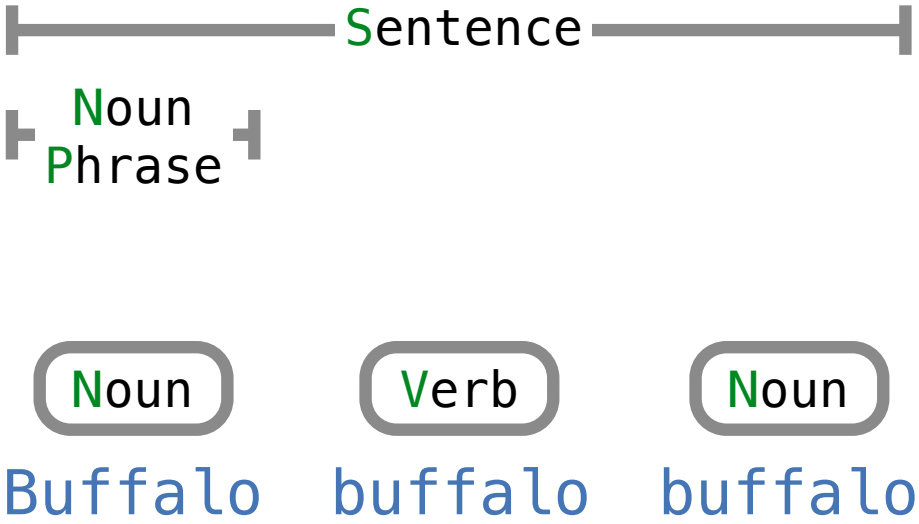
Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



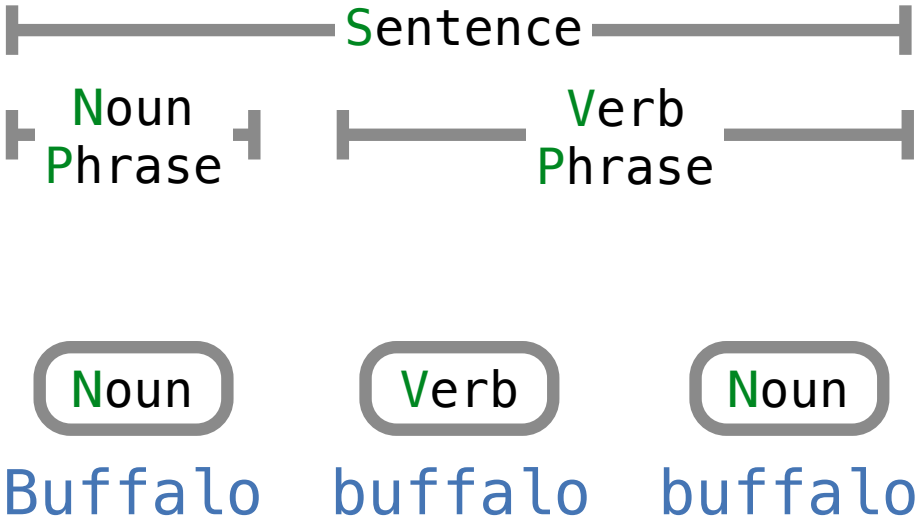
Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



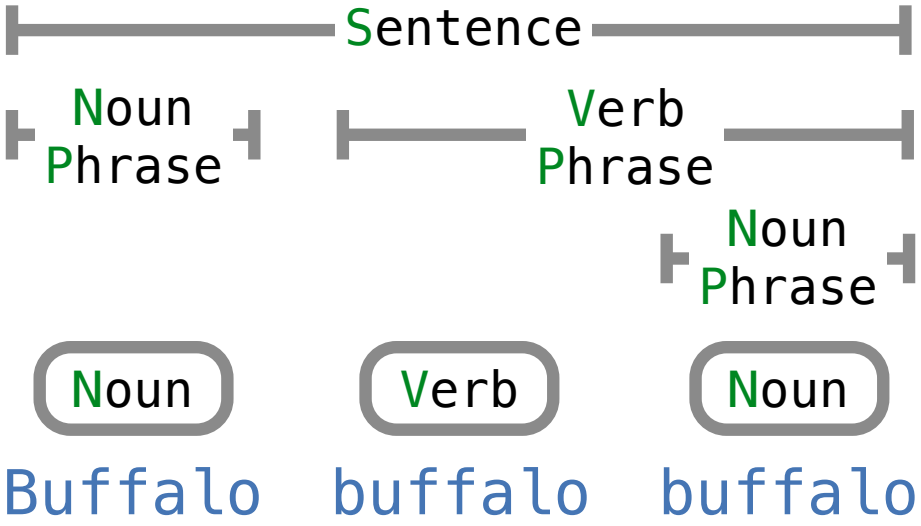
Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



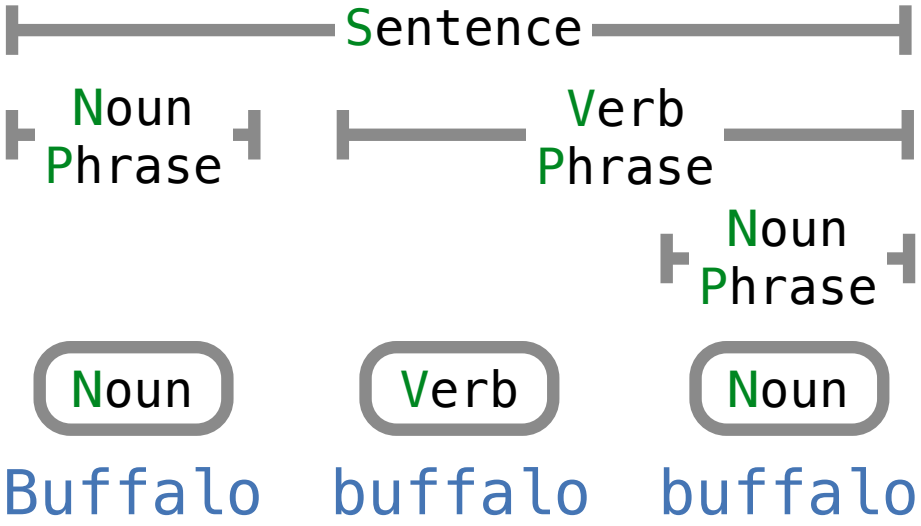
Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



Representing Syntactic Structure

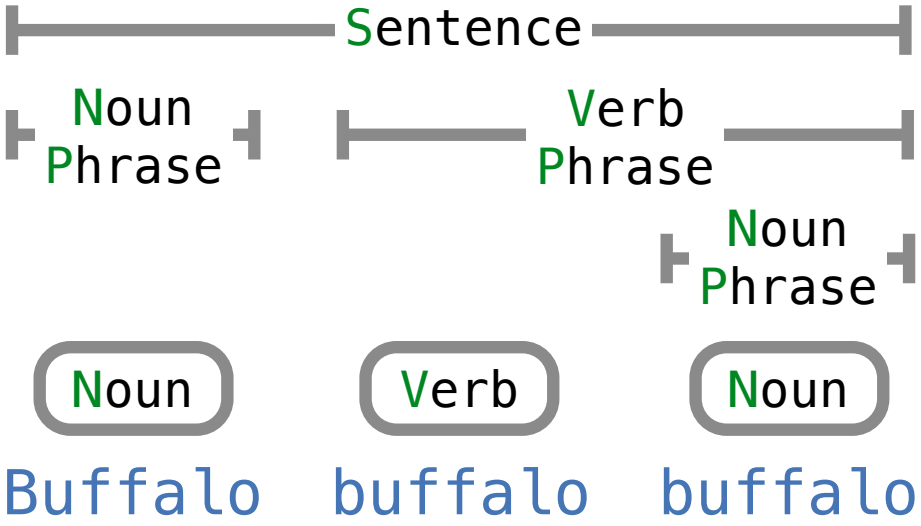
Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/

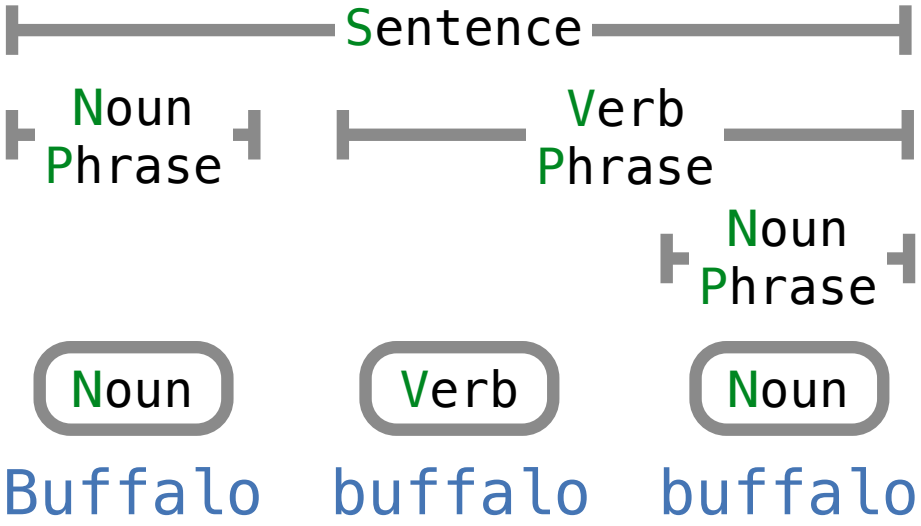


A **Tree** represents a phrase:

- **tag** -- What kind of phrase (e.g., S, NP, VP)

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/

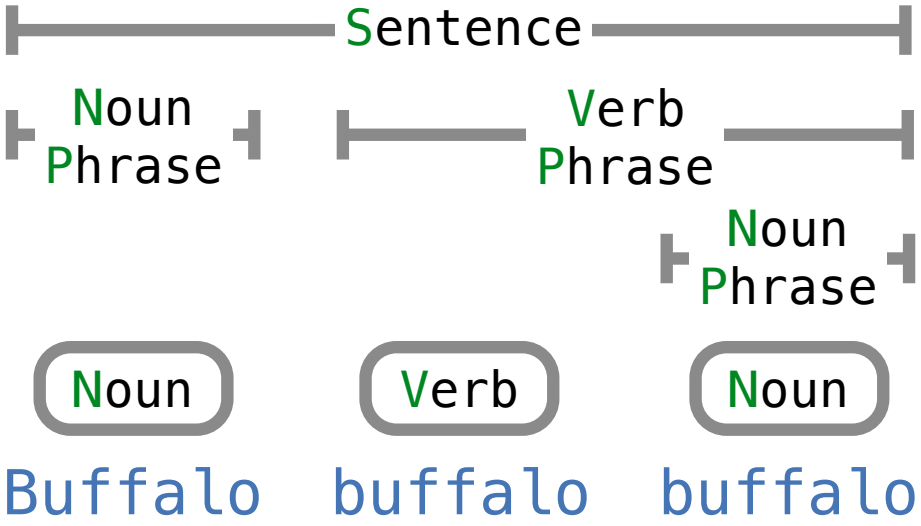


A **Tree** represents a phrase:

- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



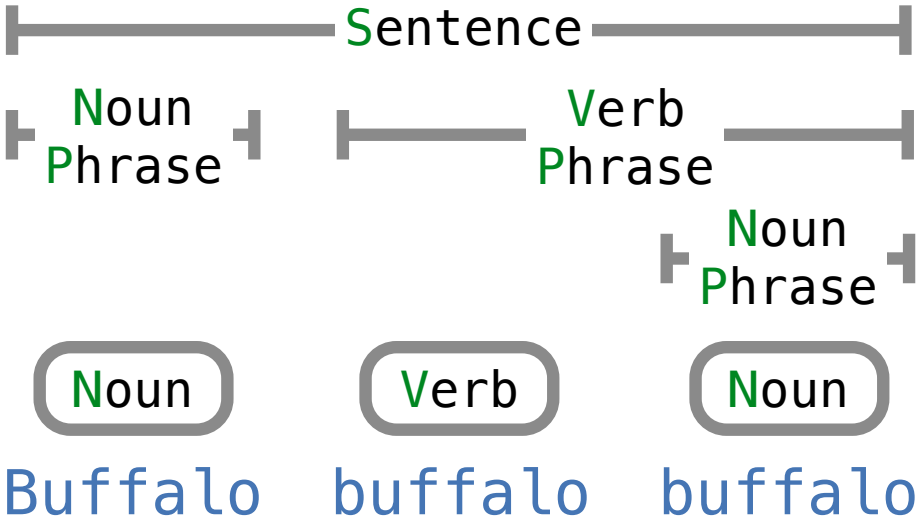
A **Tree** represents a phrase:

- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

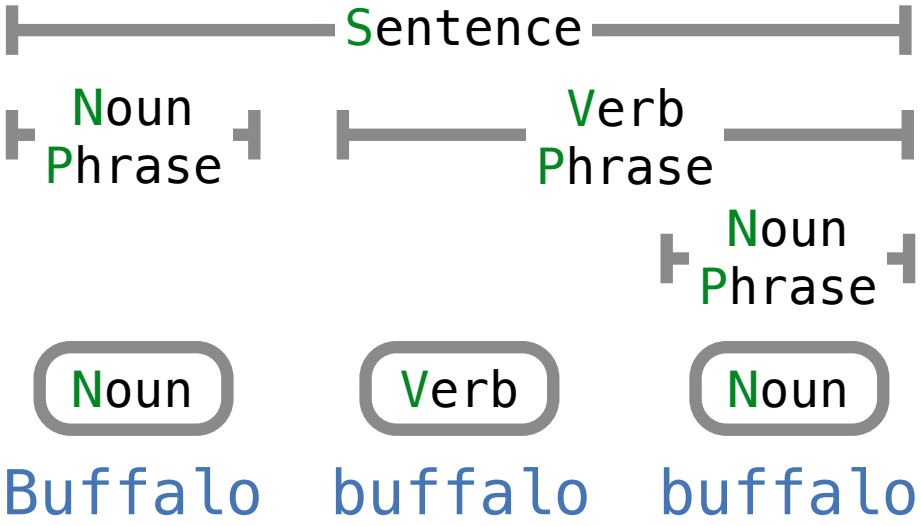
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

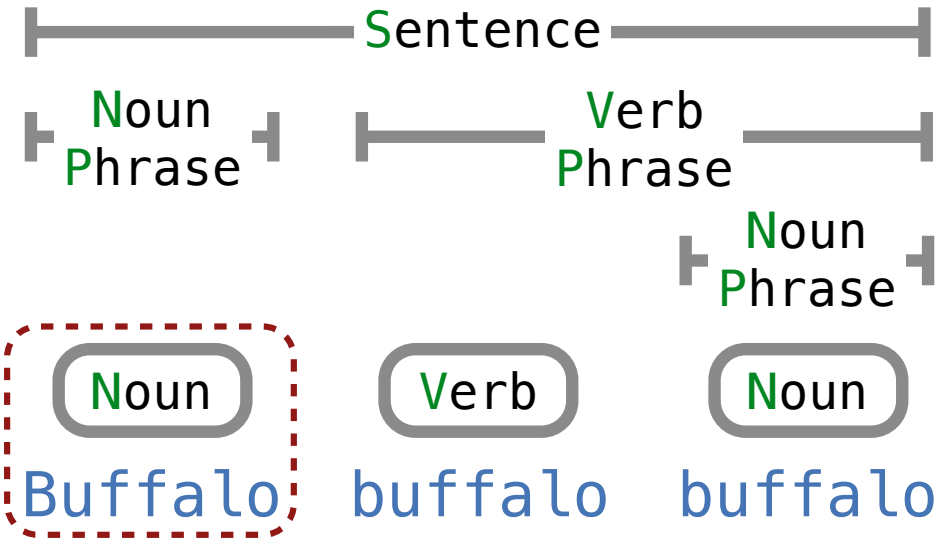
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

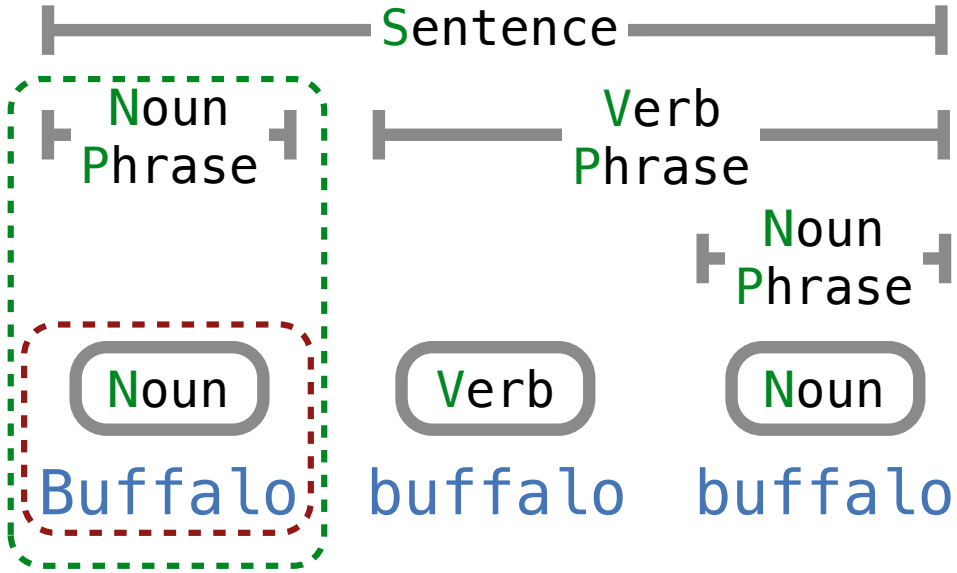
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

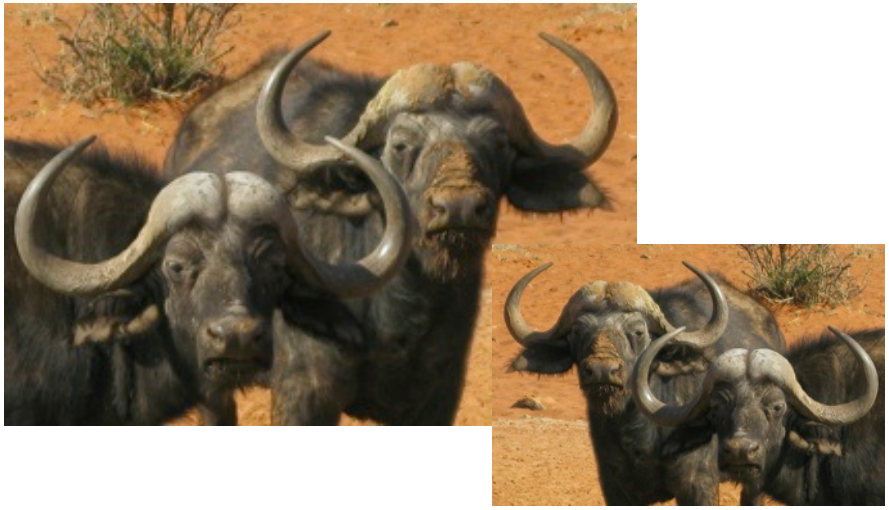
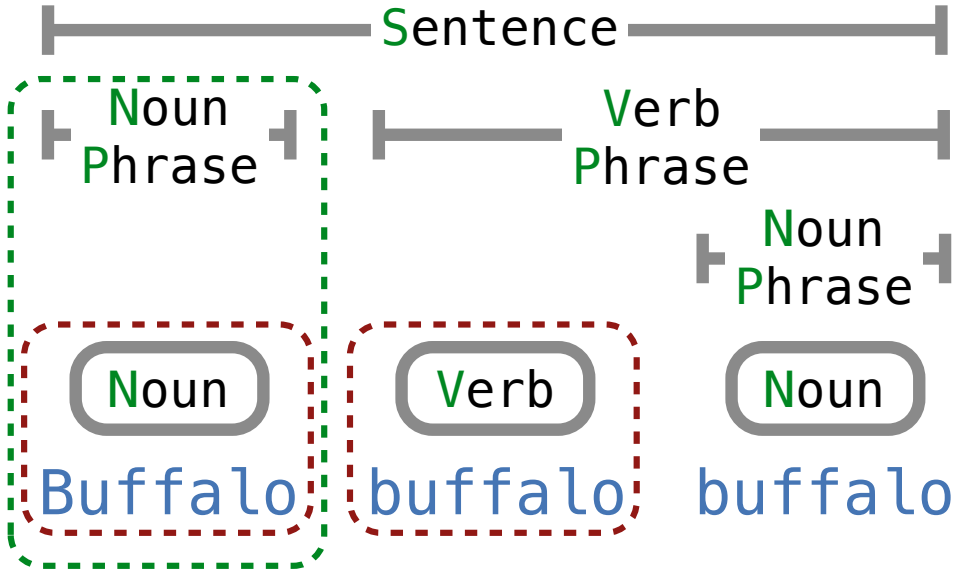


Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

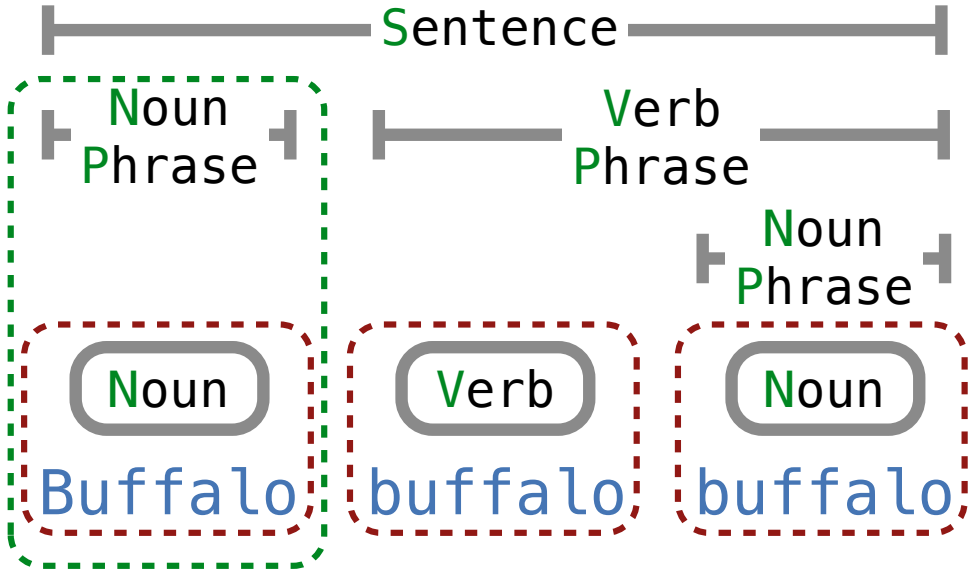
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

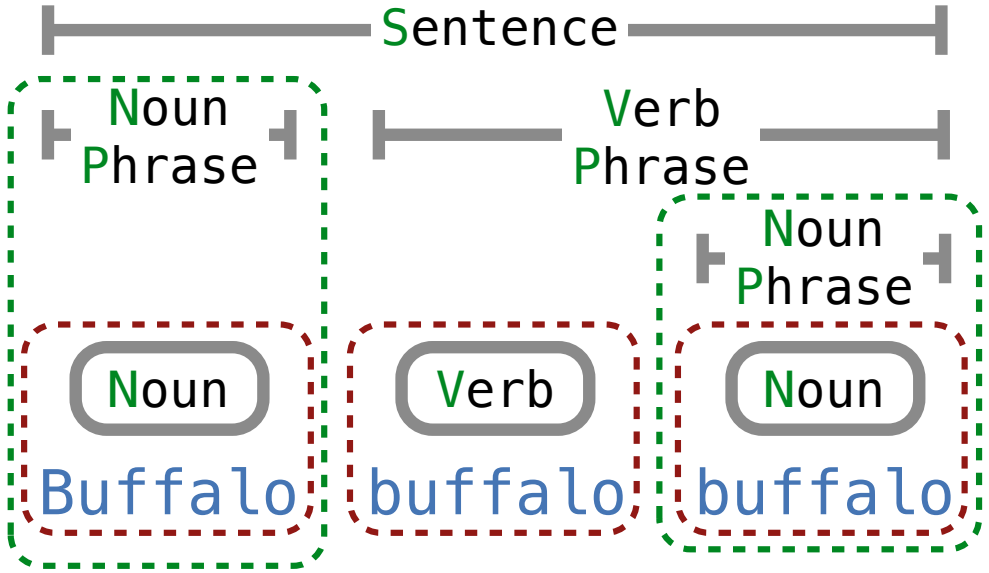
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

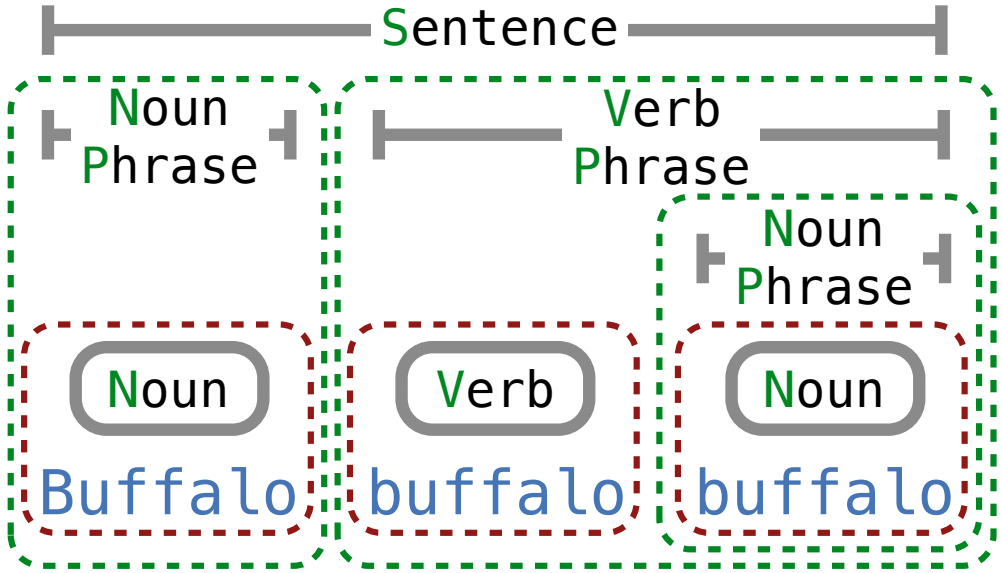
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

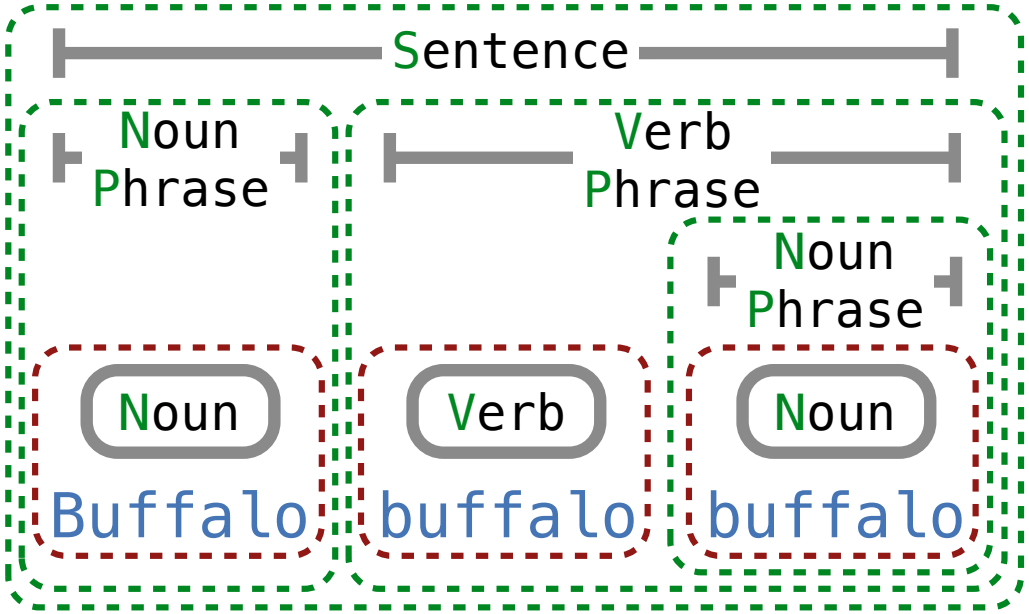
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

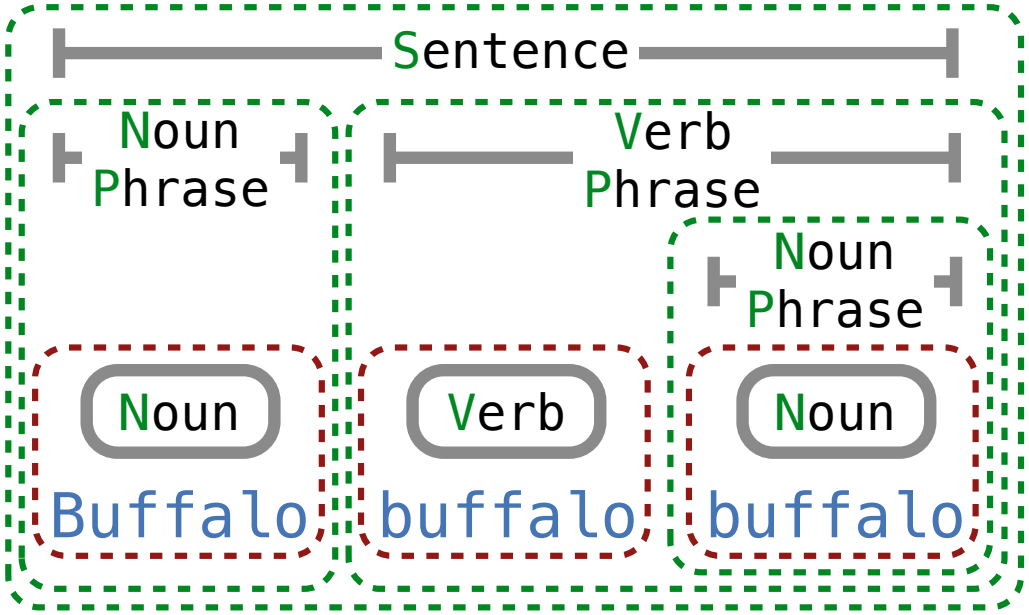
- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

Representing Syntactic Structure

Photo by Vince O'Sullivan licensed under
http://creativecommons.org/licenses/by-nc-nd/2.0/



A **Tree** represents a phrase:

- **tag** -- What kind of phrase (e.g., S, NP, VP)
- **branches** -- Sequence of Tree or Leaf components

A **Leaf** represents a single word:

- **tag** -- What kind of word (e.g., N, V)
- **word** -- The word

(Demo)

Grammars

Context-Free Grammar Rules

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

$$S \rightarrow NP\ VP$$

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

S → NP VP

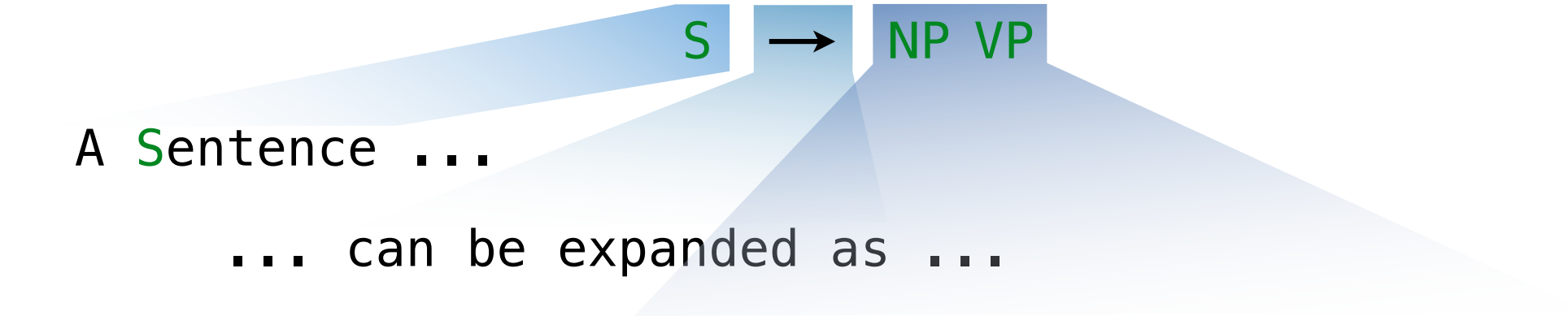
A Sentence ...

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.



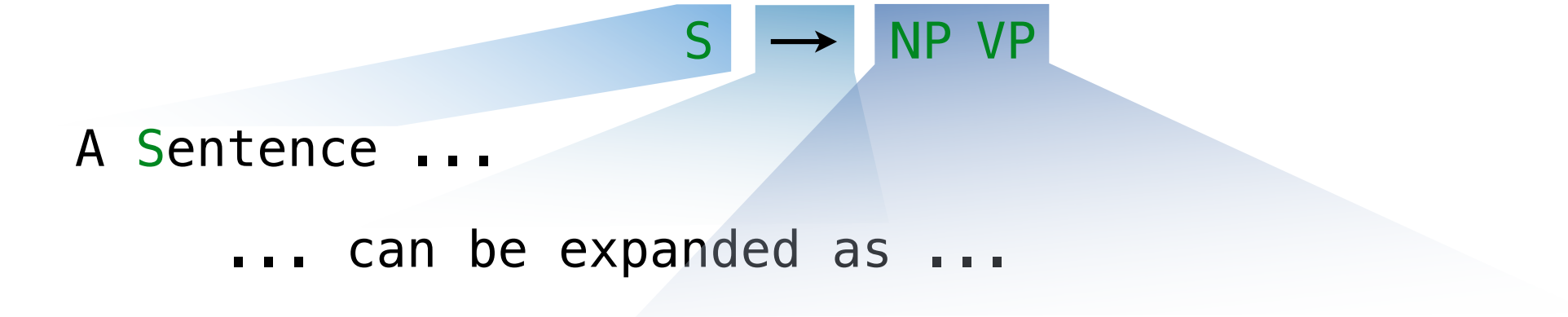
A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.



A Sentence ...

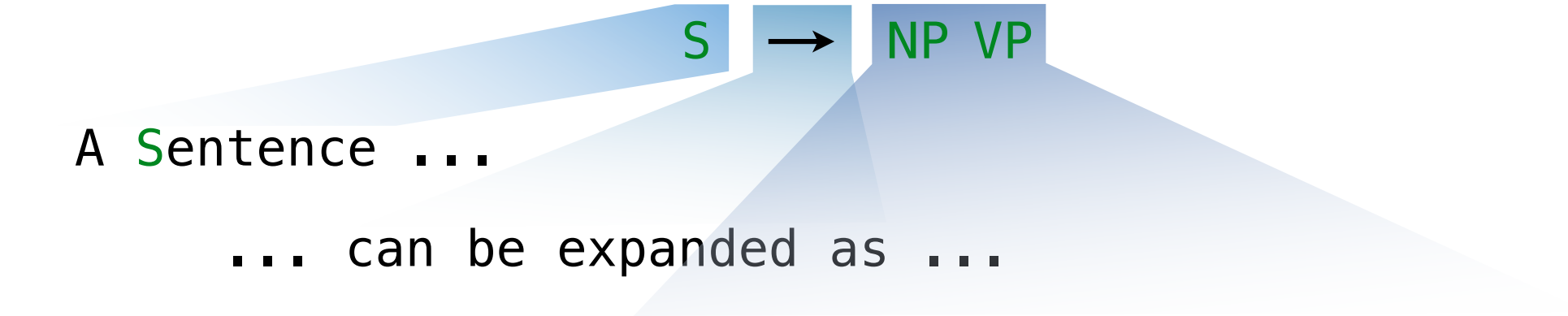
... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

S → NP VP

A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

S → NP VP

A Sentence ...

... can be expanded as ...

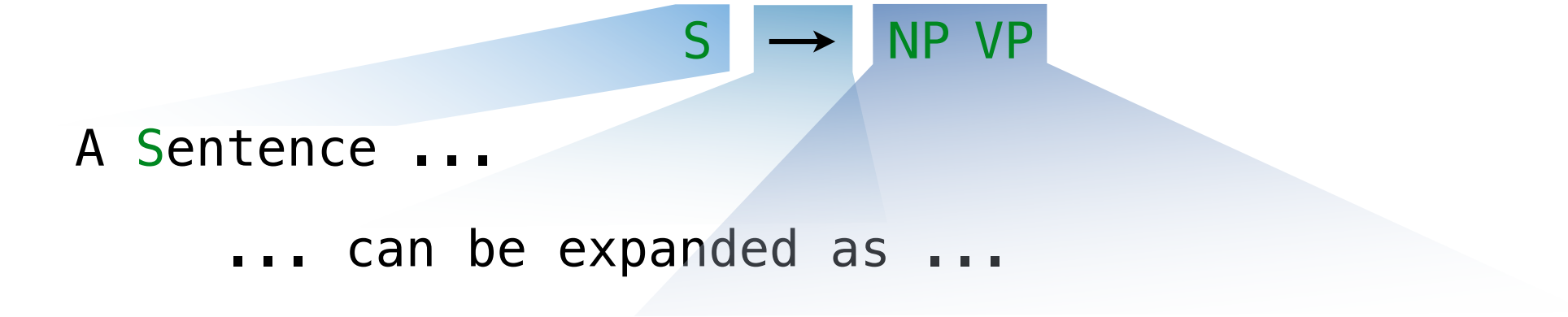
... a Noun Phrase then a Verb Phrase.

S → NP VP

NP → N

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

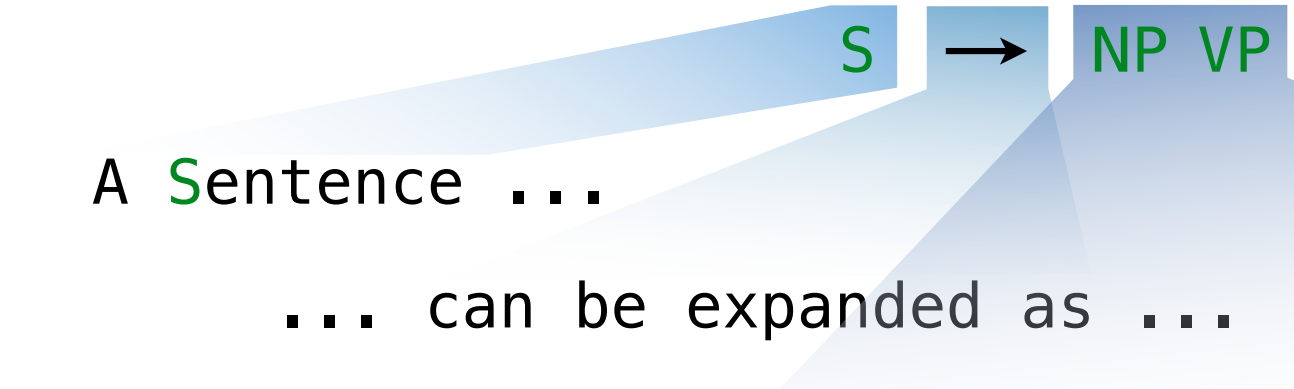


S → NP VP

NP → N

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.



A Sentence ...

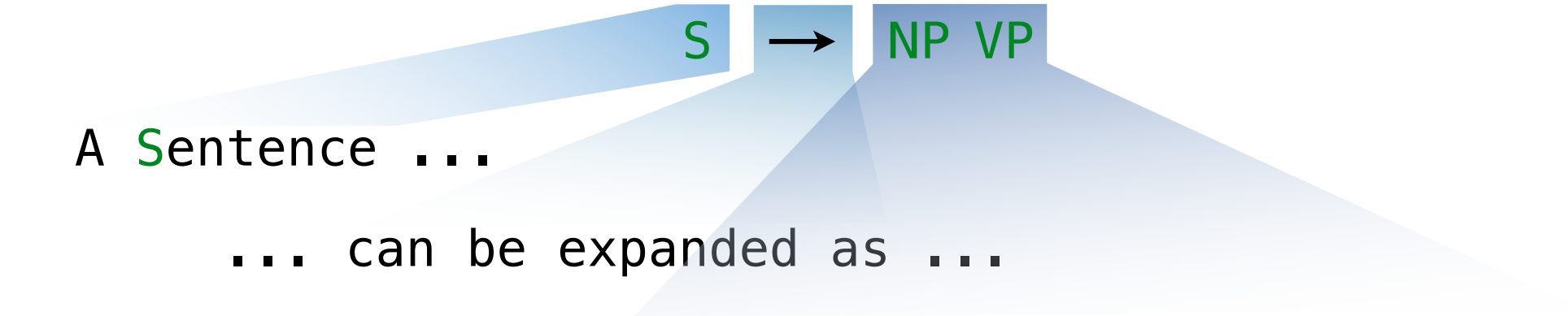
... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

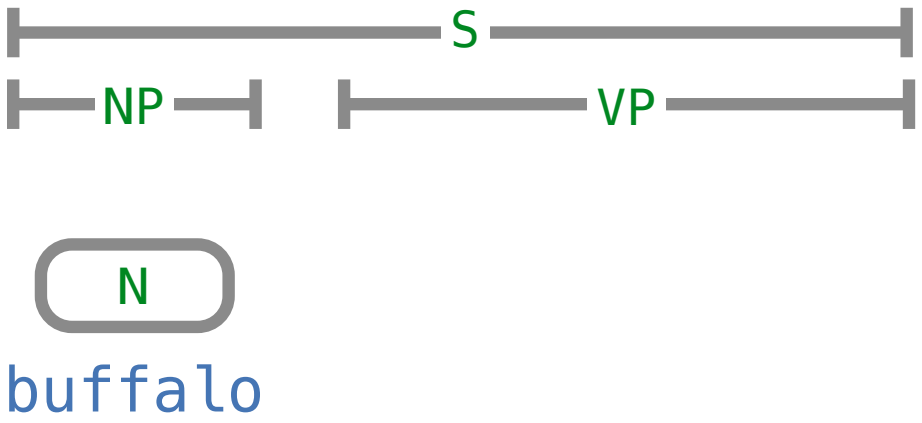
S → NP VP
NP → N
N → buffalo

Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.



S → NP VP
NP → N
N → buffalo



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

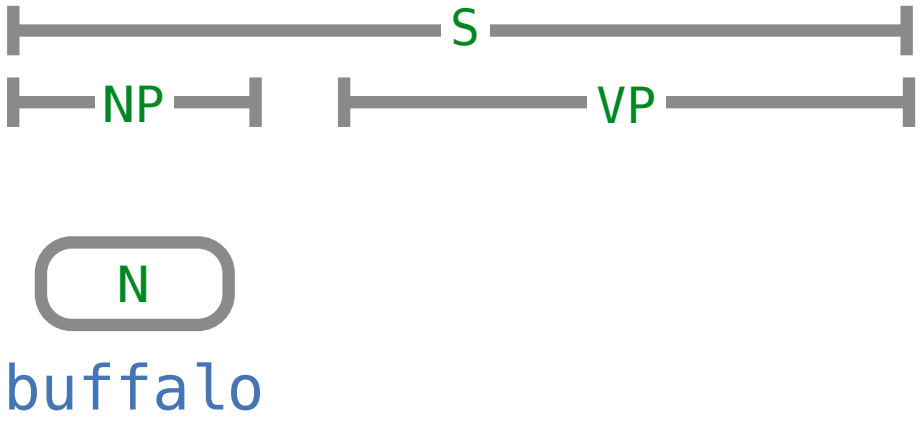
S → NP VP

A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

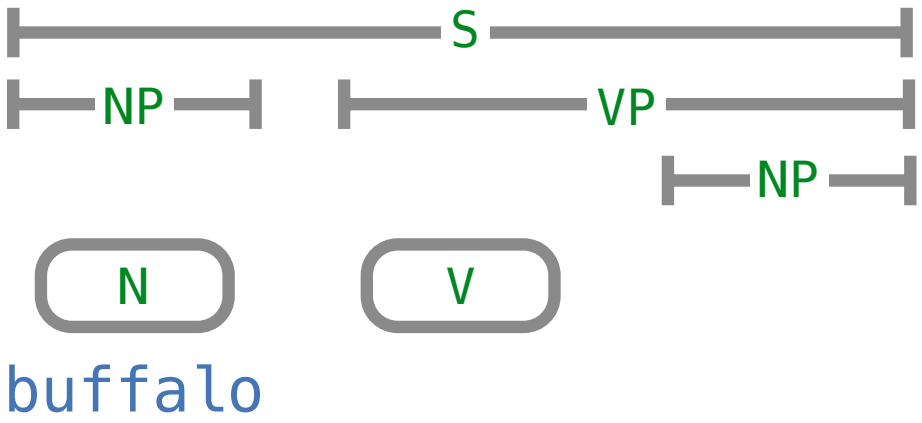
S → NP VP

A Sentence ...

... can be expanded as ...

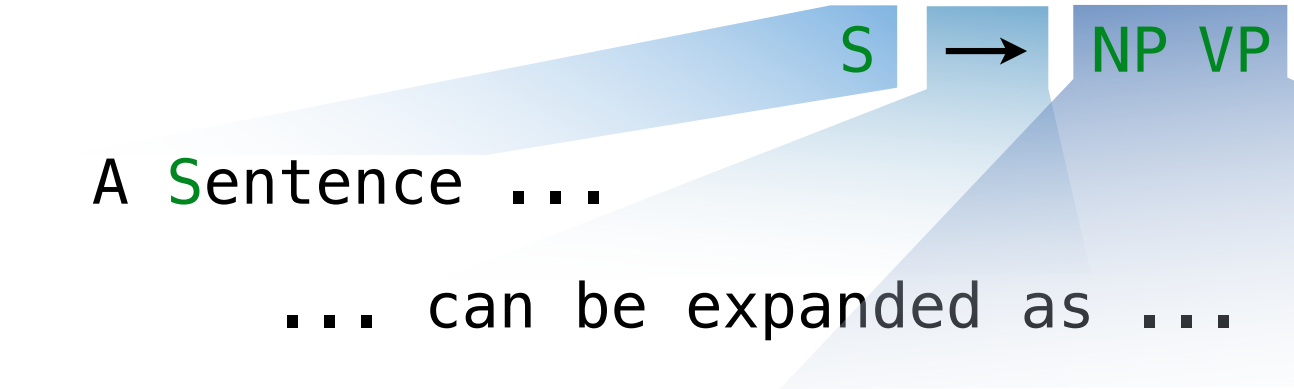
... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

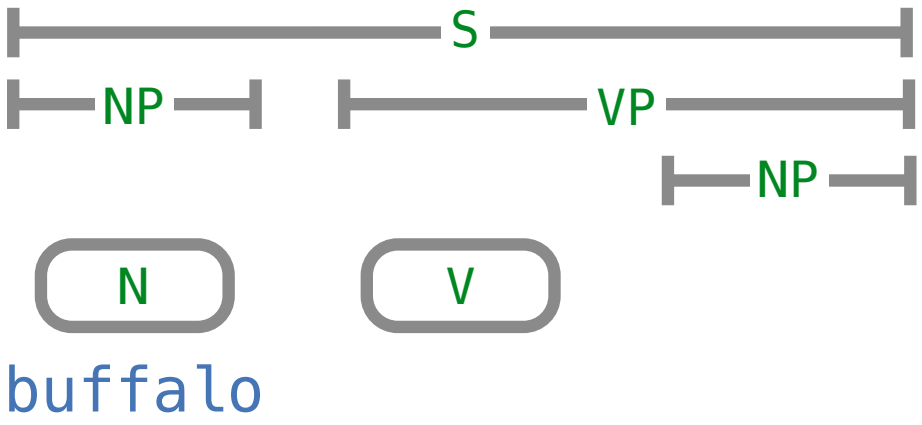


A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP
V → buffalo



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

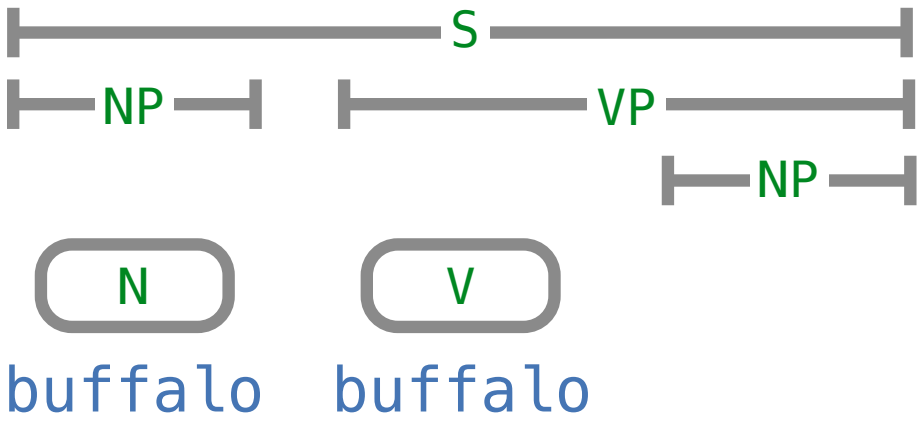
S → NP VP

A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP
V → buffalo



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

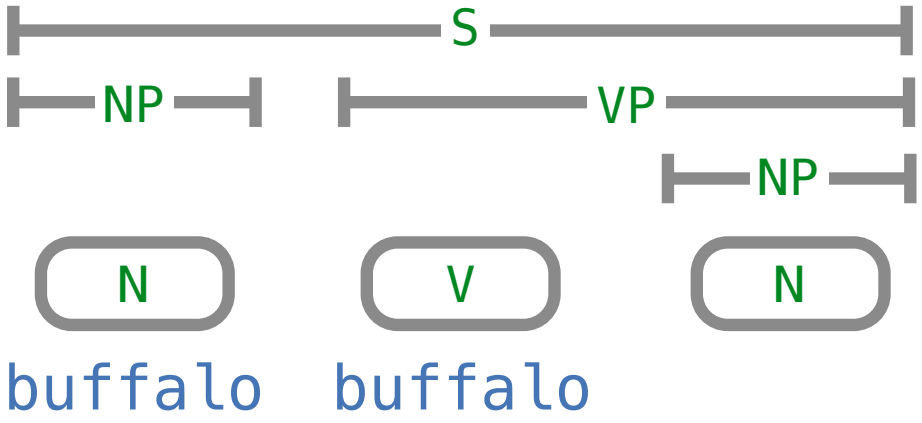
S → NP VP

A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP
V → buffalo



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

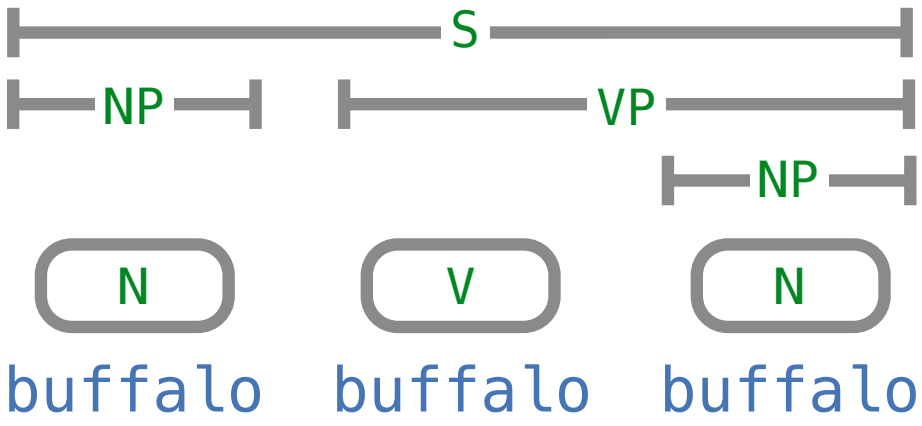
S → NP VP

A Sentence ...

... can be expanded as ...

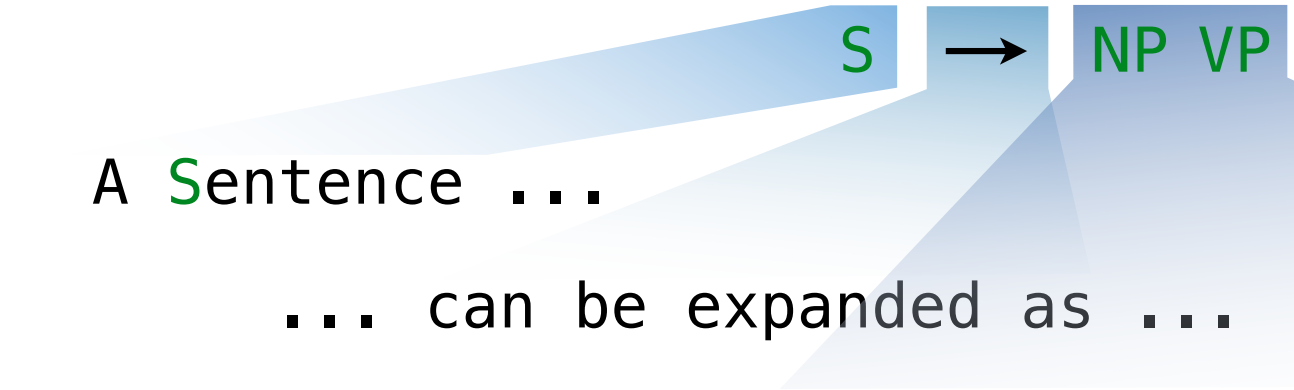
... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP
V → buffalo



Context-Free Grammar Rules

A grammar rule describes how a tag can be expanded.

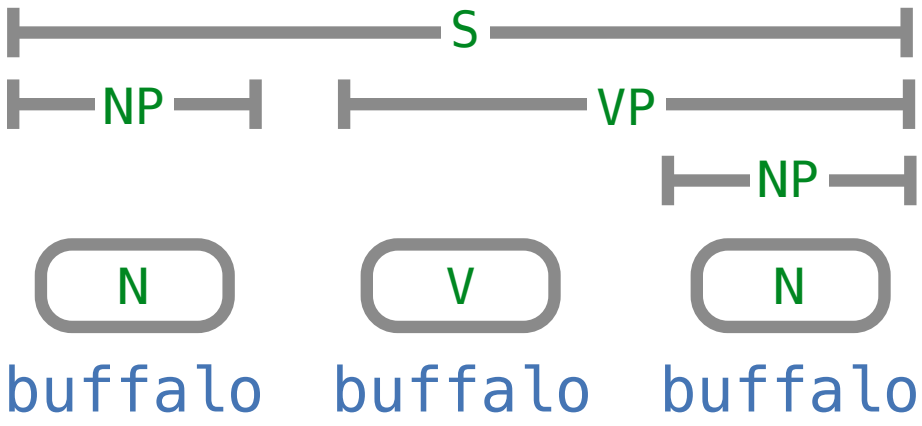


A Sentence ...

... can be expanded as ...

... a Noun Phrase then a Verb Phrase.

S → NP VP
NP → N
N → buffalo
VP → V NP
V → buffalo



(Demo)

Parsing

Exhaustive Parsing

Recursively expand, but force words to match input.

Exhaustive Parsing

Recursively expand, but force words to match input.

buffalo buffalo buffalo buffalo

Exhaustive Parsing

Recursively expand, but force words to match input.

buffalo buffalo buffalo buffalo

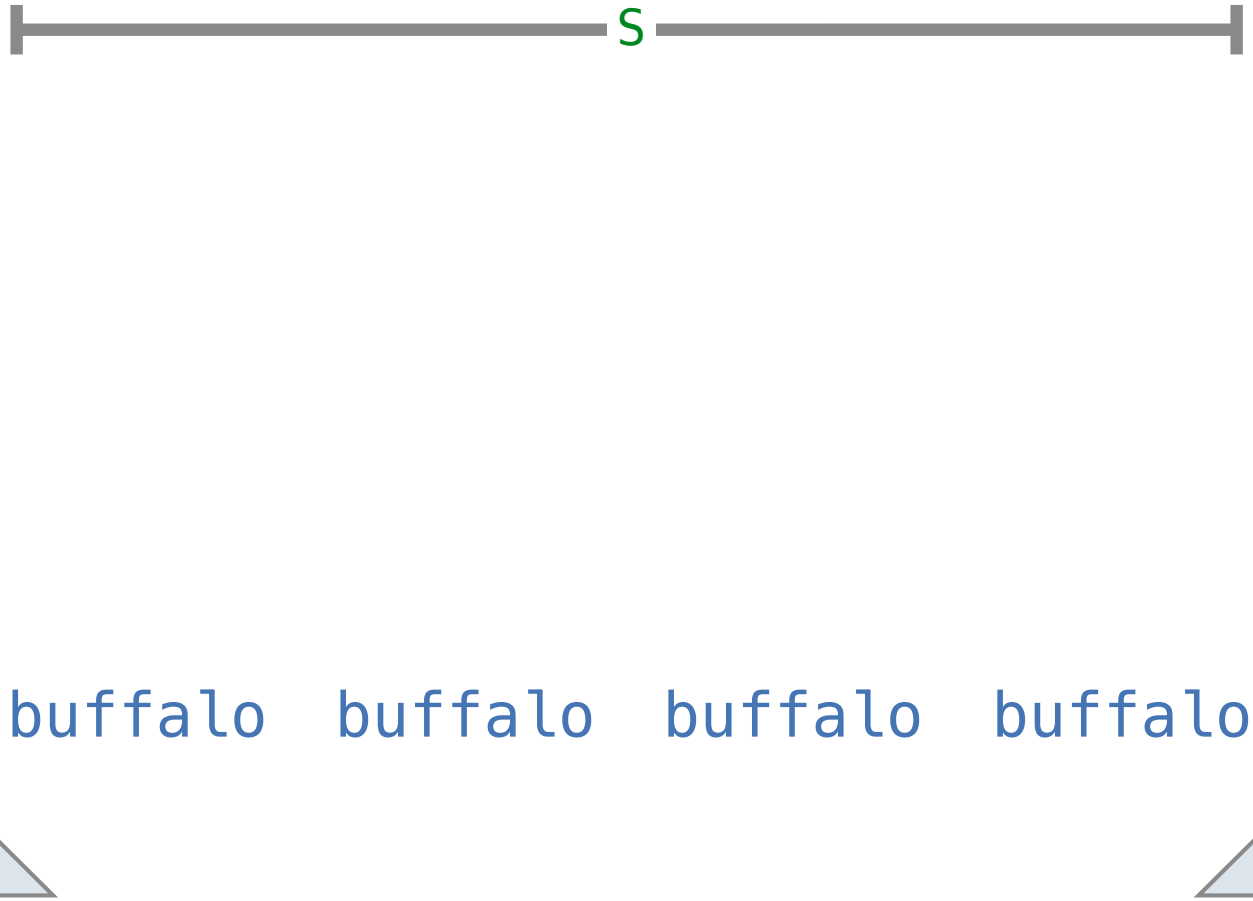
Exhaustive Parsing

Recursively expand, but force words to match input.

buffalo buffalo buffalo buffalo

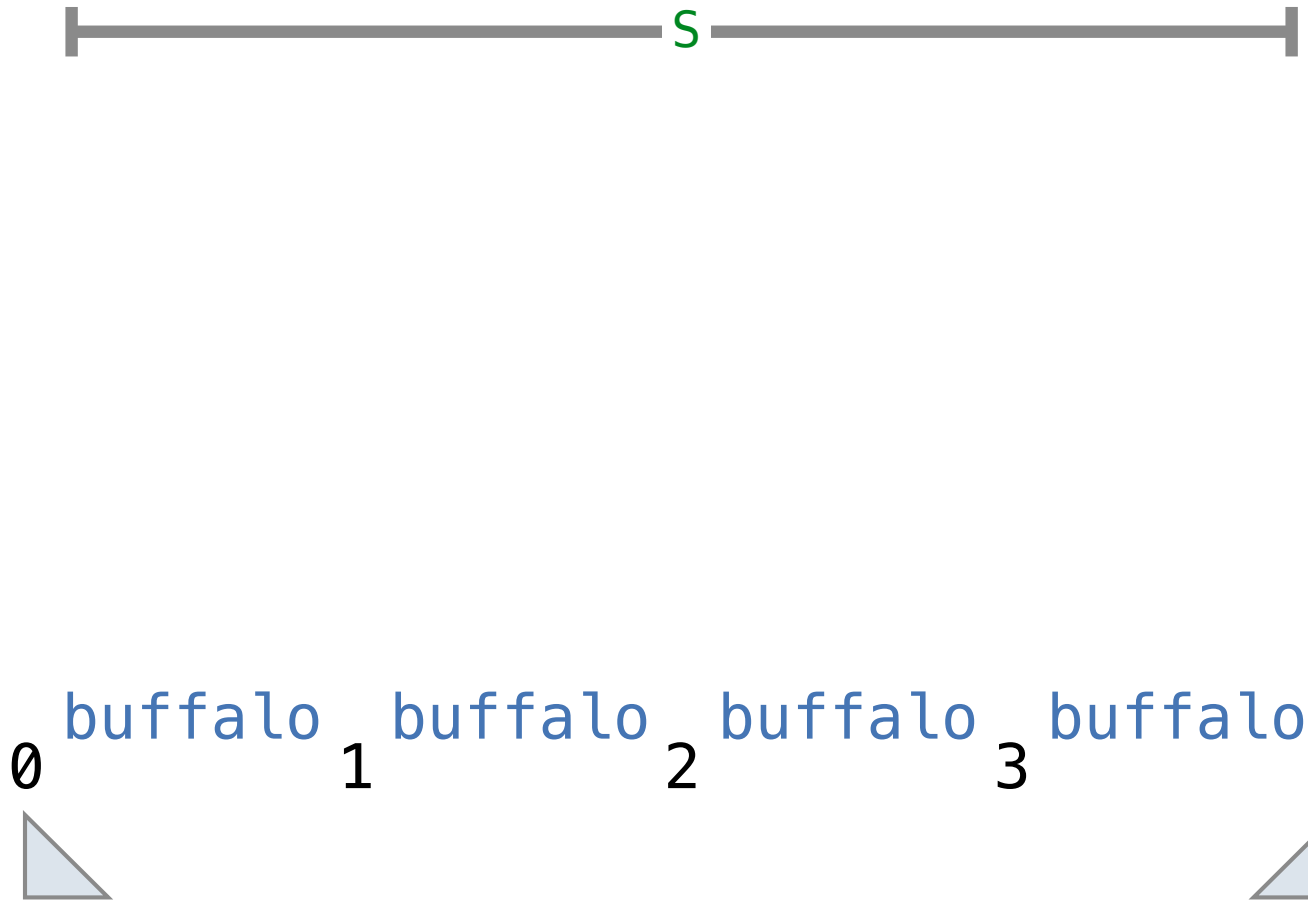
Exhaustive Parsing

Recursively expand, but force words to match input.



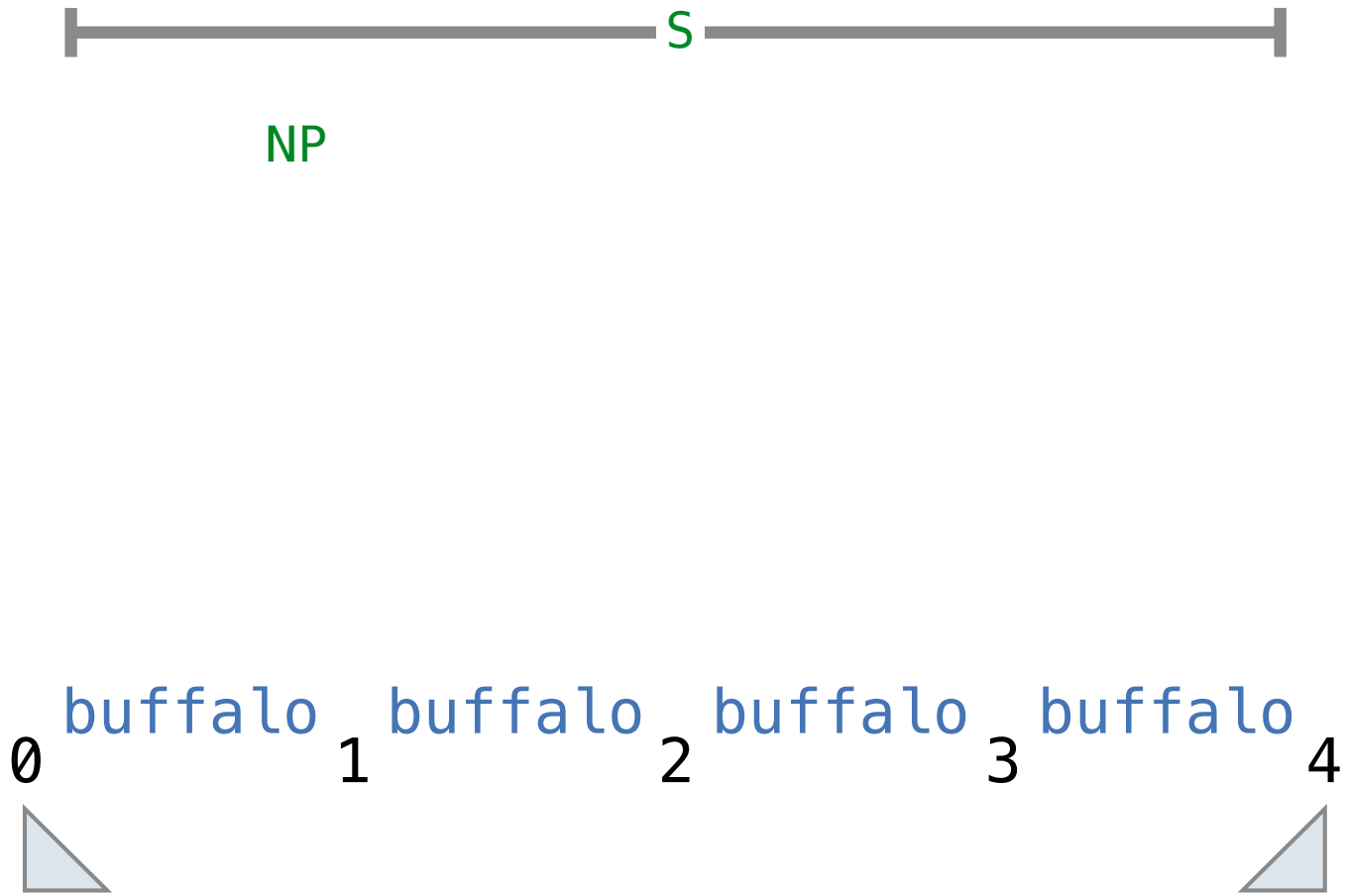
Exhaustive Parsing

Recursively expand, but force words to match input.



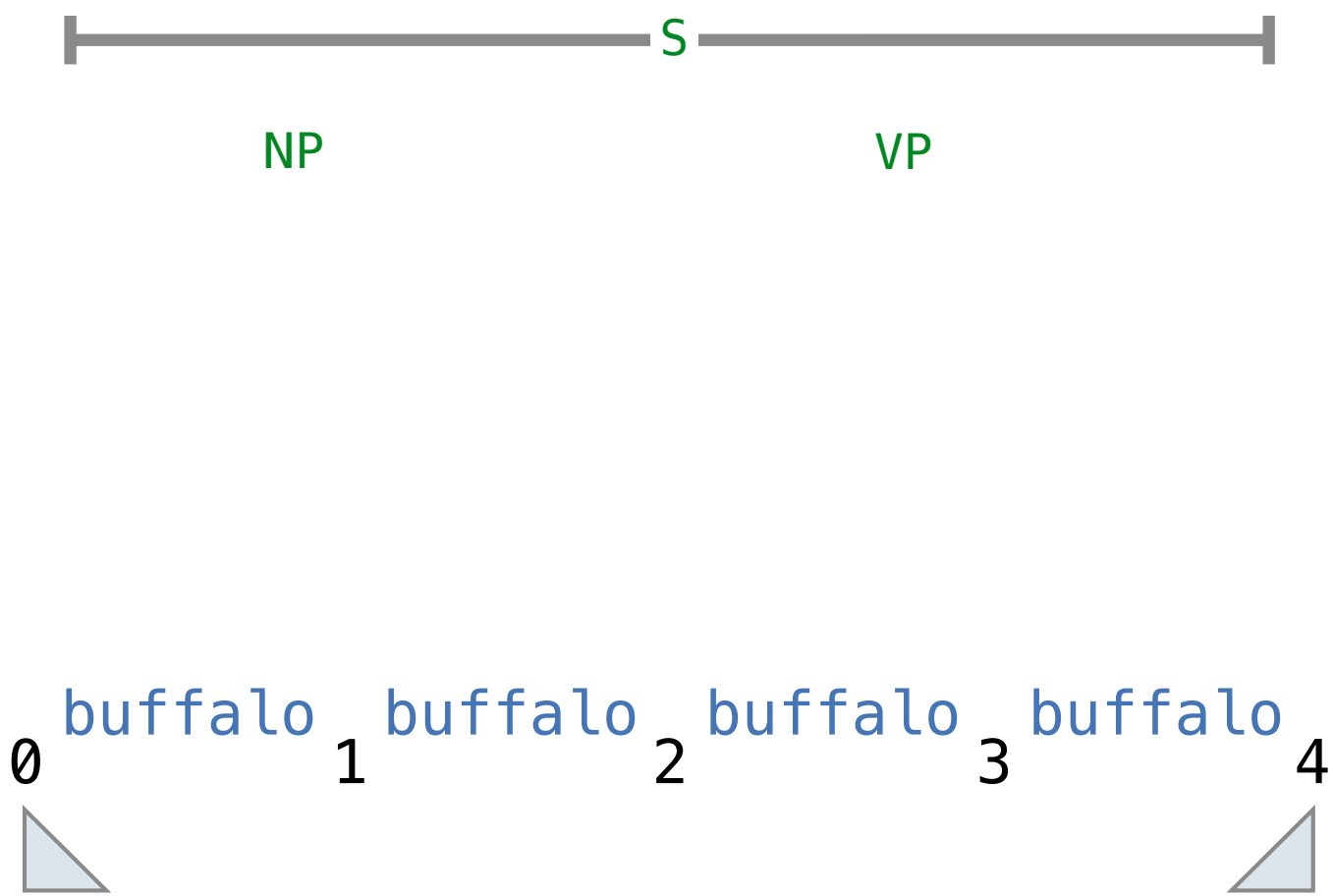
Exhaustive Parsing

Recursively expand, but force words to match input.



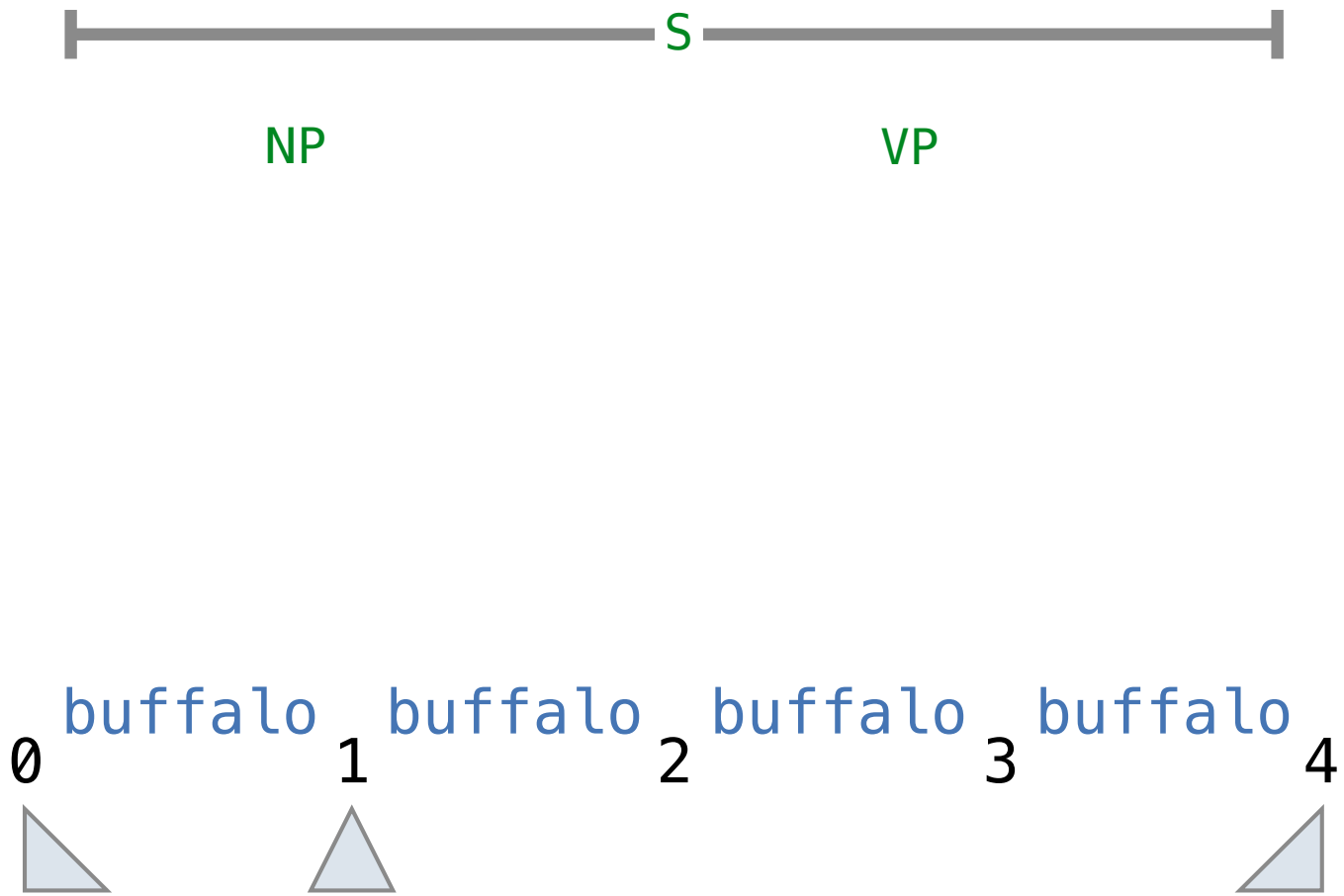
Exhaustive Parsing

Recursively expand, but force words to match input.



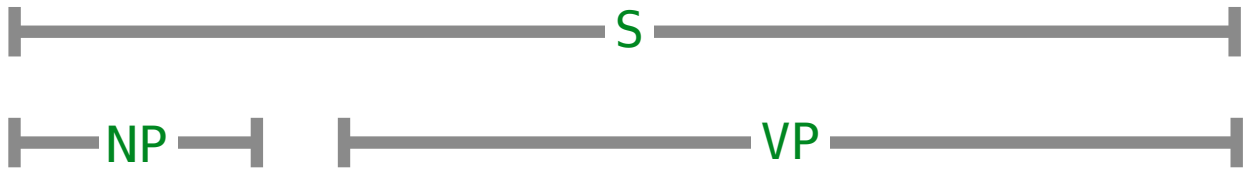
Exhaustive Parsing

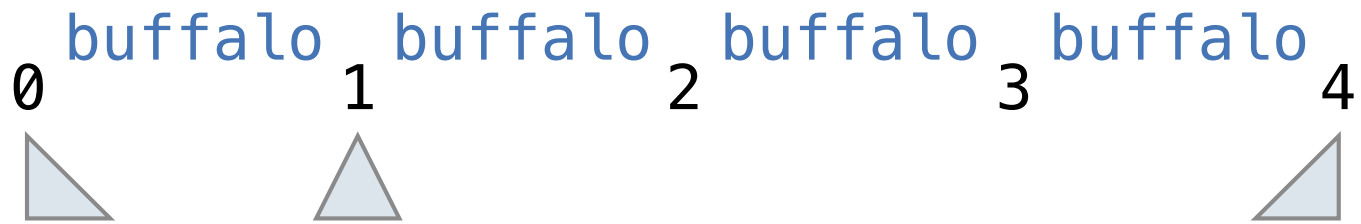
Recursively expand, but force words to match input.



Exhaustive Parsing

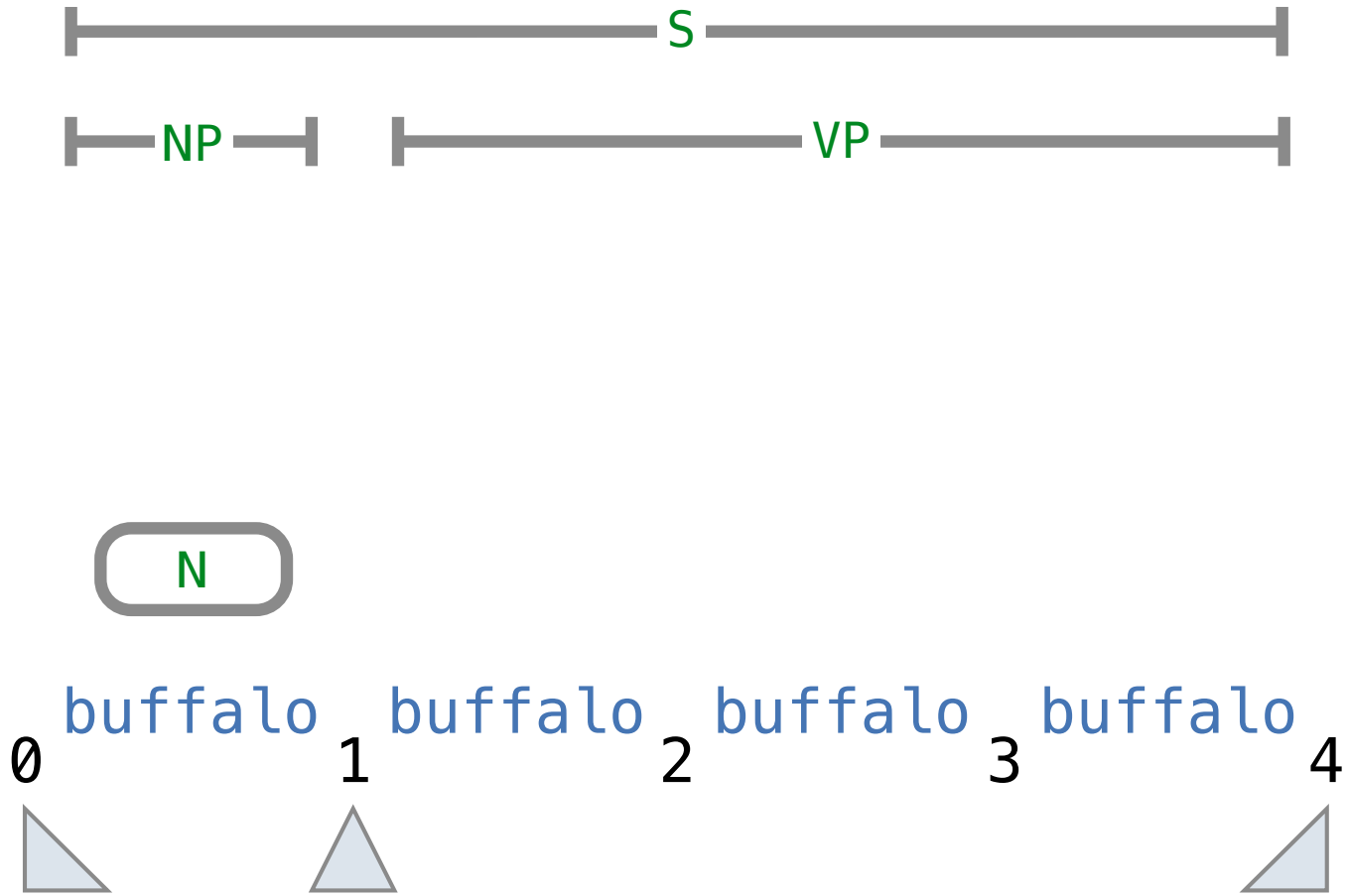
Recursively expand, but force words to match input.





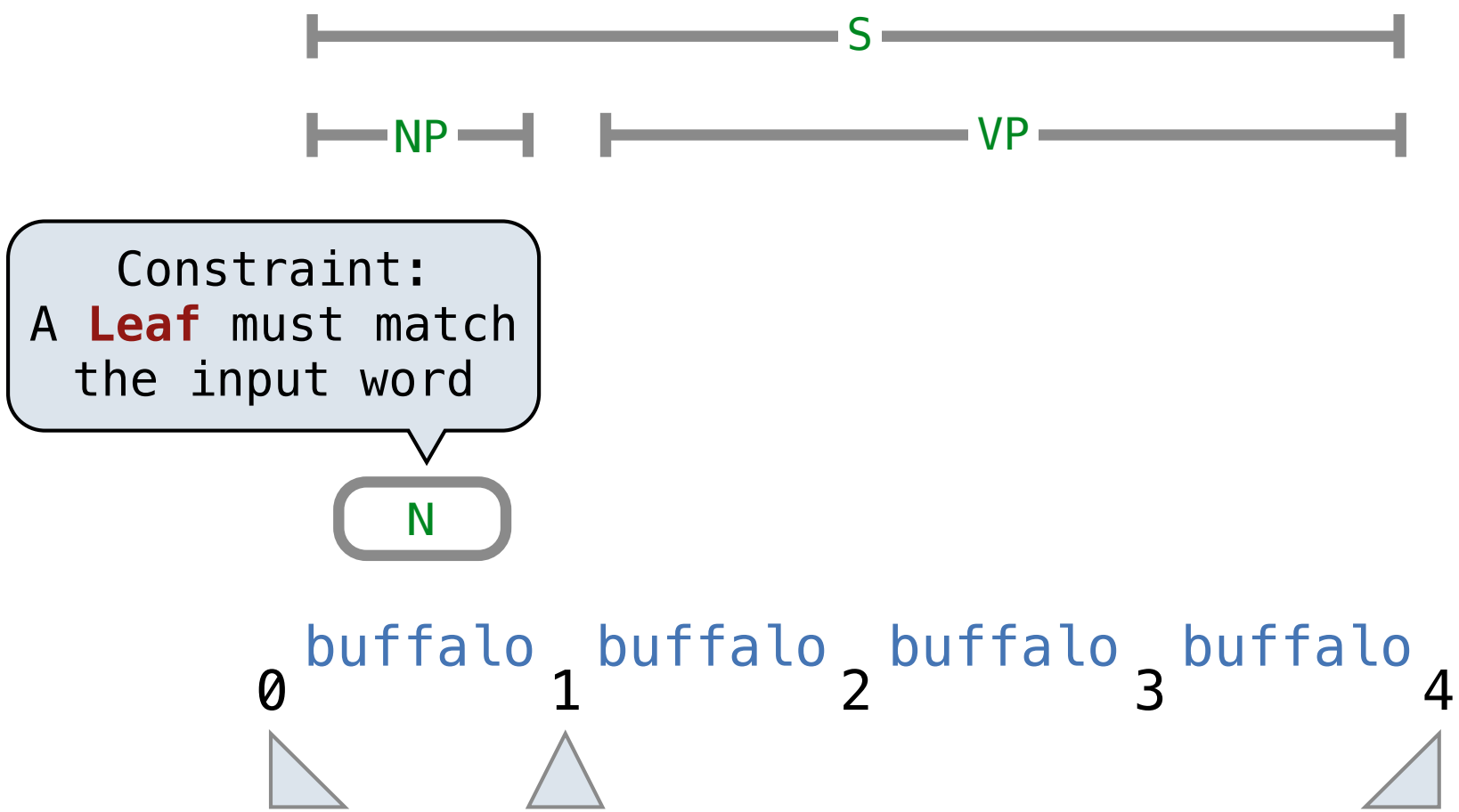
Exhaustive Parsing

Recursively expand, but force words to match input.



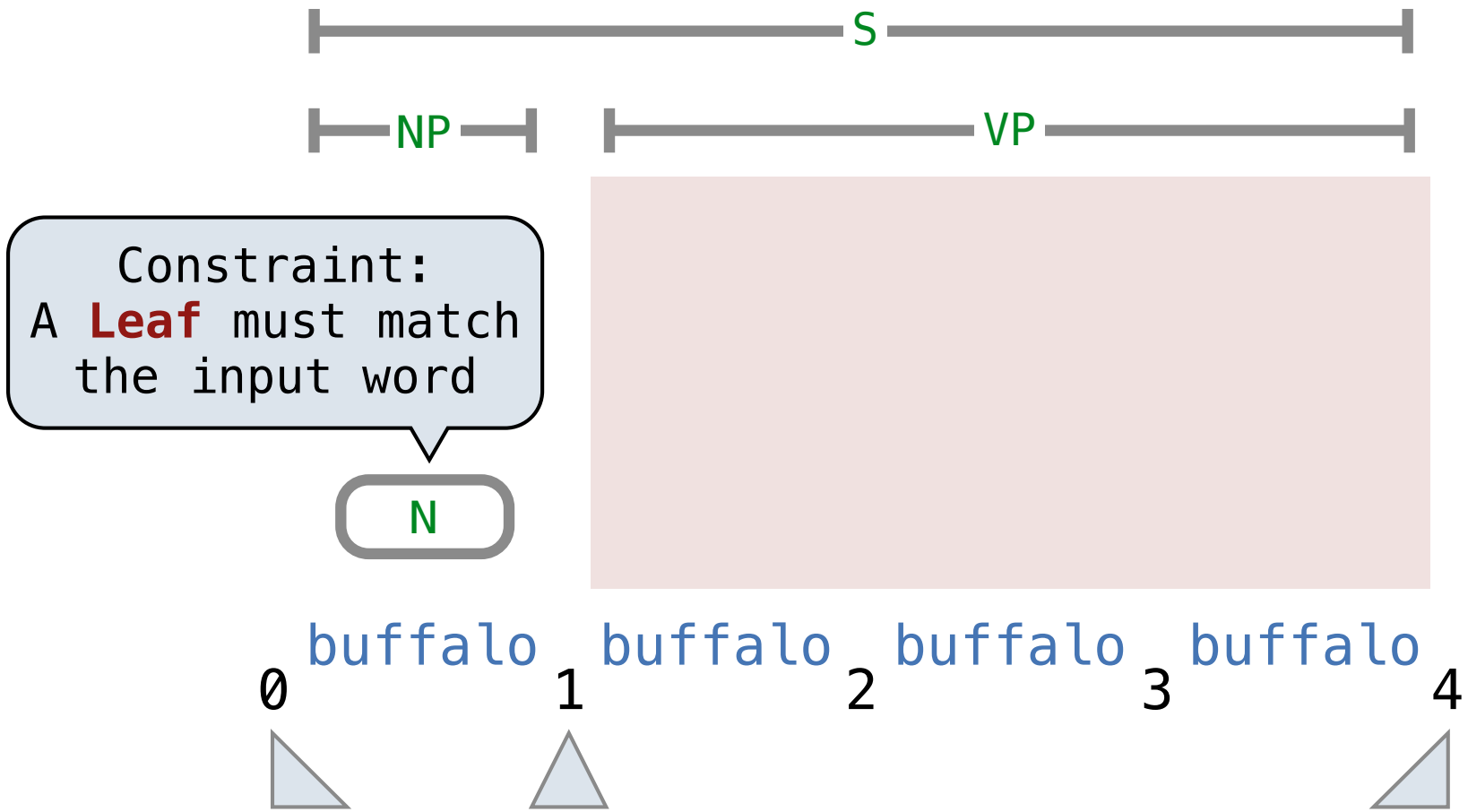
Exhaustive Parsing

Recursively expand, but force words to match input.



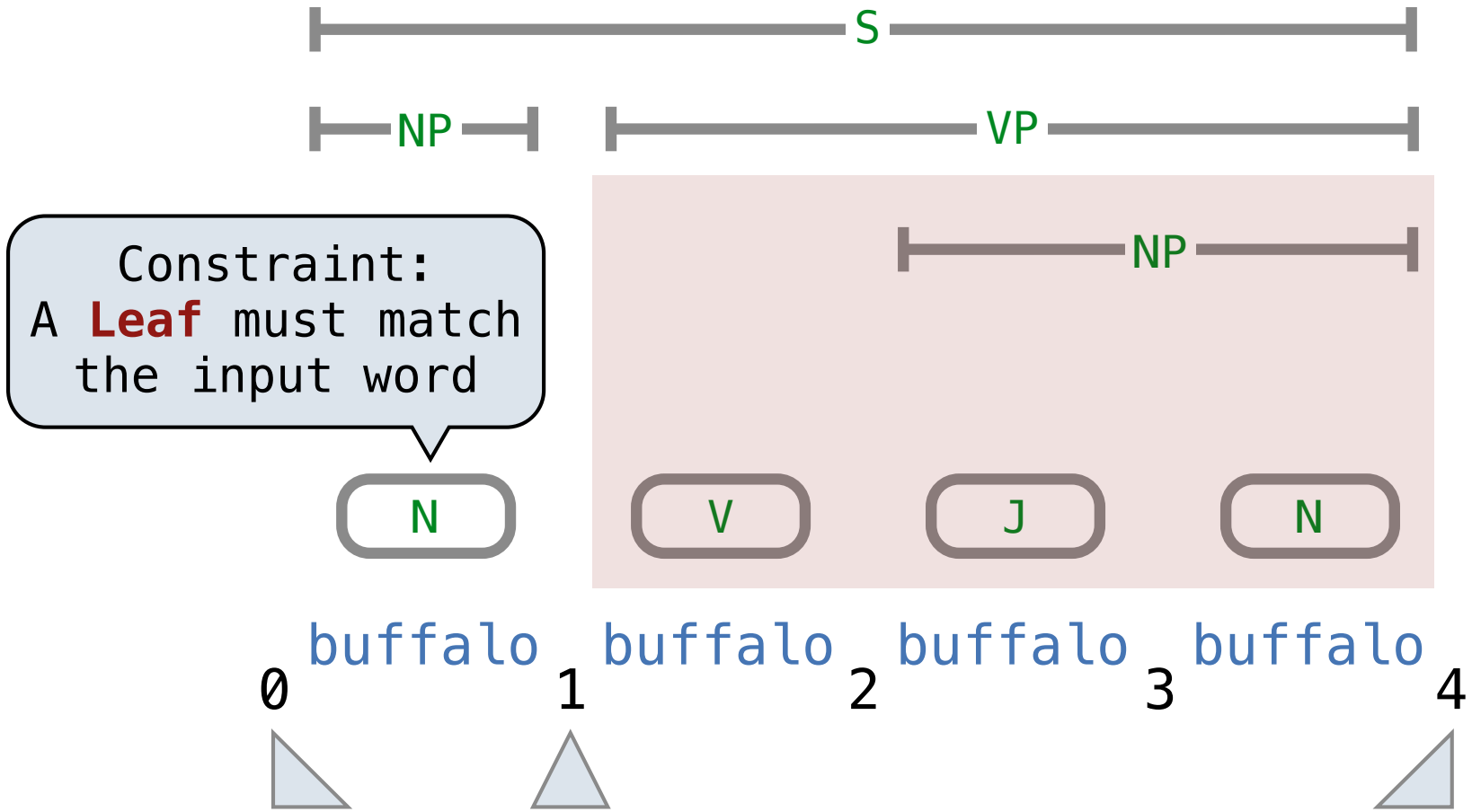
Exhaustive Parsing

Recursively expand, but force words to match input.



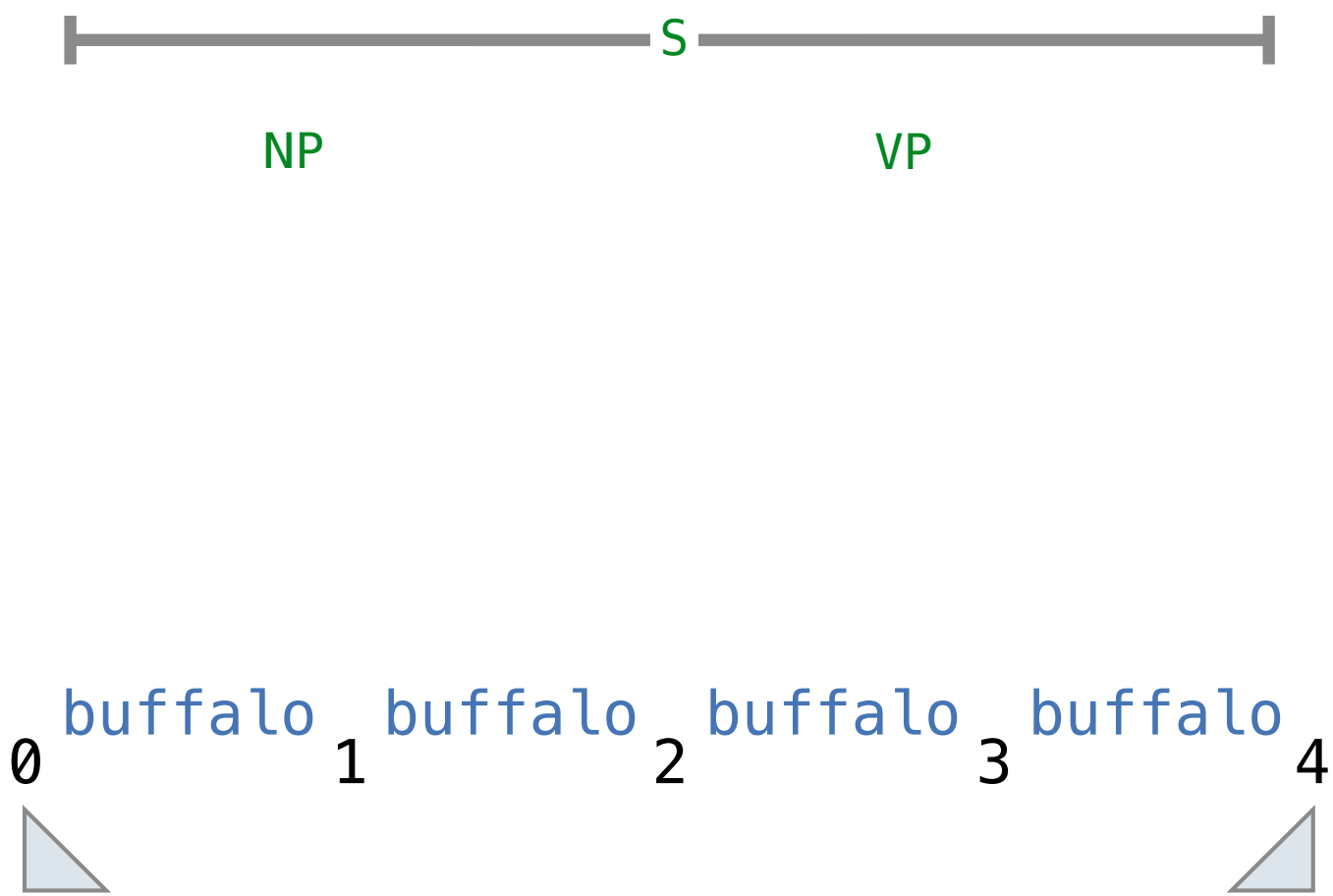
Exhaustive Parsing

Recursively expand, but force words to match input.



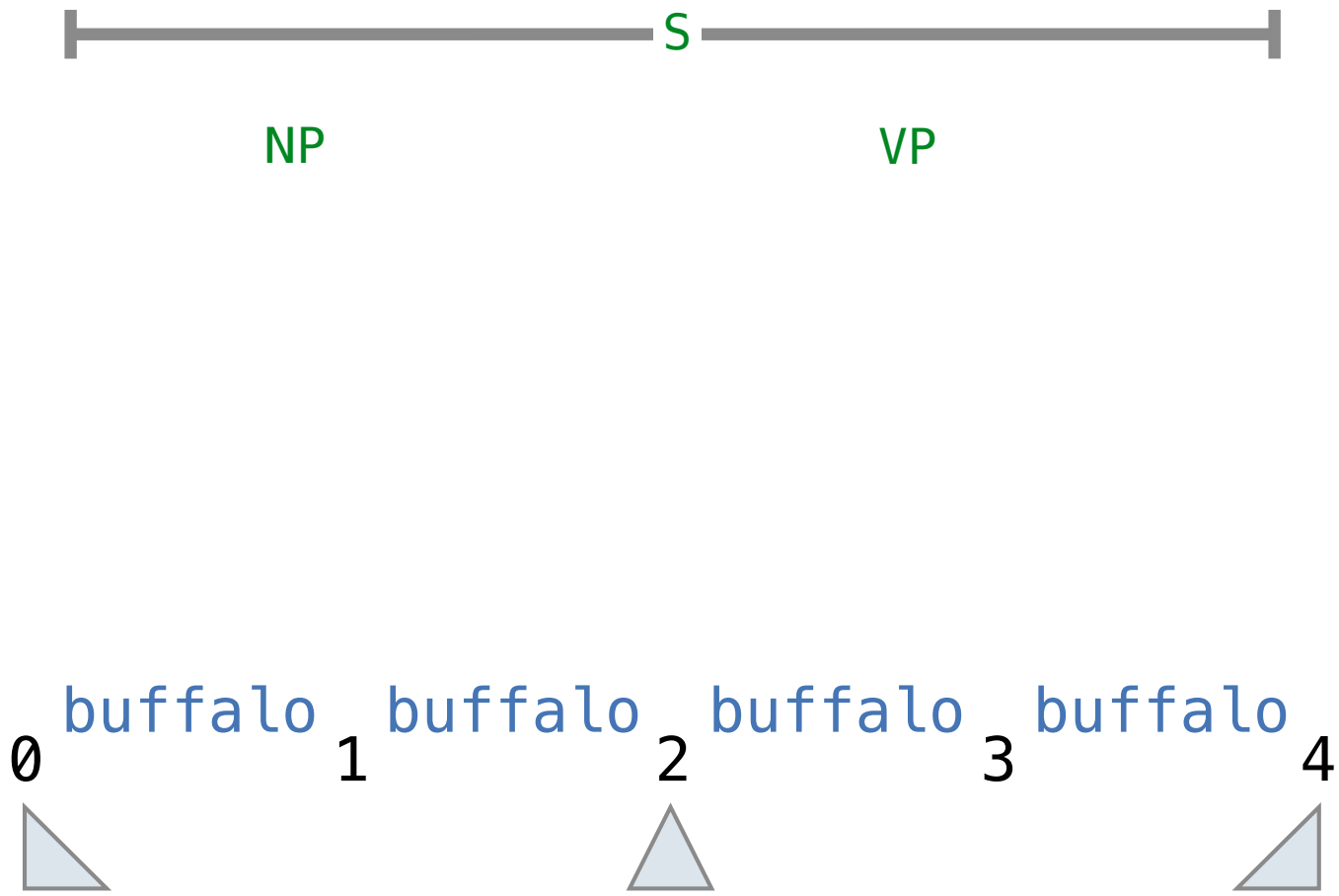
Exhaustive Parsing

Recursively expand, but force words to match input.



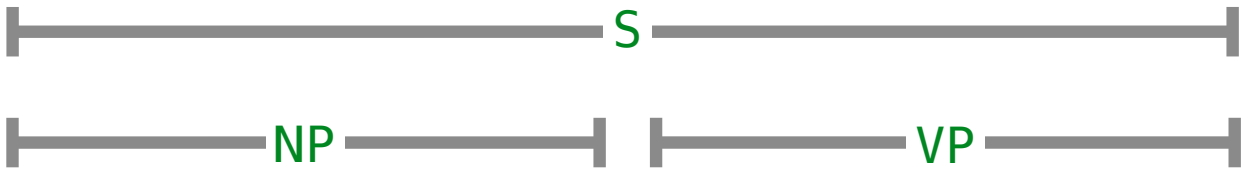
Exhaustive Parsing

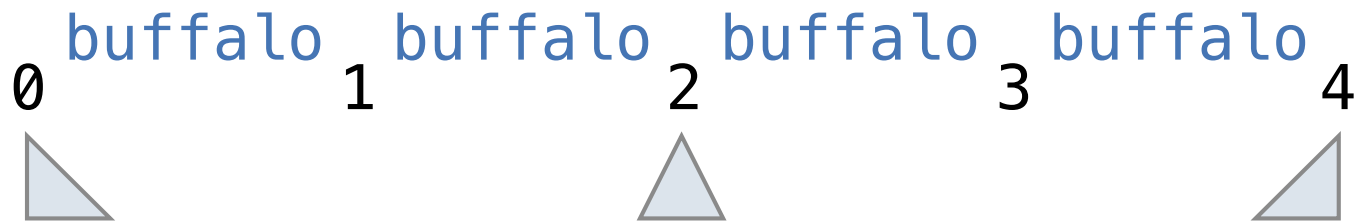
Recursively expand, but force words to match input.



Exhaustive Parsing

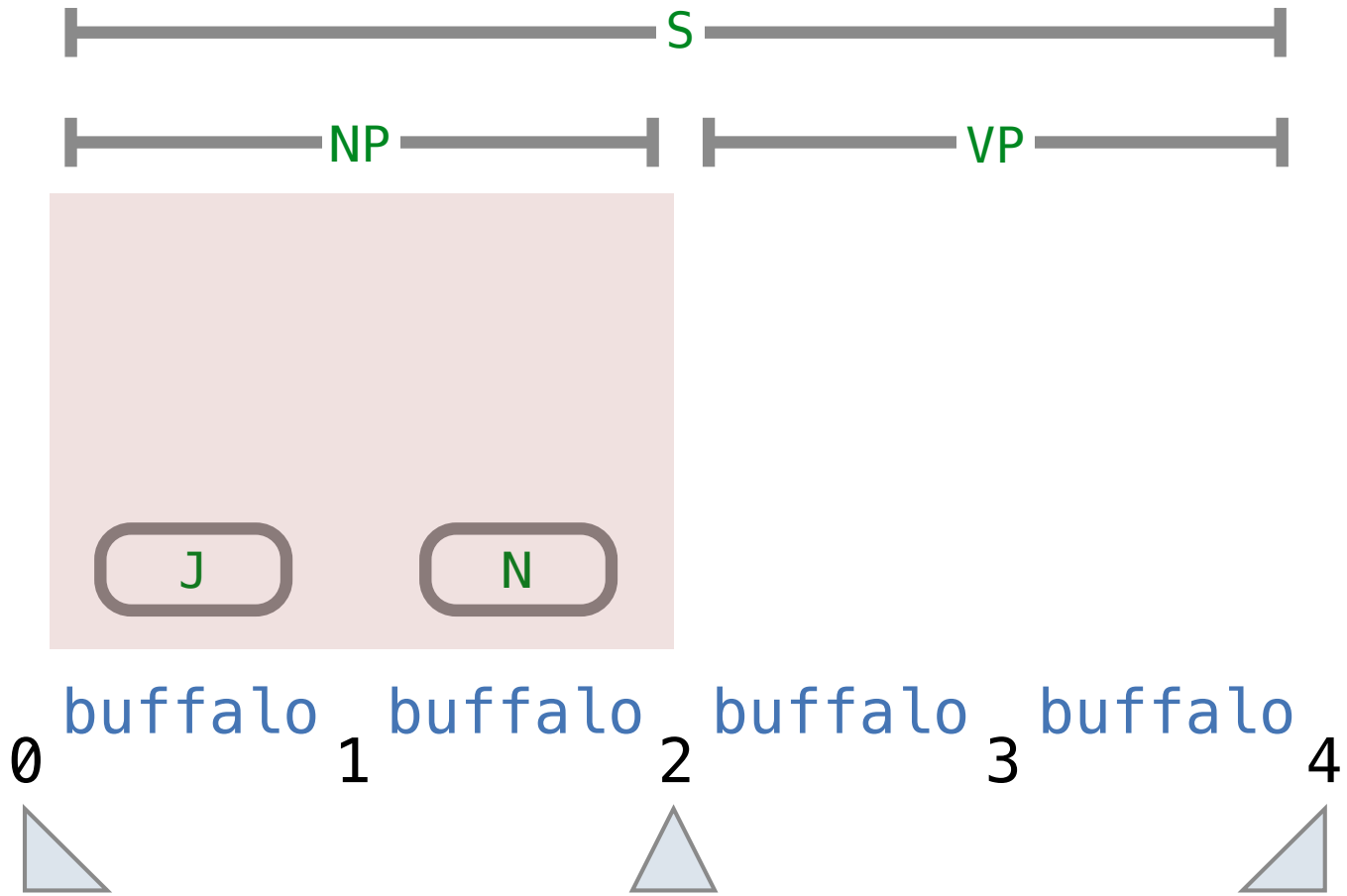
Recursively expand, but force words to match input.





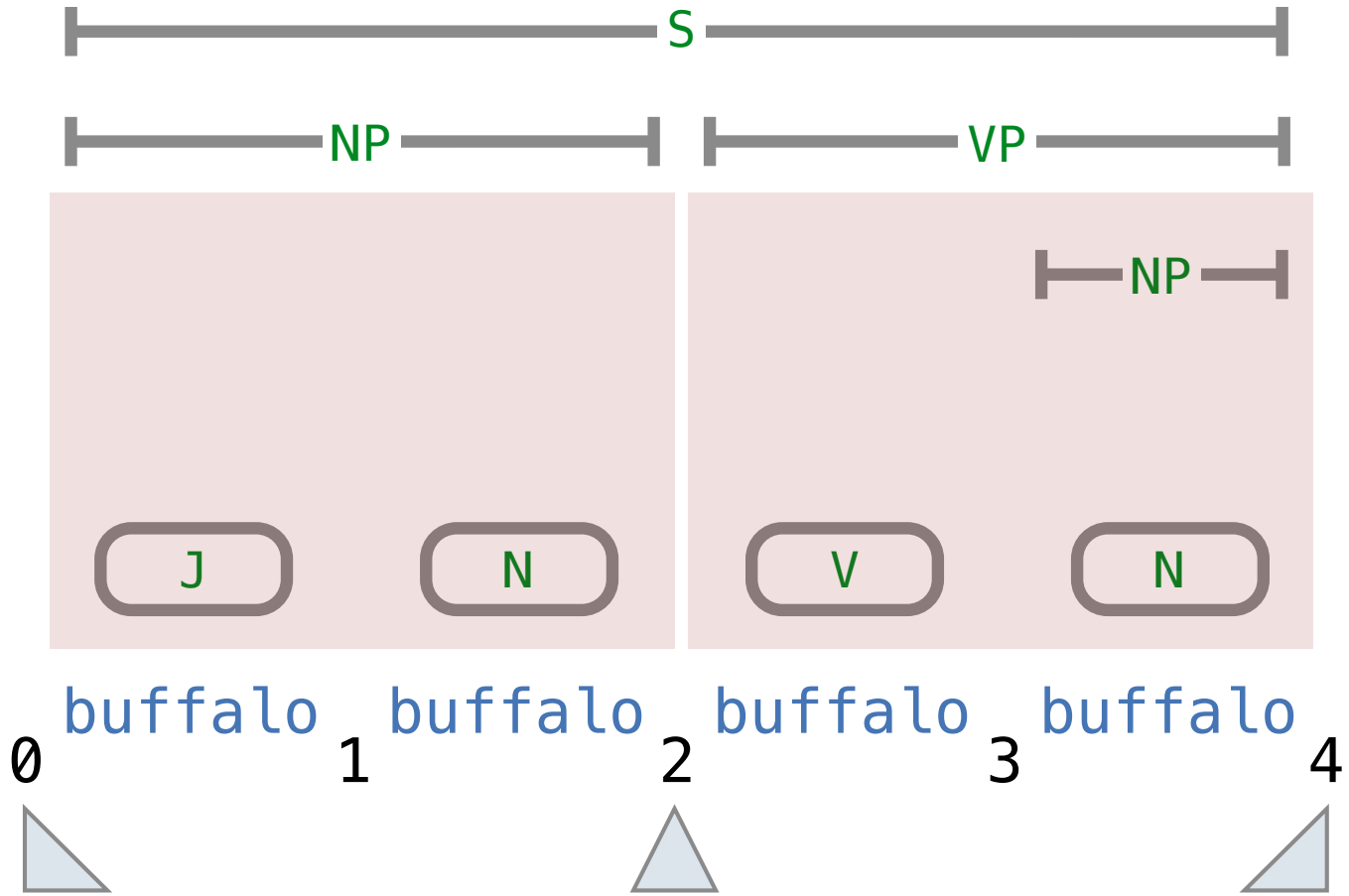
Exhaustive Parsing

Recursively expand, but force words to match input.



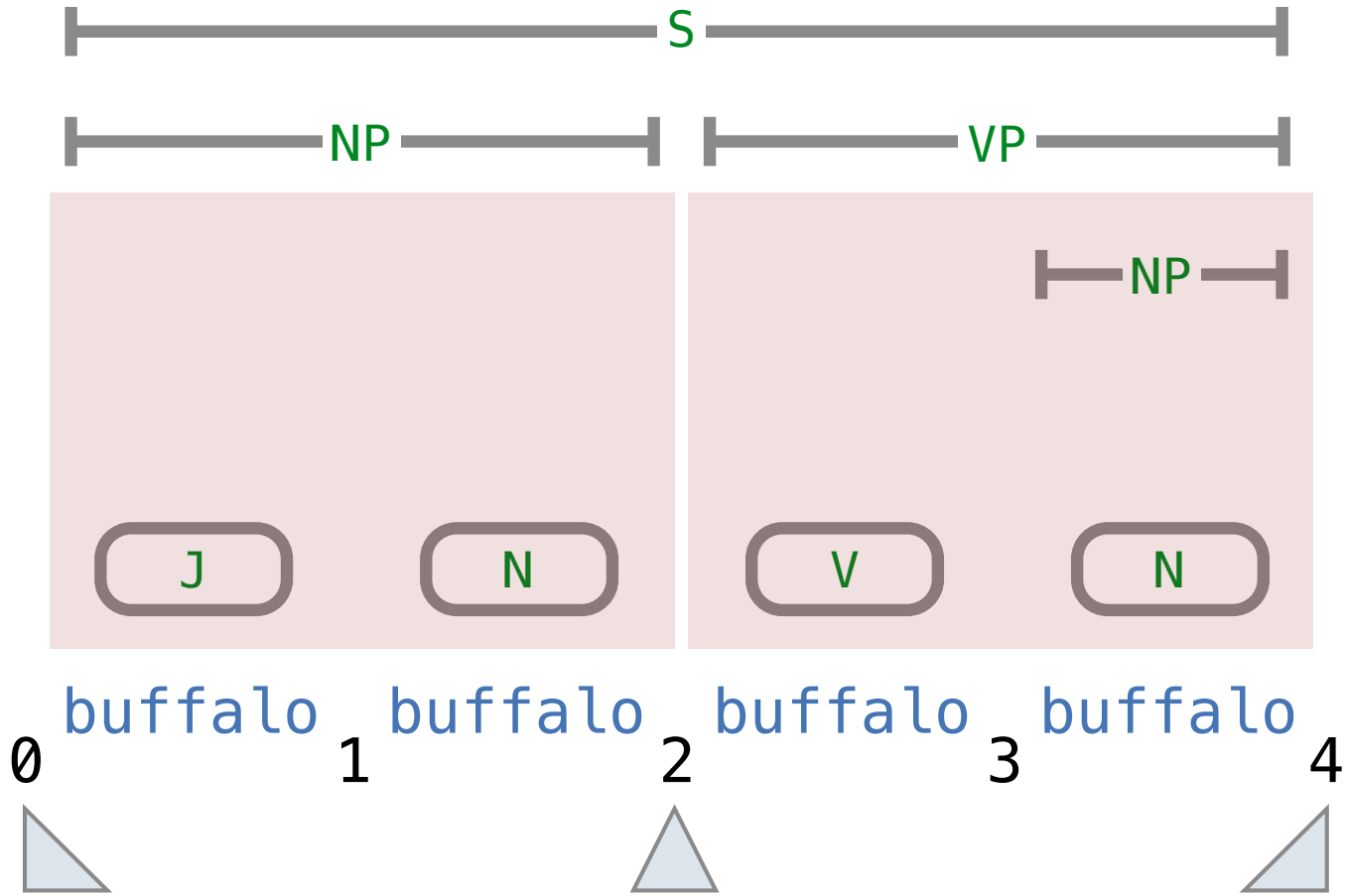
Exhaustive Parsing

Recursively expand, but force words to match input.



Exhaustive Parsing

Recursively expand, but force words to match input.



(Demo)

Learning

(Demo)

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

teacher strikes idle kids

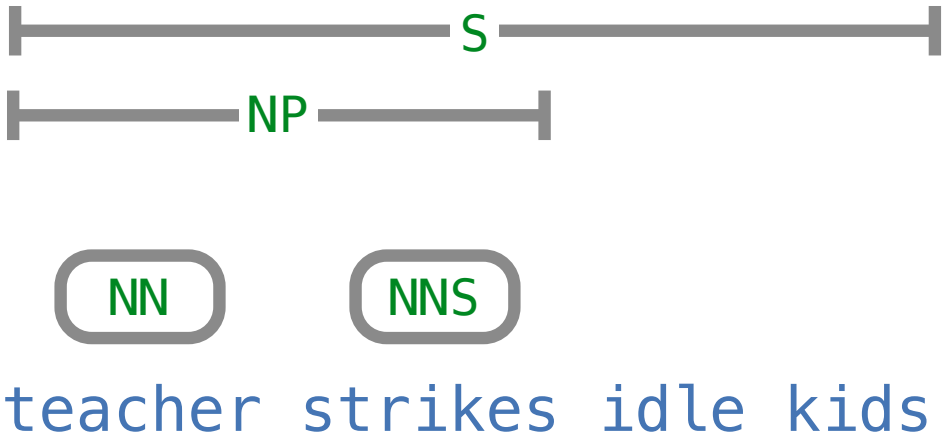
Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

teacher strikes idle kids

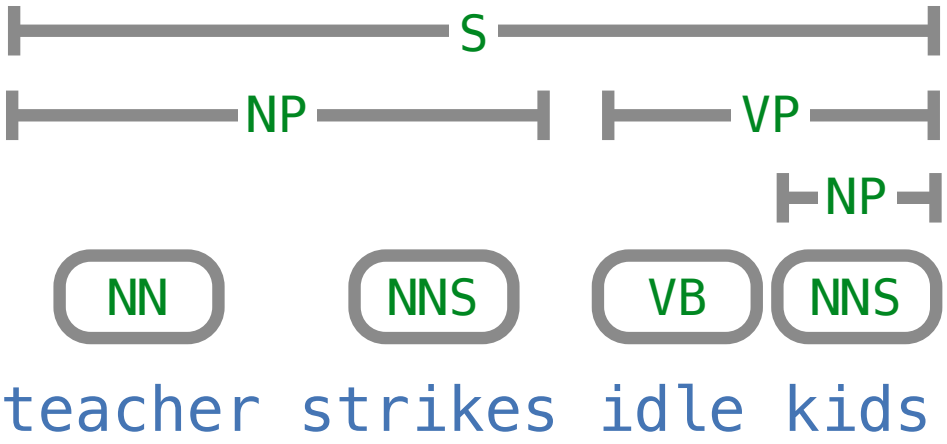
Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.



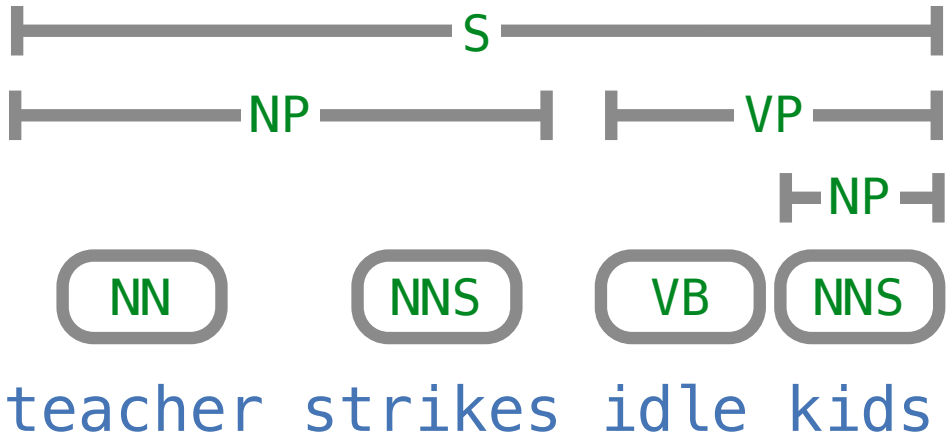
Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.



Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.



S → NP VP

NP → NN NNS

VP → VB NP

NP → NNS

NN → teacher

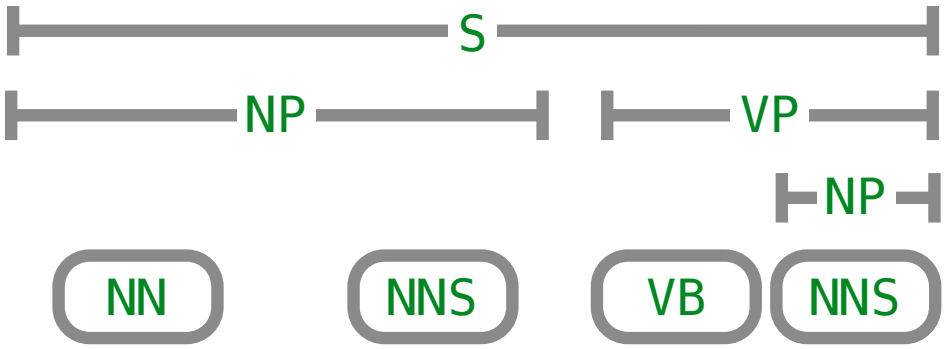
NNS → strikes

VB → idle

NNS → kids

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

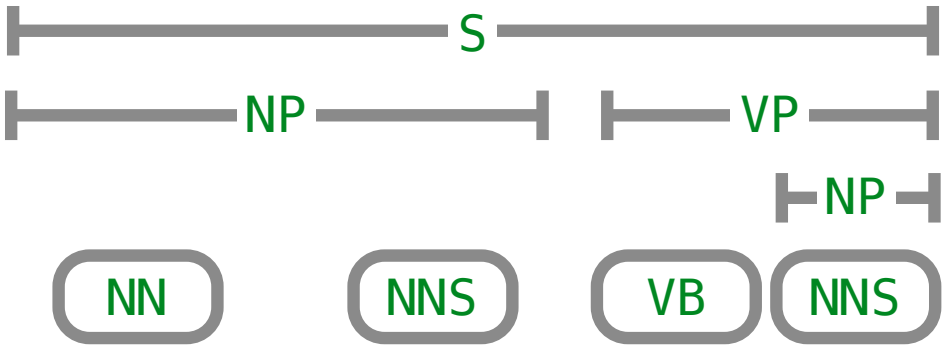


Rule frequency per 100,000 tags

S → NP VP	NN → teacher
NP → NN NNS	NNS → strikes
VP → VB NP	VB → idle
NP → NNS	NNS → kids

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

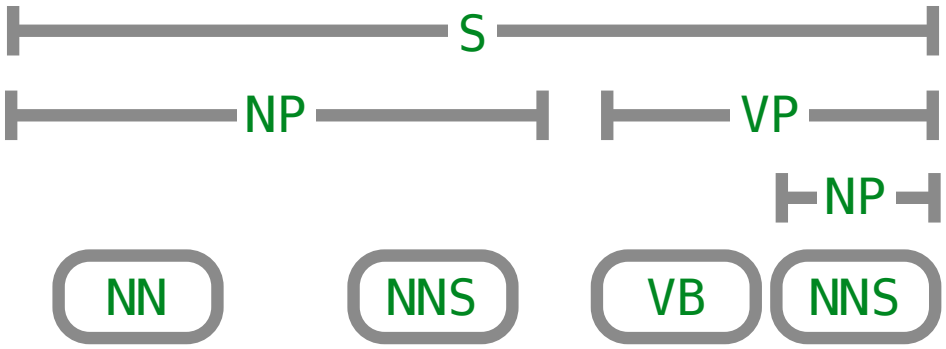


Rule frequency per 100,000 tags

S → NP VP	25372	NN →	teacher
NP → NN NNS		NNS →	strikes
VP → VB NP		VB →	idle
NP → NNS		NNS →	kids

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

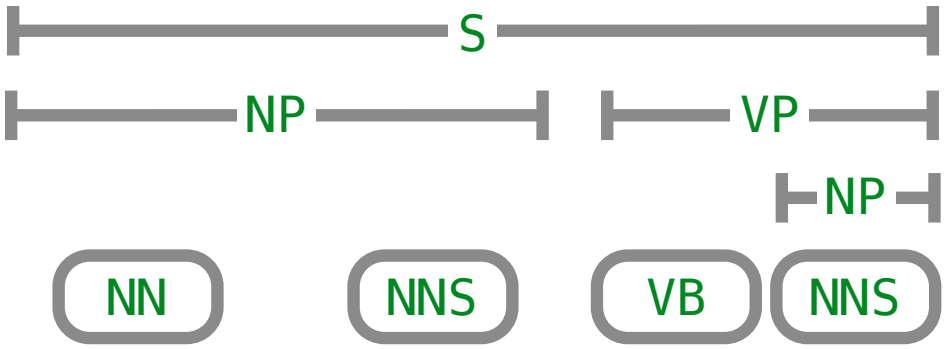


Rule frequency per 100,000 tags

S → NP VP	25372	NN →	teacher
NP → NN NNS	1335	NNS →	strikes
VP → VB NP		VB →	idle
NP → NNS		NNS →	kids

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

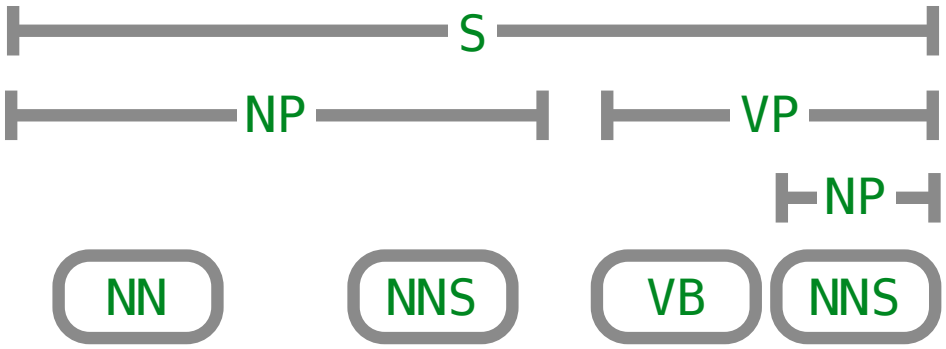


Rule frequency per 100,000 tags

S → NP VP	25372	NN → teacher	
NP → NN NNS	1335	NNS → strikes	
VP → VB NP	6679	VB → idle	
NP → NNS		NNS → kids	

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

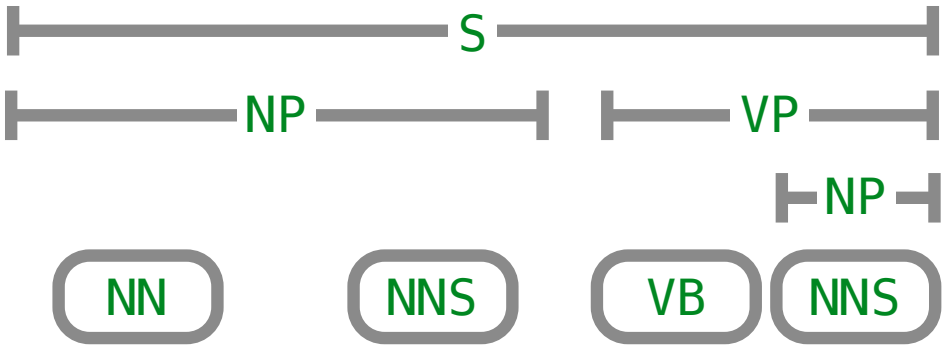


Rule frequency per 100,000 tags

S → NP VP	25372	NN → teacher	
NP → NN NNS	1335	NNS → strikes	
VP → VB NP	6679	VB → idle	
NP → NNS	4282	NNS → kids	

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

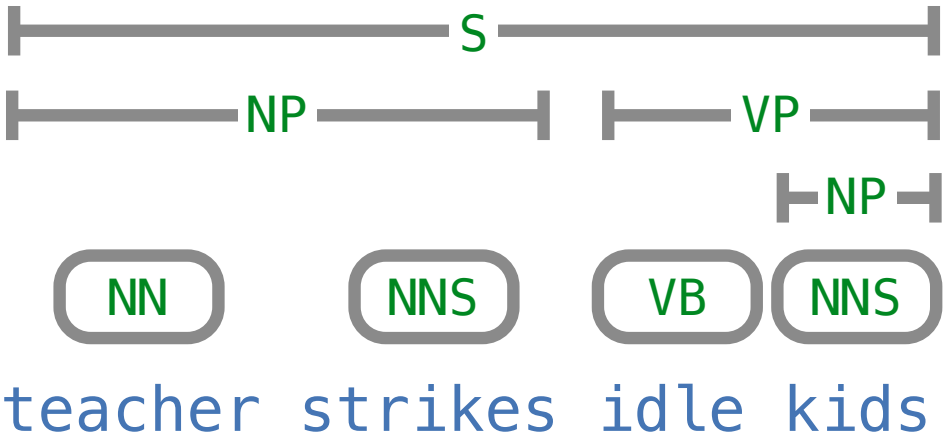


Rule frequency per 100,000 tags

S → NP VP	25372	NN → teacher	5
NP → NN NNS	1335	NNS → strikes	
VP → VB NP	6679	VB → idle	
NP → NNS	4282	NNS → kids	

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

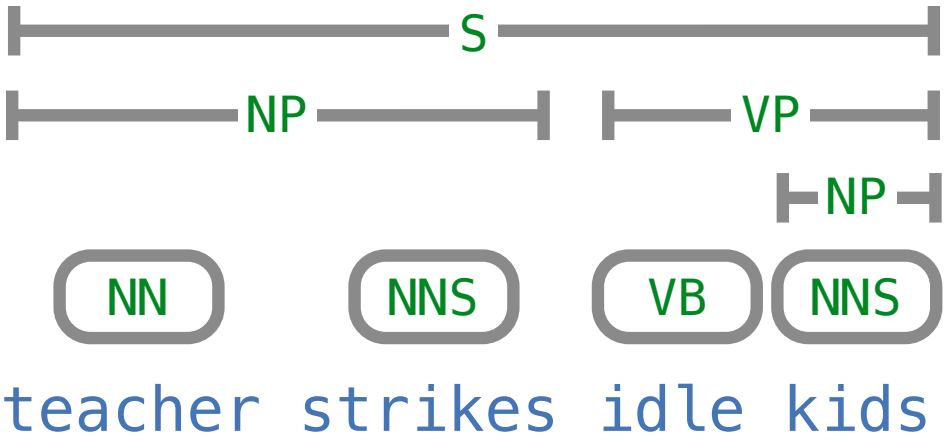


Rule frequency per 100,000 tags

S → NP VP	25372	NN → teacher	5
NP → NN NNS	1335	NNS → strikes	25
VP → VB NP	6679	VB → idle	
NP → NNS	4282	NNS → kids	

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

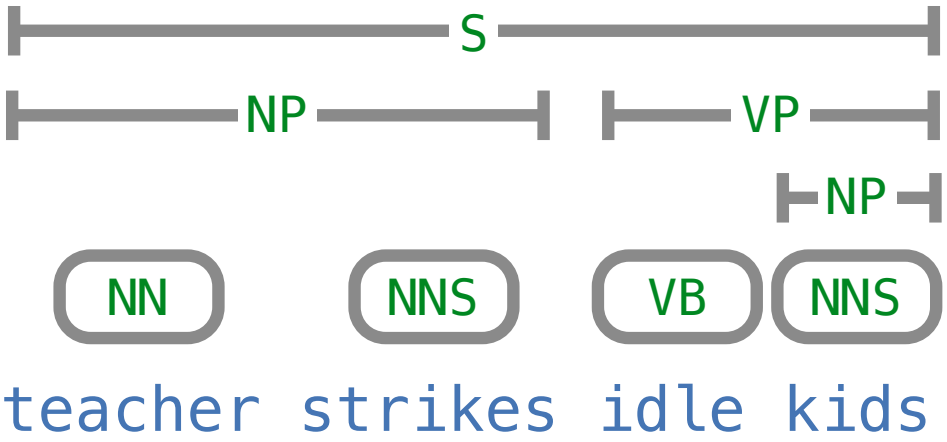


Rule frequency per 100,000 tags

S → NP VP	25372	NN → teacher	5
NP → NN NNS	1335	NNS → strikes	25
VP → VB NP	6679	VB → idle	26
NP → NNS	4282	NNS → kids	

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

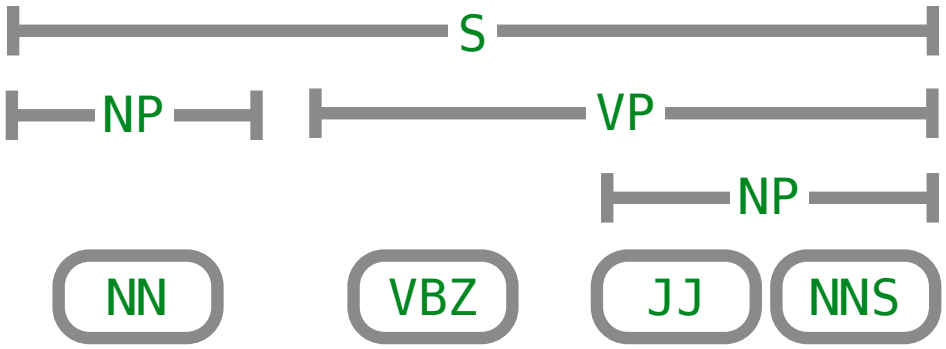


Rule frequency per 100,000 tags

S → NP VP	25372	NN → teacher	5
NP → NN NNS	1335	NNS → strikes	25
VP → VB NP	6679	VB → idle	26
NP → NNS	4282	NNS → kids	32

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.

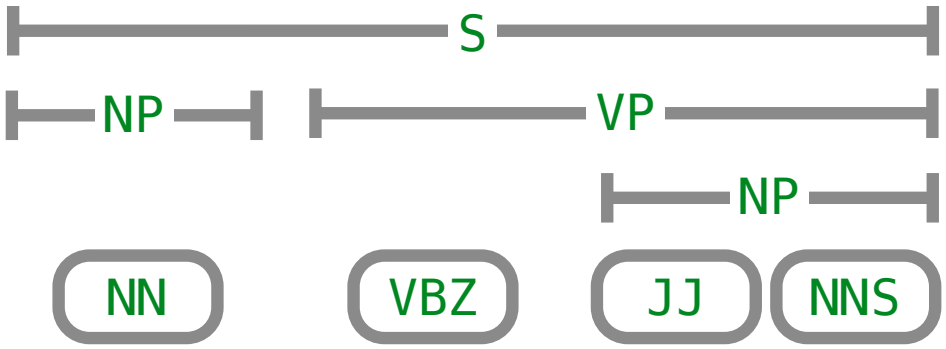


Rule frequency per 100,000 tags

S → NP VP	25372			NN → teacher	5		
NP → NN	~~1335~~	4358		VBZ → strikes	~~25~~	19	
VP → VBZ NP	~~6679~~	3160		JJ → idle	~~26~~	18	
NP → JJ NNS	~~4282~~	2526		NNS → kids	32		

Scoring a Tree Using Relative Frequencies

Not all syntactic structures are equally common.



Rule frequency per 100,000 tags

S → NP VP	25372			NN → teacher	5		
NP → NN	~~1335~~	4358		VBZ → strikes	~~25~~	19	
VP → VBZ NP	~~6679~~	3160		JJ → idle	~~26~~	18	
NP → JJ NNS	~~4282~~	2526		NNS → kids	32		

(Demo)