# LING/C SC 581: Advanced Computational Linguistics

Lecture 1

# Today's Topics

- 1. Syllabus
- 2. Scheduling
- 3. Software installation: Homework (if not already installed)
- 4. A couple of interesting things ...

### Administrivia

- No lectures scheduled on:
  - Tues Feb 27th: I'm away in Japan
  - Thurs Feb 29th: ditto.
  - Tues March 5th: Spring Break
  - Thurs March 7th: Spring Break
- Pre-recorded lecture for Feb 27th
  - Homework also given out.
- Note:
  - any other cancellation dates will be announced in due course

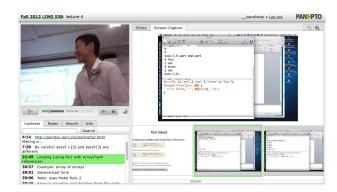
#### Course

#### • Webpage for lecture slides:

- http://sandiway.arizona.edu/#courses
- available from just before class time
  - (afterwards, look again for corrections/updates)
- in .pptx (good for animations) and .pdf formats

#### Meeting information

- Tuesday-Thursday 9:30am-10:45am.
- Modern Languages, Rm 203



#### not guaranteed not to crash!

### Accessibility and Accommodations

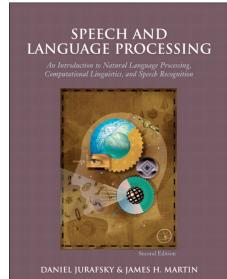
- At the University of Arizona, we strive to make learning experiences as accessible as possible.
- If you anticipate or experience barriers based on disability or pregnancy, please contact the Disability Resource Center (520-621-3268, <u>https://drc.arizona.edu/</u>) to establish reasonable accommodations.

### **Course Objectives**

- Follow-on course to LING/C SC/PSYC 538 Computational Linguistics:
  - pre-requisite: 538
  - continue with some selected material from the 538 textbook (J&M):
    - a lot of material was not covered in 438/538

#### And gain more extensive experience

- with new stuff **not in textbook**
- dealing with natural language software packages
- Installation, input data formatting
- operation
- project exercises
- useful "real-world" computational experience
- abilities gained will be of value to employers



use the v3 PDF draft

### Learning Outcomes

- Learning outcomes: by the end of the semester, you will have learnt:
  - to install and deal with natural language software packages (relates to Linguistics HLT program outcome #3)
  - to properly (re)format input data (relates to Linguistics HLT program outcome #1)
  - to complete projects, providing useful "real-world" computational experience (relates to Linguistics HLT program outcome #1)
- These and other abilities gained will be of value to employers

#### Use your own laptop/desktop

• just in case your need adminstrator rights on your machine to get things working right...

#### • Platforms

- native Windows 10/11 *maybe* possible but unsupported,
- should install Windows Subsystem for Linux: WSL2
  - (return to this topic in a couple of slides)
- MacOS
  - (see next slide)

- Platforms
  - Much research software assume Linux
  - macOS
    - Not quite Linux, some porting issues, especially with C programs.
    - Macports or Homebrew packages (bit like app-get)
    - make sure you have Xcode from the Mac App Store (free) installed
    - then make sure the command line tools component is installed



#### • Xcode on macOS:

(base) ~\$ xcode-select --version xcode-select version 2405. (base) ~\$ xcode-select xcode-select: error: no command option given Usage: xcode-select [options]

Print or change the path to the active developer directory. This directory controls which tools are used for the Xcode command line tools (for example, xcodebuild) as well as the BSD development commands (such as cc and make). Options:

-h,help	print this help message and exit
-p,print-path	print the path of the active developer directory
<pre>-s <path>,switch <path></path></path></pre>	set the path for the active developer directory
install	open a dialog for installation of the command line developer tools
-v,version	print the xcode-select version
-r,reset	reset to the default command line tools path
(base) ~\$ xcode-selectinst	all

xcode-select: note: Command line tools are already installed. Use "Software Update" in System Settings or the softwareupdate command line interface to install updates

#### • Platforms

- Windows 10/11 is possible, but you really should run some variant of Linux...
- PowerShell can be made to work, but is incompatible with Bash shell scripts
- Install WSL2 under Windows 10/11
- Why?
  - gives you a Bash shell (*with quoting consistent with the lecture slides*)
  - (simultaneously) use Linux under Windows 10 (not dual-booting)
  - can access your Windows C: drive via directory /mnt/c

## Grading

- Satisfactory completion of **all** homework tasks will result in a grade A.
- Tasks typically should be completed before the corresponding class next week.
  - email me your work (sandiway@arizona.edu).
  - also be prepared to present your work (if called upon) in class.
- Office hours (by appointment Douglass 311):
  - quick question? Hang around after the lecture.
  - longer question? Email me first. We can also meet on Zoom.

# Syllabus

#### • Homeworks

- you may discuss questions with other students
- you can use ChatGPT
- however, you must write it up yourself (*in your own words, your own code etc.*)
- cite (web) references, ChatGPT and your classmates (in the case of discussion)
- Student Code of Academic Integrity: *plagiarism* etc.
  - <u>http://deanofstudents.arizona.edu/codeofacademicintegrity</u>
- Revisions to the syllabus
  - "the information contained in the course syllabus, other than the grade and absence policies, may be subject to change with reasonable advance notice, as deemed appropriate by the instructor."

### Homework

- make sure for next time that the following software is already installed!
  - already required for 538 (last semester)
  - Python 3
  - SWI Prolog

## Windows Subsystem for Linux 2

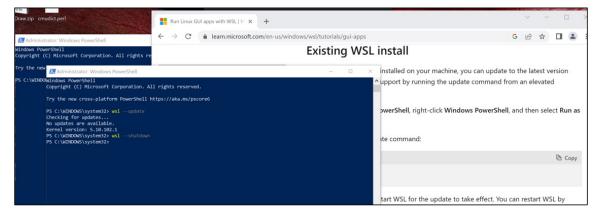


• I installed Ubuntu under WSL2

# Run Linux GUI apps on WSL2

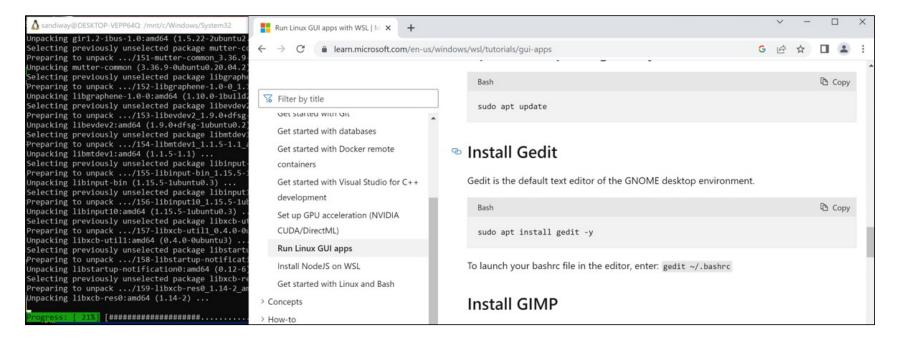
No manual Windows 10 Xserver install and configuring anymore, e.g. the DISPLAY variable!

- Link:
  - https://learn.microsoft.com/en-us/windows/wsl/tutorials/gui-apps
  - 11/28/2022
- Prerequisites
  - used to be Windows 11 only
  - Windows 10 Build 19044+ or Windows 11



### Run Linux GUI apps on WSL2

• Installing gedit sets off a storm of downloading (graphics libraries)



### Run Linux GUI apps on WSL2

• xeyes from x11-apps works: see the eyes pop up on Windows 10!

andiway@DESKTOP-VEPP64Q: /mnt/c/Windows/System32		- 0 ×	V – U
Processing triggers for dbus (1.12.16-2ubuntu2.3) Processing triggers for systemd (245.4-4ubuntu3.1) Processing triggers for sgml-base (1.29.1) sandiway@DESKTOP-VEPP64Q:/mnt/c/Windows/System32\$ Reading package lists Done Building dependency tree Reading state information Dene Reading state information Dene Reading state information Dene Reading state information Dene Reading state information Dene (7.7+8). The following package of the automatically installe	9) sudo apt install x11-apps -y	ie cross-p	G 🖻 🖈 🔲 😩
libfwupdplugin1 J Use 'sudo apt auto 0 upgraded, 0 newl e e emove and 0 n	not upgraded. sudo apt autoremove not upgraded. reed. currently installed.) 0.04.2)  ) 1 is not a symbolic link	xclock, xc	the Copy nd this is a miscellaneous collection of apps and tools calc calculator, xclipboard for cut and paste, xev for ♂ for more info.
sanutway@ueskTur-verru40./mitc/c/winuuws/systemsza	xeyes		🗅 Сору
0	Frequently Asked Questions Troubleshooting > Release Notes	sudo apt install x11-apps -y To launch, enter the name of the tool	you would like to use. For example:
	E Download PDF	<ul> <li>xcalc, xclock, xeyes</li> </ul>	

- Make sure you have Python 3 installed. Also nltk.
  - https://www.nltk.org
  - https://www.nltk.org/install.html
  - Works under macOS, Windows 10/11 and Linux.

- On WSL2:
  - sudo apt install python3 python3-pip python-is-python3
- gets you the non-graphical part.
- Both nltk and nltk\_data install fine.
  - pip3 install --user -U nltk
  - pip3 install --user -U numpy
- But:
  - sudo apt install python3-tk
- gets you the graphics so you can plot graphs from nltk

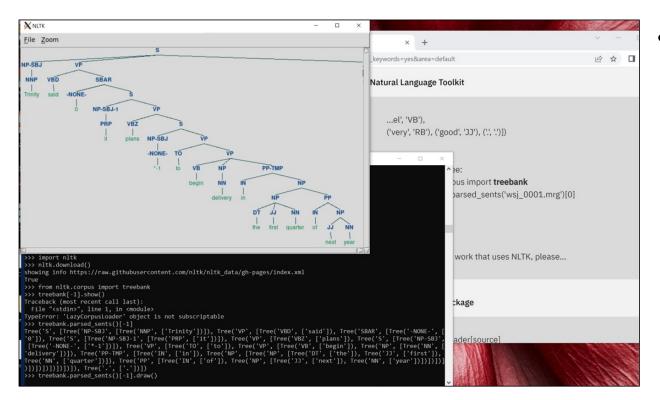
NLTK Downloader			- 🗆 X	Running Python With Graphics S × +
<u>File View Sort H</u> elp				← → C      blog.safnet.com/archive/2022/01/21/python-from-wsl/
Collections Corpora	Models All Packages			
Identifier	All packages	Size	Status 🔺	Enabling a Graphical User Interface
all-corpora all-nltk book popular tests	All the corpora All the corpora All packages available on nltk_data gh-pages branch Everything used in the NLTK Book Popular packages Packages for running tests Third-party data packages	n/a n/a n/a n/a n/a	partial partial partial partial partial installed	Executing a Python-based GUI app from WSL seems a bit odd but let's run with it, shall we? Because it is a requirement. We will need to use the tk toolkit for this class. If I understand correctly, it is included in Python 3.9+. But I have 3.8. Most likely I could find a way to upgrade to 3.9, but I don't have a compelling reason yet, and the following command will install the tk support. sudo apt install python3-tk
			1	Next: how does WSL open a GUI window in Windows 10? You need an X-Windows compatible server for that. There are several proprietary and open source options.
Download			Refresh	
Server Index: http	os://raw.githubusercontent.com/nltk/	nltk_dat	a/gh-pages/j	Chocolatey: choco install vcxsrv. this!
Download Directory: /hom	ne/sandiway/nltk_data			Once installed, you need to run it via the XLaunch command, which will be available
				Windows start menu. This Stack Overflow post has good suggestions for launching correctly. I had to read through the first few posts to get the steps right. The applicance of the steps right is the steps right.

• Inside Python:

>>> import nltk

>>> nltk.download() will now pop up a graphical menu (instead of the command line interface)

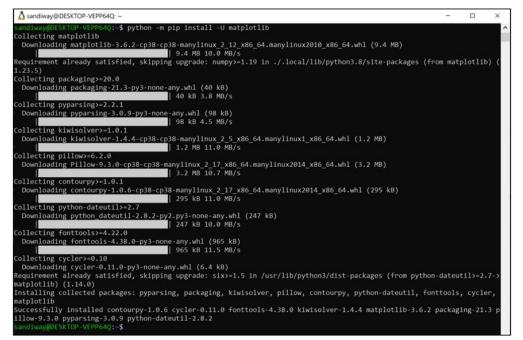
#### nltk treedraw



- The Penn Treebank sample from nltk.corpus
  - here drawing the last syntax tree in the sample

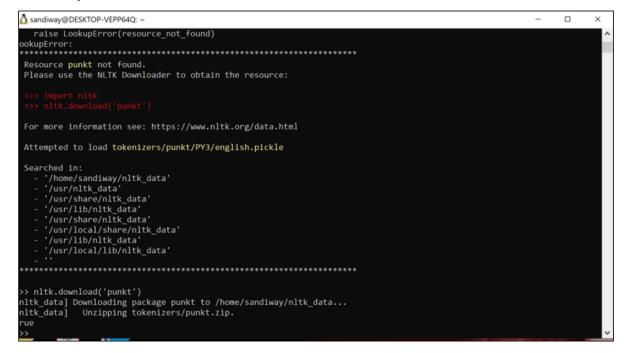
#### matplotlib

- You should install matplotlib:
  - python -m pip install -U matplotlib



#### punkt tokenizer

 Somehow the punkt tokenizer needed for nltk.sent\_tokenize() was not installed by default.



#### • On Ubuntu:

sudo add-apt-repository ppa:deadsnakes/ppa

• This PPA contains more recent Python versions packaged for Ubuntu. sudo apt install python3.9

• gets you the non-graphical part.

```
• Test:
```

```
python3.9
Python 3.9.15 (main, Oct 12 2022, 19:15:11)
[GCC 7.5.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> ^D
```

- On Ubuntu:
  - sudo apt install python3.9-tk
- gets you the Tk graphics.
- Test:

```
python3.9
Python 3.9.15 (main, Oct 12 2022, 19:15:11)
[GCC 7.5.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import tkinter
>>> tkinter._test()  # small window should show up
>>> ^D
```

#### • We also need pip3.9 to install nltk:

sudo apt install python3.9-distutils curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py % Received % Xferd Average Speed % Total Time Time Time Current Dload Upload Total Spent Left Speed 100 2509k 100 2509k 0 3790k 0 --:--:-- 3784k 0 sandiway@sandiway-XPS-15-9570:~\$ python3.9 get-pip.py --user Collecting pip Downloading pip-22.3.1-py3-none-any.whl (2.1 MB) 2.1/2.1 MB 6.1 MB/s eta 0:00:00

Installing collected packages: pip
Successfully installed pip-22.3.1

- Both nltk and nltk\_data should install fine with pip3.9.
  - pip3 install --user -U nltk
- Use:

Downloading nltk-3.7-py3-none-any.whl (1.5 MB) Collecting joblib Downloading joblib-1.2.0-py3-none-any.whl (297 kB) 298.0/298.0 kB 5.9 MB/s eta 0:00:00 Collecting tqdm Using cached tqdm-4.64.1-py2.py3-none-any.whl (78 kB)

Requirement already satisfied: click in /usr/lib/python3/dist-packages (from nltk) (6.7) Collecting regex>=2021.8.3

Downloading regex-2022.10.31-cp39-cp39manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (769 kB)

\_\_\_\_\_\_ 770.0/770.0 kB 7.4 MB/s eta 0:00:00 Installing collected packages: tqdm, regex, joblib, nltk Successfully installed joblib-1.2.0 nltk-3.7 regex-2022.10.31 tqdm-4.64.1

#### • Install numpy with pip3.9:

pip3.9 install --user -U numpy Collecting numpy Downloading numpy-1.23.5-cp39-cp39manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl (17.1 MB) ----- 17.1/17.1 MB 7.3 MB/s eta

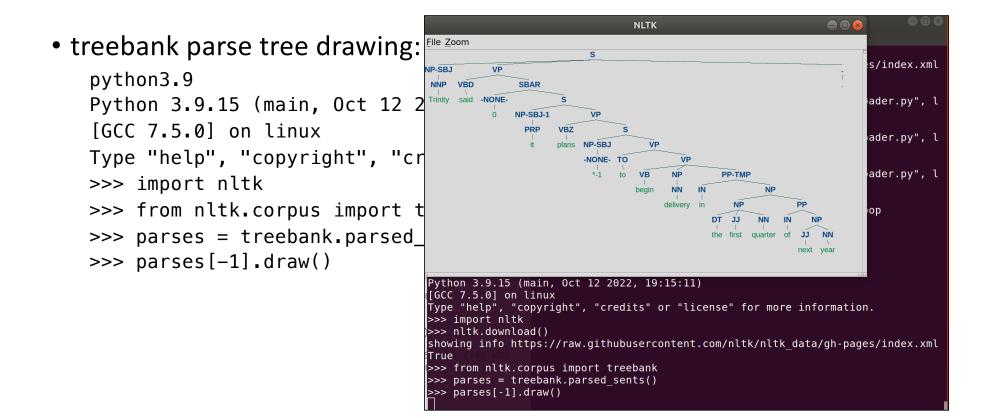
0:00:00 Installing collected packages: numpy Successfully installed numpy-1.23.5

#### • nltk.download() should pop up a window:

python3.9	
Python 3.9.15 (main, Oct 12 2022, 19:15:11)	les
[GCC 7.5.0] on linux	
Type "help", "copyright", "credits" or "license" for	Ē
>>> import nltk	
<pre>&gt;&gt;&gt; nltk.download()</pre>	
<pre>showing info https://raw.githubusercontent.com/nltk/n</pre>	
<pre>^CTraceback (most recent call last):</pre>	
File " <stdin>", line 1, in <module></module></stdin>	
File "/home/sandiway/.local/lib/python3.9/site-pack	
<pre>selfinteractive_download()</pre>	
File "/home/sandiway/.local/lib/python3.9/site-pack	ł
<pre>DownloaderGUI(self).mainloop()</pre>	
File "/home/sandiway/.local/lib/python3.9/site-pack	ł
self.top.mainloop(*args, **kwargs)	
<pre>File "/usr/lib/python3.9/tkinter/initpy", line</pre>	
<pre>self.tk.mainloop(n)</pre>	
KeyboardInterrupt	

	NLTK Downloader		
	HEIR BOWINGGER		UUU
ile <u>V</u> iew <u>S</u> ort <u>H</u> elp			
Collections Corpora	Models All Packages		
Identifier	Name	Size	Status
all	All packages	n/a	installed
all-corpora	All the corpora	n/a	installed
all-nltk	All packages available on nltk_data gh-pages branch	n/a	installed
book	Everything used in the NLTK Book	n/a	installed
popular	Popular packages	n/a	installed
tests	Packages for running tests	n/a	installed
third-party	Third-party data packages	n/a	installed
Download			Refresh
Comier Index: ht:	they / (row githubucarcontent com/pltk/pltk	data/ab n	agaa (inday
Server index: ht	tps://raw.githubusercontent.com/nltk/nltk_	_uata/gn-p	ages/index.

#### nltk treedraw

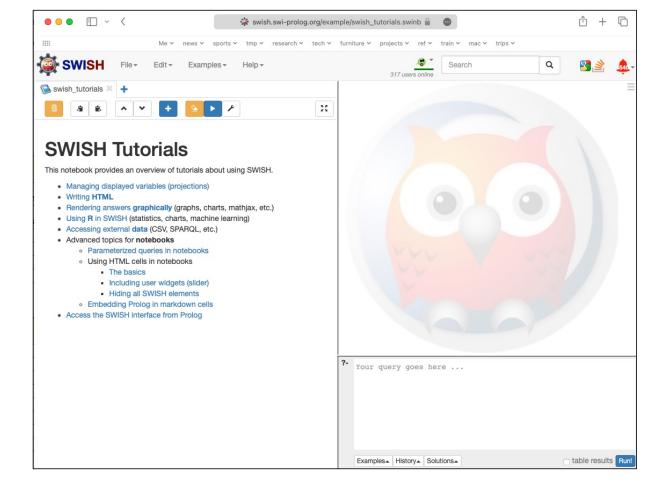


## SWI-Prolog

- <u>https://www.swi-prolog.org</u>
- Used in 538 (pre-requisite class). Make sure it's installed!
- <u>https://swish.swi-prolog.org</u>

Try SWI-Prolog online (SWISH)

# SWISH



A couple of interesting things

# *Boston Globe*: April 10<sup>th</sup> 1953

#### • Excerpt:

- I think the users will be able to see each other, if they want, as they talk.
- "Who knows but what it may actually translate from one language to another?"

#### Predicts Telephone of Future Will Be Carried Like Watch

PASADENA, Calif., April 10 (AP)—The telephone of the future? Mark R. Sullivan, San Francisco, president and director of the Pacific Telephone & Telegraph Company, said in an address last night:

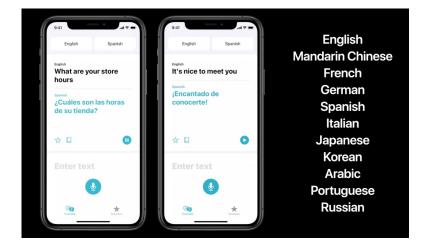
"Just what form the future telephone will take is, of course, pure speculation. Here is my prophecy: "In its final development the telephone will be carried about by the individual, perhaps as we carry a watch today. It probably will require no dial or equivalent, and I think the users will be able to see each other, if they want, as they talk.

"Who knows but what it may actually translate from one language to another?"

# Conversation Mode: Apple's Translate App

- Translate text, voice, and conversations on iPhone
  - Introduced in 2020
  - You can translate text, voice, and conversations between any of the supported languages.
  - You can download languages to translate entirely on the device
    - All translations are processed through the neural engine of the device, and as such can be used offline
  - Tap Conversation.
  - Tap 🔮 and speak

(how about on an Apple Watch?)



### Conversation Mode: Apple's Translate App

Language	<ul> <li>Date added ▼</li> </ul>
Dutch	2022
Indonesian	2022
Polish	2022
Thai	2022
Turkish	2022
Vietnamese	2022
Arabic	2020
Chinese (Mandarin - China mainland and Mandarin - Taiwan)	2020
English (UK and US)	2020
French	2020
German	2020
Italian	2020
Japanese	2020
Korean	2020
Portuguese (Brazil)	2020
Russian	2020
Spanish	2020

 This grew to 17 languages as six new languages were added in 2022, such as Turkish, Indonesian, Polish, Dutch, Thai and Vietnamese.

# Conversation Mode: Apple's Translate App

https://github.com/hollance/neural-engine

### The Neural Engine — what do we know about it?

Most new iPhones and iPads have a **Neural Engine**, a special processor that makes machine learning models really fast, but not much is publicly known about how this processor actually works.

The Apple Neural Engine (or ANE) is a type of **NPU**, which stands for Neural Processing Unit. It's like a GPU, but instead of accelerating graphics an NPU accelerates neural network operations such as convolutions and matrix multiplies.

The ANE isn't the only NPU out there — many companies besides Apple are developing their own AI accelerator chips. Besides the Neural Engine, the most famous NPU is Google's TPU (or Tensor Processing Unit).

### Apple Neural Engine (ANE)

#### M2 Max

The M2 Max uses the same 16-core Neural Engine from the M2, performing up to 15.8 trillion operations per second.

Devices:

- MacBook Pro 14-inch (2023)
- MacBook Pro 16-inch (2023)

#### A16 Bionic

The A16 has a 16-core Neural Engine, the same amount of cores as its predecessor. However, it can perform up to 17 trillion operations per second (7% faster than the previous generation).

Devices:

- iPhone 14 Pro (2022)
- iPhone 14 Pro Max (2022)
- iPhone 15 (2023)
- iPhone 15 Plus (2023)

#### A17 Pro

The A17 Pro has a 16-core Neural Engine, the same amount of cores as its predecessor. However, it can perform up to 35 trillion operations per second (up to 2x faster than the previous generation).

Devices:

- iPhone 15 Pro (2023)
- iPhone 15 Pro Max (2023)

#### https://openai.com/blog/chatgpt/

 Introducing ChatGPT research release
 Try Learn more>

 Signation
 API
 RESEARCH
 BLOG
 ABOUT

 ChatGPT: Optimizing Language Models for Dialogue
 Image: Answer of the second second

#### You

ChatGPT

What is computational linguistics?

#### ChatGPT

Computational linguistics is an interdisciplinary field that focuses on the interaction between computers and human language. It involves the development of algorithms, models, and computational methods to analyze, understand, and generate human language. This field lies at the intersection of computer science, linguistics, and cognitive science, aiming to bridge the gap between natural language and machine understanding.

Key areas within computational linguistics include:

 Natural Language Processing (NLP): NLP is a subfield of computational linguistics that deals with the interaction between computers and human language. It involves tasks such as text and speech processing, machine translation, sentiment analysis, and information retrieval. Key areas within computational linguistics include:

# ChatGPT

- 1. Natural Language Processing (NLP): NLP is a subfield of computational linguistics that deals with the interaction between computers and human language. It involves tasks such as text and speech processing, machine translation, sentiment analysis, and information retrieval.
- 2. **Speech Recognition:** This area focuses on developing algorithms and systems that can recognize and transcribe spoken language. Speech recognition has applications in voice assistants, dictation software, and more.
- 3. **Machine Translation:** Machine translation involves the development of systems that can automatically translate text or speech from one language to another. Google Translate is an example of a machine translation system.
- Information Retrieval: Information retrieval systems help retrieve relevant information from large datasets, such as search engines that provide relevant results based on user queries.

- Text Mining and Sentiment Analysis: Computational linguistics techniques are used to mine valuable information from large volumes of text data, and sentiment analysis helps determine the sentiment expressed in written or spoken language.
- 6. Syntax and Semantics Modeling: Understanding the structure (syntax) and meaning (semantics) of language is crucial for many applications. Computational linguists develop models that enable computers to understand and generate grammatically correct and semantically meaningful language.
- 7. **Dialogue Systems:** Computational linguists work on developing systems that can engage in natural language conversations with users. This includes chatbots, virtual assistants, and other conversational interfaces.
- Corpus Linguistics: Corpus linguistics involves the study of large collections of text (corpora) to extract linguistic patterns and insights. These patterns can inform the development of computational models for language understanding.

Computational linguistics plays a significant role in advancing technologies like artificial intelligence, natural language processing, and human-computer interaction. Researchers and practitioners in this field aim to create systems that can effectively communicate with humans in natural language, facilitating applications in various





what is the difference between NLP and computational linguistics?

In summary, NLP is a subfield of AI and computer science that is concerned with developing applications for natural language understanding and generation. Computational linguistics is a broader interdisciplinary field that includes the application of computational methods to linguistic research, with a focus on both theoretical and practical aspects of language analysis. NLP can be seen as a subset or application-oriented aspect of computational linguistics.

### 0 6 7 5

language understanding, language acquisition, and language evolution. While NLP is one application of computational linguistics, computational linguistics is a broader field that also covers the study of formal languages, such as programming languages.



#### ChatGPT

Natural Language Processing (NLP) and Computational Linguistics are related fields, but they have different focuses and perspectives.

#### 1. Definition:

- NLP (Natural Language Processing): NLP is a subfield of artificial intelligence (AI) and computer science that focuses on the interaction between computers and human language. It involves the development of algorithms and models to enable computers to understand, interpret, and generate human language.
- Computational Linguistics: Computational linguistics is an interdisciplinary field that encompasses the use of computational methods to study and model linguistic phenomena. It involves the application of computer science techniques to address problems in linguistics and language processing.

### 2. Scope:

- NLP: NLP has a broader scope and is primarily concerned with practical applications related to natural language, such as machine translation, sentiment analysis, speech recognition, and chatbots. NLP often emphasizes the development of systems that can perform specific language-related tasks.
- Computational Linguistics: Computational linguistics has a broader academic scope that includes the study of linguistic theories, linguistic models, and the application of computational methods to analyze and understand language. It is more focused on the intersection of linguistics and computer science.

### 3. Goals:

- NLP: The primary goal of NLP is to enable computers to process and understand human language in a way that facilitates communication and interaction. NLP often prioritizes the development of practical applications and technologies.
- Computational Linguistics: The goals of computational linguistics include both theoretical and applied aspects. It seeks to develop computational models that capture linguistic phenomena, improve our understanding of language, and contribute to the development of language technologies.

### 4. Application vs. Research:

- NLP: NLP is often associated with the practical implementation of language technologies for specific applications. It is more application-oriented, with a focus on solving real-world problems.
- Computational Linguistics: While computational linguistics also includes practical applications, it places a stronger emphasis on academic research, linguistic theory, and the development of models that contribute to our understanding of language.



sa Wonder makes us all the University of Aria	zona
---	------

\$

I'm sorry, but I am not able to browse the internet and therefore do not have information about the University of Arizona. I am a large language model trained by OpenAI and my knowledge is based solely on the text that I have been trained on, which includes a wide range of publicly available text, but does not include information specific to the University of Arizona. Can I help you with anything else?

C



#### ChatGPT 3.5 ~

Wonder makes us all the University of Arizona

#### ChatGPT

You

It seems like there might be a slight confusion or a missing context in your statement. The University of Arizona is a public research university located in Tucson, Arizona. It's known for its research and academic programs in various fields.

If you meant to express a sentiment about wonder and curiosity being universal traits that connect us all, that's a beautiful idea. Wonder and curiosity are fundamental aspects of the human experience, and they often drive exploration, learning, and innovation. Embracing a sense of wonder can lead to a lifelong pursuit of knowledge and understanding. If you have a specific question or topic in mind related to the University of Arizona or wonder, feel free to provide more details so I can better assist you!

0070

### U. of Arizona slogan

### Google Search

