

LING/C SC 581:

Advanced Computational Linguistics

Lecture 1

Today's Topics

- Syllabus
- *Scheduling*

Administrivia

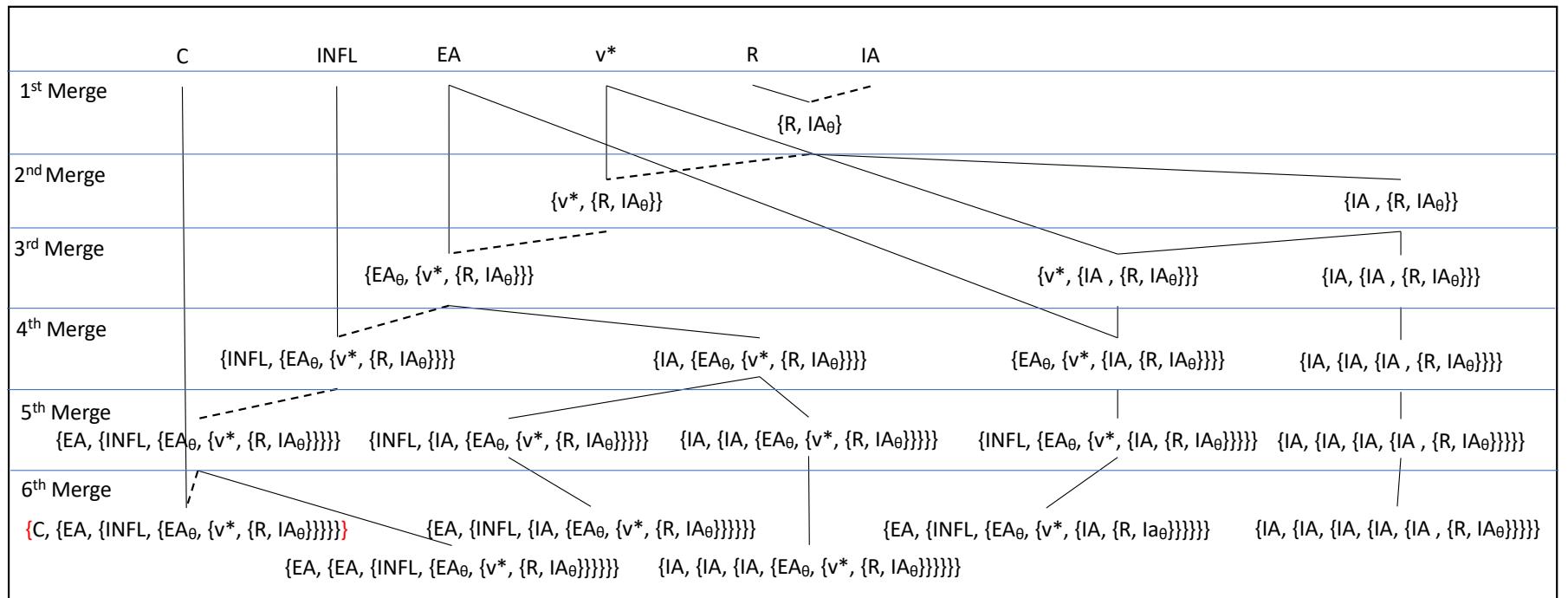
- No lectures on:
 - Tuesday March 8th Spring Break
 - Thursday March 10th Spring Break

Administrivia

- LING 506 guest lecture:
 - Advanced topic: *Computation and the Minimalist Program*
 - Time: 12:30pm-1:45pm, Douglass 216 (*tentative location*)
 - Tuesday April 19th Part 1
 - Thursday April 21st Part 2

Administrivia

- LING 506 guest lecture:
 - Reading: paper (Fong, 2022) to be handed out beforehand + lecture slides



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

Fall 2012 LING 538: lecture 4
Slides Screen Capture
isandiway > Log.out PANOPTO

4:14 http://peridoc.perl.org/perlrefut.html
Making ...
2:01 Be careful! \$aref->[3] and \$aref{3} are different
2:145 Looping (using for) with array/hash references
Example: array of arrays
39:01 Generalized form
39:06 Note: uses Make Rule 2
39:46 ...
perl -e '\$a = [1,2,3,4]; print \$a->[0];' 1
perl -e '\$a = {1,2,3,4}; print \$a->{1};' 2
perl -e '\$a = [1,2,3,4]; print \$a->{1};' 3
perl -e '\$a = {1,2,3,4}; print \$a->{1};' 4
four
1
2
3 three
4 four
hash-1.25
\$aref = [1, 'one', 2, 'two', 3, 'three', 4, 'four'];
foreach \$item (@\$aref) {
 print \$item, "\n";
}
1
one
2
two
3
three
4
four
Perl Week
Looping (using for) with array/hash references
Example: array of arrays
39:01 Generalized form
39:06 Note: uses Make Rule 2
39:46 ...
perl -e '\$a = [1,2,3,4]; print \$a->[0];' 1
perl -e '\$a = {1,2,3,4}; print \$a->{1};' 2
perl -e '\$a = [1,2,3,4]; print \$a->{1};' 3
perl -e '\$a = {1,2,3,4}; print \$a->{1};' 4
four
1
2
3 three
4 four
hash-1.25
\$aref = [1, 'one', 2, 'two', 3, 'three', 4, 'four'];
foreach \$item (@\$aref) {
 print \$item, "\n";
}
1
one
2
two
3
three
4
four

not guaranteed not to crash!

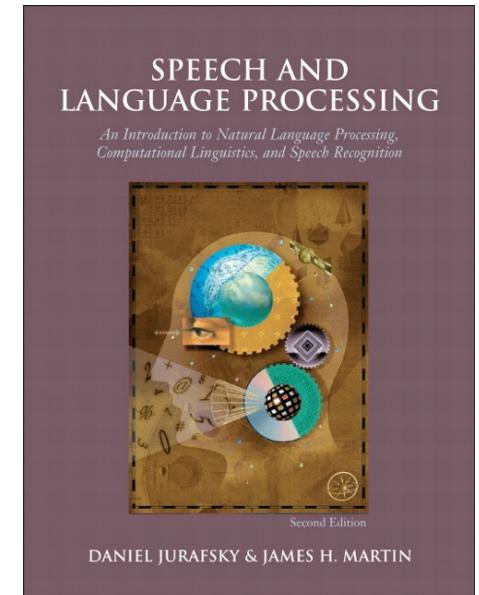
- **Meeting information**
 - Tuesday 3:30pm-4:45pm.
 - Education Building, Rm. 441A
 - *until further notice, COVID-19 measures may change this to online*

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, <https://drc.arizona.edu/>) 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



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

Computational Requirements

- **Use your own laptop/desktop**
 - just in case you need administrator rights on your machine to get things working right...
- **Platforms**
 - Windows 10 *is* possible but you really should run some variant of Linux...
 - **Windows Subsystem for Linux: WSL2**
 - *(return to this topic in a couple of slides)*
 - Linux on VirtualBox (available for free, Windows 10 or MacOS)
 - virtualization software (possible performance issues; Intel CPU must support VT-x)
 - MacOS
 - *(see next slide)*

Computational Requirements

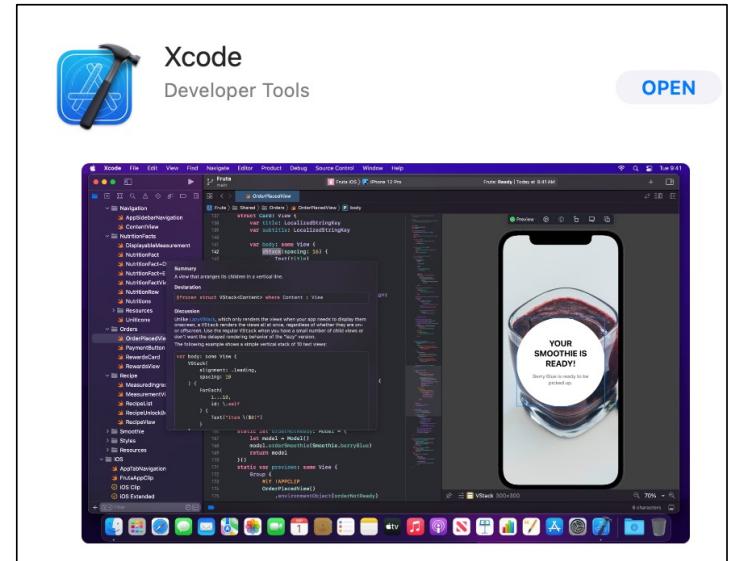
- **Platforms**

- 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-select --version  
xcode-select version 2384.  
~$ xcode-select --install
```

```
xcode-select: error: command line tools are already installed, use "Software Update" to install updates  
~$
```

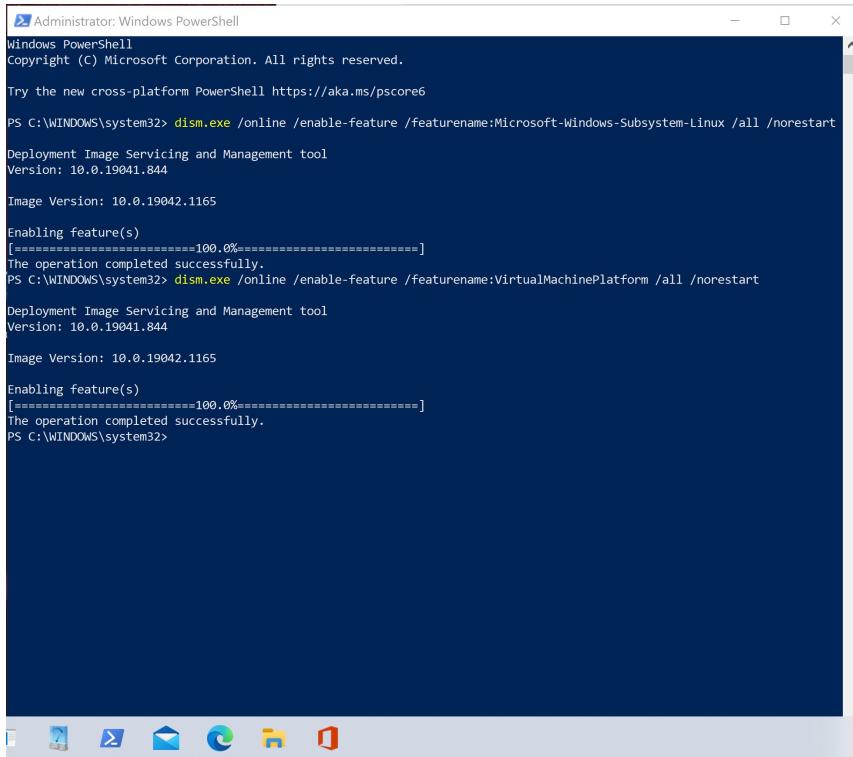


Computational Requirements

- **Platforms**

- Windows 10 *is* possible but you really should run some variant of Linux...
- Consider installing WSL2
 - <https://docs.microsoft.com/en-us/windows/wsl/install-win10>
- 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
- Wanna stick with PowerShell?
 - https://www.powershelladmin.com/wiki/Running_perl_one-liners_and_scripts_from_powershell
 - <https://www.johndcook.com/blog/2020/12/14/shells-quoting-and-one-liners/>
 - perl -e 'print qq{hello\n}'
 - q{} and qq{} can be used for single and doubling quoting.

Windows Subsystem for Linux 2



```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\WINDOWS\system32> dism.exe /online /enable-feature /featurename:Microsoft-Windows-Subsystem-Linux /all /norestart
Deployment Image Servicing and Management tool
Version: 10.0.19041.844

Image Version: 10.0.19042.1165

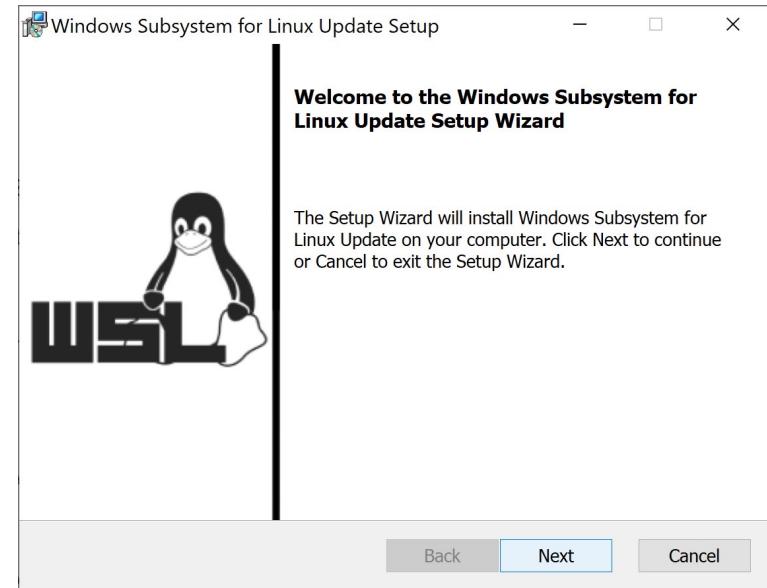
Enabling feature(s)
[=====100.0%=====]
The operation completed successfully.

PS C:\WINDOWS\system32> dism.exe /online /enable-feature /featurename:VirtualMachinePlatform /all /norestart
Deployment Image Servicing and Management tool
Version: 10.0.19041.844

Image Version: 10.0.19042.1165

Enabling feature(s)
[=====100.0%=====]
The operation completed successfully.

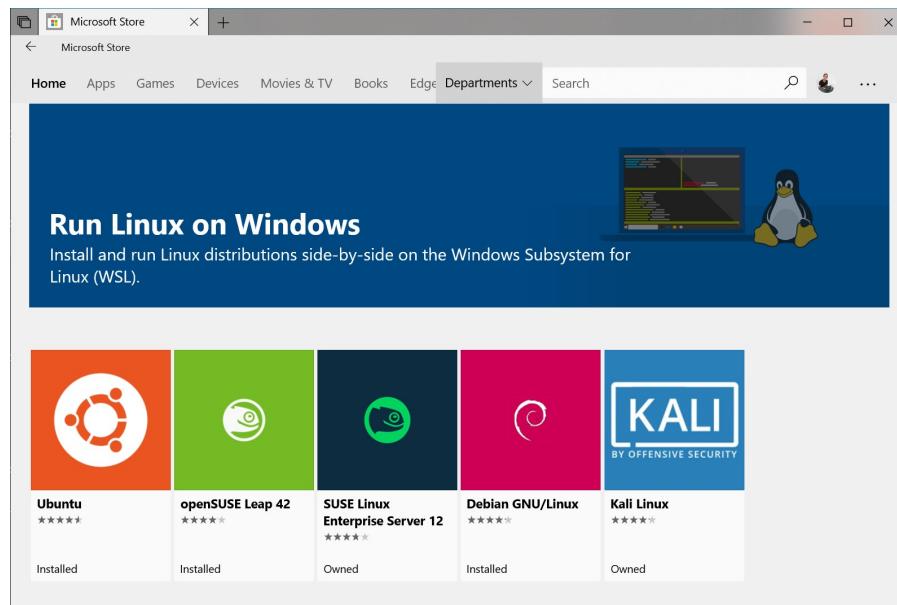
PS C:\WINDOWS\system32>
```



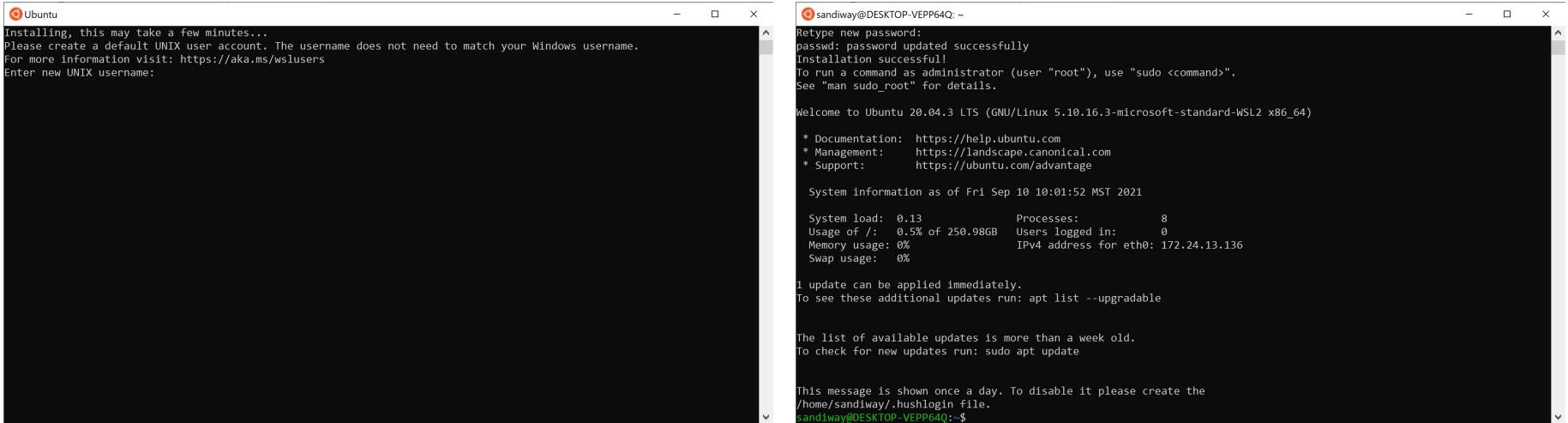
Then reboot your machine back into Windows 10

Windows Subsystem for Linux 2

- I installed Ubuntu



Windows Subsystem for Linux 2



The image shows two side-by-side terminal windows for the Windows Subsystem for Linux (WSL2). Both windows have a dark theme.

Left Terminal:

- Icon: Ubuntu
- Title bar: Ubuntu
- Text:

```
Installing, this may take a few minutes...
Please create a default UNIX user account. The username does not need to match your Windows username.
For more information visit: https://aka.ms/wslusers
Enter new UNIX username:
```

Right Terminal:

- Icon: sandiway@DESKTOP-VEPP64Q
- Title bar: sandiway@DESKTOP-VEPP64Q: ~
- Text:

```
sandiway@DESKTOP-VEPP64Q: ~
Retype new password:
passwd: password updated successfully
Installation successful!
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.10.16.3-microsoft-standard-WSL2 x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Fri Sep 10 10:01:52 MST 2021

System load: 0.13           Processes:          8
Usage of /: 0.5% of 250.98GB  Users logged in:  0
Memory usage: 0%              IPv4 address for eth0: 172.24.13.136
Swap usage: 0%

1 update can be applied immediately.
To see these additional updates run: apt list --upgradable

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

This message is shown once a day. To disable it please create the
/home/sandiway/.hushlogin file.
sandiway@DESKTOP-VEPP64Q:~$
```

Windows Subsystem for Linux 2

```
sudo apt update  
sudo apt upgrade
```

```
sandeway@DESKTOP-VEPP64Q:~  
Unpacking alsa-ucm-conf (1:2.2-1ubuntu0.10) over (1:2.2-1ubuntu0.9) ...  
Preparing to unpack .../08-libssh-4_0.9.3-2ubuntu2.2_amd64.deb ...  
Unpacking libssh-4:amd64 (0.9.3-2ubuntu2.2) over (0.9.3-2ubuntu2.1) ...  
Preparing to unpack .../09-squashfs-tools_1x3a4.4-1ubuntu0.1_amd64.deb ...  
Unpacking squashfs-tools (1:4.4-1) over (1:4.4-1) ...  
Preparing to unpack .../10-open-vm-tools_2%3a11.2.5-2ubuntu1~ubuntu20.04.1_amd64.deb ...  
Unpacking open-vm-tools (2:11.2.5-2ubuntu1~ubuntu20.04.1) over (2:11.0.5-4) ...  
Setting up cpio (2.13+dfsg-2ubuntu0.3) ...  
Setting up squashfs-tools (1:4.4-1ubuntu0.1) ...  
Setting up alsa-ucm-conf (1:2.2-1ubuntu0.10) ...  
Setting up libss11.1:amd64 (1.1.1f-1ubuntu0.8) ...  
Setting up libntfs-3g883 (1:2017.3.23AR.3-3ubuntu1.1) ...  
Setting up ntfs-3g (1:2017.3.23AR.3-3ubuntu1.1) ...  
Setting up libssh-4:amd64 (0.9.3-2ubuntu2.2) ...  
Setting up python-apt-common (2.0.0ubuntu0.20.04.6) ...  
Setting up open-vm-tools (2:11.2.5-2ubuntu1~ubuntu20.04.1) ...  
Installing new version of config file '/etc/vmware-tools/tools.conf.example' ...  
Installing new version of config file '/etc/vmware-tools/vgauth.conf' ...  
invoke-rc.d: could not determine current runlevel  
Setting up openssl (1.1.1f-1ubuntu2.8) ...  
Setting up python3-apt (2.0.0ubuntu0.20.04.6) ...  
Processing triggers for systemd (245.4-4ubuntu3.11) ...  
Processing triggers for man-db (2.9.1-1) ...  
Processing triggers for initramfs-tools (0.136ubuntu6.6) ...  
Processing triggers for libc-bin (2.31-0ubuntu9.2) ...  
sandeway@DESKTOP-VEPP64Q:~$ which perl  
/usr/bin/perl  
sandeway@DESKTOP-VEPP64Q:~$ which cpan  
/usr/bin/cpan  
sandeway@DESKTOP-VEPP64Q:~$ sudo cpan Lingua::EN::CMUDict
```

make is not there!
sudo apt install make

```
sandeway@DESKTOP-VEPP64Q:~  
LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz  
Has already been prepared  
Running make for L/LM/OMETCALF/Lingua-EN-CMUDict-0.07.tar.gz  
Warning: Prerequisite 'File::ShareDir -> 1.03' for 'LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz' failed when processing 'REHSACK/File::ShareDir-1.11.tar.gz' with 'make => NO'. Continuing, but chances to succeed are limited.  
LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz  
make -- NOT OK  
No such file or directory  
sandeway@DESKTOP-VEPP64Q:~$ sudo apt install make  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
Suggested packages:  
  make-doc  
The following NEW packages will be installed:  
  make  
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.  
Need to get 162 kB of archives.  
After this operation, 393 kB of additional disk space will be used.  
Get:1 http://archive.ubuntu.com/ubuntu focal/main amd64 make amd64 4.2.1-1.2 [162 kB]  
Fetched 162 kB in 1s (205 kB/s)  
Selecting previously unselected package make.  
(Reading database ... 32236 files and directories currently installed.)  
Preparing to unpack .../make_4.2.1-1.2_amd64.deb ...  
Unpacking make (4.2.1-1.2) ...  
Setting up make (4.2.1-1.2) ...  
Processing triggers for man-db (2.9.1-1) ...  
sandeway@DESKTOP-VEPP64Q:~$
```

Windows Subsystem for Linux 2

```
sandiway@DESKTOP-VEPP64Q: ~
Has already been prepared
Running make for L/LM/LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
cp lib/Lingua/EN/CMUDict.pm blib/lib/Lingua/EN/CMUDict.pm
Manifying 1 pod document
  LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
  make -- OK
Running make test for LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
PERL_DL_NONLAZY=1 "/usr/bin/perl" "-MExtUtils::Command::MM" "-MTest::Harness" "-e" "undef *Test::Harness::Switches; test_harness(0, 'inc', 'blib/lib', 'blib/arch')" t/*
t/Lingua-EN-CMUDict.t .. ok
All tests successful.
Files=1, Tests=1, 0 wallclock secs ( 0.02 usr  0.00 sys +  0.05 cusr  0.00 csys =  0.07 CPU)
Result: PASS
  LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
  make test -- OK
Running make install for LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
Manifying 1 pod document
Installing /usr/local/share/perl/5.30.0/auto/share/dist/Lingua-EN-CMUDict/cmusyldict.db
Installing /usr/local/share/perl/5.30.0/Lingua/EN/CMUDict.pm
Installing /usr/local/man/man3/Lingua::EN::CMUDict.3pm
Appending installation info to /usr/local/lib/x86_64-linux-gnu/perl/5.30.0/perllocal.pod
  LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
  make install -- OK
sandiway@DESKTOP-VEPP64Q:~$ which perl
/usr/bin/perl
sandiway@DESKTOP-VEPP64Q:~$ which cpan
/usr/bin/cpan
sandiway@DESKTOP-VEPP64Q:~$ perl cmudict.perl apple
apple: 2
sandiway@DESKTOP-VEPP64Q:~$
```

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@email.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
 - however, you must write it up yourself (*in your own words, your own code etc.*)
 - cite (web) references and your classmates (*in the case of discussion*)
 - Student Code of Academic Integrity: *plagiarism* etc.
 - <http://deanofstudents.arizona.edu/codeofacademicintegrity>
- 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.”

Boston Globe: April 10th 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:

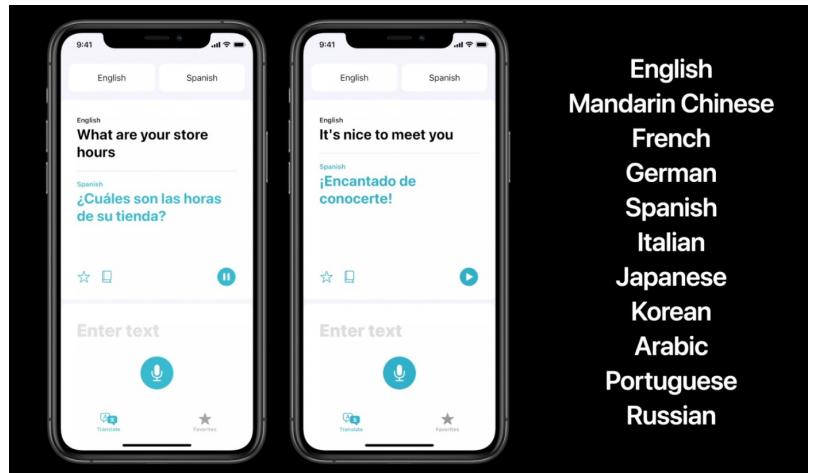
"Just what form the future telephone will take is, of course, pure speculation. Here is my prophecy:

"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?"

Apple's iOS Translate App

- Translate text, voice, and conversations on iPhone
- **is there an iWatch app?**
 - You can translate text, **voice**, and conversations between any of the supported languages.
 - You can **download languages to translate entirely on device**, even without an internet connection.
 - Tap Conversation.
 - Tap  and speak



Python3 and nltk

- Next time, we'll start with Wordnet using nltk and python 3
- Please check and make sure your system is working.
- Test:
 - import nltk
 - from nltk.corpus import wordnet as wn

If you want to get ahead a bit, make sure these modules are installed and import correctly.

- import networkx as nx
- import plotly.graph_objects as go
- import plotly.io as pio
- from networkx.drawing.nx_agraph import graphviz_layout
- import matplotlib.pyplot as plt
- from textwrap import wrap

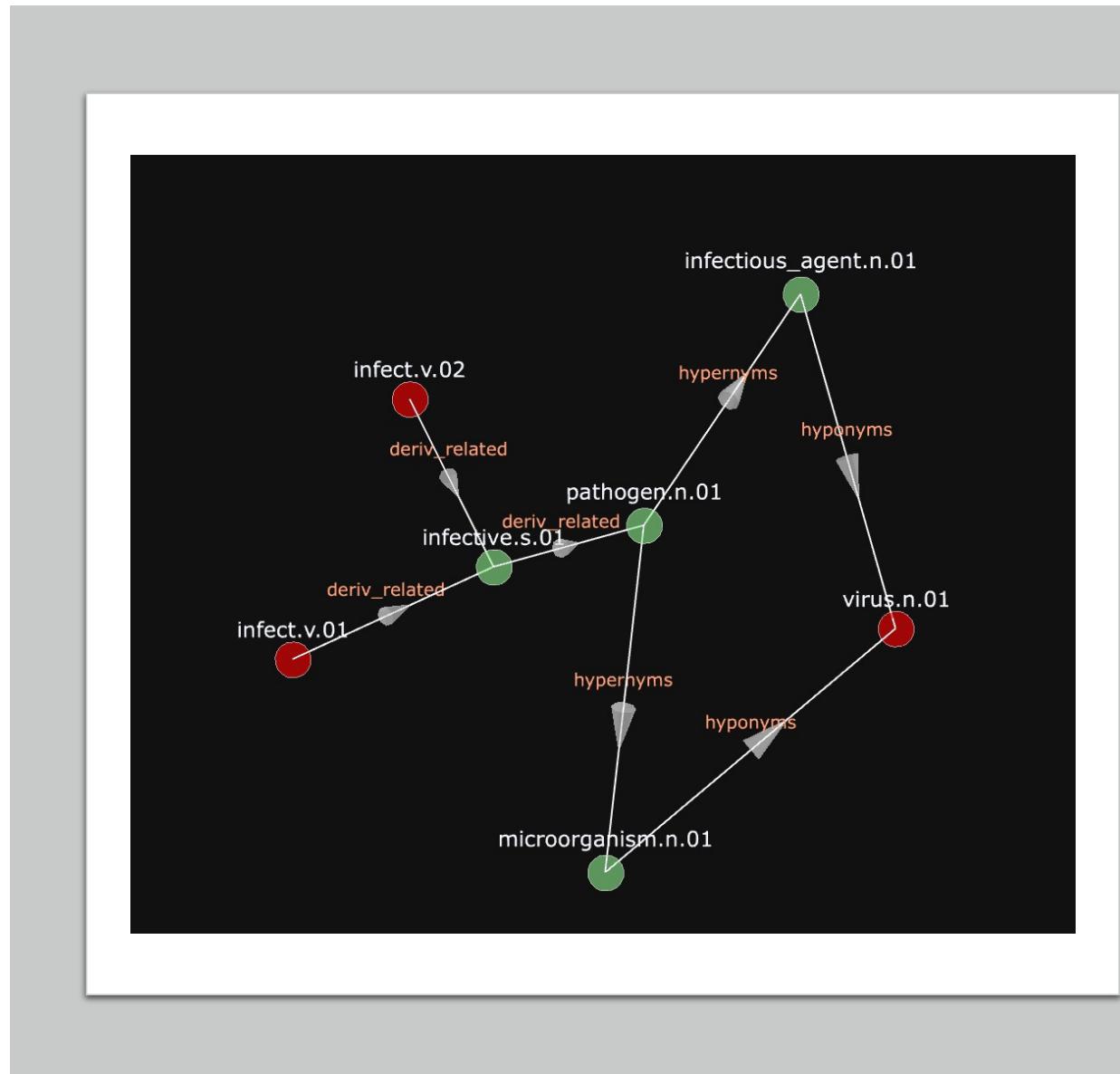
Python3 and nltk

- Example:

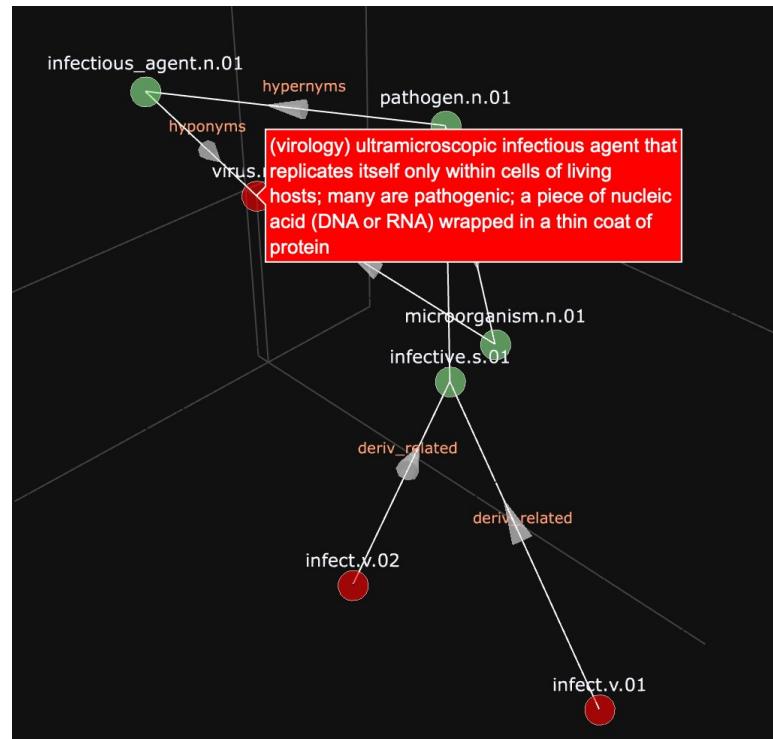
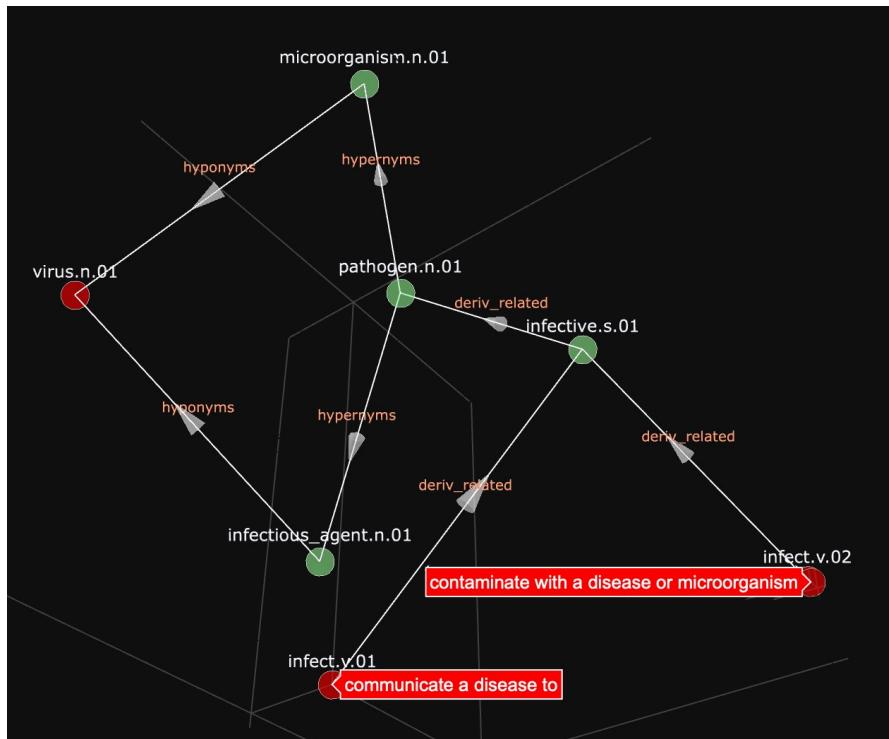
```
~$ python3
Python 3.8.3 (v3.8.3:6f8c8320e9, May 13 2020, 16:29:34)
[Clang 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more
information.
>>> import nltk
>>> from nltk.corpus import wordnet as wn
>>> wn.synset('motorbike.n.1')
Synset('minibike.n.01')
>>> wn.synset('motorbike.n.1').lemmas()
[Lemma('minibike.n.01.minibike'), Lemma('minibike.n.01.motorbike')]
```

Example: a preview

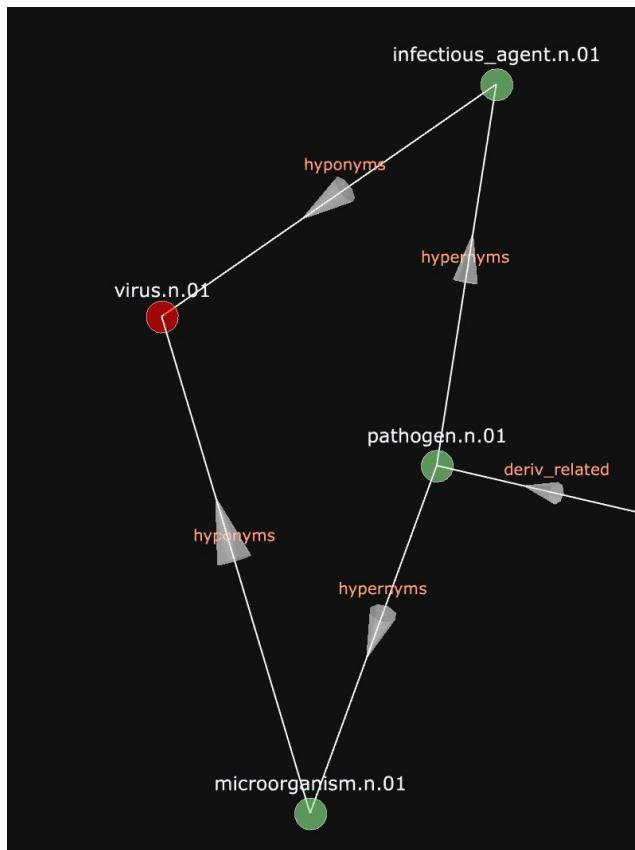
- Click on this 3D graph:
 - <https://sandiway.arizona.edu/ling581-22/wordnet3d-2.html>



Example: *infect* and *virus*



Example: two hypernyms of pathogen



Word to search for: Search WordNet

Display Options: Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Display options for word: word#sense number

Noun

- S: (n) **pathogen#1** (any disease-producing agent (especially a virus or bacterium or other microorganism))
 - *direct hypernym* / *inherited hypernym* / *sister term*
 - S: (n) [microorganism#1](#), [micro-organism#1](#) (any organism of microscopic size)
 - S: (n) [infectious agent#1](#), [infective agent#1](#) (an agent capable of producing infection)
 - *derivationally related form*

1. We can infer a pathogen is an instance of an **organism** (of a particular size). This is one kind of **hypernymy**.
2. We can also infer a pathogen inherits a **telic role of infection**. This is another kind of **hypernymy**.