Introduction to Computational Linguistics

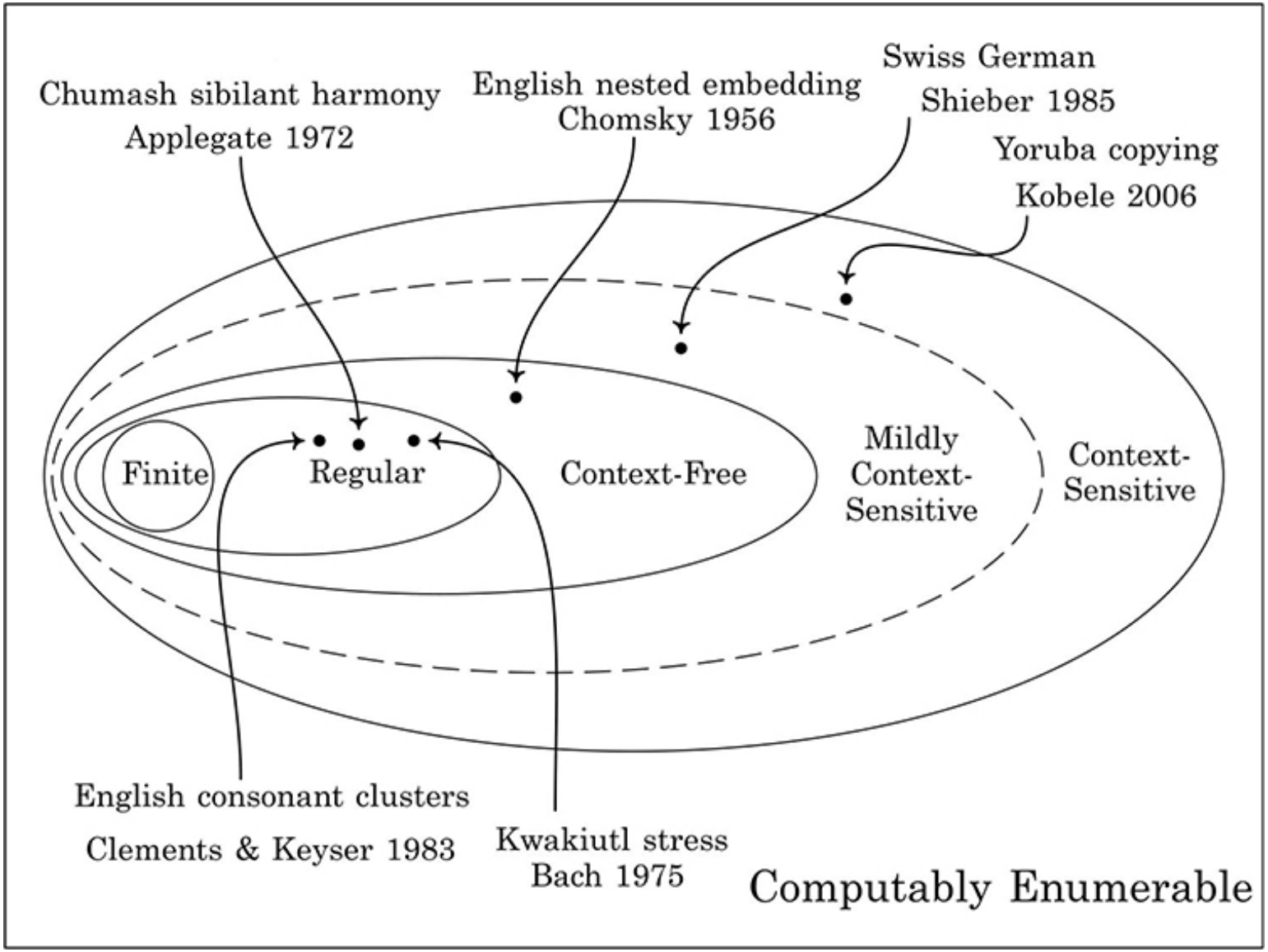
Section

Olga Zamaraeva
University of Washington
April 10, 2020

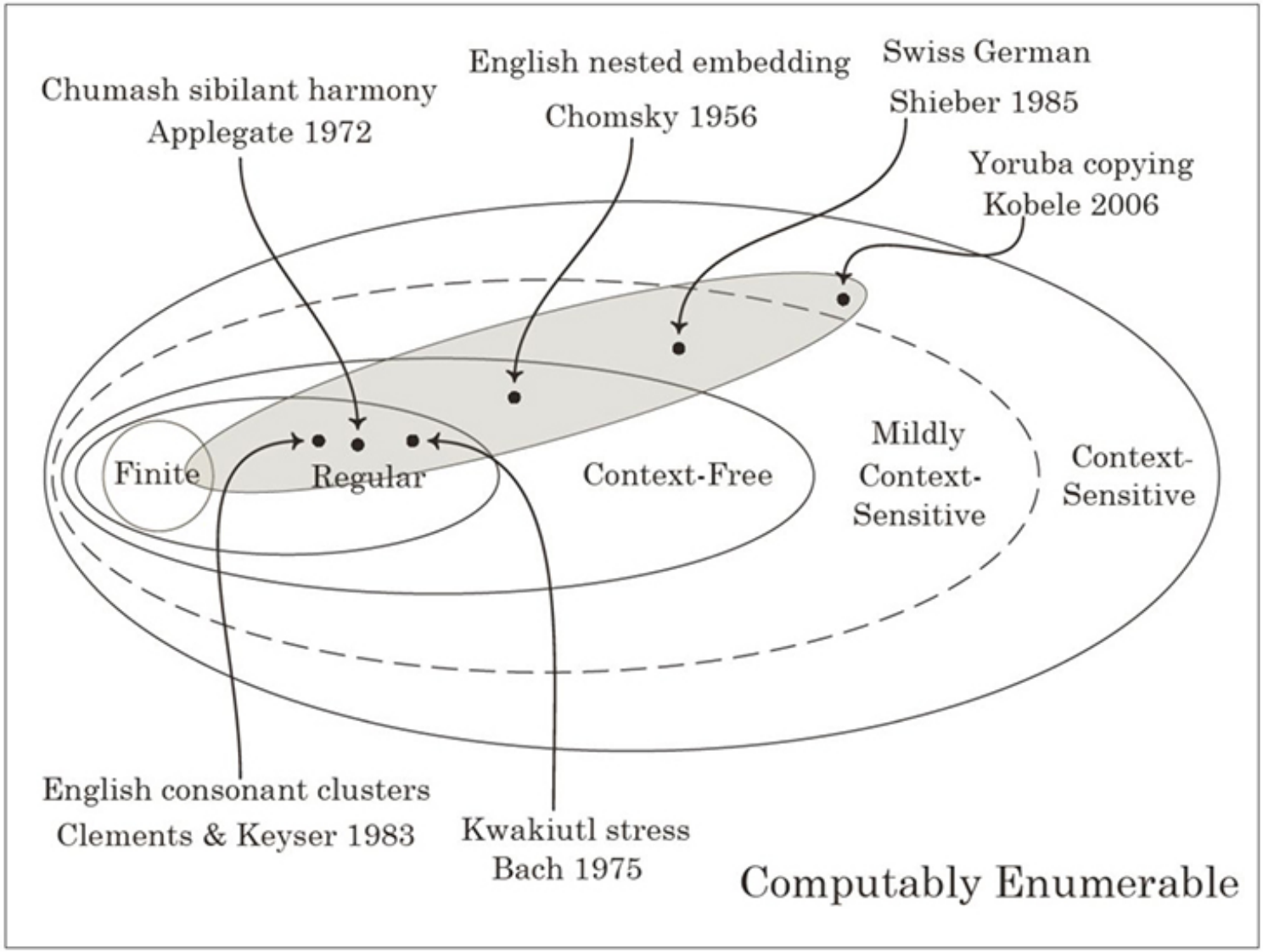
Plan for today

- My notes on language theory
- Assignment 1 final notes
- Assignment 2 overview

Language Theory: Why the Chomsky Hierarchy?



Language Theory: What Does the Learner Need to Learn?



Difficulties, Complaints, and Questions

- Introductory programming
  - I have no idea what to do with this tool/program!!!
- Introductory linguistics
  - The assignment is unclear!!!
- Why is intro to programming not a prerequisite?
- Why is basic programming worth learning?
- Can't you just give us the full list of sentences/words our program is supposed to handle?

My advice

- Work in a **systematic** manner
- Analyse the tasks into small components
- Make sure you understand each component separately before combining things

My advice

- Work in a **systematic** manner
- Analyse the tasks into small components
- Make sure you understand each component separately before combining things
- Example: What does this line of code mean?
  - `str = re.sub(pattern,replacement,str)`

My advice

- Work in a **systematic** manner
- Analyse the tasks into small components
- Make sure you understand each component separately before combining things
- Example: What does this line of code mean?
  - `str = re.sub(pattern,replacement,str)`
  - Look up variables, assignment, functions, arguments
  - Look up re.sub() definition in python3 docs
  - Debug your program step by step
  - Pay attention to the program state

My advice

- Work in a **systematic** manner
- Analyse the tasks into small components
- Make sure you understand each component separately before combining things
- Example: What does this line of code mean?
  - `str = re.sub(pattern,replacement,str)`
  - Look up variables, assignment, functions, arguments
  - Look up re.sub() definition in python3 docs
  - Debug your program step by step
  - Pay attention to the program state
- Example: What does the ask mean?
  - *"sentences.txt must include the subject-verb agreement paradigm of the verb "to be" in 1st and 2nd person"*

My advice

- Work in a **systematic** manner
- Analyse the tasks into small components
- Make sure you understand each component separately before combining things
- Example: What does this line of code mean?
  - `str = re.sub(pattern,replacement,str)`
  - Look up variables, assignment, functions, arguments
  - Look up re.sub() definition in python3 docs
  - Debug your program step by step
  - Pay attention to the program state
- Example: What does the ask mean?
  - *"sentences.txt must include the subject-verb agreement paradigm of the verb "to be" in 1st and 2nd person"*
  - Look up subject-verb agreement in a linguistics text
  - Get access to the full paradigm in e.g. ESL materials
  - Define and/or cite things in your write up

Debugging the program

- Shown in the section video from April 3, starting at about minute 34

Assignment 2

- Excellent tutorial, do read (a few times if necessary)
- Technical component
  - Entirely in command line
    - visualization of FSA possible using a program called “graphviz” (not required)
  - No python
  - A special language/formalism called “foma”
  - The off-the-shelf program “foma” which understands that language/formalism
- Linguistic component
  - Basic (but fairly extensive!) morphology; phonology
  - Linguistic description of an unfamiliar language

Working on patas (if you have Windows)

- Connect to patas via ssh
- Type `/opt/foma/bin/foma`
- (This is also in your instructions for Assign.2)
- To exit: `Ctrl+D`
- Note that you can still edit your files locally and copy them to patas to run `foma` on them

Working locally (recommended for OSX/linux)

- Install foma following the instructions for Assign.2
- Copy the file “foma” into your `/usr/local/bin`
  - In terminal, from the directory where you put the new fles: `cp foma /usr/local/bin/`

foma example

- example.lexc and example.foma
- Representing a trivial example with English morphology and phonology
  - The tutorial has another, more complete one
- Files are available to you on Canvas/Files
- Start with the lexc file (morphology)
- Start foma, read the lexicon FSA in, examine the output (`pairs`)
  - How do I read the lexicon FSA in???
- Continue with the foma file (phonology)
- if you have `graphviz`, try `view`
  - graphviz may or may not install on your machine
  - get it from their website rather than pip