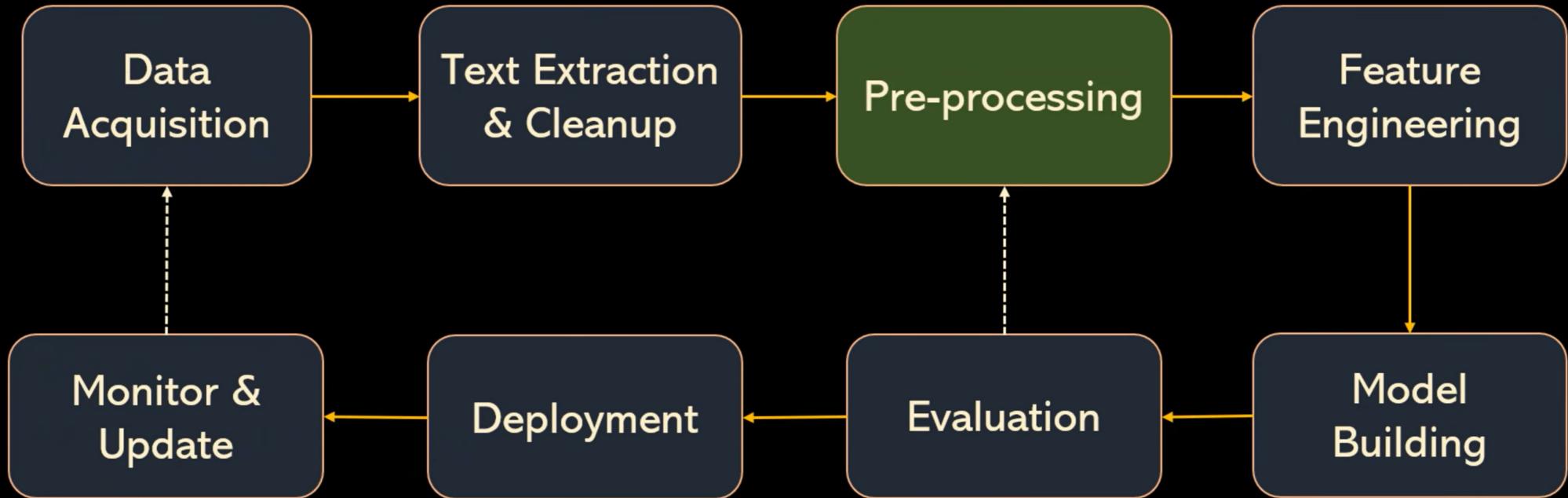


# Tokenization in Spacy



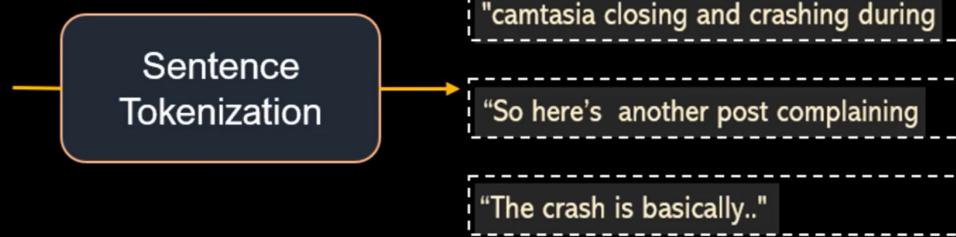


Credits: Pr



# Pre-Processing

"camtasia closing and crashing during ba  
another post complaining about somethi  
times before: During video production (t  
production as that's what I usually use),  
The crash is basically.."



# Pre-Processing

"camtasia closing and crashing during ba  
another post complaining about somethi  
times before: During video production (th  
production as that's what I usually use),  
The crash is basically.."

Sentence  
Tokenization

"camtasia closing and crashing during"  
"So here's another post complaining"  
"The crash is basically.."

Word  
Tokenization

camtasia

closing

crashing



# Pre-Processing

"camtasia closing and crashing during ba  
another post complaining about somethi  
times before: During video production (th  
production as that's what I usually use),  
The crash is basically.."

Sentence  
Tokenization

"camtasia closing and crashing during"  
"So here's another post complaining"  
"The crash is basically.."

Word  
Tokenization

camtasia      closing      crashing

Stemming,  
Lemmatization

camtasia      clo



Tokenization is a process of  
splitting text into  
meaningful segments



**WHAT'S THE BIG DEAL IN**

**TOKENIZATI**

imgflip.com



Dr. strange ordered samosas, ravioli etc. for his lunch.



Dr. strange ordered samosas, ravioli etc. for his lunch.

Sentence # 1

Dr

Sentence # 2

strange ordered samosas, ravioli etc

Sentence # 3

for his lunch



Dr. strange ordered samosas, ravioli etc. for his lunch.

Sentence # 1

Dr

Sentence # 2

strange ordered samosas, ravioli etc

Sentence # 3

for his lunch

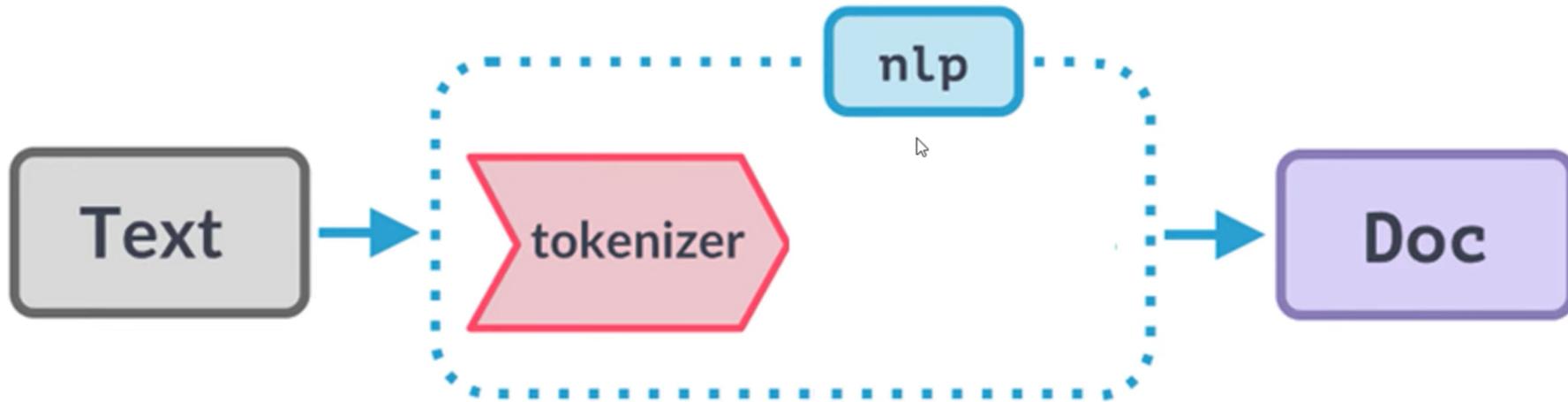
You see, it is not just about splitting a sentence by .

Or

Splitting words in a sentence by spaces



```
nlp = spacy.blank("en")
```



# “Let’s go to N.Y!”



# “Let’s go to N.Y!”

Split the words by spaces. i.e.  
`text.split(' ')`

“Let’s

go

to

N.Y.”



“Let's

go

to

N.Y.!”

PREFIX

“

Let's

go

to

N.Y.!”



“Let's

go

to

N.Y.!”

“

Let's

go

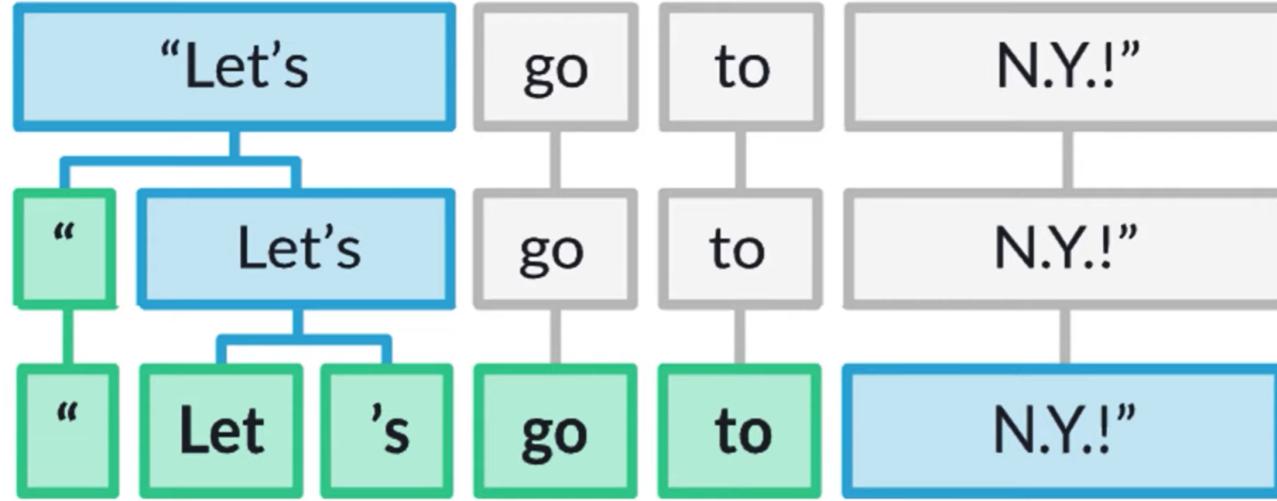
to

N.Y.!”

PREFIX

\$ ( “



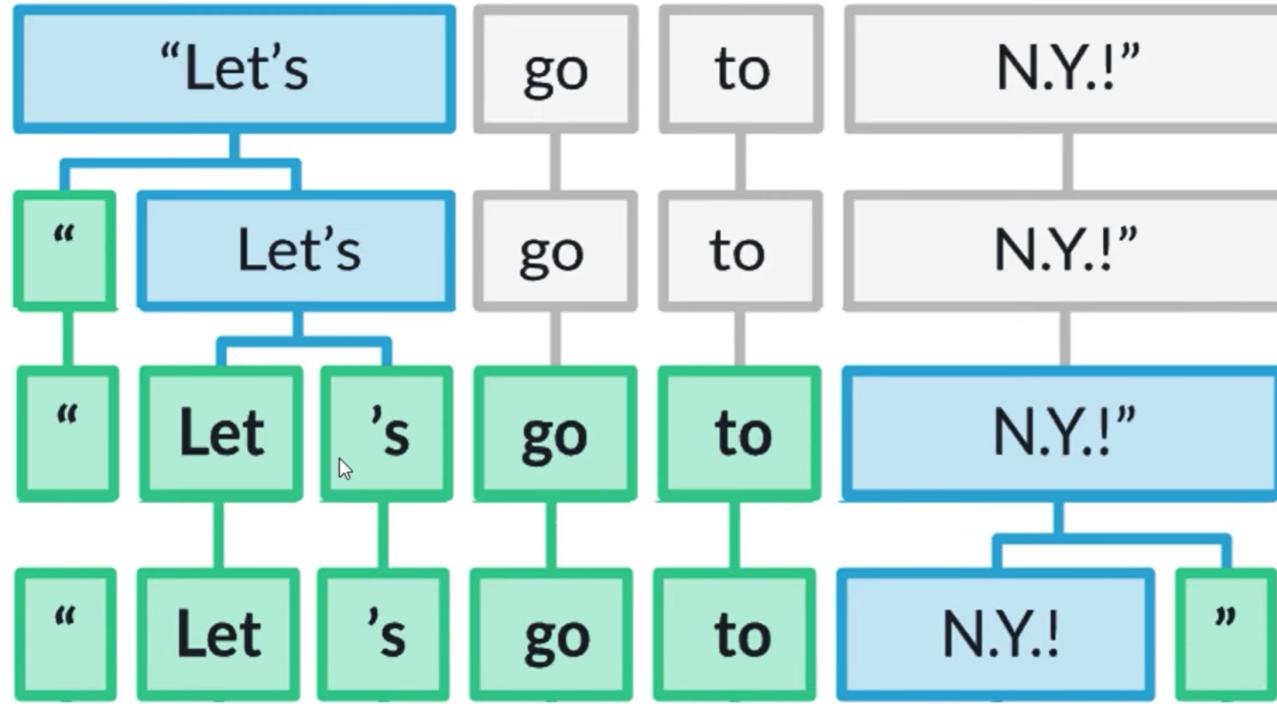


PREFIX

\$ ( “

EXCEPTION





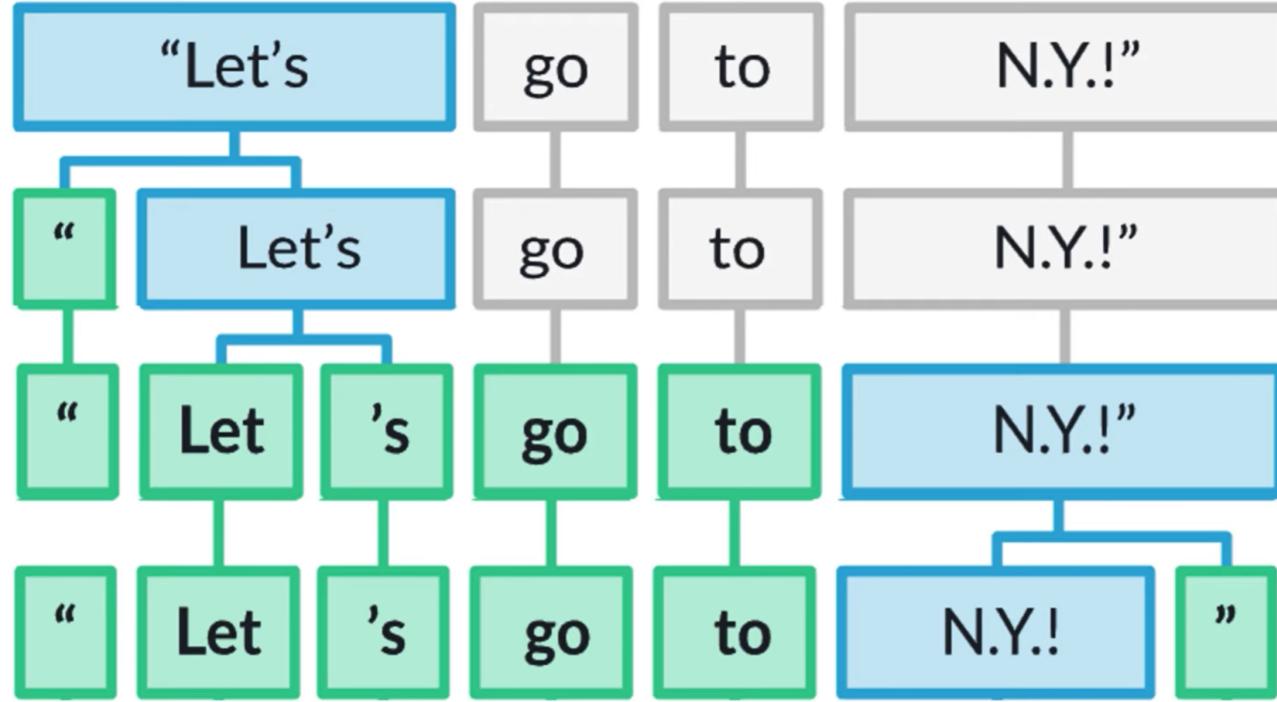
PREFIX

\$ ( “

EXCEPTION

SUFFIX





PREFIX

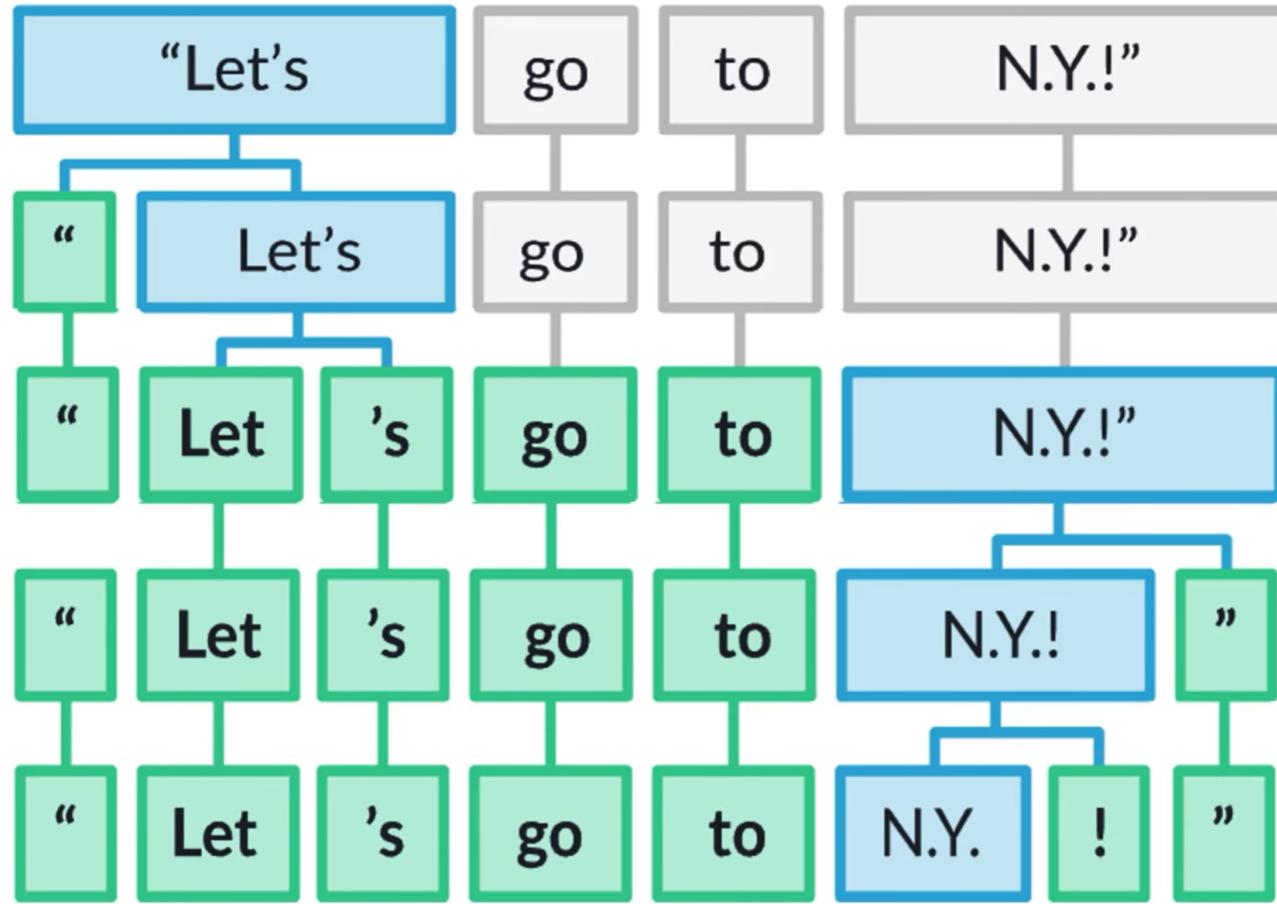
\$ ( “

EXCEPTION

SUFFIX

km ) ! “





PREFIX

\$ ( “

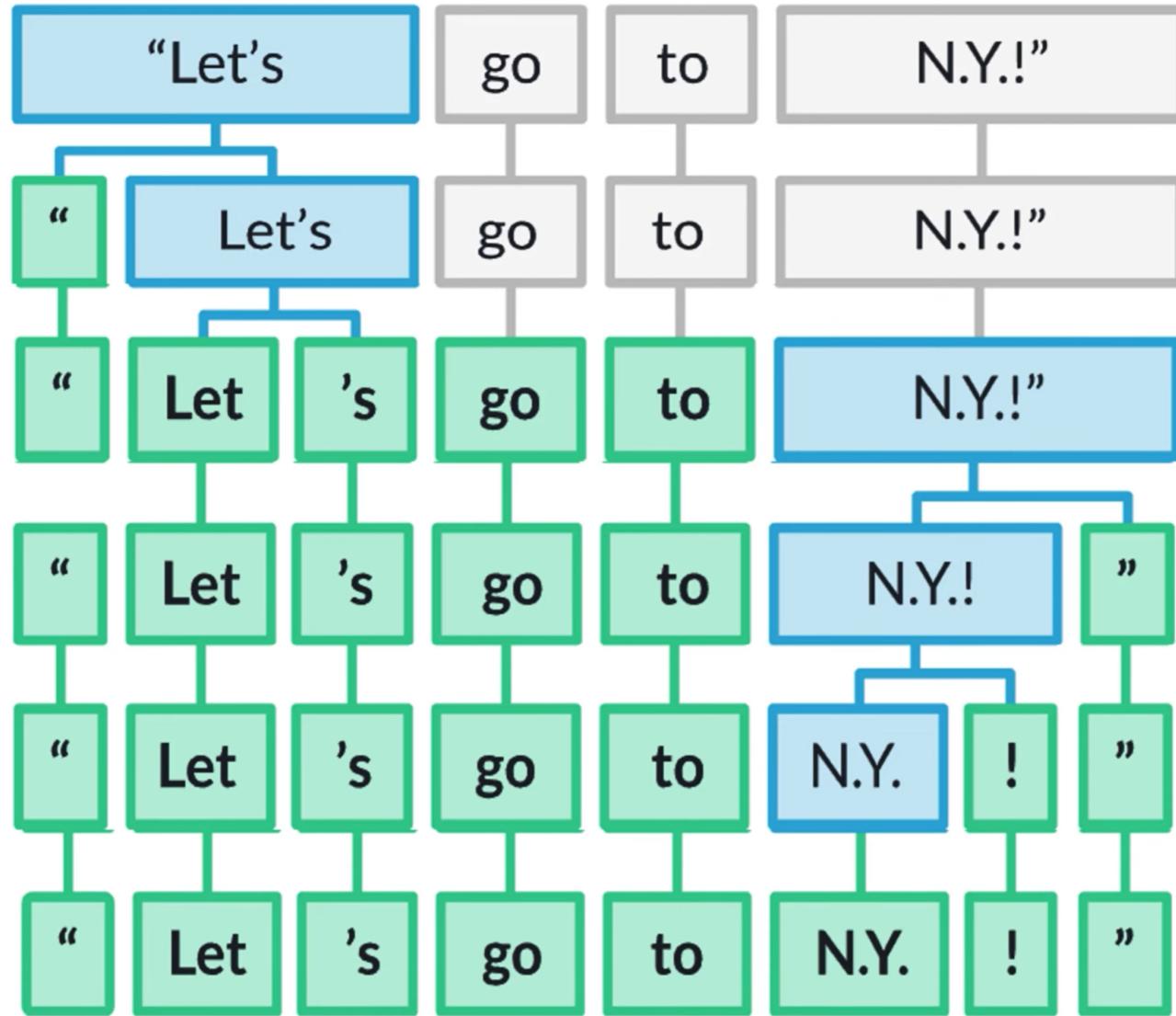
EXCEPTION

km ) ! “

SUFFIX

SUFFIX





PREFIX

\$ ( “

EXCEPTION

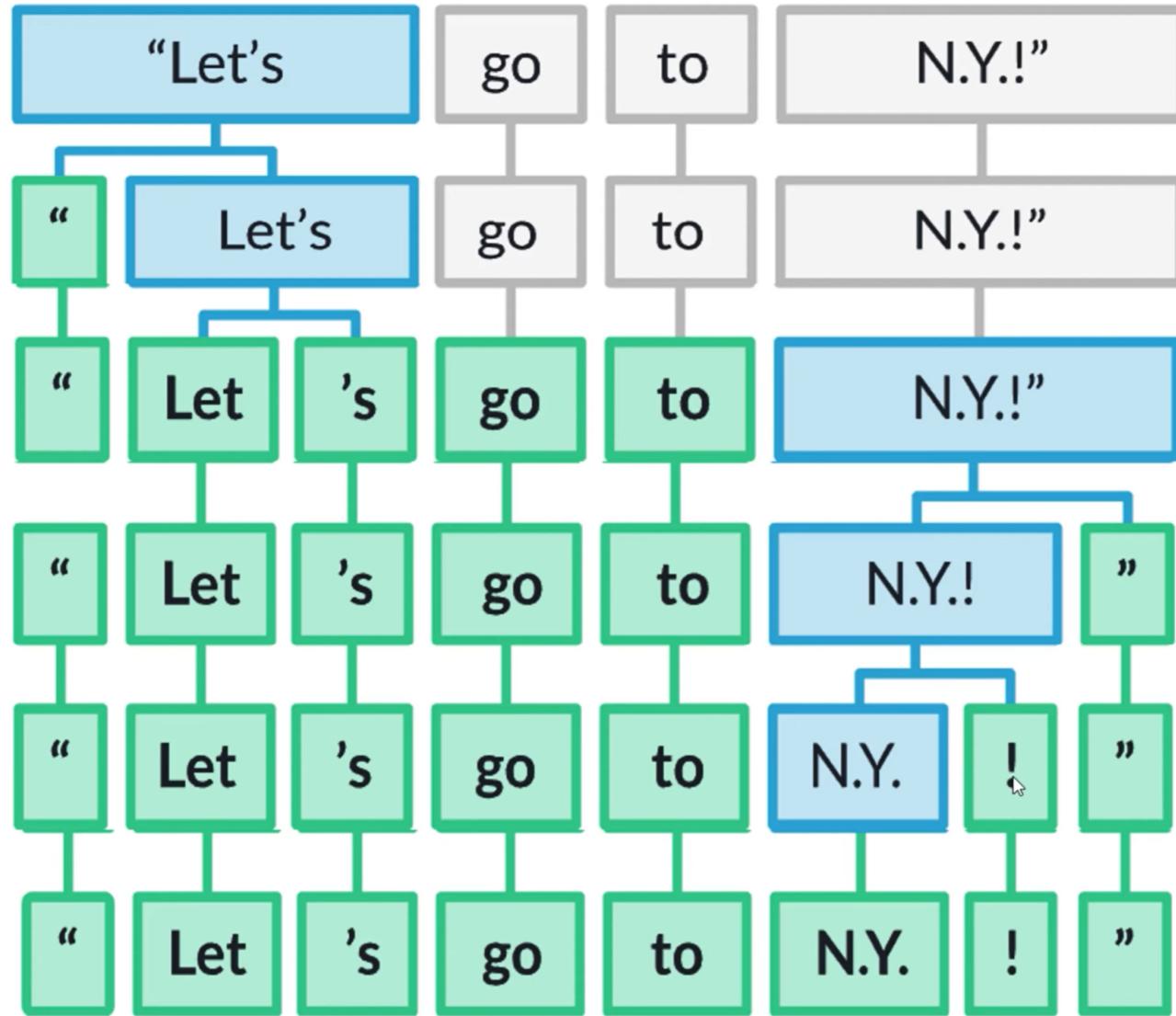
km ) ! “

SUFFIX

SUFFIX

EXCEPTION





PREFIX

\$ ( “

EXCEPTION

km ) ! “

SUFFIX

SUFFIX

EXCEPTION

DONE

