

LING/C SC/PSYC 438/538

Lecture 1

Sandiway Fong

Contents

- The Big Picture
- Syllabus
- Questions about the Syllabus
- Homeworks 1 and 2 (*possibly the null homework*)
- Some real intro next time!

Questions for Computation and Language

- We assume language can be modeled as a formal system

Research Question:

- What is the nature of this formal system for human language?
- How is it different from other (mathematical) formal systems?

- What we study here has much to do with formal systems

Examples:

- programming
- regular expressions (regex), finite state automata (FSA)
- context-free (CFG) and context-sensitive grammars
- parse trees, word dependencies, sets

Questions for Computation and Language

- More fundamental questions:
 - Why only us?
 - 50 billion species since life emerged on earth almost 4 billion years ago (Mayr, 1995)
 - Modern humans a recent arrival (2-3TYA) through a remarkable series of accidents
- Language and thought is a **species-specific** property (Chomsky)
 - closest relatives to modern humans have about the same auditory system
 - only a human infant reflexively develops complex systems for constructing and expressing thought
- Language shows peculiar properties
 - (rules) **not arbitrary**
 - **Research question:** why is it the way it is, and not some other way?

Questions for Computation and Language

- Wanna know more?
 - <https://youtu.be/X4F9NSVVuw>
 - https://www.ls-japan.org/modules/documents/LSJpapers/journals/160_chomsky.pdf

[Special Contribution]

Minimalism:

Where Are We Now, and Where Can We Hope to Go*

NOAM CHOMSKY

University of Arizona/MIT

Syllabus

Description of Course

- An introductory level course at the advanced level for computational linguistics. Required core course for the Master's Human Language Technology (HLT) program.

Course Pre-requisites

- 438: LING 388 or familiarity with one or more of the following: formal languages, syntax, data structures, or compilers.
- 538: no formal pre-requisites.

Syllabus

Instructor and Contact Information

- Instructor: Sandiway Fong
- Contact email: sandiway@arizona.edu (all homework to be submitted here)
- Homepage: sandiway.arizona.edu
- Dept. of Linguistics Office: Douglass 311

Hours:

- make appointments by email or try my office

Meet:

- Rm 204, Psychology Building: 3:30-4:45pm Tuesdays/Thursdays
- *another good time to ask questions is right after class!*

Syllabus

Course Format and Teaching Methods

- Lecture with slides. Panopto videos (when available) for lecture review.
- All homeworks will be introduced and **reviewed** in class.

Course Objectives

Topics covered include:

- Introductory programming relevant to computational linguistics in two or more programming languages. We will use Perl, Python and Prolog this semester.
- Introduction to a range of topics in computational linguistics, see detailed list of topics later below.

Syllabus

Course Learning Outcomes

After completing this course, students will:

1. Have acquired the ability to read and write programs in two or more programming languages.
 - Relates to Linguistics Department HLT program outcome #1.
2. Be familiar with basic concepts, techniques and applications in computational linguistics.
 - Relates to Linguistics Department HLT program outcome #2 and Linguistics Department Undergraduate program outcome #1.
3. **538-only:** be able to present and explain advanced concepts in computational linguistics. (See chapter presentation requirement.)
4. Be equipped to take more advanced classes in computational linguistics, e.g. 581 (Spring) or 439/539 (Statistical NLP).

Syllabus

Absence and Class Participation Policy

- I expect you to attend lectures (attendance will not be taken).
- The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at: <http://catalog.arizona.edu/policy/class-attendance-participation-and-administrative-drop>.
- Tell me ahead of time so we can make alternative arrangements in the case of missed homeworks. **No homework will be accepted late. Explained below.**
- Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored. See: <https://deanofstudents.arizona.edu/absences>.
- The UA policy regarding absences for any sincerely held religious belief, observance or practice will be accommodated where reasonable, <http://policy.arizona.edu/human-resources/religious-accommodation-policy>.

Syllabus

Required Text

- 438: None.
- 538 Presentations: *Speech and Language Processing*, Jurafsky & Martin, draft 3rd edition (PDF available).

Required or Special Materials

- All necessary software will be available online at no cost to the student.
- However, students are expected to either have a laptop/desktop capable of handling homework and classwork, or make use of UA lab computers (?)
- Mac, PC (Windows 10) or Linux.

Syllabus

Assignments and Examinations: Schedule/Due Dates

- All homeworks will be introduced **and reviewed** in class.
- Homework submissions by email to me only.
- Late homework will be not accepted since all homeworks will be solved/reviewed in class.
- **Quick homeworks** are normally due at midnight before the next class, and are generally assigned in class on a **Tuesday** and due **Wednesday** midnight (before **Thursday** class).
- Homeworks not categorized as quick are normally assigned in class on a **Thursday** and due the following **Monday** midnight (before next **Tuesday's** class). (Some longer homeworks may have an extended due date.)
- Students can expect a total of 10-14 homeworks over the course.

Syllabus

Final Examination or Project

- No examinations, e.g. mid-term or final, are scheduled for this course.

Grading Scale and Policies

- **438:**
 - 100% of the grade comes from the homework assignments.
- **538:**
 - 75% of the grade comes from the homework assignments (possibly a superset of the 438 assignments), 25% of the grade comes from a textbook chapter presentation.
- Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, which are available at <http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete> and <http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal> respectively.

Syllabus

Scheduled Topics/Activities

- Topics will be drawn from the following:
 - – Programming Languages: Perl and Python
 - – Regular Expressions (Theoretical and practical)
 - – Automata (Finite State) and Transducers (Finite State)
 - – Programming Language: Prolog (definite clause grammars)
 - – NLTK (Natural Language Toolkit)
 - – Part of Speech (POS) Tagging
 - – Stemming (Morphology)
 - – Edit Distance (Spelling)
 - – Grammars (Regular, Context-free)
 - – Parsing (Syntax trees, dependency trees, algorithms)
 - – *and more ...*

Syllabus

Code of Academic Integrity

- You may discuss homework questions with anyone (or anything).
- You may look things up on the web and use answers found therein; however, you must write it up yourself (in your own words/own code *etc.*).
- You must cite all (web) references including ChatGPT, and your classmates (in the case of shared discussion).
- Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials.
- However, graded work/exercises must be the product of independent effort unless otherwise instructed.
- Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: <http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity>.

Syllabus

UA Nondiscrimination and Anti-harassment Policy

- The University is committed to creating and maintaining an environment free of discrimination; see <http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>.

Subject to Change Statement

- Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.



Syllabus

- Questions?

Course website

- Download lecture slides from my homepage
 - <http://sandiway.arizona.edu/ling538-23>
 - available from just before class time
 - (afterwards, please look again for updates and corrections)
 - in .pptx (good for animations) and .pdf formats

Course website

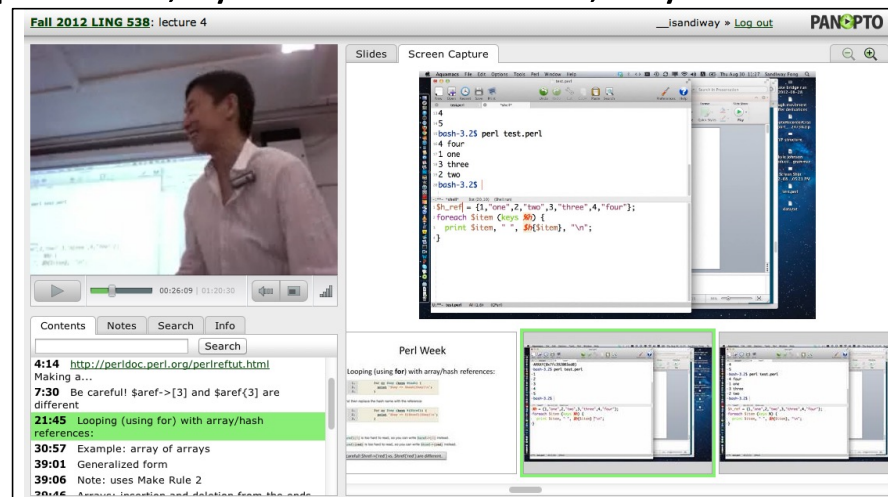
1st hit

Google search results for 'sandway'. The search bar shows 'sandway' and buttons for 'Maps', 'Images', and 'Directions'. Below the search bar, it says 'About 458,000 results (0.38 seconds)'. The first result is for 'Sandway Fong' with the URL 'https://sandway.arizona.edu'. A blue arrow points from the text '1st hit' to this result.

Screenshot of the website 'Sandway Fong'. The navigation menu includes 'About me', 'Research', 'Software', 'Classes', 'UA-related', 'Cycling', 'Running', and 'Trains'. The 'Classes' menu item is circled in red. Below the navigation is a photo of Sandway Fong wearing a face mask, with the caption '(Riding the Tucson streetcar, free during the pandemic.)'. Below the photo is a paragraph of text: 'My work intersects computer science and linguistics. Principally, I'm interested in the computational modeling of linguistic theory. I work at the University of Arizona. I direct and teach in the in-person Master's Program in Human Language Technology'.

Panopto

- Lectures will be recorded using the Panopto system
 - accessible via the course webpage and your browser
 - **sometimes crashes**
 - (video, laptop screen, synchronized slides, keyword search)



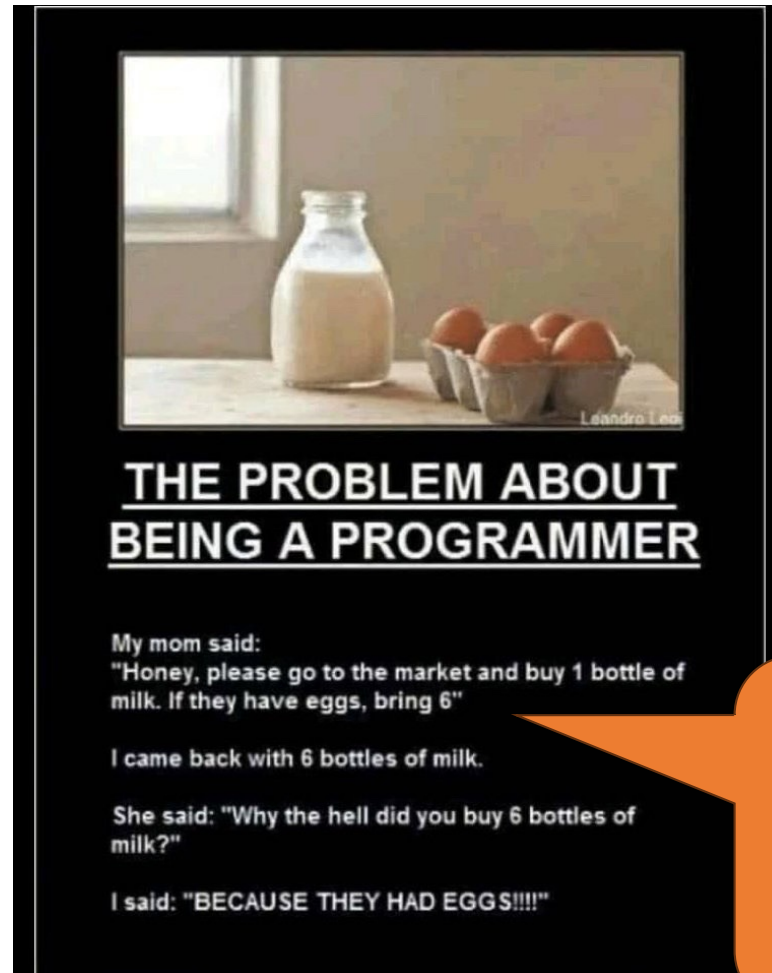


Syllabus

- I'm gonna assume you don't know how to program at all (*yet*)
 - we're going to use **Perl** and **Python**
 - good to learn both ...
 - good to be polyvalent, you'll also get some **Prolog**

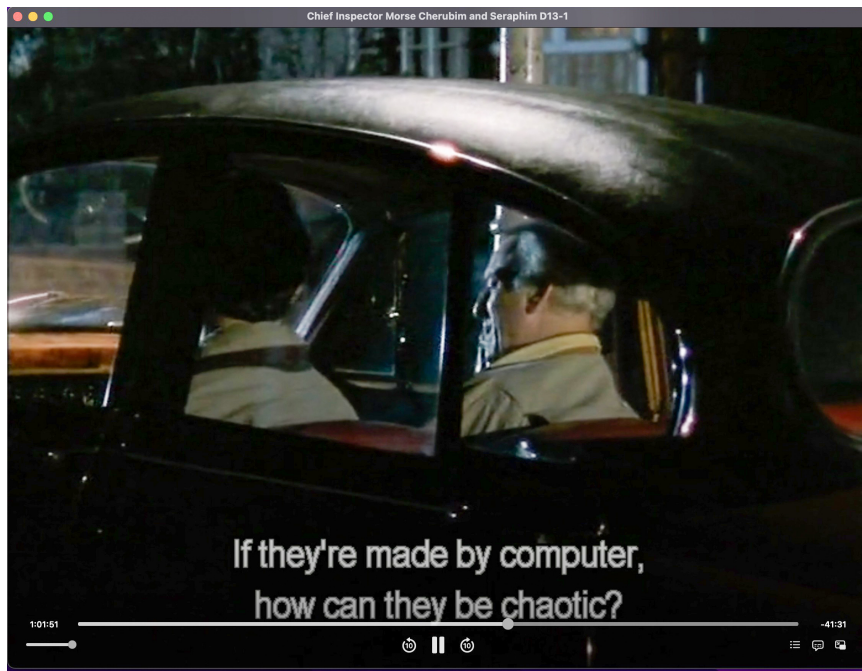
Being Precise

- Always check to see if you typed something in exactly right.
- *saw this on my Facebook feed*
- **Do you understand the joke?**



*if condition:
do
command*

Chief Inspector Morse



Homework: Reading

Homework 1:

- Chapter 1 from JM (2nd ed)
 - **READ IT before next time!**
 - **in-class Quiz on Thursday**
- available on course webpage as:
<https://sandiway.arizona.edu/ling538-23/1.pdf>

Chapter 1 Introduction

*Dave Bowman: Open the pod bay doors, HAL.
HAL: I'm sorry Dave, I'm afraid I can't do that.
Stanley Kubrick and Arthur C. Clarke,
screenplay of 2001: A Space Odyssey*

The idea of giving computers the ability to process human language is as old as the idea of computers themselves. This book is about the implementation and implications of that exciting idea. We introduce a vibrant interdisciplinary field with many names corresponding to its many facets, names like **speech and language processing**, **human language technology**, **natural language processing**, **computational linguistics**, and **speech recognition and synthesis**. The goal of this new field is to get computers to perform useful tasks involving human language, tasks like enabling human-machine communication, improving human-human communication, or simply doing useful processing of text or speech.

Conversational agent

One example of a useful such task is a **conversational agent**. The HAL 9000 computer in Stanley Kubrick's film *2001: A Space Odyssey* is one of the most recognizable characters in 20th century cinema. HAL is an artificial agent capable of such advanced language behavior as speaking and understanding English, and at a crucial moment in the plot, even reading lips. It is now clear that HAL's creator, Arthur C. Clarke, was a little optimistic in predicting when an artificial agent such as HAL would be available. But just how far off was he? What would it take to create at least the language-related parts of HAL? We call programs like HAL that converse with humans in natural language **conversational agents** or **dialogue systems**. In this text we study the various components that make up modern conversational agents, including language input (**automatic speech recognition** and **natural language understanding**) and language output (dialogue and response planning and **speech synthesis**).

Dialogue system

Let's turn to another useful language-related task, that of making available to non-English-speaking readers the vast amount of scientific information on the Web in English. Or translating for English speakers the hundreds of millions of Web pages written in other languages like Chinese. The goal of **machine translation** is to automatically translate a document from one language to another. We introduce the algorithms and mathematical tools needed to understand how modern machine translation works. Machine translation is far from a solved problem; we cover the algorithms currently used in the field, as well as important component tasks.

Machine translation

Many other language processing tasks are also related to the Web. Another such task is **Web-based question answering**. This is a generalization of simple Web search, where instead of just typing keywords, a user might ask complete questions, ranging from easy to hard, like the following:

Question answering

- What does "divergent" mean?
- What year was Abraham Lincoln born?
- How many states were in the United States that year?

Homework 2

Could be the *null* homework for many of you:







- Install Perl, and
- Install Python (version 3.X, **not** 2.7)

Homework 2: Install Perl

- Install Perl on your computer
 - pre-installed on macs and Linux, check your machine from the Terminal/command line
 - on Windows PCs, if you don't already have it, it's freely available here
 - <https://www.perl.org/get.html>

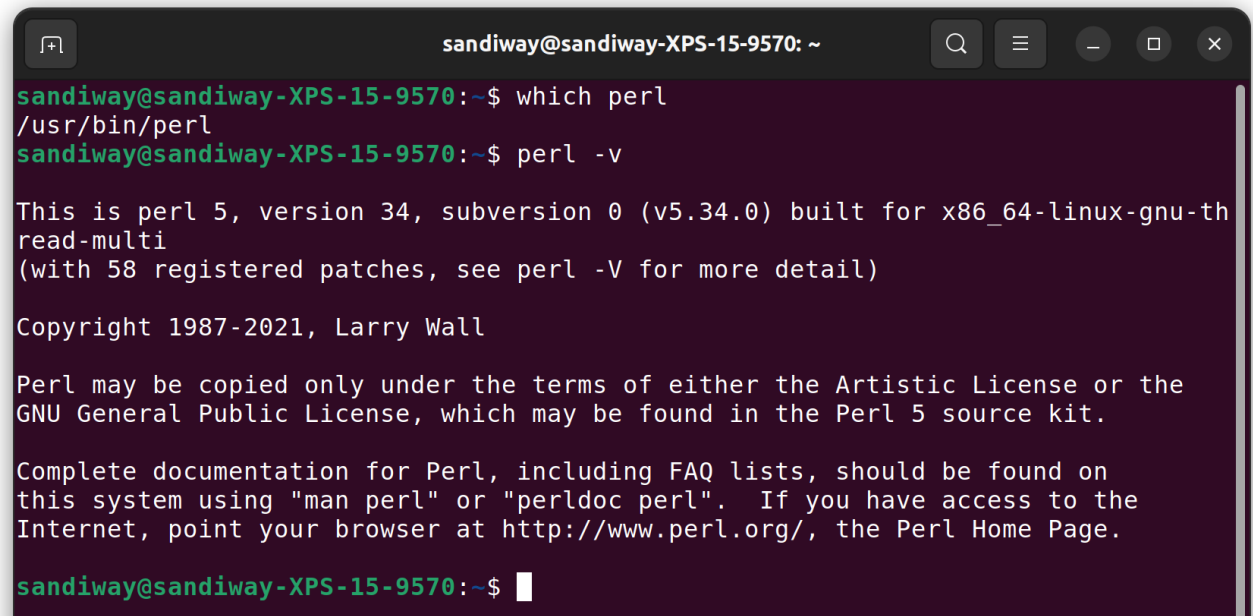
Perl runs on over 100 platforms!

We recommend that you always run the latest stable version, currently 5.38.0. If you're running a version older than 5.8.3, you may find that the latest version of CPAN modules will not work.

Unix/Linux	macOS	Windows
 Included (may not be latest)	 Included (may not be latest)	 Windows Strawberry Perl & ActiveState Perl
 GET STARTED	 GET STARTED	 GET STARTED

Homework 2: Install Perl

- How to check?
 - `which perl`
 - `perl -v`
- Ubuntu (Terminal):



```
sandiway@sandiway-XPS-15-9570: ~
sandiway@sandiway-XPS-15-9570:~$ which perl
/usr/bin/perl
sandiway@sandiway-XPS-15-9570:~$ perl -v

This is perl 5, version 34, subversion 0 (v5.34.0) built for x86_64-linux-gnu-th
read-multi
(with 58 registered patches, see perl -V for more detail)

Copyright 1987-2021, Larry Wall

Perl may be copied only under the terms of either the Artistic License or the
GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on
this system using "man perl" or "perldoc perl".  If you have access to the
Internet, point your browser at http://www.perl.org/, the Perl Home Page.

sandiway@sandiway-XPS-15-9570:~$
```

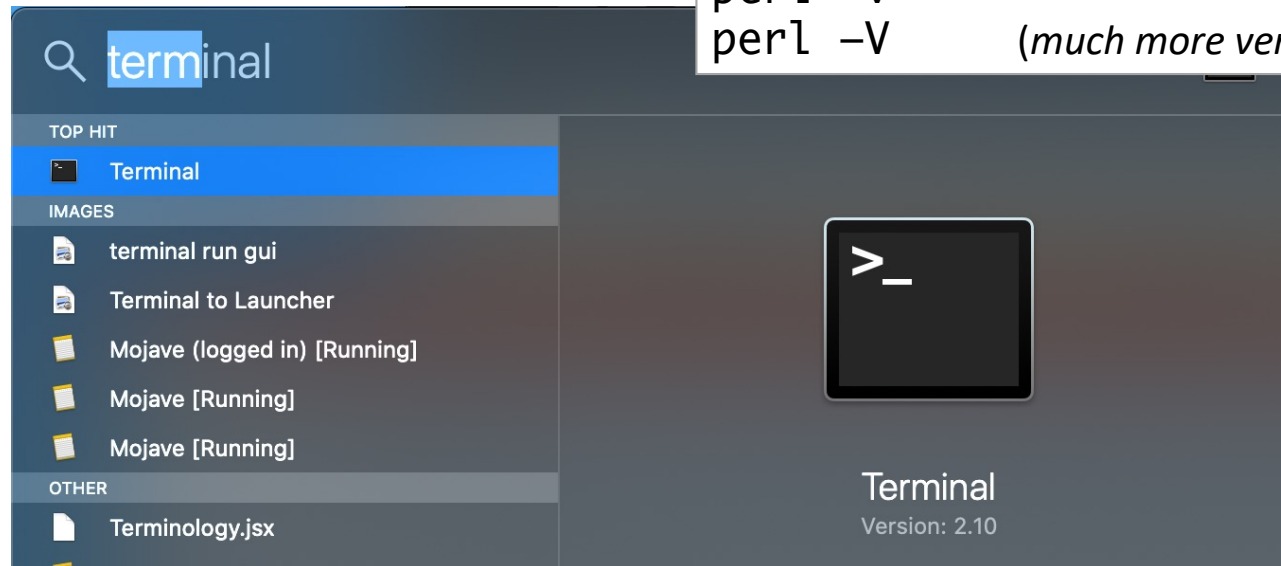
Homework 2: Install Perl

- Mac (Terminal): (*complete path specified here*)
 - `/usr/bin/perl`

commands:

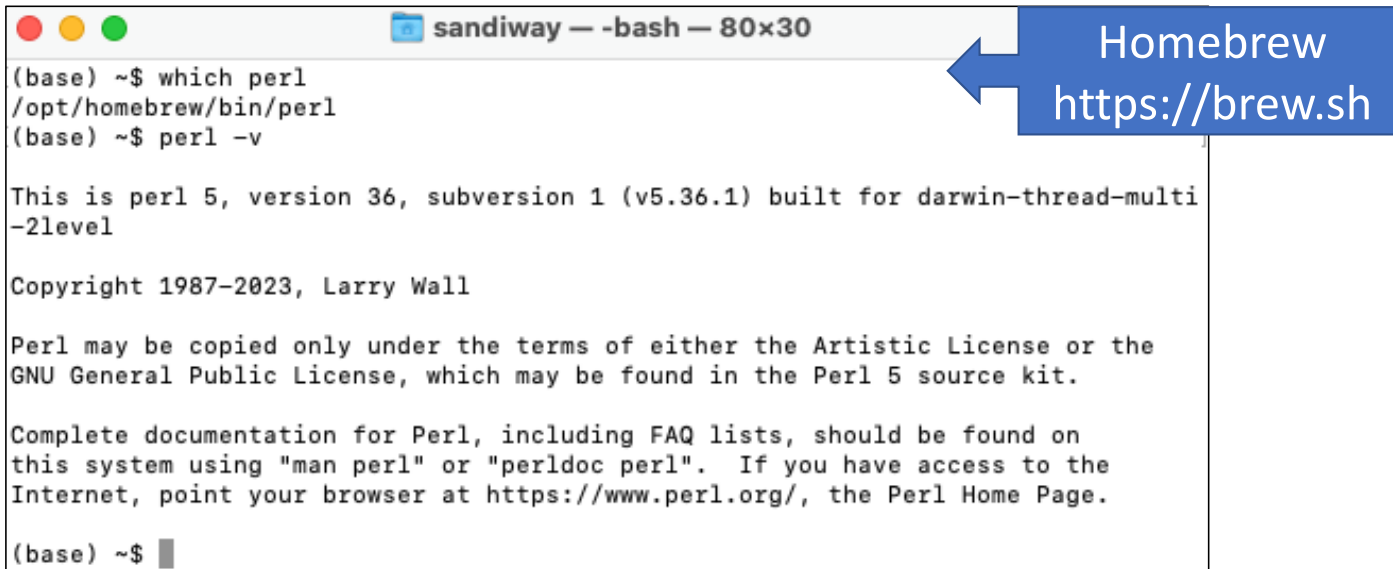
```
perl -v
```

```
perl -V      (much more verbose)
```



Homework 2: Install Perl

- Actually, on my Mac laptop, typing perl at the Terminal gives ...



The image shows a terminal window titled "sandiway — -bash — 80x30". The terminal output is as follows:

```
(base) ~$ which perl
/opt/homebrew/bin/perl
(base) ~$ perl -v

This is perl 5, version 36, subversion 1 (v5.36.1) built for darwin-thread-multi-2level

Copyright 1987-2023, Larry Wall

Perl may be copied only under the terms of either the Artistic License or the GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on this system using "man perl" or "perldoc perl".  If you have access to the Internet, point your browser at https://www.perl.org/, the Perl Home Page.

(base) ~$ █
```

A blue callout box with a white arrow pointing to the terminal output contains the text: Homebrew
<https://brew.sh>

Homework 2: Install Perl

- On my Mac, typing `PATH/usr/bin/perl -v` at the Terminal gives ...

```
sandiway — -bash — 80x30
(base) ~$ /usr/bin/perl -v

This is perl 5, version 30, subversion 3 (v5.30.3) built for darwin-thread-multi-2level
(with 2 registered patches, see perl -V for more detail)

Copyright 1987-2020, Larry Wall

Perl may be copied only under the terms of either the Artistic License or the
GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on
this system using "man perl" or "perldoc perl".  If you have access to the
Internet, point your browser at http://www.perl.org/, the Perl Home Page.

(base) ~$ █
```

means I have more than one version of Perl on my mac!

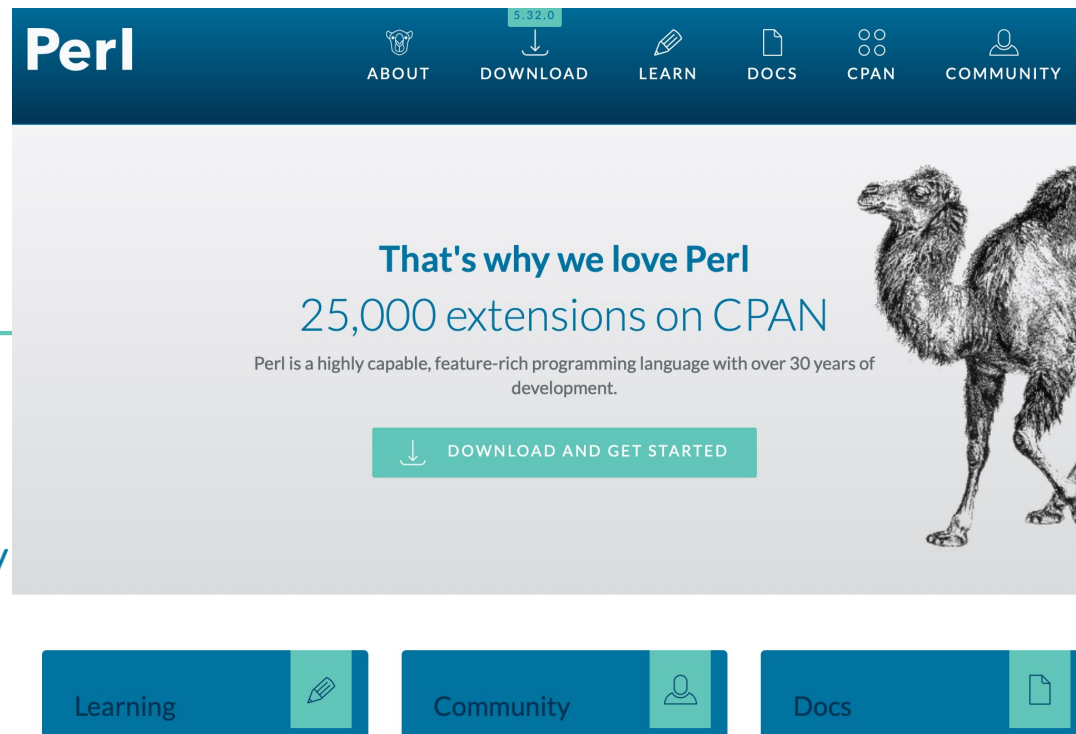
Homework 2: Install Perl

<https://www.perl.org/learn.html>

Get Started

start here
soon

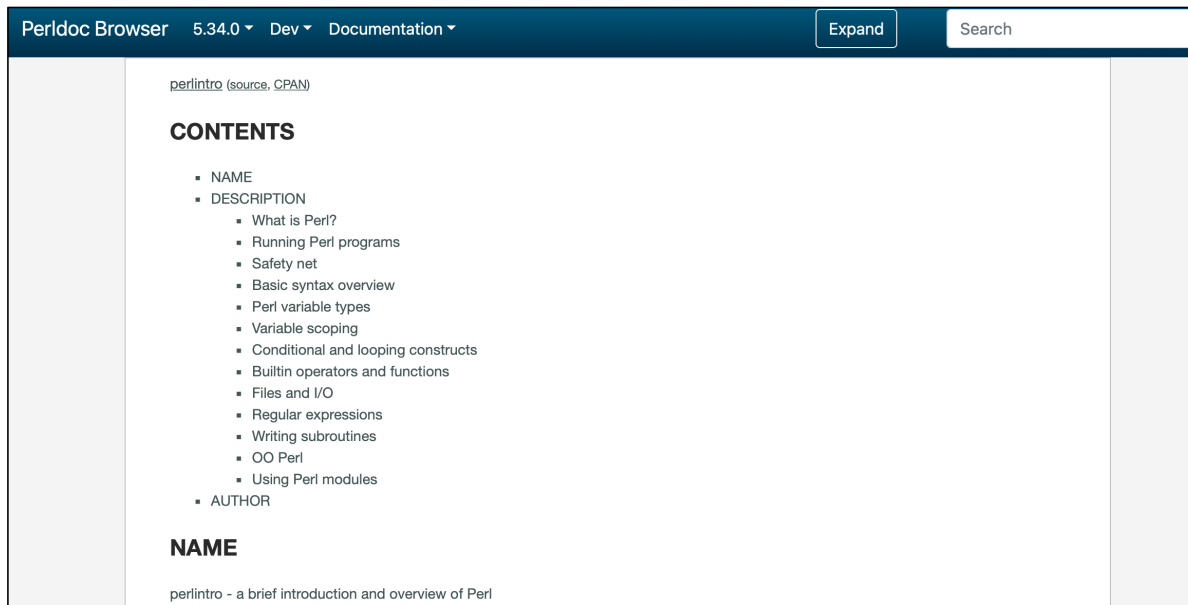
- learn.perl.org
- A brief introduction
- Free online Perl books
- Join your local community
- Books and More



The screenshot shows the Perl website homepage. At the top, there is a dark blue navigation bar with the Perl logo on the left and several menu items: ABOUT, DOWNLOAD (with a version number 5.32.0 above it), LEARN, DOCS, CPAN, and COMMUNITY. Below the navigation bar is a large light gray area with the text "That's why we love Perl" and "25,000 extensions on CPAN". Below this text is a teal button with a download icon and the text "DOWNLOAD AND GET STARTED". To the right of the text is a detailed illustration of a camel. At the bottom of the page, there are three teal buttons: "Learning" with a pencil icon, "Community" with a person icon, and "Docs" with a document icon.

Learning Perl

- Learn Perl
 - <https://perldoc.perl.org/perlintro.html>



The screenshot shows the Perl documentation page for 'perlintro'. The page has a dark blue header with 'Perldoc Browser 5.34.0 Dev Documentation' and a search bar. The main content area is white and contains the following text:

`perlintro` (source, CPAN)

CONTENTS

- NAME
- DESCRIPTION
 - What is Perl?
 - Running Perl programs
 - Safety net
 - Basic syntax overview
 - Perl variable types
 - Variable scoping
 - Conditional and looping constructs
 - Builtin operators and functions
 - Files and I/O
 - Regular expressions
 - Writing subroutines
 - OO Perl
 - Using Perl modules
- AUTHOR

NAME

`perlintro` - a brief introduction and overview of Perl



Homework 2: Install Python

About

Downloads

Documentation

Community

```
# Python 3: Fibonacci sequence
>>> def fib(n):
>>>     a, b = 0, 1
>>>     while a < n:
>>>         print(a, end=' ')
>>>         a, b = b, a+b
>>>     print()
>>> fib(1000)
0 1 1 2 3 5 8 ... 987
```

All releases

Source code

Windows

macOS

Other Platforms

License

Alternative Implementations

Download for macOS

Python 3.11.4

Not the OS you are looking for?

Python is available for many operating systems and architectures.

[View the full list of downloads.](#)

- www.python.org

- **Note:** 3.x is not backwards compatible with Python 2.7!

Python is a programming language that
and integrate systems more effectively

Homework 2: Install Python

- On my Mac laptop, I have the Anaconda version:

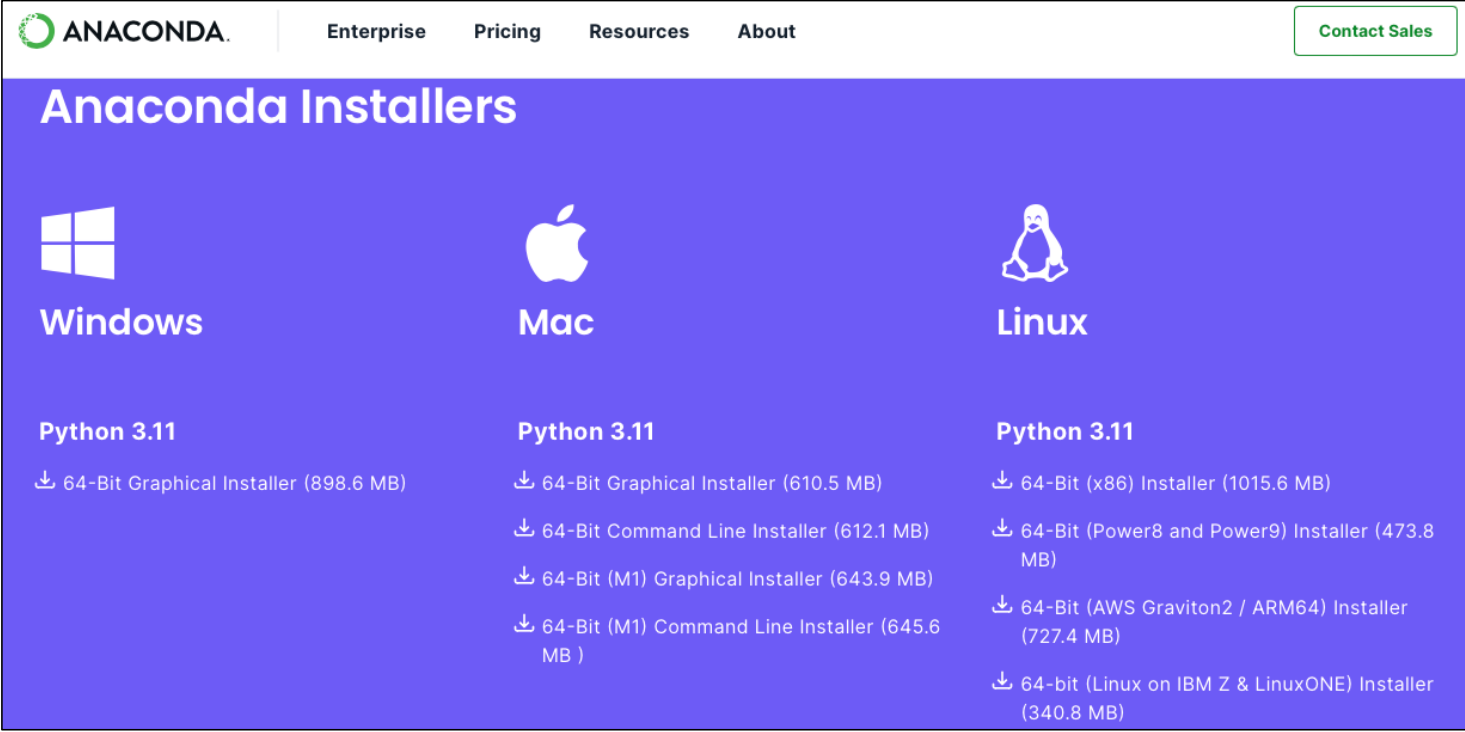
```
sandiway — -bash — 80x30
(base) ~$ which python
/Users/sandiway/opt/anaconda3/bin/python
(base) ~$ python
Python 3.9.16 | packaged by conda-forge | (main, Feb 1 2023, 21:38:11)
[Clang 14.0.6 ] on darwin
Type "help", "copyright", "credits" or "license" for more information.
->
(base) ~$ /usr/bin/python
bash: /usr/bin/python: No such file or directory
(base) ~$
```

*Why?
many pre-
built
packages*

*python 2.7 no
longer included
with macOS*

Homework 2: Install Python

<https://www.anaconda.com/download>



The screenshot shows the Anaconda website's download page. The header includes the Anaconda logo, navigation links for Enterprise, Pricing, Resources, and About, and a Contact Sales button. The main content area is titled "Anaconda Installers" and is divided into three columns for Windows, Mac, and Linux. Each column lists Python 3.11 installers with their respective file sizes.

Operating System	Python Version	Installer Type	File Size
Windows	Python 3.11	64-Bit Graphical Installer	898.6 MB
		64-Bit Command Line Installer	612.1 MB
Mac	Python 3.11	64-Bit Graphical Installer	610.5 MB
		64-Bit (M1) Graphical Installer	643.9 MB
		64-Bit (M1) Command Line Installer	645.6 MB
Linux	Python 3.11	64-Bit (x86) Installer	1015.6 MB
		64-Bit (Power8 and Power9) Installer	473.8 MB
		64-Bit (AWS Graviton2 / ARM64) Installer	727.4 MB
		64-bit (Linux on IBM Z & LinuxONE) Installer	340.8 MB



Which language is easier?

A subjective question ...

- All good programmers know more than one programming language, always an advantage to be versatile.
- In NLP, Python is overwhelmingly popular, but we will do both Perl and Python. AND for writing grammars, Prolog.