

# Ling 555 — Programming for Linguists

Python — Linguistic Examples and Practice

Robert Albert Felty

Speech Research Laboratory  
Indiana University

Oct. 13, 2008

## Prereqs

theory

Practice

homework

Linguistic  
Examples

Resources

### While we are waiting

Please download the courseBackground.txt file from:

[http://robfelty.com/teaching/L555Fall2008/  
resources/sapir.txt](http://robfelty.com/teaching/L555Fall2008/resources/sapir.txt)

```
curl -o courseBackground.txt
```

[http://robfelty.com/teaching/L555Fall2008/  
resources/sapir.txt](http://robfelty.com/teaching/L555Fall2008/resources/sapir.txt)

### For Wednesday:

Read Chapter 6 on Abstraction (Functions)

# Outline

theory

① Theory

- Sub-problems
- Class specific advice

Practice

② Practice

- Headers
- Debugging
- Double for loops

homework

③ Homework questions

Linguistic  
Examples

④ Linguistic Examples

- frequency
- co-occurrence

Resources

⑤ Resources

# Sub-problems

theory

sub-problems

L555

Practice

homework

Linguistic  
Examples

Resources

- Start out with well known idioms, like how to open and read a file
- Compare the current problem to other problems you have solved
- Break the problem down into chunks, and solve a bit at a time
  - If you know you need to use a loop, write the inside of the loop first, with one item, then write the loop

# Class specific advice

theory

sub-problems

L555

Practice

homework

Linguistic  
Examples

Resources

## Don't fall behind!

This class is cumulative. That is, we will continue to build on our knowledge throughout the course. If you don't understand a particular concept, come talk to me before we move on.

## How not to fall behind

- Read my comments in your homework
- Look over my solutions to the homework.
  - If you don't understand it, e-mail or talk to me.
- Read the assigned readings before class, and try out some of the examples in the book as you read

# Using Headers

theory

Practice

headers

debug

forfor

homework

Linguistic  
Examples

Resources

## Example

Instead of simply ignoring headers, we can use them to structure our data

```
for line in fileObject:
    line=line.rstrip('\n')
    if firstline==true:
        firstline=false
        headers=line.split('\t')
        continue
    columns=line.split('\t')
    # store second and third column
    contentsDict[columns[0]]=
        {headers[1]: columns[1],
         headers[2]: columns[2]}
```

# Debugging

theory

Practice

headers

debug

forfor

homework

Linguistic

Examples

Resources

## Definition

Roughly speaking, debugging is the process of checking to see that your code is actually doing what you expect it to do.

## method

- 1 Print statements
- 2 Debugger

# Debugging with print statements

theory

Practice

headers

debug

forfor

homework

Linguistic  
Examples

Resources

## Example

```
mylist=list()
for line in file:
    columns=line.split()
    mylist.append(columns)
pprint.pprint(mylist)
```



# Debugging with a debugger

theory

Many programming languages come with a debugger.

Practice

Many IDEs also come with a debugger.

headers

debug

forfor

homework

Linguistic  
Examples

Resources

## Using the IDLE debugger

- 1 Open a python script you have written
- 2 From the Python shell window, select Debug > debugger
  - This opens a Debugger control window
- 3 Check local, globals, and stack
- 4 Now run your script by hitting F5
- 5 Instead of running the whole script, the debug control window should come to the forefront (if not, click on that window)

# IDLE debugger continued

theory

Practice

headers

debug

forfor

homework

Linguistic  
Examples

Resources

## Controlling program flow

- Go** Run up until the next breakpoint or the end of the program
- Step** Go through the program line by line by clicking on step. Note that this will also step through any functions or modules
- Over** Go through program line by line, but don't go into functions or modules
- Out** Run through the end of the current function (or, if not inside a function, until the end of the script)

# Breakpoints

theory

Practice

headers

debug

forfor

homework

Linguistic  
Examples

Resources

## Example

Your program returns an error on line 32. If you set a breakpoint at line 32, you can run your program up to that point, and see what might be wrong

- To set a breakpoint in IDLE, right-click on a line in your script file(or Control-click if you only have one mouse button on your Mac)
- Select “add breakpoint”
- The line should be highlighted
- Now run the program again, and select “Go” in the debug control window
- Your program will be executed up to (but not including) the line where you set a breakpoint

# Double for loops

theory

Practice

headers

debug

forfor

homework

Linguistic

Examples

Resources

## Example

Just like we can nest conditionals, we can also nest loops

```
lol=[[10,20,30],['a','b','c'],['foo','bar']]  
for row in range(len(lol)):  
    for column in  
        range(len(lol[row])):  
        print row, column,  
            lol[row][column]
```

More examples in [forfor.py](#)

# Homework questions?

**theory**

**Practice**

**homework**

**Linguistic  
Examples**

**Resources**

# frequency

theory

Practice

homework

Linguistic  
Examples

frequency

co-occurrence

Resources

## What to do with a dict?

You could construct a dictionary of a corpus's word frequencies. Each word would be a key, and the corresponding value would be its frequency.

theory

Practice

homework

Linguistic  
Examples

frequency

co-occurrence

Resources

## Definition

One way to measure semantic relatedness is by co-occurrence frequency. If two words frequently occur close to each other in text, they probably share a semantic relationship.

## Different kinds of frequency

- Some words might be very frequent, and relatively context independent. That is, they occur with many other words
- Other words might be frequent, but only occur with a few other words
- These types of words are probably stored and accessed differently in the brain

## Resources for you:

theory

Practice

homework

Linguistic  
Examples

Resources

Some of the examples we have covered in class today can be found on the website at:

<http://robfelty.com/teaching/L555Fall2008/resources/py>

The files from today are:

- 1 useHeaders.py
- 2 forfor.py
- 3 co-occurrence.py