Stuttgart Fall School in CL
Class 1: (Formal) Languages, Regular Languages

Dr. Meaghan Fowlie

September 15, 2019

Course Aims

By the end of the course, I want you to:

- Be able to think computationally about language and linguistic theory
- Have a basic understanding of how parsers behave and why
- Be able to situate human language in the broad spectrum of logically possible languages

Course Aims

By the end of the course, I want you to:

- Be able to think computationally about language and linguistic theory
- Have a basic understanding of how parsers behave and why
- Be able to situate human language in the broad spectrum of logically possible languages

Course Aims

By the end of the course, I want you to:

- Be able to think computationally about language and linguistic theory
- Have a basic understanding of how parsers behave and why
- Be able to situate human language in the broad spectrum of logically possible languages

Opening Exercise

Query: *What is human language?*

Formal Language Theory

Definition (Alphabet)

A (usually finite) set of symbols (commonly written Σ)

Definition (Language)

A set of sequences of elements from an alphabet ($L \subseteq \Sigma^*$)

Definition (Grammar)

A finite set of rules defining a language

Notation (ϵ)

Use ϵ to stand for the empty string (the sequence of symbols with nothing in it). (Also common is λ, but linguists tend to avoid it because a lot of linguists use lambda calculus!)

Query: *What are we missing with these definitions?*

Formal Language Theory

Definition (Alphabet)

A (usually finite) set of symbols (commonly written Σ)

Definition (Language)

A set of sequences of elements from an alphabet ($L \subseteq \Sigma^*$)

Definition (Grammar)

A finite set of rules defining a language

Notation (ϵ)

Use ϵ to stand for the empty string (the sequence of symbols with nothing in it). (Also common is λ, but linguists tend to avoid it because a lot of linguists use lambda calculus!)

Query: *What are we missing with these definitions?*

Formal Language Theory

Definition (Alphabet)

A (usually finite) set of symbols (commonly written Σ)

Definition (Language)

A set of sequences of elements from an alphabet ($L \subseteq \Sigma^*$)

Definition (Grammar)

A finite set of rules defining a language

Notation (ϵ)

Use ϵ to stand for the empty string (the sequence of symbols with nothing in it). (Also common is λ, but linguists tend to avoid it because a lot of linguists use lambda calculus!)

Query: *What are we missing with these definitions?*

Formal Language Theory

Definition (Alphabet)

A (usually finite) set of symbols (commonly written Σ)

Definition (Language)

A set of sequences of elements from an alphabet ($L \subseteq \Sigma^*$)

Definition (Grammar)

A finite set of rules defining a language

Notation (ϵ)

Use ϵ to stand for the empty string (the sequence of symbols with nothing in it). (Also common is λ, but linguists tend to avoid it because a lot of linguists use lambda calculus!)

Query: *What are we missing with these definitions?*

Formal Language Theory

Definition (Alphabet)

A (usually finite) set of symbols (commonly written Σ)

Definition (Language)

A set of sequences of elements from an alphabet ($L \subseteq \Sigma^*$)

Definition (Grammar)

A finite set of rules defining a language

Notation (ϵ)

Use ϵ to stand for the empty string (the sequence of symbols with nothing in it). (Also common is λ, but linguists tend to avoid it because a lot of linguists use lambda calculus!)

Query: *What are we missing with these definitions?*

Human Language

- a natural object
- lives in the minds of speakers
- implicit knowledge
- Can be spoken, signed, arguably also written
- Complete knowledge of a language seems to only be possible with (strong, immersive) exposure in childhood. Call this L1/native language/mother tongue/first language(s)
- Linguistic question: what do we know when we know a language?

Human Language

- a natural object
- lives in the minds of speakers
- implicit knowledge
- Can be spoken, signed, arguably also written
- Complete knowledge of a language seems to only be possible with (strong, immersive) exposure in childhood. Call this L1/native language/mother tongue/first language(s)
- Linguistic question: what do we know when we know a language?

Human Language

- a natural object
- lives in the minds of speakers
- implicit knowledge
- Can be spoken, signed, arguably also written
- Complete knowledge of a language seems to only be possible with (strong, immersive) exposure in childhood. Call this L1/native language/mother tongue/first language(s)
- Linguistic question: what do we know when we know a language?

Human Language

- a natural object
- lives in the minds of speakers
- implicit knowledge
- Can be spoken, signed, arguably also written
- Complete knowledge of a language seems to only be possible with (strong, immersive) exposure in childhood. Call this L1/native language/mother tongue/first language(s)
- Linguistic question: what do we know when we know a language?

Human Language

- a natural object
- lives in the minds of speakers
- implicit knowledge
- Can be spoken, signed, arguably also written
- Complete knowledge of a language seems to only be possible with (strong, immersive) exposure in childhood. Call this L1/native language/mother tongue/first language(s)
- Linguistic question: what do we know when we know a language?

Human Language

- a natural object
- lives in the minds of speakers
- implicit knowledge
- Can be spoken, signed, arguably also written
- Complete knowledge of a language seems to only be possible with (strong, immersive) exposure in childhood. Call this L1/native language/mother tongue/first language(s)
- Linguistic question: what do we know when we know a language?

Is the man who is tall happy?

(1) a. He is happy
b. Is he happy?

(2) a. The man who is tall is happy
b. Is the man who is tall happy?
c. *Man the who is tall happy?
d. *Is the man who tall is happy?

Bias: language is hierarchically structured, and rules about it depend on that structure

Is the man who is tall happy?

(1) a. He is happy
b. Is he happy?

(2) a. The man who is tall is happy
b. Is the man who is tall happy?
c. *Man the who is tall happy?
d. *Is the man who tall is happy?

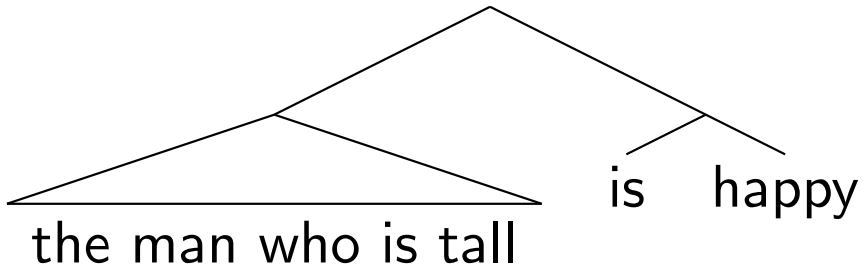
Bias: language is hierarchically structured, and rules about it depend on that structure

Is the man who is tall happy?

(1) a. He is happy
b. Is he happy?

(2) a. The man who is tall is happy
b. Is the man who is tall happy?
c. *Man the who is tall happy?
d. *Is the man who tall is happy?

Bias: language is hierarchically structured, and rules about it depend on that structure



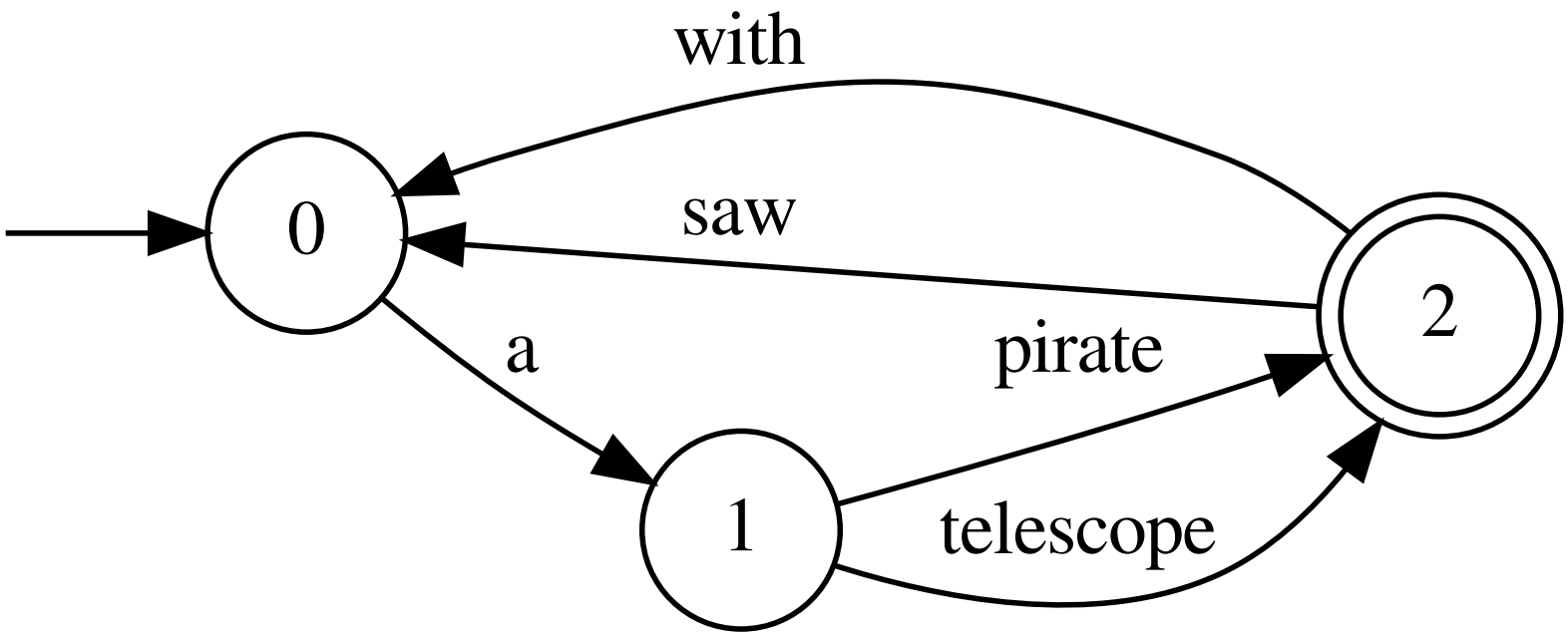
Pirate Language

Let $\Sigma_p = \{\text{a, pirate, say, telescope, with}\}$
Let $L_{\text{pirate}} = \Sigma_p^* \cap \text{English}$

Exercise

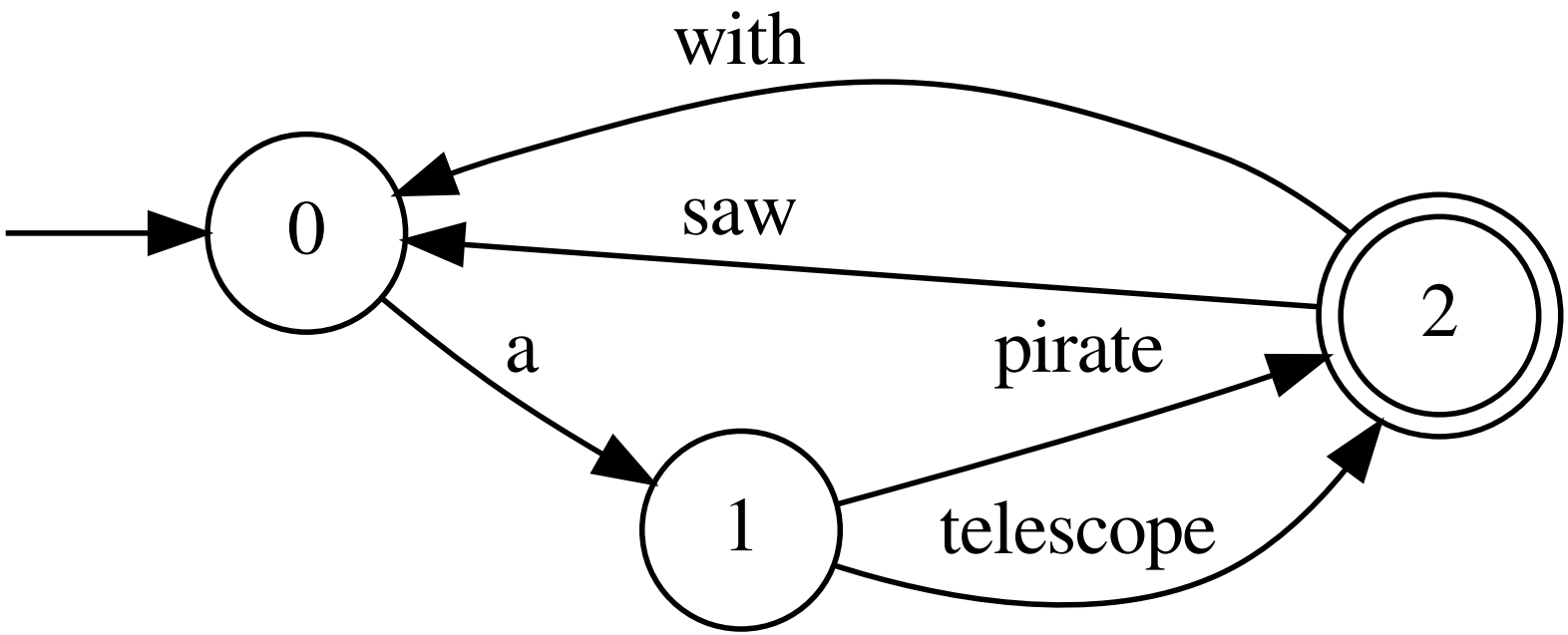
Try to write down L_{pirate}

Finite State Automata



Query: *Does this FSA generate* L_{pirate}*?*

Finite State Automata



Query: *Does this FSA generate* L_{pirate}*?*

Finite State Automata