

Language & Technology

Wrapping Up

Thomas Graf

Stony Brook University
`lin120@thomasgraf.net`

Be Skeptical

- ▶ Humans overestimate technology.
Loebner Prize/Turing test
- ▶ Don't buy the hype.
deep learning, artificial intelligence, culturomics, neuroscientism, ...
- ▶ Technological advances also come with downsides.
unemployment, rising job requirements, mass surveillance, ...

The State-of-the-Art in Language Technology

- ▶ Almost everything involves **n-grams models**.
 - ▶ word completion/prediction
 - ▶ stylistic analysis
 - ▶ OCR
 - ▶ spell checking
 - ▶ web search
 - ▶ word meaning
 - ▶ machine translation
- ▶ Frequency information from corpora/tree banks is used to add **probabilities** to the models.

Language is Complex

- ▶ Language comes easy to humans, but it is **incredibly complex**.
 - ▶ culture-specific rules of turn taking
 - ▶ hidden structure of sentences
 - ▶ how that hidden structure is inferred
 - ▶ building sentence meaning from word meaning
 - ▶ tons of variation across languages and dialects
- ▶ We have to deal with this complexity.
- ▶ Probabilities won't cut it because of **Zipf's law**.

Related Linguistics Courses

- ▶ LIN 101 *Human Language*
- ▶ LIN 110 *Anatomy of English Words*
- ▶ LIN 200 *Language in the USA*
- ▶ as background for sentence models:
 - LIN 311 *Syntax*
 - LIN 346 *Language & Meaning*
- ▶ as background for speech recognition:
 - LIN 201 *Phonetics*
 - LIN 301 *Phonology*

Planned Course for Fall 2018

- ▶ LIN 425 *Computational Data Analysis*
- ▶ computer-assisted statistics and data analysis
- ▶ useful for both computational and experimental linguistics

Related Linguistics Courses

- ▶ LIN 101 *Human Language*
- ▶ LIN 110 *Anatomy of English Words*
- ▶ LIN 200 *Language in the USA*
- ▶ as background for sentence models:
 - LIN 311 *Syntax*
 - LIN 346 *Language & Meaning*
- ▶ as background for speech recognition:
 - LIN 201 *Phonetics*
 - LIN 301 *Phonology*

Planned Course for Fall 2018

- ▶ LIN 425 *Computational Data Analysis*
- ▶ computer-assisted statistics and data analysis
- ▶ useful for both computational and experimental linguistics

My Computational Courses This Coming Spring

1 LIN 488 Internship

Help me adapt my Jupyter notebooks on mathematics in linguistics for an undergraduate audience.

2 LIN 630 Parsing & Processing

Computational parsing techniques and how they compare to human sentence processing.

Prerequisite: LIN 311 Syntax

Two Course Recommendations Outside Linguistics

- 1 SOC 330 *Media and Society*
(Jason J. Jones; no programming, but a bit of computational sociology and big data)
- 2 AMS 103 *Applied Math in Technology*
(Matthew Reuter; includes a live cookie baking session)

► **Universal**

- strings, integers, lists
- variables
- print, input
- conditionals
- while loops
- for loops
- custom functions
- regular expressions

► **Python-Specific**

- counters
- positions and slices
- built-in functions like `len`, `str.lower`, ...

Sticking with Python

- ▶ You invested lots of time and energy on picking up some programming skills. They **should not go to waste!**
- ▶ The most important thing: find yourself a **project!**

Project suggestions

- ▶ Computational analysis in a book report
- ▶ Automatic menu generator for New York hipster restaurants
- ▶ Facebook bot to like all your friends' posts and leave a short reply
- ▶ Script to delete all unread emails that are older than 30 days
- ▶ Reminder to take a break after 2h in front of the screen
- ▶ Automating something you frequently do manually

Some Project Ideas from Reddit

Check out the Reddit thread:

Our department keeps a calendar of activities in excel. I use python to import that spreadsheet, break the activities into itemized tasks, put due dates on each task, and spit out those tasks for each teammate.

I stitched together a video encoding script, a mp3 tag reading script and a YouTube uploading script to automatically upload all of the old music I'd recorded to YouTube. Saved me 100s of hours of manual work.

Some Project Ideas from Reddit

Check out the Reddit thread:

Our department keeps a calendar of activities in excel. I use python to import that spreadsheet, break the activities into itemized tasks, put due dates on each task, and spit out those tasks for each teammate.

I stitched together a video encoding script, a mp3 tag reading script and a YouTube uploading script to automatically upload all of the old music I'd recorded to YouTube. Saved me 100s of hours of manual work.

Some Project Ideas from Reddit

Check out the Reddit thread:

Our department keeps a calendar of activities in excel. I use python to import that spreadsheet, break the activities into itemized tasks, put due dates on each task, and spit out those tasks for each teammate.

I stitched together a video encoding script, a mp3 tag reading script and a YouTube uploading script to automatically upload all of the old music I'd recorded to YouTube. Saved me 100s of hours of manual work.

Some Project Ideas from Reddit [cont.]

This is a pretty silly usage, but the autopsy package is really great for making auto-clickers for various web-games.

I have a script that checks the Powerball and Megamillions jackpot size and calculates the expectation value of a play. If either is positive, it sends me a text to buy a ticket.

Not complete yet, but I am using python on a RasPi to automate the temperature, humidity, and lighting for my two pythons' cages

Some Project Ideas from Reddit [cont.]

This is a pretty silly usage, but the autopsy package is really great for making auto-clickers for various web-games.

I have a script that checks the Powerball and Megamillions jackpot size and calculates the expectation value of a play. If either is positive, it sends me a text to buy a ticket.

Not complete yet, but I am using python on a RasPi to automate the temperature, humidity, and lighting for my two pythons' cages

Some Project Ideas from Reddit [cont.]

This is a pretty silly usage, but the autopsy package is really great for making auto-clickers for various web-games.

I have a script that checks the Powerball and Megamillions jackpot size and calculates the expectation value of a play. If either is positive, it sends me a text to buy a ticket.

Not complete yet, but I am using python on a RasPi to automate the temperature, humidity, and lighting for my two pythons' cages

Some Project Ideas from Reddit [cont.]

My brother in law has a "habit" of going to jail. All. The. Time. My wife wants to know when it happens, so I setup a script to poll the local jail inmate rosters, and send me an email when his name shows up. It works pretty well.

Other Fun Stuff with Python

- ▶ Python coding for *Minecraft*
<http://www.instructables.com/id/Python-coding-for-Minecraft/>
- ▶ Modding *The Sims 4*
http://simswiki.info/wiki.php?title=Tutorials:TS4_General_Modding
- ▶ Creating hi-res texture mods (e.g. for Resident Evil 4 HD)
https://youtu.be/u_8l3vaVrfE
- ▶ *Gray Hat Python* book on hacking with Python

Code Combat & Code Wars



If you liked Code Combat, check out Code Wars!

How much Python Do I Know?

- ▶ We covered enough Python for 90% of all problems.
- ▶ But Python can do a lot more:
 - ▶ strides
 - ▶ dictionaries (of which Counters are a special case)
 - ▶ unpacking
 - ▶ zipping
 - ▶ recursive functions
 - ▶ object oriented programming with classes
 - ▶ generators
 - ▶ decorators
- ▶ And there's many powerful libraries:
 - ▶ pyplot
 - ▶ pandas
 - ▶ nltk
 - ▶ scipy
 - ▶ scikit-learn
- ▶ Don't worry about any of this for now.

Additional Teaching Materials

If at some point you need to know more Python for your project, check out some of these:

- ▶ The remainder of our textbook
Automate the Boring Stuff With Python
- ▶ The programming historian
<http://programminghistorian.org/>
- ▶ Python programming for the humanities
<http://www.karsdorp.io/python-course/>
- ▶ Natural language processing with Python
<http://www.nltk.org/book/>
- ▶ Mark Summerfield's book *Programming in Python 3*
- ▶ For more theory, check out the book *Language and Computers*

Don't do it Alone!

- ▶ Form learning/reading groups with your classmates.
- ▶ Check out SBU undergraduate clubs and the events they host.
- ▶ Join online communities.
- ▶ Need advice? Send me an email or come to my office hours!

- ▶ We have launched an M.A. in Computational Linguistics.
- ▶ And we are developing a B.A., too.
- ▶ So expect more UG CompLing courses in the future!
- ▶ **CompLing Info Session:** Wednesday, December 13

Additional Announcements

I will post one or two items to Blackboard this month that might be of interest to you.