What's New in Mathematical Linguistics

Jeffrey Heinz

V-NYI
2023/01/20

Thesis

Far from being a fossil from a former era, mathematical thinking about language

1. continues to play an essential role in understanding natural languages and
2. continues to make critical contributions to our understanding of how things which compute language—both humans and machines—can learn.

Outline

- What is mathematical linguistics?
- Past
- Present
- Future

Part I

What is mathematical linguistics?

What is mathematics?

Marcus Kracht (Los Angeles circa 2005)

"It is a way of thinking."

Eugenia Cheng *How to Bake* π (2015) : 8

"Math, like recipes, has both ingredients and method. …In math, the method is probably even more important than the ingredients."

Abstraction

Eugenia Cheng *How to Bake* π (2015) : 16/22

- "Math is there to make things simpler, by finding things that look the same if you ignore some small details."
- "Abstraction can appear to take you further and further away from reality, but really you're getting closer and closer to the heart of the matter."

Abstraction

Eugenia Cheng *How to Bake π* (2015) : 16/22

- "Math is there to make things simpler, by finding things that look the same if you ignore some small details."
- "Abstraction can appear to take you further and further away from reality, but really you're getting closer and closer to the heart of the matter."

Noam Chomsky *The Minimalist Program* (1995) : 6

"*Idealization*, it should be noted, is a misleading term for the only reasonable way to approach a grasp of reality."

I disagree with the word 'only,' but I do think abstraction is underappreciated.

Abstraction

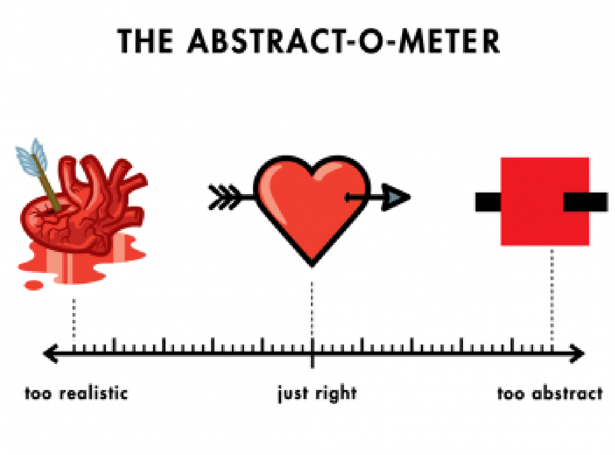
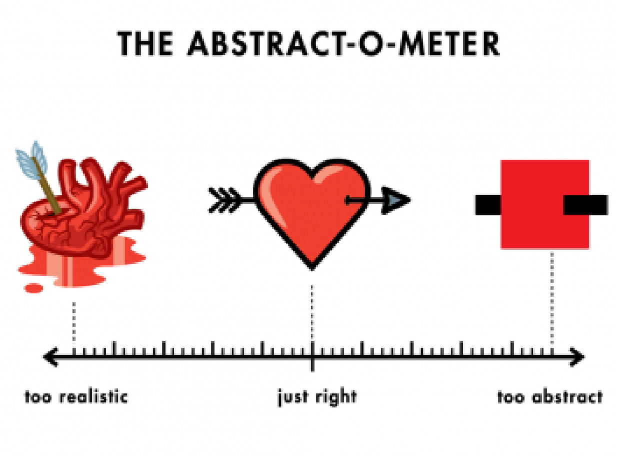


image credit: https://computersciencewiki.org/index.php/Abstraction



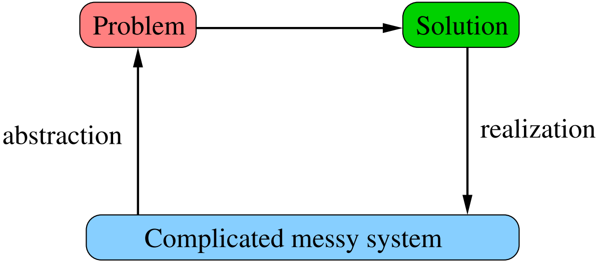
Many things were at one time considered to be "too abstract": 0, real numbers, $\sqrt{-1}$, uncountable infinity, number theory, …

image credit: `https://computersciencewiki.org/index.php/Abstraction`

Deducing consequences from premises.

Advantages:

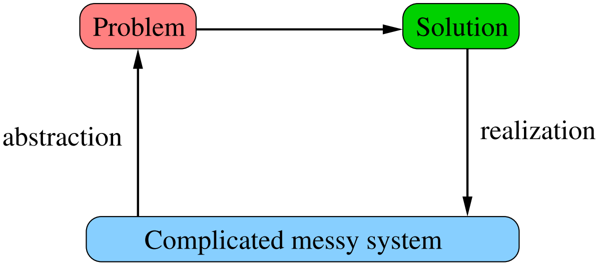
1. Can provide complete, verifiable, interpretable & understandable *solutions* to *problems*.
2. Can provide fresh insight into reality.
3. Truth is timeless.



Deducing consequences from premises.

Disadvantages:

1. The 'abstraction' and 'realization' steps take additional work and time.



Part II

The 20th Century (Past)

Timeframe

- Of course much is owed to Boole and Frege of the 19th century
- Of course much is owed to Russell, Church, Turing, Post, the Polish school of Logic, and others
- Of course there are many others I omit (Montague, Bach, Mönnich, Savitch, Johnson, ...)
- I am going to focus on specific contributions in the latter half of the 20th century which directly contributed to my own interests.

Conservativity (Keenan and Stavi 1986)

The theory of Generalized Quantifiers addresses determiner expressions like

- *every, all, some, not one, more than three, fewer than twenty, most, how many, which, more male than female, less than half, ...*

in utterances like

> ____ birds fly south for the winter.

1. What are the (possible) denotations of these expressions?
2. How arbitrary can they be?

Conservativity (Keenan and Stavi 1986)

All birds fly south for the winter.

Denotation of **all**

- ALL P Q is true iff P$\subseteq$ Q

Conservativity (Keenan and Stavi 1986)

All birds fly south for the winter.

Denotation of **all**

- ALL P Q is true iff P$\subseteq$ Q

Conservativity

D is **conservative** iff D P Q = D P R whenever P$\cap$Q = P$\cap$R.

Informally, this means that in evaluating D P Q we ignore the elements of Q which do not lie in P.

CONSERVATIVITY (KEENAN AND STAVI 1986)

All birds fly south for the winter.

Denotation of **all**

- ALL P Q is true iff $P \subseteq Q$

Conservativity

D is **conservative** iff D P Q = D P R whenever $P \cap Q = P \cap R$.

Informally, this means that in evaluating D P Q we ignore the elements of Q which do not lie in P.

An example of non-conservative D:

- HÄRTIG P Q is true iff $|P| = |Q|$

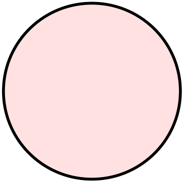
Logically Possible Generalized Quantifiers

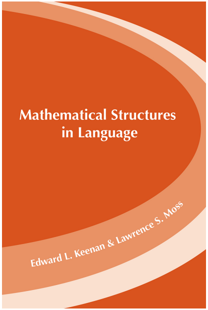
GQs satisfying conservativity

Conservativity (Keenan and Stavi 1986)

Logically Possible Generalized Quantifiers

GQs satisfying
conservativity





Mathematics of Sequences

- Language unfolds over time.
- We observe sequences of linguistic events.
- What is the mathematics of sequences?
- What is the mathematics of other relational structures like trees and graphs?

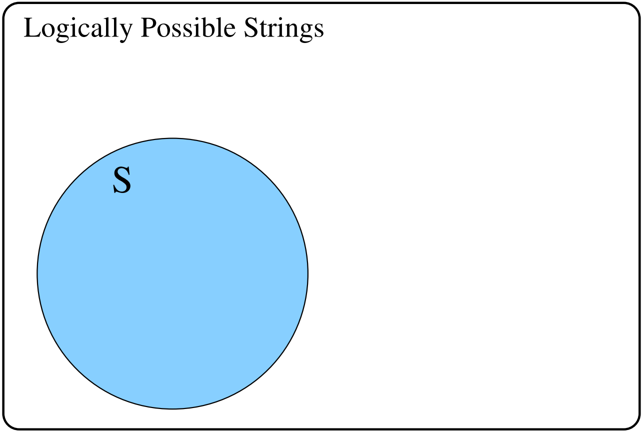
Mathematics of Sequences

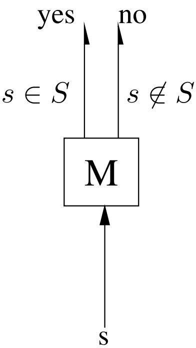
- Language unfolds over time.
- We observe sequences of linguistic events.
- What is the mathematics of sequences?
- What is the mathematics of other relational structures like trees and graphs?

Knowledge of language includes knowledge of which sequences are licit and which are not.

- John laughed and laughed. ✓
- John and laughed. ✗

A Membership problem





Variations thereof

Functions on the string domain ...

Function Type	Output Type
$\Sigma^* \rightarrow \{T, F\}$	Booleans
$\Sigma^* \rightarrow \Sigma^*$	Strings
$\Sigma^* \rightarrow \mathbb{N}$	Natural Numbers
$\Sigma^* \rightarrow [0, 1]$	Reals in the Unit Interval
$\Sigma^* \rightarrow P(\Sigma^*)$	Stringsets
...	

Variations thereof

Functions on the string domain ...

Function Type	Output Type
$\Sigma^* \to \{T, F\}$	Booleans
$\Sigma^* \to \Sigma^*$	Strings
$\Sigma^* \to \mathbb{N}$	Natural Numbers
$\Sigma^* \to [0, 1]$	Reals in the Unit Interval
$\Sigma^* \to P(\Sigma^*)$	Stringsets
...	

Mathematics classifies **numerical** functions according to general properties: linear, polynomial, trigonometric, logarithmic, ...

Variations thereof

Functions on the string domain ...

Function Type	Output Type
$\Sigma^* \to \{T, F\}$	Booleans
$\Sigma^* \to \Sigma^*$	Strings
$\Sigma^* \to \mathbb{N}$	Natural Numbers
$\Sigma^* \to [0, 1]$	Reals in the Unit Interval
$\Sigma^* \to P(\Sigma^*)$	Stringsets
...	

Mathematics classifies **numerical** functions according to general properties: linear, polynomial, trigonometric, logarithmic, ...

How can we classify functions like those above?

Classifying Membership Problems

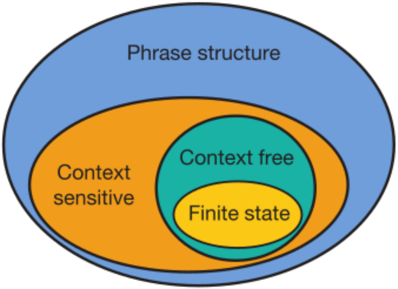
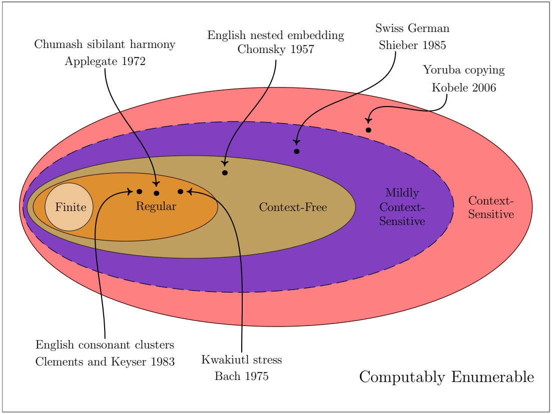


Figure 3 The Chomsky hierarchy and the logical necessity of universal grammar. Finite-state grammars are a subset of context-free grammars, which are a subset of context-sensitive grammars, which are a subset of phrase-structure grammars, which represent all possible grammars. Natural languages are considered to be more powerful than regular languages. The crucial result of learning theory is that there exists no procedure that could learn an unrestricted set of languages; in most approaches, even the class of regular languages is not learnable. The human brain has a procedure for learning language, but this procedure can only learn a restricted set of languages. Universal grammar is the theory of this restricted set.

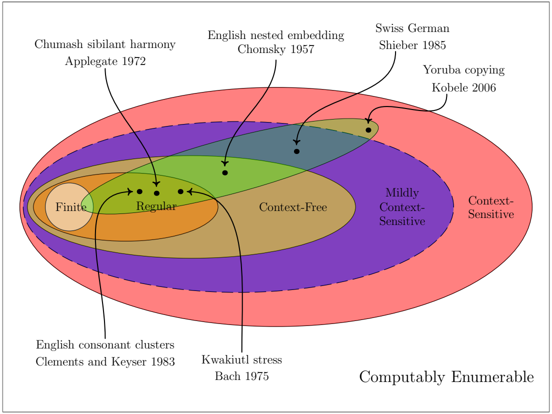
Nowak et al. 2002, *Nature*

Where is natural language?



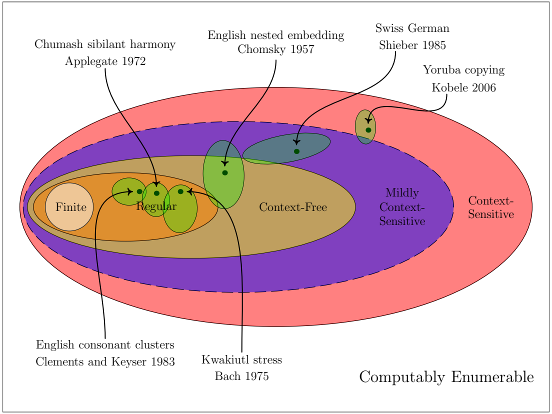
1. Morphology/phonology is regular with the exception of total reduplication (Johnson 1972, Kaplan and Kay 1994, Roark and Sproat 2007, a.o.)
2. Syntax not regular and not even context-free (Joshi 1984, Schieber 1985, Joshi, Vijay-Shanker & Weir 1991, Stabler 1997, a.o.)

Where is natural language?



1. Morphology/phonology is regular with the exception of total reduplication (Johnson 1972, Kaplan and Kay 1994, Roark and Sproat 2007, a.o.)
2. Syntax not regular and not even context-free (Joshi 1984, Schieber 1985, Joshi, Vijay-Shanker & Weir 1991, Stabler 1997, a.o.)

Where is natural language?



1. Morphology/phonology is regular with the exception of total reduplication (Johnson 1972, Kaplan and Kay 1994, Roark and Sproat 2007, a.o.)
2. Syntax not regular and not even context-free (Joshi 1984, Schieber 1985, Joshi, Vijay-Shanker & Weir 1991, Stabler 1997, a.o.)

Part III

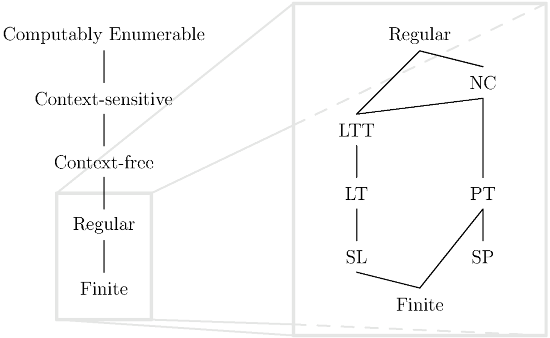
The 21st Century (Present)

Advances

1. Granularity of Characterizations
   1. Sets of sequences: $\Sigma^* \to \{0, 1\}$
   2. Sequence to sequence transformations: $\Sigma^* \to \Delta^*$
2. Connecting These Characterizations to Computational Learning Theory.
3. Unifying Computational Analysis of Syntax/Phonology/Morphology via Representations (Trees and Sequences).

Granularity of Characterizations

- The Regular/Context-Free/Context-Sensitive distinctions are **course-grained**.
- We now have much **finer-grained** characterizations.



- Strikingly, these subregular classes, when applied to trees, let us probe the non-regular stringsets!

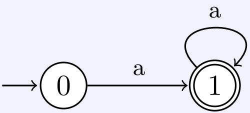
Logical formula

$(\forall x)[\mathtt{a}(x)] \wedge (\exists x)[\mathtt{a}(x)]$

Regular expression

aa^*

Finite-state acceptor



For each type of grammar, there is an effective procedure which checks whether any string satisfies the expression.

$\acute{\sigma}\sigma$	$\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}$	
$\acute{\sigma}\sigma\grave{\sigma}$	$\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma$	
$\acute{\sigma}\sigma\grave{\sigma}\sigma$	$\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}$	$\ldots$

$$\begin{aligned}
\texttt{first}(x) &\stackrel{\text{def}}{=} \neg(\exists y)[y \triangleleft x] \\
\texttt{stress}(x) &\stackrel{\text{def}}{=} \acute{\sigma}(x) \vee \grave{\sigma}(x) \\
\rtimes\acute{\sigma} &\stackrel{\text{def}}{=} (\exists x)[\acute{\sigma}(x) \wedge \texttt{first}(x)] \\
\texttt{x}\acute{\sigma} &\stackrel{\text{def}}{=} (\exists x)[\acute{\sigma}(x) \wedge \neg\texttt{first}(x)] \\
\texttt{lapse} &\stackrel{\text{def}}{=} (\exists x, y)[x \triangleleft y \wedge \sigma(x) \wedge \sigma(y)] \\
\texttt{clash} &\stackrel{\text{def}}{=} (\exists x, y)[x \triangleleft y \wedge \texttt{stress}(x) \wedge \texttt{stress}(y)]
\end{aligned}$$

Logical formula

$$\texttt{Maranungku} \stackrel{\text{def}}{=} \rtimes\acute{\sigma} \wedge \neg\texttt{x}\acute{\sigma} \wedge \neg\texttt{lapse} \wedge \neg\texttt{clash}$$

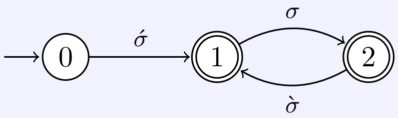
Chandlee and Heinz (2017)

Maranguku Stress

Regular expression

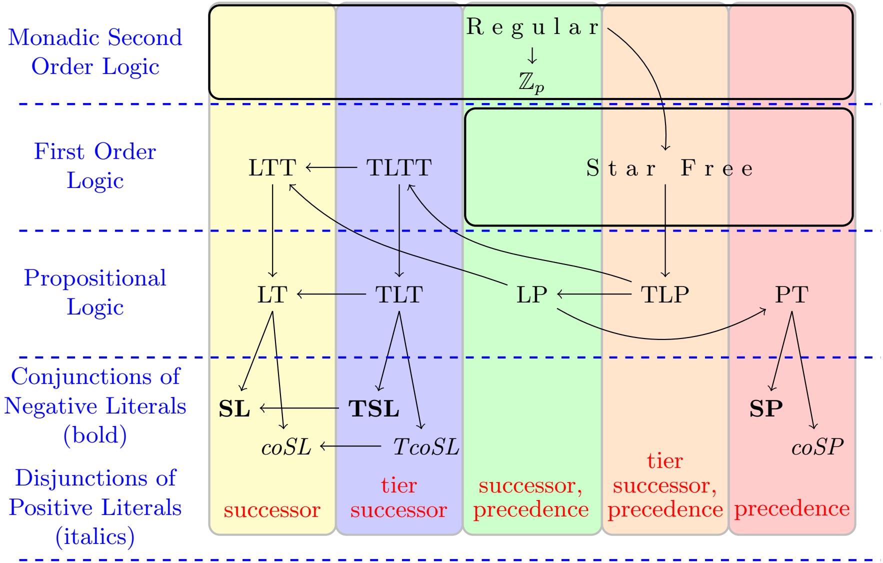
$$\acute{\sigma}(\sigma\grave{\sigma})^* + \acute{\sigma}(\sigma\grave{\sigma})^*\sigma$$

Finite-state acceptor



Chandlee and Heinz (2017)

Subregular Sets of Sequences



Phonotactic patterns

1. Nearly all are at the bottom!

   SL, coSL, TSL, TcoSL, SP, coSP

2. And are consquently describable with simple logical languages and well-behaved automata.
3. These classes are “local” in particular, precise ways.
4. These classes are sufficiently structured to facilitate their learnability (with a parameter k).

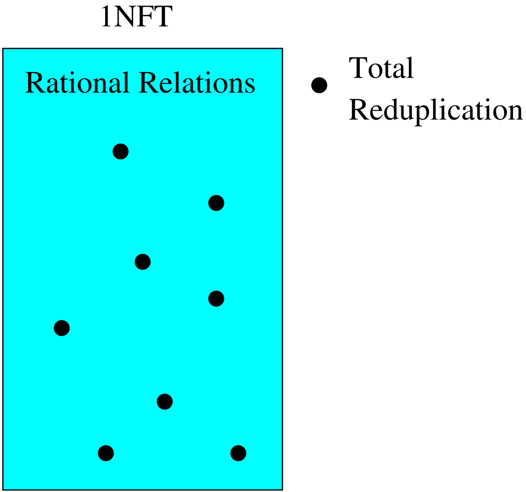
(Heinz 2018, Rogers and Lambert 2019, Lambert et al. 2021)

String-to-String Transformations

The established, foundational view (Roark and Sproat 2007)

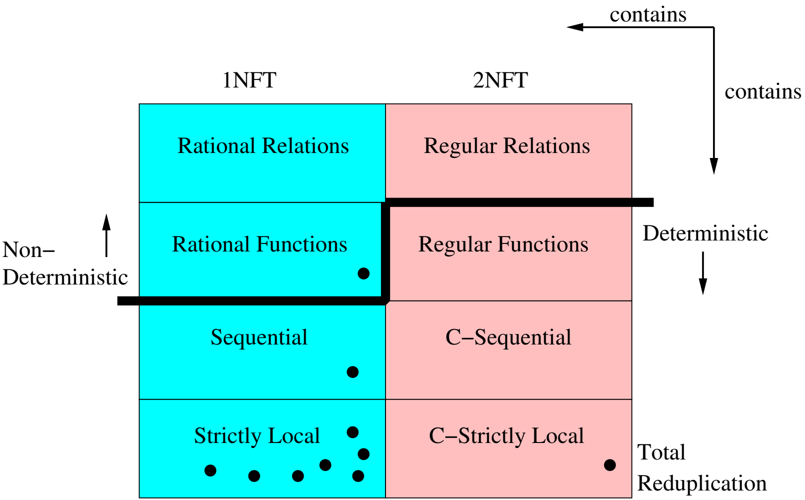
Rational Relations	non-Rational Relations
Prefixation	Total Reduplication
Suffixation	
Circumfixation	
Infixation	
Truncation	
Root and pattern	
Umlaut/Ablaut	
Partial Reduplication	
Assimilation	
Harmony	
Epenthesis	
Deletion	
...	

This basic view pictorially



String-to-String Transformations

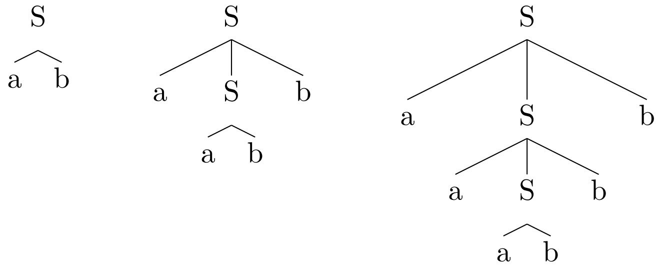
A more articulated view



(Chandlee 2017, Dolatian and Heinz 2020)

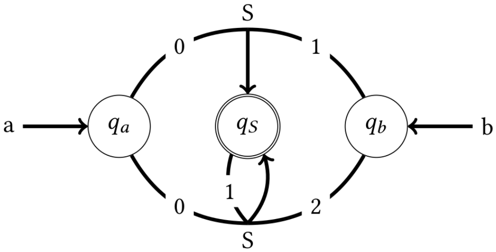
What about Syntax?

Tree structures are a foundation to distinct theories of syntax (Stabler 2019).



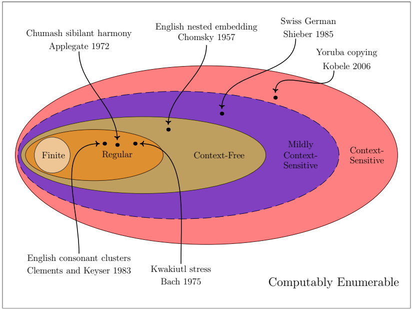
...

Logic and Finite-State Automata handle trees too!



(image: Lambert 2022)

Mildly Context Sensitive Stringsets



1. Take a set of trees T recognized by a finite-state tree acceptor.

$$T$$

2. Transform the trees in T with a finite-state tree transducer F.

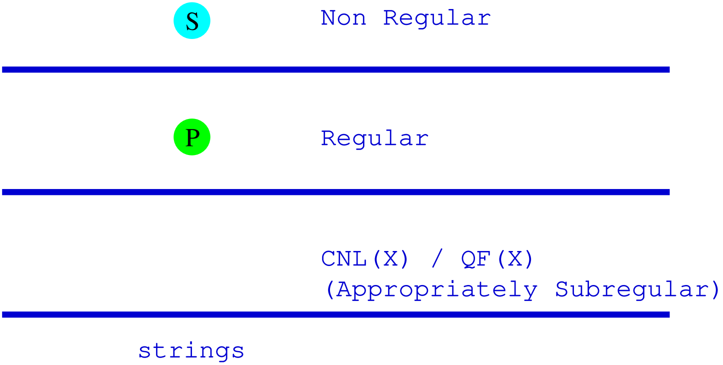
$$F(T)$$

3. Spell-out those trees to yield a set of strings.

$$Y(F(T))$$

4. That set of strings is mildly context-sensitive!

Subregular Syntax



20th century view

S Non Regular

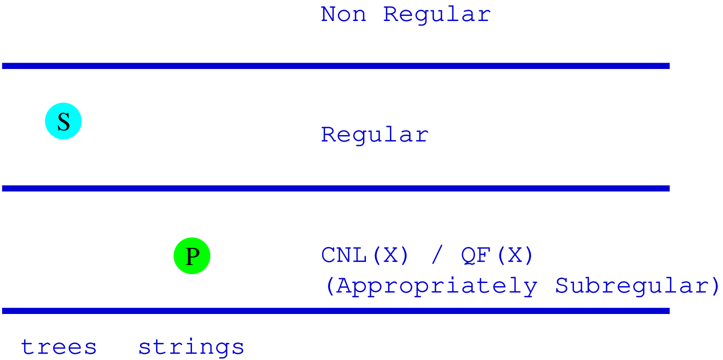
Regular

P CNL(X) / QF(X)
(Appropriately Subregular)

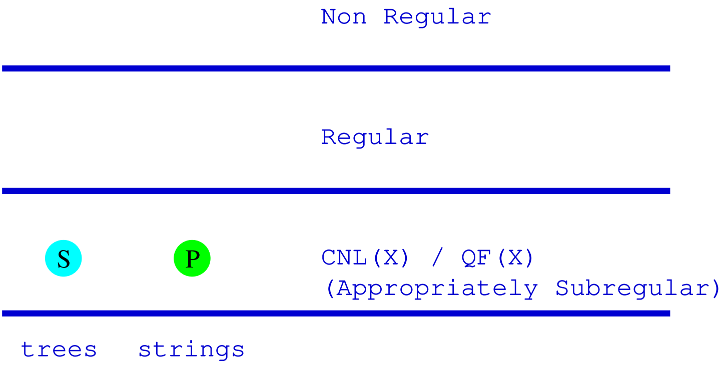
---

strings

(Heinz 2018)



(Stabler 2019)


Non Regular
Regular
S
P
CNL(X) / QF(X)
(Appropriately Subregular)
trees
strings

(Graf 2022)

Part IV

The Future (Conclusion)

The future is now

Main points

1. Mathematical characterizations of linguistic structure teach us that many, if not all, linguistic generalizations are subregular
2. This means for example that they can be described with first order logic or even weaker logics.
3. These characterizations provide insight into locality and other conditions on linguistic structures.
4. They also help us understand language learnability (not much discussed in this talk)

The future is now

Future

1. Continue to study other linguistic representations (autosegmental, prosodic, features)
2. Continue to connect mathematical characterizations to learnability and acquisition
3. Continue to draw more explicit parallels between disparate modules (e.g. wh-movement in syntax and consonant harmony in phonology)
4. Continue to develop applications and scale them up

THANK YOU

I ♥ V-NYI