CFGs

David Kauchak
CS30 – Spring 2016

Grammars

What is a grammar?

Grammars

Language view:

A grammar is a set of structural rules that govern the composition of sentences, phrases and words.

Computational view:

A grammar (often called a "formal grammar") is a set of rules that describe what strings are valid in a formal language.

Grammars

What types of (formal) grammars have you heard of before?

Lots of different kinds of grammars:

- regular
- context-free
- context-sensitive
- recursively enumerable
- transformation grammars

Context Free Grammars (CFG)

How many people have heard of them?

What do you know about them?

Where are they used?

CFG production rules

S → NP VP

left hand side (single symbol)

right hand side (one or more symbols)

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

S

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

S

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

A B C

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

A B C

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

A really C

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

A really C

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

A really like cs

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

A really like cs

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

I really like cs

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

Grammars "generate" or "derive" strings:

S → A B C
A → I
B → really
B → really, B
C → like cs

I really like cs

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

No more rules apply, so we're done!

CFG example

S → A B C
A → I
B → really
B → really, B
C → like cs

Grammars "generate" or "derive" strings:

I really like cs

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

Is this the only string that can be derived?

CFG example

S → A B C
A → I
B → really
B → really, B
C → like cs

Grammars "generate" or "derive" strings:

A really, B C

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

S → A B C
A → I
B → really
B → really, B
C → like cs

Grammars "generate" or "derive" strings:

A really, really, B C

We can apply a rule by substituting the symbol on the left hand side with the symbols on the right

CFG example

S → A B C
A → I
B → really
B → really, B
C → like cs

Grammars describe a language, i.e. the strings (aka sentences) that are part of that language

I really, really, … like cs

CFGs formally

G = (**NT**, **T**, **P**, **S**)

NT: finite set of nonterminal symbols

T: finite set of terminal symbols, **NT** and **T** are disjoint

P: finite set of productions of the form
$A \rightarrow \alpha$, $A \in NT$ and $\alpha \in (T \cup NT)^*$

S $\in$ NT: start symbol

What language does this represent?

S → a S
S → E
E → b E
E → b

What language does this represent?

S → aS
S → E
E → bE
E → b

Two options

S

What language does this represent?

S → aS
S → E
E → bE
E → b

S
↓
aS

What language does this represent?

S → aS
S → E
E → bE
E → b

aS
↓
aaS

What language does this represent?

S → aS
S → E
E → bE
E → b

aaS
↓
aaaS

- Can do this as many times as we want
- Keeps adding more a's to the front

What language does this represent?

S → aS
S → E
E → bE
E → b

aaaS
↓
aaaE

Eventually, apply second rule

What language does this represent?

S → aS
S → E
E → bE
E → b

Two options

aaaE

What language does this represent?

S → aS
S → E
E → bE
E → b

aaaE
↓
aaabE

What language does this represent?

S → aS
S → E
E → bE
E → b

aaabE
↓
aaabbE

What language does this represent?

S → aS
S → E
E → bE
E → b

aaabbE
↓
aaabbbE

What language does this represent?

S → aS
S → E
E → bE
E → b

aaabbE
↓
aaabb…bE

- Can do this as many times as we want
- Keeps adding more b's to the end

What language does this represent?

S → aS
S → E
E → bE
E → b

aaabb…bE
↓
aaabb…bb

Eventually, apply second rule

What language does this represent?

S → aS
S → E
E → bE
E → b

aaabb…bE
↓
aaabb…bb

Grammar represents all strings with zero or more a's followed by one or more b's