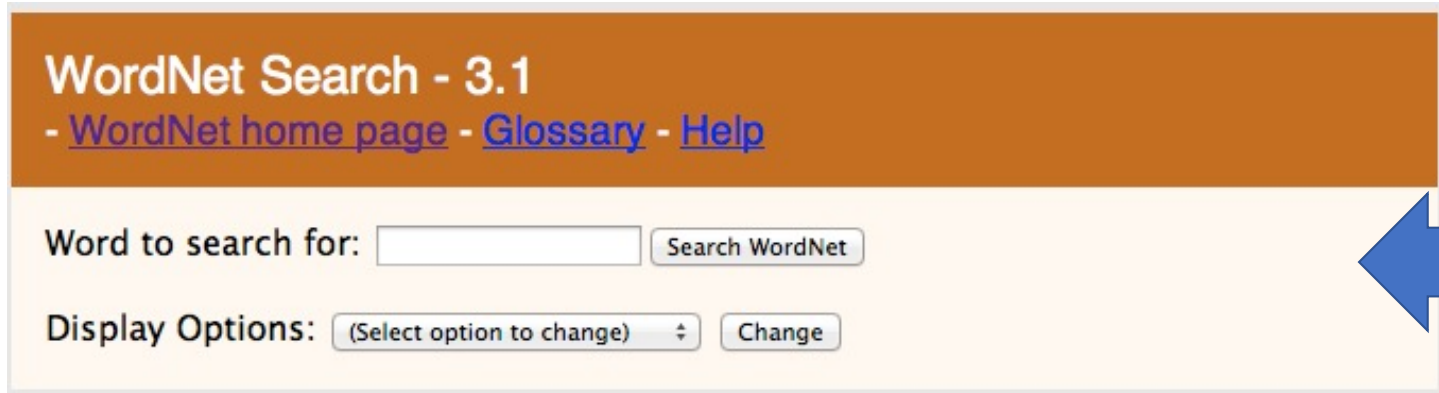# LING/C SC 581:
Advanced Computational Linguistics

Lecture 2

Today's Topic

- WordNet 3.0
  - (3.1 the latest version but only online or the database files only)
  - http://wordnetweb.princeton.edu/perl/webwn
  - **Homework**: try the installations and use the programs in this lecture…

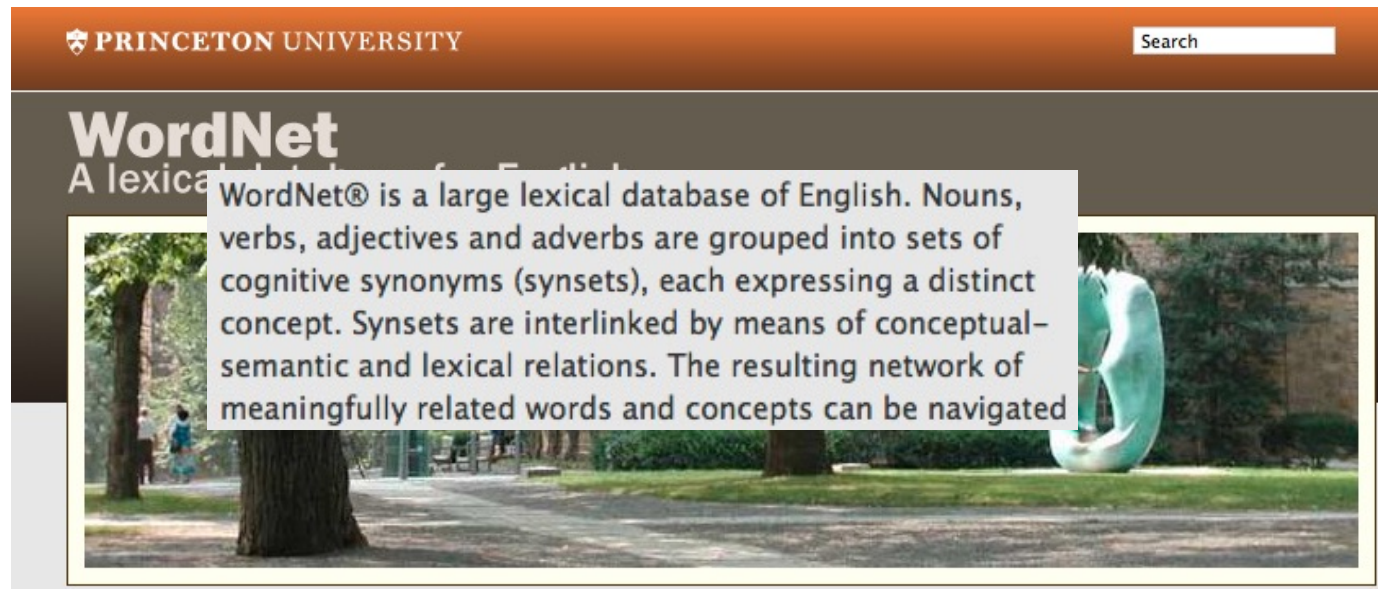online version of the WordNet browser!

Today's Topic

sad state of software today ...

- backwards compatibility not assured,
- people don't test well enough

WordNet 3.0

- https://wordnet.princeton.edu

WordNet 3.1 Demo

- Relations between word senses grouped into synonym sets (**synsets**)

Relations

The most frequently encoded relation among synsets is the super-subordinate relation (also called hyperonymy, hyponymy or ISA relation). It links more general synsets like {furniture, piece_of_furniture} to increasingly specific ones like {bed} and {bunkbed}. Thus, WordNet states that the category furniture includes bed, which in turn includes bunkbed; conversely, concepts like bed and bunkbed make up the category furniture. All noun hierarchies ultimately go up the root node {entity}. Hyponymy relation is transitive: if an armchair is a

WordNet 3.1 Demo

Meronymy, the part-whole relation holds between synsets like {chair} and {back, backrest}, {seat} and {leg}. Parts are inherited from their superordinates: if a chair has legs, then an armchair has legs as well. Parts are not inherited "upward" as they may be characteristic only of specific kinds of things rather than the class as a whole: chairs and kinds of chairs have legs, but not all kinds of furniture have legs.

Verb synsets are arranged into hierarchies as well; verbs towards the bottom of the trees (troponyms) express increasingly specific manners characterizing an event, as in {communicate}-{talk}-{whisper}. The specific manner expressed depends on the semantic field; volume (as in the example above) is just one dimension along which verbs can be elaborated. Others are speed (move-jog-run) or intensity of emotion (like-love-idolize). Verbs describing events that necessarily and unidirectionally entail one another are linked: {buy}-{pay}, {succeed}-{try}, {show}-{see}, etc.

WordNet 3.1 Demo

Adjectives are organized in terms of antonymy. Pairs of "direct" antonyms like wet-dry and young-old reflect the strong semantic contract of their members. Each of these polar adjectives in turn is linked to a number of "semantically similar" ones: dry is linked to parched, arid, dessicated and bone-dry and wet to soggy, waterlogged, etc. Semantically similar adjectives are "indirect antonyms" of the contral member of the opposite pole. Relational adjectives ("pertainyms") point to the nouns they are derived from (criminal-crime).

antonym

NLTK and WordNet

Sample usage for wordnet

WordNet Interface

WordNet is just another NLTK corpus reader, and can be imported like this:

```
>>> from nltk.corpus import wordnet
```

For more compact code, we recommend:

```
>>> from nltk.corpus import wordnet as wn
```

Words

Look up a word using `synsets()`; this function has an optional `pos` argument which lets you constrain the part of speech of the word:

```
>>> wn.synsets('dog')
[Synset('dog.n.01'), Synset('frump.n.01'), Synset('dog.n.03'), Synset('cad.n.01'),
Synset('frank.n.02'), Synset('pawl.n.01'), Synset('andiron.n.01'), Synset('chase.v.01')]
>>> wn.synsets('dog', pos=wn.VERB)
[Synset('chase.v.01')]
```

The other parts of speech are `NOUN`, `ADJ` and `ADV`. A synset is identified with a 3-part name of the form: word.pos.nn:

NLTK and WordNet

Test your nltk:
```
>>> from nltk.corpus import wordnet as wn
>>> wn.synsets('cat')
[Synset('cat.n.01'),
Synset('guy.n.01'),
Synset('cat.n.03'),
Synset('kat.n.01'),
Synset('cat-o'-nine-tails.n.01'),
Synset('caterpillar.n.02'),
Synset('big_cat.n.01'),
Synset('computerized_tomography.n.01'),
Synset('cat.v.01'),
Synset('vomit.v.01')]
```

```
>>> s = wn.synsets('cat')
>>> s[6]
Synset('big_cat.n.01')
>>> s[6].lemma_names()
['big_cat', 'cat']
>>> s[6].lemma_names('fra')
['chat', 'fauve', 'félin']
>>> s[6].hypernyms()
[Synset('feline.n.01')]
>>> s[6].hypernyms()[0].hypernyms()
[Synset('carnivore.n.01')]
>>>
s[6].hypernyms()[0].hypernyms()[0].hypernyms()
[Synset('placental.n.01')]
```

Open Multilingual WordNet

NLTK and WordNet

- Interlingua is English WordNet senses

The WordNet corpus reader gives access to the Open Multilingual WordNet, using ISO-639 language codes.

```
>>> sorted(wn.langs())
['als', 'arb', 'bul', 'cat', 'cmn', 'dan', 'ell', 'eng', 'eus',
'fin', 'fra', 'glg', 'heb', 'hrv', 'ind', 'isl', 'ita', 'ita_iwn',
'jpn', 'lit', 'nld', 'nno', 'nob', 'pol', 'por', 'ron', 'slk',
'slv', 'spa', 'swe', 'tha', 'zsm']
>>> wn.synsets(b'\xe7\x8a\xac'.decode('utf-8'), lang='jpn')
[Synset('dog.n.01'), Synset('spy.n.01')]
```

WordNet 3.0

- Download: https://wordnet.princeton.edu/download/current-version

WordNet 2.1 for Windows

WordNet browser, command-line tool, and database files with InstallShield self-extracting installer:

Download: WordNet-2.1.exe

WordNet 3.0 for UNIX-like systems (including: Linux, Mac OS X, Solaris)

Prolog version of WordNet 3.0

ANSI Prolog version of the WordNet database.

Download: WNprolog-3.0.tar.gz

2.1 is obsolete, don't want that!

WordNet 3.0

- Steps:

  1. Download ...

Name	Date Modified	Size	Kind
WordNet-3.0.tar	1:05 PM	39.5 MB	tar archive

Name	Date Modified	Size	Kind
▶ 📁 WordNet-3.0	8/22/07	--	Folder
WordNet-3.0.tar	1:05 PM	39.5 MB	tar archive

  2. **Please read the README and INSTALL**
     - The WordNet browser makes use of the open source Tcl and Tk packages. Tcl and Tk must be installed BEFORE you compile WordNet.
     - You must also have a C compiler before installing Tcl/Tk or WordNet. WordNet has been built and tested with the GNU gcc compiler. (*Xcode*)
     - If you're running OS X and installed the Aqua Tcl/Tk package from the web site above, use the following settings with configure: `--with-tcl=/Library/Frameworks/Tcl.framework --with-tk=/Library/Frameworks/Tk.framework`
     - TK_LIBRARY - on OS X, may need to be set to the directory that contains the `tk.tcl` file (usually a subidrectory of where the Tk library is installed).

  3. `./configure` (*GNU autoconf*)

# WordNet 3.0 on macOS

use the following settings:
```
--with-tcl=/Library/Frameworks/Tcl.framework
--with-tk=/Library/Frameworks/Tk.framework
```

```
~$ ls /System/Library/Frameworks/T*
/System/Library/Frameworks/TWAIN.framework:
Resources        Versions

/System/Library/Frameworks/Tcl.framework:
Resources        Versions

/System/Library/Frameworks/Tk.framework:
Resources        Versions        tkConfig.sh
~$
```

- ./configure may fail:
  - `checking for Tcl configuration... configure: error: /System/Library/Frameworks/Tcl.framework directory doesn't contain tclConfig.sh`

- Xcode versions:
  - `./configure --with-tcl=/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX.sdk/System/Library/Frameworks/Tcl.framework --with-tk=/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX.sdk/System/Library/Frameworks/Tk.framework`

WordNet 3.0 on macOS

Homebrew (a package manager for macOS)
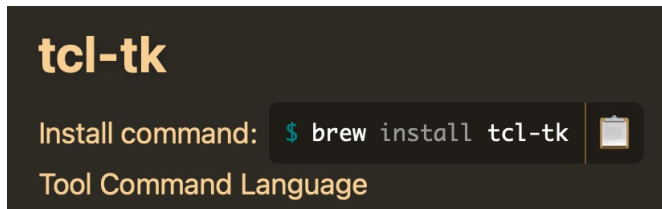

```
WordNet-3.0$ ./configure
checking for gcc... gcc
checking for C compiler default output file name... a.out
checking whether the C compiler works... yes
checking whether we are cross compiling... no
checking for suffix of executables...
checking for suffix of object files... o
checking whether we are using the GNU C compiler... yes
checking whether gcc accepts -g... yes
…
checking for Tcl configuration... found /usr/local/lib/tclConfig.sh
checking for Tk configuration... found /usr/local/lib/tkConfig.sh
checking for existence of /usr/local/lib/tclConfig.sh... loading
checking for existence of /usr/local/lib/tkConfig.sh... loading
configure: creating ./config.status
config.status: creating Makefile
```

WordNet 3.0 on MacOS

```
WordNet is now configured
  Installation directory:                 /usr/local/WordNet-3.0

  To build and install WordNet:

    make
    make install
  To run, environment variables should be set as follows:


    PATH - include ${exec_prefix}/bin
    WNHOME - if not using default installation location, set to /usr/local/WordNet-3.0

    See INSTALL file for details and additional environment variables
    which may need to be set on your system.
```

WordNet 3.0 on macOS

```
config.status: executing default commands

WordNet is now configured

  Installation directory:                    /usr/local/WordNet-3.0

  To build and install WordNet:

    make
    make install

  To run, environment variables should be set as follows:

    PATH - include ${exec_prefix}/bin
    WNHOME - if not using default installation location, set to /usr/local/WordNet-3.0


    See INSTALL file for details and additional environment variables
    which may need to be set on your system.
```

WordNet 3.0 on macOS

- `make` could generate errors, e.g.

```
5 warnings generated.
gcc  -g -O2   -o wn  wn-wn.o -L../lib -lWN
if gcc -DHAVE_CONFIG_H -I. -I. -I.. -I.. -I../include -iwithsysroot /System/Library/Frameworks/Tcl
.framework/Versions/8.5/Headers  -I/usr/include -I.. -I../include -iwithsysroot /System/Library/Fr
ameworks/Tcl.framework/Versions/8.5/Headers  -I/usr/include   -g -O2 -MT wishwn-tkAppInit.o -MD -
MP -MF ".deps/wishwn-tkAppInit.Tpo" -c -o wishwn-tkAppInit.o `test -f 'tkAppInit.c' || echo './'`t
kAppInit.c; \
        then mv -f ".deps/wishwn-tkAppInit.Tpo" ".deps/wishwn-tkAppInit.Po"; else rm -f ".deps/wis
hwn-tkAppInit.Tpo"; exit 1; fi
In file included from tkAppInit.c:16:
/Applications/Xcode.app/Contents/Developer/Platforms/MacOSX.platform/Developer/SDKs/MacOSX.sdk/usr
/include/tk.h:86:11: fatal error:
    'X11/Xlib.h' file not found
#       include <X11/Xlib.h>
                ^~~~~~~~~~~~
1 error generated.
make[2]: *** [wishwn-tkAppInit.o] Error 1
make[1]: *** [all-recursive] Error 1
make: *** [all] Error 2
WordNet-3.0$
```

> **a .h file means a C compiler header file**
> X11 is now under directory `/opt/X11`,
> include files are in `/opt/X11/include`

WordNet 3.0

- On Macs, make sure directory /opt/X11 exists already.

- If not, obviously `CFLAGS=-I/opt/X11/include` won't work

- Install XQuartz from https://www.xquartz.org

XQuartz

Home	The XQuartz project is an open-source effort to develop a version of the X.Org X Window System that runs on OS X. Together with supporting libraries and applications, it forms the X11.app that Apple shipped with OS X versions 10.5 through 10.7.
Releases	

Quick Download

Download	Version	Released	Info
XQuartz-2.7.11.dmg	2.7.11	2016-10-29	For OS X 10.6.3 or later

Support
Contributing
Bug Reporting
GitHub

License Info

WordNet 3.0 on macOS

- So we need to configure again with X11 directory specified before we retry `make`:
  - `WordNet-3.0$` `./configure` `CFLAGS=-I/opt/X11/include`

WordNet 3.0 on macOS

- `make` could also generate this error:

```
          then mv -f ".deps/wishwn-stubs.Tpo" ".deps/wishwn-stubs.Po"; else rm -f
".deps/wishwn-stubs.Tpo"; exit 1; fi
stubs.c:43:17: error: no member named 'result' in 'struct Tcl_Interp'
        interp -> result =
        ~~~~~~      ^
stubs.c:55:14: error: no member named 'result' in 'struct Tcl_Interp'
    interp -> result = bitfieldstr;
    ~~~~~~      ^
```

Fixing it properly

Each of those places really ought to be changed to use `Tcl_SetResult`, i.e. from:

```
interp->result = "usage: glosses [1 | 0]";
```

to

```
Tcl_SetResult(interp, "usage: glosses [1 | 0]", TCL_DYNAMIC);
```

stack overflow

answered May 14 '14 at 18:10

Donal Fellows
124k ●18 ●134 ●202

WordNet 3.0 on macOS

```
diff src/stubs.orig.c src/stubs.c
```

```
183c194,195
<     interp -> result = helptext[pos][searchtype];
---
>     //interp -> result = helptext[pos][searchtype];
>     Tcl_SetResult(interp, helptext[pos][searchtype], TCL_DYNAMIC);
193c205,206
<         interp -> result = "usage: reopendb";
---
>         //interp -> result = "usage: reopendb";
>         Tcl_SetResult(interp, "usage: reopendb", TCL_DYNAMIC);
207c220,221
<         interp -> result = "usage: abortsearch";
---
>         //interp -> result = "usage: abortsearch";
>         Tcl_SetResult(interp, "usage: abortsearch", TCL_DYNAMIC);
src$ 
```

WordNet 3.0 on macOS

- make (*again*)

…

5 warnings generated.

gcc -g -O2   -o wn wn-wn.o -L../lib -lWN

if gcc -DHAVE_CONFIG_H -I. -I. -I.. -I.. -I../include -I/usr/local/include -I/opt/X11/include -I/usr/local/include -I.. -I../include -I/usr/local/include -I/opt/X11/include -I/usr/local/include -g -O2 -MT wishwn-tkAppInit.o -MD -MP -MF ".deps/wishwn-tkAppInit.Tpo" -c -o wishwn-tkAppInit.o `test -f 'tkAppInit.c' || echo './'`tkAppInit.c; \

then mv -f ".deps/wishwn-tkAppInit.Tpo" ".deps/wishwn-tkAppInit.Po"; else rm -f ".deps/wishwn-tkAppInit.Tpo"; exit 1; fi

if gcc -DHAVE_CONFIG_H -I. -I. -I.. -I.. -I../include -I/usr/local/include -I/opt/X11/include -I/usr/local/include -I.. -I../include -I/usr/local/include -I/opt/X11/include -I/usr/local/include -g -O2 -MT wishwn-stubs.o -MD -MP -MF ".deps/wishwn-stubs.Tpo" -c -o wishwn-stubs.o `test -f 'stubs.c' || echo './'`stubs.c; \

then mv -f ".deps/wishwn-stubs.Tpo" ".deps/wishwn-stubs.Po"; else rm -f ".deps/wishwn-stubs.Tpo"; exit 1; fi

gcc -g -O2   -o wishwn wishwn-tkAppInit.o wishwn-stubs.o -L../lib -lWN -L/usr/local/lib -ltk8.6 -L/usr/local/lib -ltcl8.6 -L/opt/X11/lib -lX11 -Wl,-weak-lXss -lXext -L/opt/X11/lib -lXft -lfontconfig -lpthread -lz -lpthread

make[2]: Nothing to be done for `all-am'.

WordNet 3.0 on macOS

- **(ONE WAY) just run it without make install**
  - `cd src`
  - `export PATH=`pwd`:$PATH`
  - `./wnb`

    still works!

    (code was last updated a long, long time ago – 2007!!!)

- **Sigh!** menu fonts are screwed up MacOS Catalina, an annoying Apple bug – sign of incredibly poor QA there, but pulldown menus and main display still work.
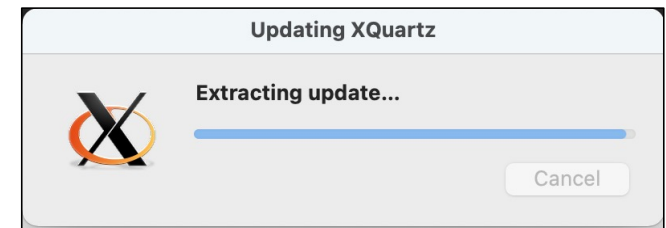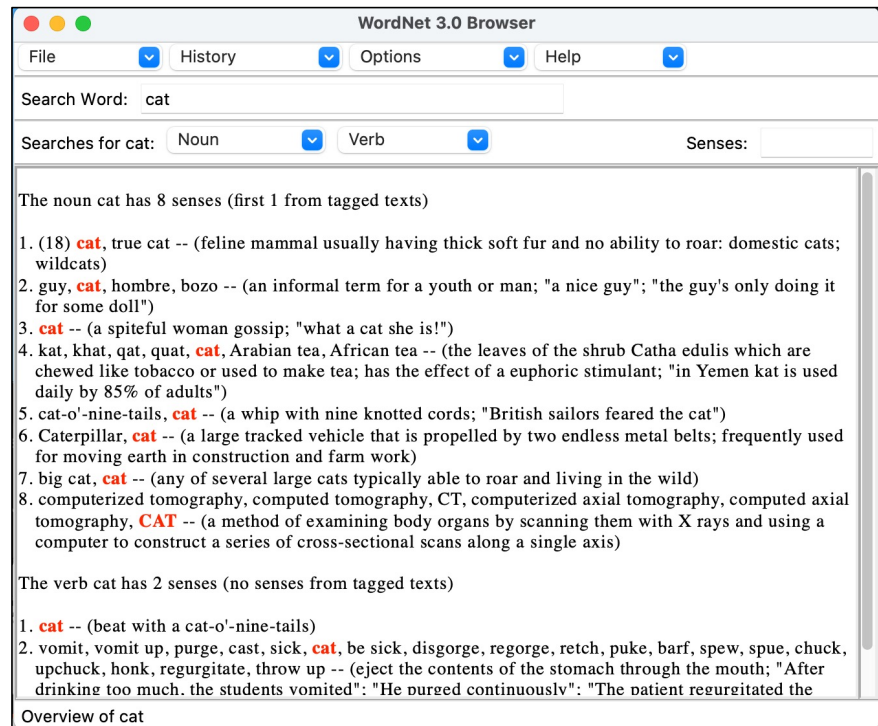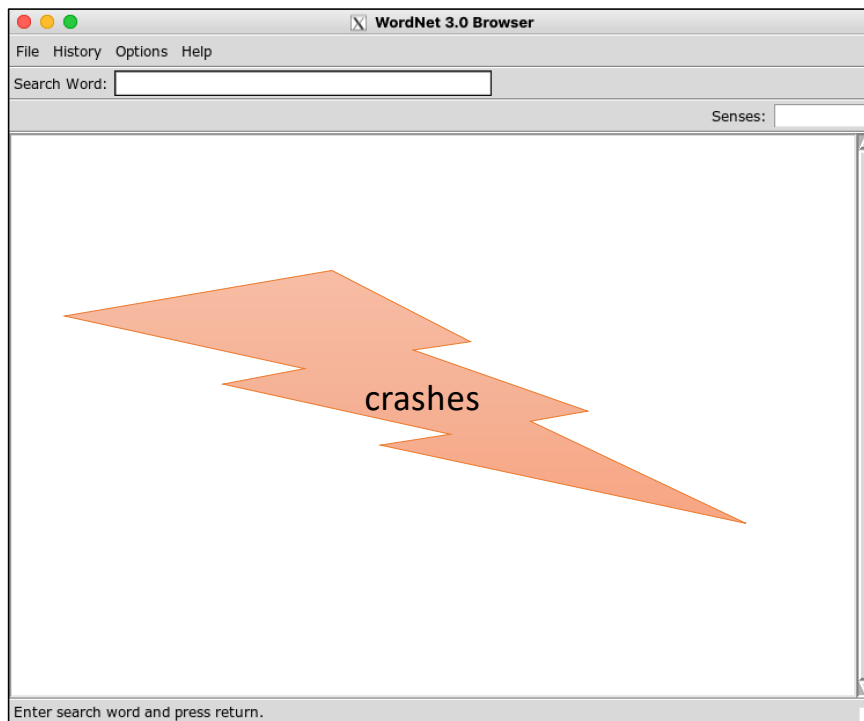


```
[ling581-20$ wish
DEPRECATION WARNING: The system version of Tk is deprecated and may be removed i
n a future release. Please don't rely on it. Set TK_SILENCE_DEPRECATION=1 to sup
press this warning.
% pack [text .t]
2020-04-07 13:13:08.275 Wish[66332:3921184] CoreText note: Client requested name
 ".SFNSMono-Regular", it will get Times-Roman rather than the intended font. All
 system UI font access should be through proper APIs such as CTFontCreateUIFontF
orLanguage() or +[NSFont systemFontOfSize:].
2020-04-07 13:13:08.275 Wish[66332:3921184] CoreText note: Set a breakpoint on C
TFontLogSystemFontNameRequest to debug.
2020-04-07 13:13:08.322 Wish[66332:3921184] CoreText note: Client requested name
 ".SF NS Mono", it will get Times-Roman rather than the intended font. All syste
m UI font access should be through proper APIs such as CTFontCreateUIFontForLang
uage() or +[NSFont systemFontOfSize:].
%
```

wish interpreter
running tcl/tk

WordNet 3.0 on macOS

compiled in 2018

WordNet 3.0 Browser

File History Options Help

Search Word:

Senses:

crashes

Enter search word and press return.

WordNet 3.0 Browser

File   History   Options   Help

Search Word: cat

Searches for cat: Noun   Verb   Senses:

The noun cat has 8 senses (first 1 from tagged texts)

1. (18) cat, true cat -- (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
2. guy, cat, hombre, bozo -- (an informal term for a youth or man; "a nice guy"; "the guy's only doing it for some doll")
3. cat -- (a spiteful woman gossip; "what a cat she is!")
4. kat, khat, qat, quat, cat, Arabian tea, African tea -- (the leaves of the shrub Catha edulis which are chewed like tobacco or used to make tea; has the effect of a euphoric stimulant; "in Yemen kat is used daily by 85% of adults")
5. cat-o'-nine-tails, cat -- (a whip with nine knotted cords; "British sailors feared the cat")
6. Caterpillar, cat -- (a large tracked vehicle that is propelled by two endless metal belts; frequently used for moving earth in construction and farm work)
7. big cat, cat -- (any of several large cats typically able to roar and living in the wild)
8. computerized tomography, computed tomography, CT, computerized axial tomography, computed axial tomography, CAT -- (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

The verb cat has 2 senses (no senses from tagged texts)

1. cat -- (beat with a cat-o'-nine-tails)
2. vomit, vomit up, purge, cast, sick, cat, be sick, disgorge, regorge, retch, puke, barf, spew, spue, chuck, upchuck, honk, regurgitate, throw up -- (eject the contents of the stomach through the mouth; "After drinking too much, the students vomited"; "He purged continuously"; "The patient regurgitated the

Overview of cat

WordNet 3.0

- **(official WAY)** `sudo make install`

  ```
  …
  Making install in src
  test -z "/usr/local/WordNet-3.0/bin" || /Users/sandiway/Downloads/WordNet-
  3.0/install-sh -d "/usr/local/WordNet-3.0/bin"
    /usr/bin/install -c 'wn' '/usr/local/WordNet-3.0/bin/wn'
    /usr/bin/install -c 'wishwn' '/usr/local/WordNet-3.0/bin/wishwn'
  test -z "/usr/local/WordNet-3.0/bin" || /Users/sandiway/Downloads/WordNet-
  3.0/install-sh -d "/usr/local/WordNet-3.0/bin"
   /usr/bin/install -c 'wnb' '/usr/local/WordNet-3.0/bin/wnb'
  ```

- Running the WordNet browser:

  ```
  /usr/local/WordNet-3.0/bin/wnb
  /usr/local/WordNet-3.0/bin/wnb: line 3: wishwn: command not found
  ```

WordNet 3.0

• PATH to wnb (WordNet browser)
export PATH=/usr/local/WordNet–3.0/bin:$PATH
which wnb
• **/usr/local/WordNet–3.0/bin/wnb**

wnb
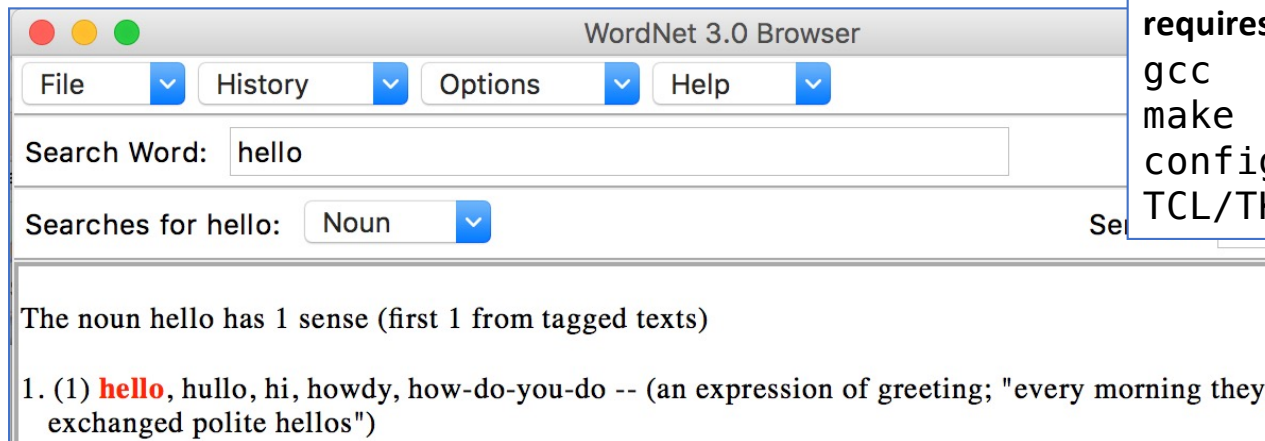
Put PATH modification in your startup shell script, e.g.
`.profile`
`.login`
`.bash_profile`

requires packages:
`gcc`
`make`
`configure` (GNU autoconf)
`TCL/TK`

WordNet 3.0 Browser

File History Options Help

Search Word: hello

Searches for hello: Noun

The noun hello has 1 sense (first 1 from tagged texts)

1. (1) **hello**, hullo, hi, howdy, how-do-you-do -- (an expression of greeting; "every morning they exchanged polite hellos")

WordNet 3.1 vs 3.0

Perl Interface to WordNet 3.0

- http://www.cpan.org/

look for WordNet QueryData

Perl Interface to WordNet 3.0

https://metacpan.org/pod/release/JRENNIE/WordNet-QueryData-1.49/QueryData.pm

Perl Interface to WordNet 3.0

- http://www.cpan.org/modules/INSTALL.html

Perl Interface to WordNet 3.0

- Use the cpanm command or manually (see next slides)

sandiway$ sudo cpanm WordNet::QueryData

--> Working on WordNet::QueryData

Fetching http://www.cpan.org/authors/id/J/JR/JRENNIE/WordNet-QueryData-1.49.tar.gz ... OK

Configuring WordNet-QueryData-1.49 ... OK

Building and testing WordNet-QueryData-1.49 ... OK

Successfully installed WordNet-QueryData-1.49

1 distribution installed

Perl Interface to WordNet 3.0

Perl Interface to WordNet 3.0

`cd WordNet-QueryData-1.49`

`perl Makefile.PL`

- Checking if your kit is complete...
- Looks good
- Writing Makefile for WordNet::QueryData

make

- cp QueryData.pm blib/lib/WordNet/QueryData.pm

- Manifying blib/man3/WordNet::QueryData.3pm

---

make test

```
PERL_DL_NONLAZY=1 /usr/bin/perl "-Iblib/lib" "-Iblib/arch" test.pl
ok 1
ok 2
ok 3
ok 4
ok 5
…
ok 104
ok 105
ok 106
```

---

sudo make install

Installing /Library/Perl/**5.18**/WordNet/QueryData.pm

Installing /usr/local/share/man/man3/WordNet::QueryData.3pm

Appending installation info to /Library/Perl/Updates/5.16.2/darwin-thread-multi-2level/perllocal.pod

Perl Interface to WordNet 3.0

NAME ⬆

WordNet::QueryData - direct perl interface to WordNet database

SYNOPSIS ⬆

```perl
use WordNet::QueryData;

my $wn = WordNet::QueryData->new( noload => 1);

print "Synset: ", join(", ", $wn->querySense("cat#n#7", "syns")), "\n";
print "Hyponyms: ", join(", ", $wn->querySense("cat#n#1", "hypo")), "\n";
print "Parts of Speech: ", join(", ", $wn->querySense("run")), "\n";
print "Senses: ", join(", ", $wn->querySense("run#v")), "\n";
print "Forms: ", join(", ", $wn->validForms("lay down#v")), "\n";
print "Noun count: ", scalar($wn->listAllWords("noun")), "\n";
print "Antonyms: ", join(", ", $wn->queryWord("dark#n#1", "ants")), "\n";
```

Perl Interface to WordNet 3.0

- It requires knowledge of where you installed your WordNet-3.0 directory, can be specified as parameter to new()

```perl
my $wn = WordNet::QueryData->new("/usr/local/wordnet/dict");
```

- A Perl one-liner to test this (word-wrapped in Powerpoint):

```
perl -le 'use WordNet::QueryData; print WordNet::QueryData-
>new("/Users/sandiway/Downloads/WordNet-3.0/dict")-
>querySense("cat#n#1","syns")'
cat#n#1true_cat#n#1
```