

LING/C SC 581:

Advanced Computational Linguistics

Lecture 1

Prof. Sandiway Fong

Course Details

- Course Website:
 - sandiway.arizona.edu > Classes (*Tab*) > 581 (Spring 2026 *link*)
- Lectures:
 - slides / **Panopto** (not guaranteed!)
 - come to class!
- Syllabus:
 - see course website
 - similar to 438/538
 - but homeworks only

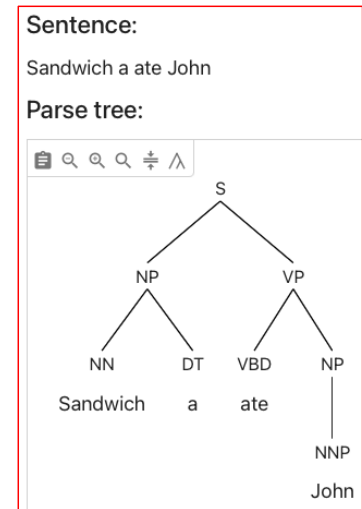
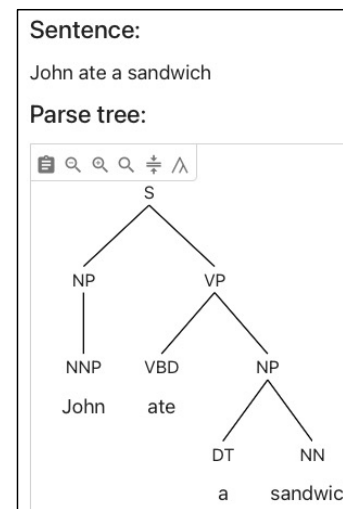
Homework 1

- Answer syntax review questions: Q1, Q2, Q3
 - Due date: Saturday midnight
 - Email to: sandiway@arizona.edu
 - Title: 581 HW1: **YOUR NAME**
- Copy the full **Penn Treebank** (PTB) corpus
 - instructions given out in Panopto and in class (*not on class slides!*)
- install tregex (see windows/Ubuntu/macos notes)
 - <https://nlp.stanford.edu/software/tregex.shtml>
 - we'll be using the software called tregex to search the PTB
 - it's written in Java and requires a Java runtime environment

581 assumes 438/538

HW 1: Q1: what rule of English syntax is violated?

- Knowledge assumed:
 - **Syntax:** ability to understand a parse
 - parser.kitaev.io
 - phase structure tree
 - see Penn tagsets (*this lecture*)
 - dependency-based diagram
 - point out what's wrong!
 - argue coherently for one parse over another!



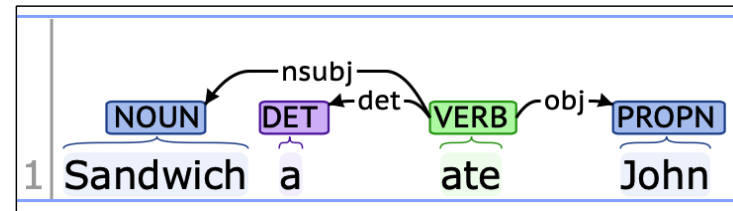
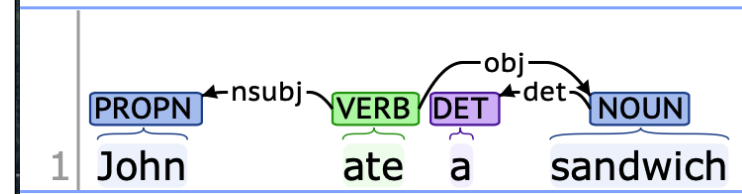
- *John ate a sandwich*
- **Sandwich a ate John*

(pg.113, Chomsky, 1956)

581 assumes 438/538

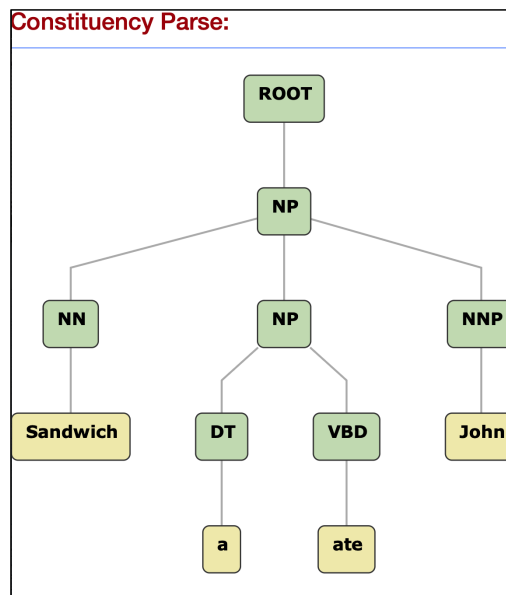
- Universal Dependencies:
 - <https://universaldependencies.org>
- Dependency parse:
 - (stanza.run)
 - <https://stanza.stanford.edu>
 - *John ate a sandwich*
 - **Sandwich a ate John*
(pg.113, Chomsky, 1956)

Universal Dependencies:

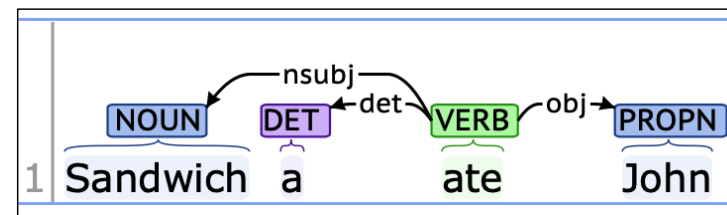


HW 1: Q2: what's wrong here? Also compare with Q1.

Dependency -> Constituent Tree Conversion



Suppose a "**phrase**" = group of more than one word



HW 1: Q3: What's wrong with the **phrase** labels built by the conversion?

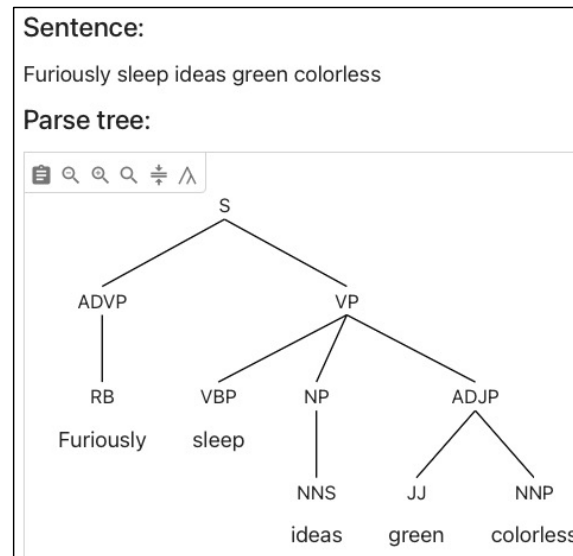
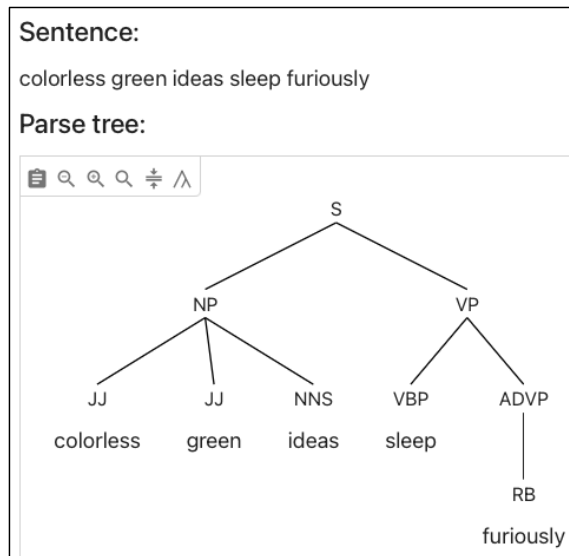


Berkeley Neural Parser

parser.kitaev.io

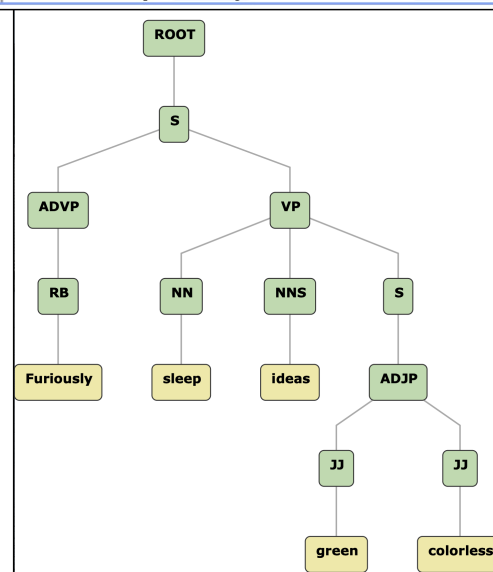
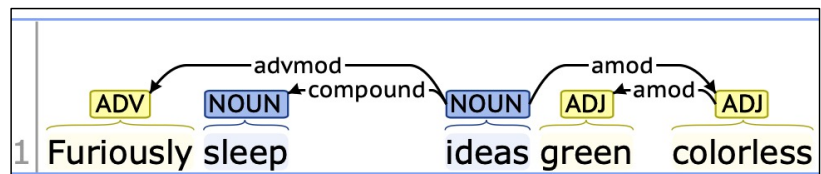
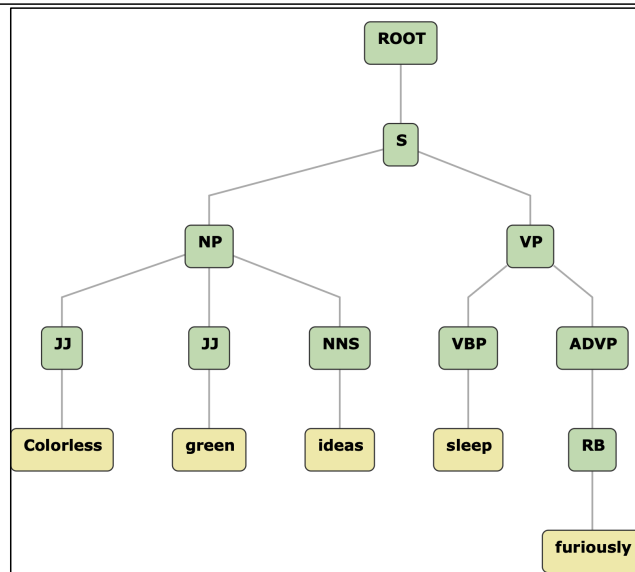
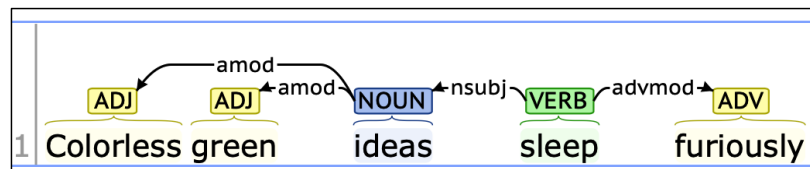
*Colorless green ideas sleep furiously / *Furiously sleep ideas green colorless*

(pg.116, Chomsky, 1956)



Statistical systems
can accept nearly
anything, including
unnatural languages.
*Is that a feature
or a bug?*

Stanza.run



581 assumes 438/538

- Write a phrase structure grammar
 - we will use **Prolog** (DCG rules)
- Examples:
 - `np(np(a,N)) --> [a], nn(N).`
 - `s(s(NP,VP)) --> np(NP), vp(VP).`
- Query (*adds two lists*):
 - `?- np(X, [a, man], []).`
 - `[a, man]` is the list of words passed to `np` rule.
 - `[]` is the list of words remaining after the `np` rule (succeeds).
 - for what values of `X` is query true?

Format:

- LHS `-->` RHS.
- LHS: a **term**
 - term **functor**: nonterminal (`xp`)
 - + term **arguments**
 - `xp(xp(...), Parameters)`
- RHS: , -separated terminals/terms
 - **terminal**: `[a]`
- Note on Prolog case-sensitivity:
 - uppercase: variable
- Note on Prolog syntax:
 - a period ends every statement/query

581 assumes 438/538

- Assume you have `nltk` (nltk.org) and Python working

nlTK treebank

nlTK: corpus treebank

- *a 100,000 word subset of the PTB (not the full thing!)*

```
>>> import nlTK
```

```
>>> from nlTK.corpus import treebank
```

```
>>> t = treebank.parsed_sents()
```

```
>>> t[0]
```

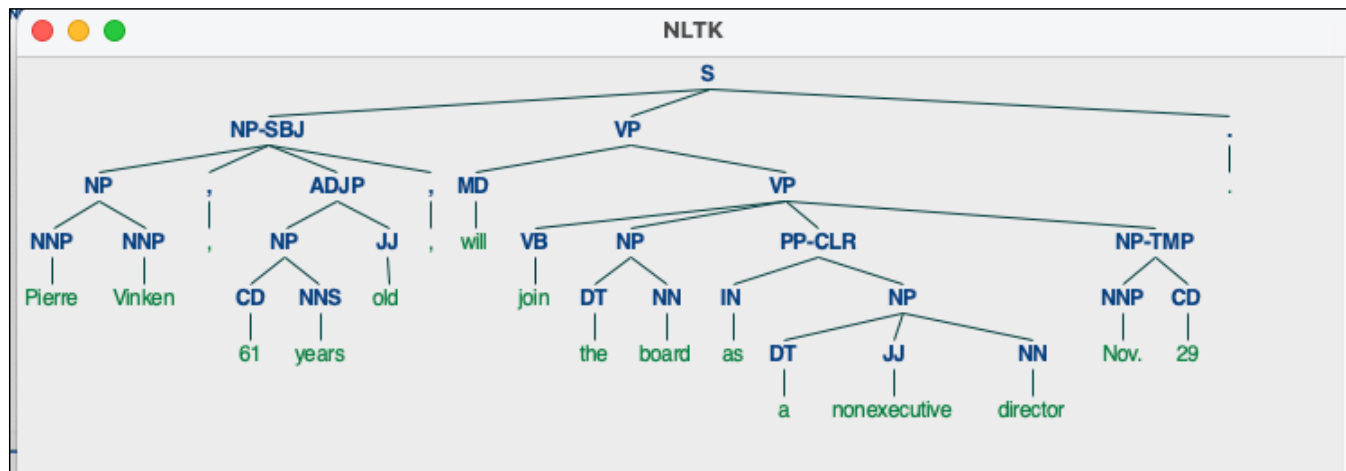
```
Tree('S', [Tree('NP-SBJ', [Tree('NP', [Tree('NNP', ['Pierre']), Tree('NNP',  
['Vinken'])])), Tree(',', ['']), Tree('ADJP', [Tree('NP', [Tree('CD', ['61']),  
Tree('NNS', ['years'])])), Tree('JJ', ['old'])]), Tree(',', ['']), Tree('VP',  
[Tree('MD', ['will']), Tree('VP', [Tree('VB', ['join']), Tree('NP', [Tree('DT',  
['the']), Tree('NN', ['board'])]), Tree('PP-CLR', [Tree('IN', ['as']), Tree('NP',  
[Tree('DT', ['a']), Tree('JJ', ['nonexecutive']), Tree('NN', ['director'])])])),  
Tree('NP-TMP', [Tree('NNP', ['Nov.']), Tree('CD', ['29'])])])])), Tree('.', ['.'])])])
```

```
>>> t[0].draw()
```

```
>>> len(t)
```

```
3914
```

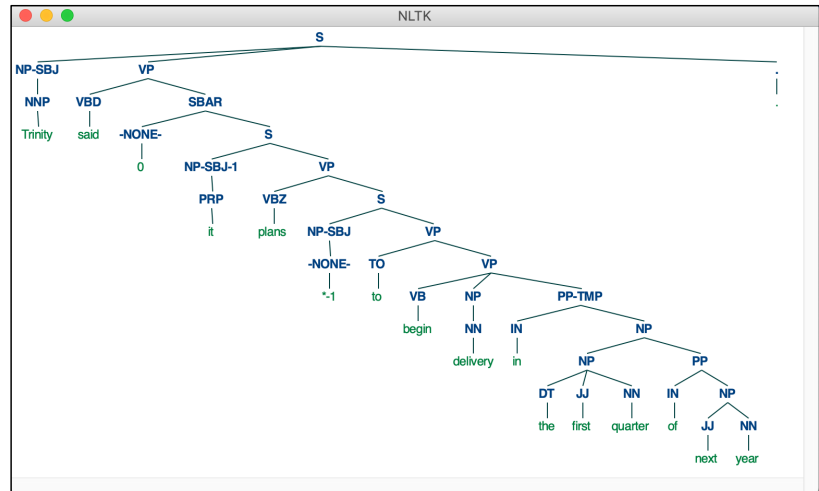
nltk treebank



```
Tree('S', [Tree('NP-SBJ', [Tree('NP', [Tree('NNP', ['Pierre']), Tree('NNP', ['Vinken'])]), Tree(',', ['']), Tree('ADJP', [Tree('NP', [Tree('CD', ['61']), Tree('NNS', ['years'])]), Tree('JJ', ['old'])]), Tree(',', [''])]), Tree('VP', [Tree('MD', ['will']), Tree('VP', [Tree('VB', ['join']), Tree('NP', [Tree('DT', ['the']), Tree('NN', ['board'])]), Tree('PP-CLR', [Tree('IN', ['as']), Tree('NP', [Tree('DT', ['a']), Tree('JJ', ['nonexecutive']), Tree('NN', ['director'])])]), Tree('NP-TMP', [Tree('NNP', ['Nov.']), Tree('CD', ['29'])])]), Tree('.', ['.'])])])])])])])
```

nltk treebank

- `>>> t[-1].draw()`
- The *sample* of the well-known Penn Treebank (PTB) Wall Street Journal (WSJ) corpus includes:
 - 3,914 parsed sentences
 - 49,000+ parsed sentences in the full corpus



nlTK treebank

- Words:

```
>>> w = treebank.words()
```

```
>>> len(w)
```

```
100676
```

```
>>> w
```

```
['Pierre', 'Vinken', ',', '61', 'years', 'old', ',', ...]
```

```
>>> tw = treebank.tagged_words()
```

```
>>> tw
```

```
[('Pierre', 'NNP'), ('Vinken', 'NNP'), (',', ' '), ...]
```

```
>>> tw[:10]
```

```
[('Pierre', 'NNP'), ('Vinken', 'NNP'), (',', ' '), ('61', 'CD'), ('years',  
'NNS'), ('old', 'JJ'), (',', ' '), ('will', 'MD'), ('join', 'VB'), ('the',  
'DT')]
```

nlk: Corpus Readers

- <http://www.nltk.org/howto/corpus.html#parsed-corpora>
 - The NLTK data package includes a 10% sample of the Penn Treebank (in `treebank`), as well as the Sinica Treebank (in `sinica_treebank`).
- Reading the Penn Treebank (Wall Street Journal sample):

```
>>> from nltk.corpus import treebank
1. treebank.fileids()
2. treebank.words(fileid)
3. treebank.tagged_words(fileid)
4. treebank.parsed_sents(fileid)
```

nlk: Corpus Readers

`treebank.words(fileid)`

```
>>> len(treebank.words('wsj_0003.mrg'))
782
>>> treebank.words('wsj_0003.mrg')
['A', 'form', 'of', 'asbestos', 'once', 'used', '*', ...]
>>> list(treebank.words('wsj_0003.mrg'))[:200]
['A', 'form', 'of', 'asbestos', 'once', 'used', '*', '*', 'to', 'make', 'Kent', 'cigarette', 'filters', 'has', 'caused',
', 'a', 'high', 'percentage', 'of', 'cancer', 'deaths', 'among', 'a', 'group', 'of', 'workers', 'exposed', '*', 'to',
'it', 'more', 'than', '30', 'years', 'ago', ',', 'researchers', 'reported', '0', '*T*-1', '.', 'The', 'asbestos', 'fib',
er', ',', 'crocidolite', ',', 'is', 'unusually', 'resilient', 'once', 'it', 'enters', 'the', 'lungs', ',', 'with', 'ev',
en', 'brief', 'exposures', 'to', 'it', 'causing', 'symptoms', 'that', '*T*-1', 'show', 'up', 'decades', 'later', ',',
'researchers', 'said', '0', '*T*-2', '.', 'Lorillard', 'Inc.', ',', 'the', 'unit', 'of', 'New', 'York-based', 'Loews',
'Corp.', 'that', '*T*-2', 'makes', 'Kent', 'cigarettes', ',', 'stopped', 'using', 'crocidolite', 'in', 'its', 'Micron',
ite', 'cigarette', 'filters', 'in', '1956', '.', 'Although', 'preliminary', 'findings', 'were', 'reported', '*-2', 'mo',
re', 'than', 'a', 'year', 'ago', ',', 'the', 'latest', 'results', 'appear', 'in', 'today', '"s', 'New', 'England', 'Jo',
urnal', 'of', 'Medicine', ',', 'a', 'forum', 'likely', '*', 'to', 'bring', 'new', 'attention', 'to', 'the', 'problem',
',', 'A', 'Lorillard', 'spokewoman', 'said', ',', '``', 'This', 'is', 'an', 'old', 'story', '.', 'We', '"re', 'talkin',
g', 'about', 'years', 'ago', 'before', 'anyone', 'heard', 'of', 'asbestos', 'having', 'any', 'questionable', 'properti',
es', '.', 'There', 'is', 'no', 'asbestos', 'in', 'our', 'products', 'now', '.', '"', 'Neither', 'Lorillard', 'nor', 't',
he', 'researchers', 'who', '*T*-3', 'studied', 'the', 'workers', 'were', 'aware', 'of', 'any', 'research', 'on', 'smo',
kers', 'of', 'the', 'Kent', 'cigarettes', '.']
>>>
```

nlTK: Corpus Readers

```
treebank.fileids()
```

[illegible]

```
>>> len(treebank.fileids())
```

199

nlk: Corpus Readers

treebank.tagged_words(*fileid*)

```
>>> >>> list(treebank.tagged_words('wsj_0003.mrg'))[:100]
[('A', 'DT'), ('form', 'NN'), ('of', 'IN'), ('asbestos', 'NN'), ('once', 'RB'), ('used', 'VBN'), (*, '-NONE-'), (*, '-NONE-'), ('to', 'TO'), ('make', 'VB'), ('Kent', 'NNP'), ('cigarette', 'NN'), ('filters', 'NNS'), ('has', 'VBZ'), ('caused', 'VBN'), ('a', 'DT'), ('high', 'JJ'), ('percentage', 'NN'), ('of', 'IN'), ('cancer', 'NN'), ('deaths', 'NNS'), ('among', 'IN'), ('a', 'DT'), ('group', 'NN'), ('of', 'IN'), ('workers', 'NNS'), ('exposed', 'VBN'), (*, '-NONE-'), ('to', 'TO'), ('it', 'PRP'), ('more', 'RBR'), ('than', 'IN'), ('30', 'CD'), ('years', 'NNS'), ('ago', 'IN'), (',', ','), ('researchers', 'NNS'), ('reported', 'VBD'), ('0', '-NONE-'), (*T*-1, '-NONE-'), ('.', '.'), ('The', 'DT'), ('asbestos', 'NN'), ('fiber', 'NN'), (',', ','), ('crocidolite', 'NN'), (',', ','), ('is', 'VBZ'), ('unusually', 'RB'), ('resilient', 'JJ'), ('once', 'IN'), ('it', 'PRP'), ('enters', 'VBZ'), ('the', 'DT'), ('lungs', 'NNS'), (',', ','), ('with', 'IN'), ('even', 'RB'), ('brief', 'JJ'), ('exposures', 'NNS'), ('to', 'TO'), ('it', 'PRP'), ('causing', 'VBG'), ('symptoms', 'NNS'), ('that', 'WDT'), (*T*-1, '-NONE-'), ('show', 'VBP'), ('up', 'RP'), ('decades', 'NNS'), ('later', 'JJ'), (',', ','), ('researchers', 'NNS'), ('said', 'VBD'), ('0', '-NONE-'), (*T*-2, '-NONE-'), ('.', '.'), ('Lorillard', 'NNP'), ('Inc.', 'NNP'), (',', ','), ('the', 'DT'), ('unit', 'NN'), ('of', 'IN'), ('New', 'JJ'), ('York-based', 'JJ'), ('Loews', 'NNP'), ('Corp.', 'NNP'), ('that', 'WDT'), (*T*-2, '-NONE-'), ('makes', 'VBZ'), ('Kent', 'NNP'), ('cigarettes', 'NNS'), (',', ','), ('stopped', 'VBD'), ('using', 'VBG'), ('crocidolite', 'NN'), ('in', 'IN'), ('its', 'PRP$'), ('Micronite', 'NN'), ('cigarette', 'NN'), ('filters', 'NNS')]
>>> █
```

nltk: Corpus Readers

`treebank.parsed_sents(fileid)`

```
>>> len(treebank.parsed_sents('wsj_0003.mrg'))
30
>>> treebank.parsed_sents('wsj_0003.mrg')[0]
Tree('S', [Tree('S-TPC-1', [Tree('NP-SBJ', [Tree('NP', [Tree('NP', [Tree('DT', ['A']), Tree('NN', ['form'])]), Tree('P
P', [Tree('IN', ['of']), Tree('NP', [Tree('NN', ['asbestos'])])])]), Tree('RRC', [Tree('ADVP-TMP', [Tree('RB', ['once'
])), Tree('VP', [Tree('VBN', ['used']), Tree('NP', [Tree('-NONE-', ['*'])]), Tree('S-CLR', [Tree('NP-SBJ', [Tree('-NO
NE-', ['*'])]), Tree('VP', [Tree('TO', ['to']), Tree('VP', [Tree('VB', ['make']), Tree('NP', [Tree('NNP', ['Kent']), T
ree('NN', ['cigarette']), Tree('NNS', ['filters'])])])])]), Tree('VP', [Tree('VBZ', ['has']), Tree('VP', [Tree('
VBN', ['caused']), Tree('NP', [Tree('NP', [Tree('DT', ['a']), Tree('JJ', ['high']), Tree('NN', ['percentage'])]), Tree
('PP', [Tree('IN', ['of']), Tree('NP', [Tree('NN', ['cancer']), Tree('NNS', ['deaths'])])]), Tree('PP-LOC', [Tree('IN'
, ['among']), Tree('NP', [Tree('NP', [Tree('DT', ['a']), Tree('NN', ['group'])]), Tree('PP', [Tree('IN', ['of']), Tree
('NP', [Tree('NP', [Tree('NNS', ['workers'])]), Tree('RRC', [Tree('VP', [Tree('VBN', ['exposed']), Tree('NP', [Tree('-
NONE-', ['*'])]), Tree('PP-CLR', [Tree('TO', ['to']), Tree('NP', [Tree('PRP', ['it'])])]), Tree('ADVP-TMP', [Tree('NP'
, [Tree('QP', [Tree('RBR', ['more']), Tree('IN', ['than']), Tree('CD', ['30'])]), Tree('NNS', ['years'])]), Tree('IN',
['ago'])])])])])])])]), Tree(',', [''], Tree('NP-SBJ', [Tree('NNS', ['researchers'])]), Tree('VP', [Tree('V
BD', ['reported']), Tree('SBAR', [Tree('-NONE-', ['0']), Tree('S', [Tree('-NONE-', ['*T*-1'])])])]), Tree('.', ['.'])]
)
>>>
```

Penn Treebank

- Download instructions in class.
- Linguistic Data Consortium (LDC)
 - U. of Arizona is a member, we have access
- Website:
 - <https://catalog.ldc.upenn.edu/LDC99T42>

Item Name: Treebank-3
Author(s): Mitchell P. Marcus, Beatrice Santorini, Mary Ann Marcinkiewicz, Ann Taylor
LDC Catalog No.: LDC99T42

This 1999 release contains the following Treebank-2 Material:

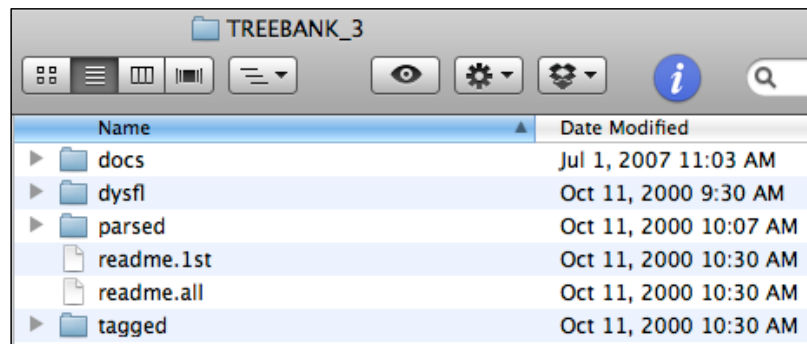
- One million words of 1989 Wall Street Journal material annotated in Treebank II style.
- A small sample of ATIS-3 material annotated in Treebank II style.
- A fully tagged version of the Brown Corpus.

and the following new material:

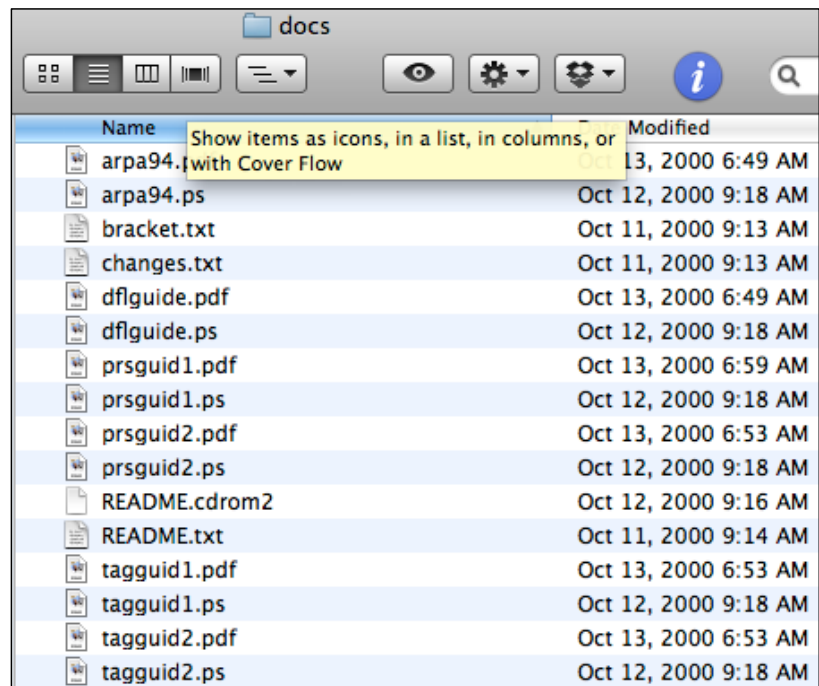
- Switchboard tagged, dysfluency-annotated, and parsed text
- Brown parsed text

Treebank Guides

- Tagging Guide
- arpa94 paper
- Parse Guide



Name	Date Modified
▶ docs	Jul 1, 2007 11:03 AM
▶ dysfl	Oct 11, 2000 9:30 AM
▶ parsed	Oct 11, 2000 10:07 AM
readme.1st	Oct 11, 2000 10:30 AM
readme.all	Oct 11, 2000 10:30 AM
▶ tagged	Oct 11, 2000 10:30 AM



Name	Modified
arpa94.t	Oct 13, 2000 6:49 AM
arpa94.ps	Oct 12, 2000 9:18 AM
bracket.txt	Oct 11, 2000 9:13 AM
changes.txt	Oct 11, 2000 9:13 AM
dflguide.pdf	Oct 13, 2000 6:49 AM
dflguide.ps	Oct 12, 2000 9:18 AM
prsguid1.pdf	Oct 13, 2000 6:59 AM
prsguid1.ps	Oct 12, 2000 9:18 AM
prsguid2.pdf	Oct 13, 2000 6:53 AM
prsguid2.ps	Oct 12, 2000 9:18 AM
README.cdrom2	Oct 12, 2000 9:16 AM
README.txt	Oct 11, 2000 9:14 AM
tagguid1.pdf	Oct 13, 2000 6:53 AM
tagguid1.ps	Oct 12, 2000 9:18 AM
tagguid2.pdf	Oct 13, 2000 6:53 AM
tagguid2.ps	Oct 12, 2000 9:18 AM

Penn POS Tagset

POS Tag	Description	Example
CC	coordinating conjunction	and
CD	cardinal number	1, third
DT	determiner	the
EX	existential there	<i>there is</i>
FW	foreign word	d'hoevre
IN	preposition/subordinating conjunction	in, of, like
JJ	adjective	green
JJR	adjective, comparative	greener
JJS	adjective, superlative	greenest
LS	list marker	1)
MD	modal	could, will
NN	noun, singular or mass	table
NNS	noun plural	tables
NNP	proper noun, singular	John
NNPS	proper noun, plural	Vikings

Penn POS Tagset

PDT	predeterminer	<i>both</i> the boys
POS	possessive ending	friend's
PRP	personal pronoun	I, he, it
PRP\$	possessive pronoun	my, his
RB	adverb	however, usually, naturally, here, good
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	<i>give up</i>
TO	to	<i>to</i> go, <i>to</i> him
UH	interjection	uhhuhhuhh
VB	verb, base form	take
VBD	verb, past tense	took
VBG	verb, gerund/present participle	taking
VBN	verb, past participle	taken
VBP	verb, sing. present, non-3d	take
VBZ	verb, 3rd person sing. present	takes

Penn POS Tagset

WDT	wh-determiner	which
WP	wh-pronoun	who, what
WP\$	possessive wh-pronoun	whose
WRB	wh-abverb	where, when

Penn Syntax Tagset

- (from The Penn Treebank: An overview, Taylor, Marcus & Santorini)

Table 1.2. The Penn Treebank syntactic tagset

ADJP	Adjective phrase
ADVP	Adverb phrase
NP	Noun phrase
PP	Prepositional phrase
S	Simple declarative clause
SBAR	Subordinate clause
SBARQ	Direct question introduced by <i>wh</i> -element
SINV	Declarative sentence with subject-aux inversion
SQ	Yes/no questions and subconstituent of SBARQ excluding <i>wh</i> -element
VP	Verb phrase
WHADVP	Wh-adverb phrase
WHNP	Wh-noun phrase
WHPP	Wh-prepositional phrase
X	Constituent of unknown or uncertain category
*	“Understood” subject of infinitive or imperative
0	Zero variant of <i>that</i> in subordinate clauses
T	Trace of <i>wh</i> -Constituent

Penn Syntax Tagset

- (from The Penn Treebank: An overview, Taylor, Marcus & Santorini)

<i>Table 1.3. Functional Tags</i>	
<i>Text Categories</i>	
-HLN	headlines and datelines
-LST	list markers
-TTL	titles
<i>Grammatical Functions</i>	
-CLF	true clefts
-NOM	non NPs that function as NPs
-ADV	clausal and NP adverbials
-LGS	logical subjects in passives
-PRD	non VP predicates
-SBJ	surface subject
-TPC	topicalized and fronted constituents
-CLR	closely related - see text
<i>Semantic Roles</i>	
-VOC	vocatives
-DIR	direction & trajectory
-LOC	location
-MNR	manner
-PRP	purpose and reason
-TMP	temporal phrases

Treebank Guides

- Parts-of-speech (POS) Tagging Guide, tagguid1.pdf (34 pages):

Part-of-Speech Tagging Guidelines
for the Penn Treebank Project
(3rd Revision, 2nd printing)

Beatrice Santorini

June 1990 ¹

tagguid2.pdf: addendum, see POS tag 'TO'

Treebank Guides

- Parsing guide 1, prsguid1.pdf (318 pages):

**Bracketing Guidelines for Treebank II Style
Penn Treebank Project ¹**

Principal authors:

Ann Bies, Mark Ferguson, Karen Katz, and Robert MacIntyre

Major contributors:

Victoria Tredinnick, Grace Kim, Mary Ann Marcinkiewicz, Britta Schasberger ²

January 1995

prsguid2.pdf: addendum for the Switchboard corpus

Tregex

- URL: <https://nlp.stanford.edu/software/tregex.shtml>

Contents

The download is a 9 Mb zip file. It contains:

1. README-tregex.txt -- Basic information about the distribution, including a "quickstart" guide.
2. README-tsurgeon.txt -- information about Tsurgeon.
3. README-gui.txt -- information about using the graphical interface
4. LICENSE -- Tregex is licensed under the Gnu General Public License.
5. stanford-tregex.jar -- This is a JAR file containing all the Stanford classes necessary to run tregex.
6. src directory -- a directory with the source files for Tregex and Tsurgeon
7. lib directory -- library files required for recompiling the distribution (with Mac OS X customization; see [lib/ABOUT-AppleJavaExtensions.txt](#) for removing this dependency)
8. build.xml, Makefile -- files for recompiling (with ant or make) the distribution
9. javadoc -- Javadocs for the distribution
10. tregex.sh, tsurgeon.sh -- sample scripts for running Tregex and Tsurgeon from the command line
11. run-tregex-gui.command, run-tregex-gui.bat -- shell script for running the graphical interface for Tregex with more memory for searching larger treebanks; can be double-clicked to open on a Mac or PC, respectively
12. examples directory -- example files for Tregex and Tsurgeon

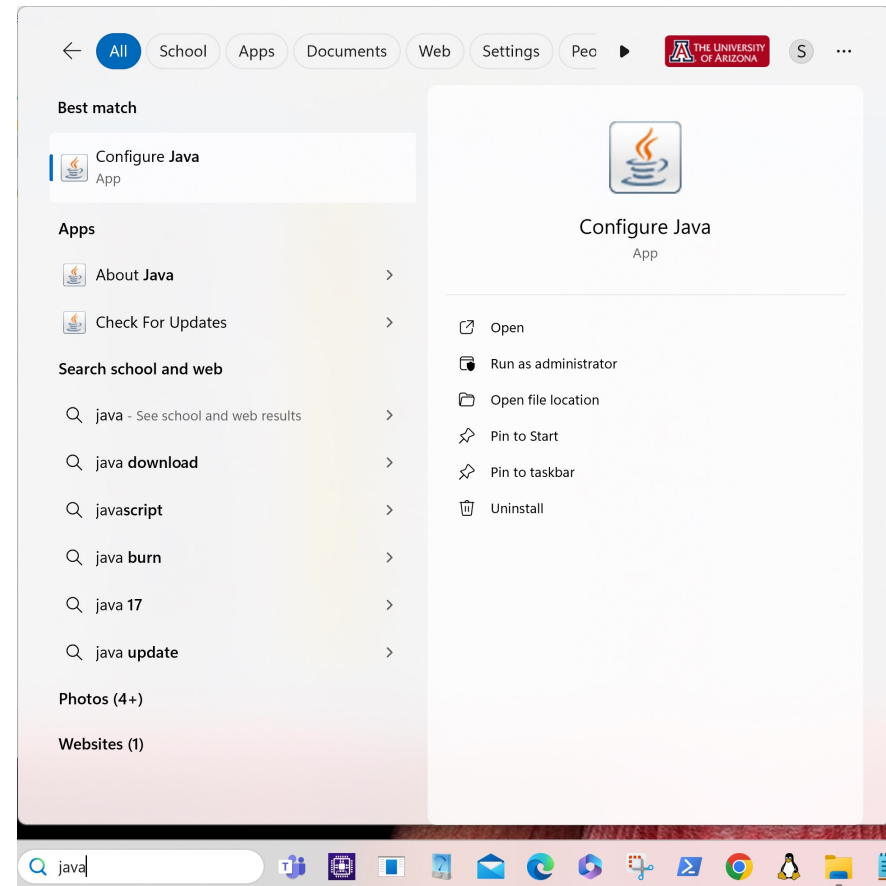
Download

Download Tregex version 4.2.0 (source and executables for all platforms)

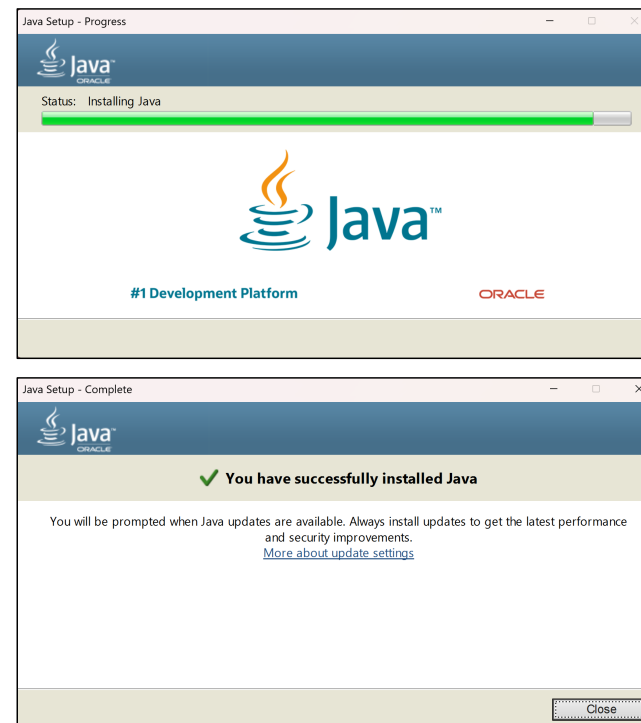
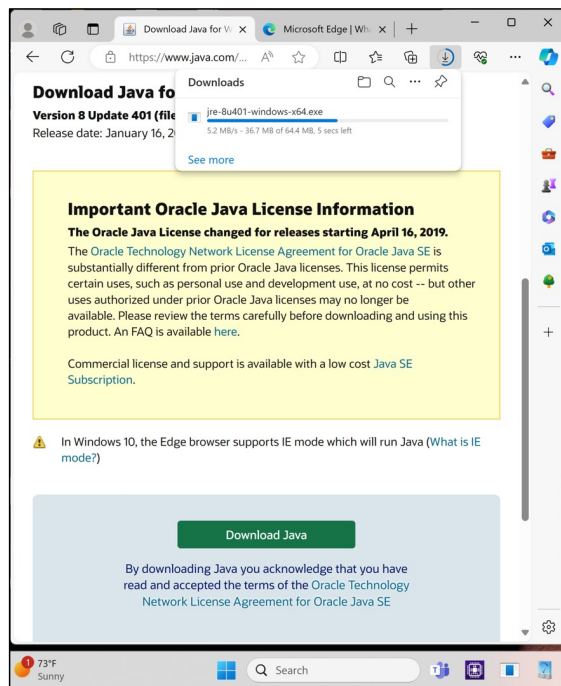
[Download Tregex version 3.4 Mac OS X disk image](#) (GUI packaged as Mac application; Java 1.7 runtime included)

Windows 11: jre 8 update 401 January 2024

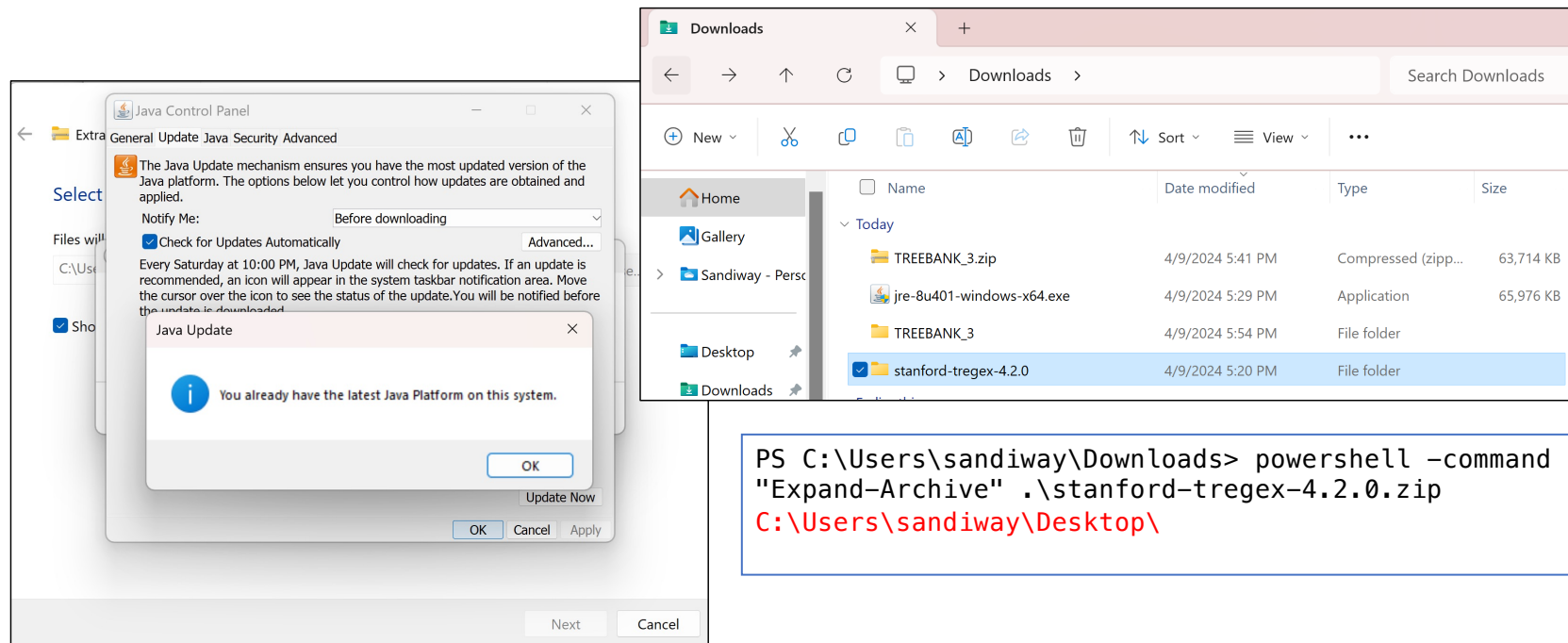
Make sure Java's run-time
environment is installed



jre-8u401-windows-x86.exe

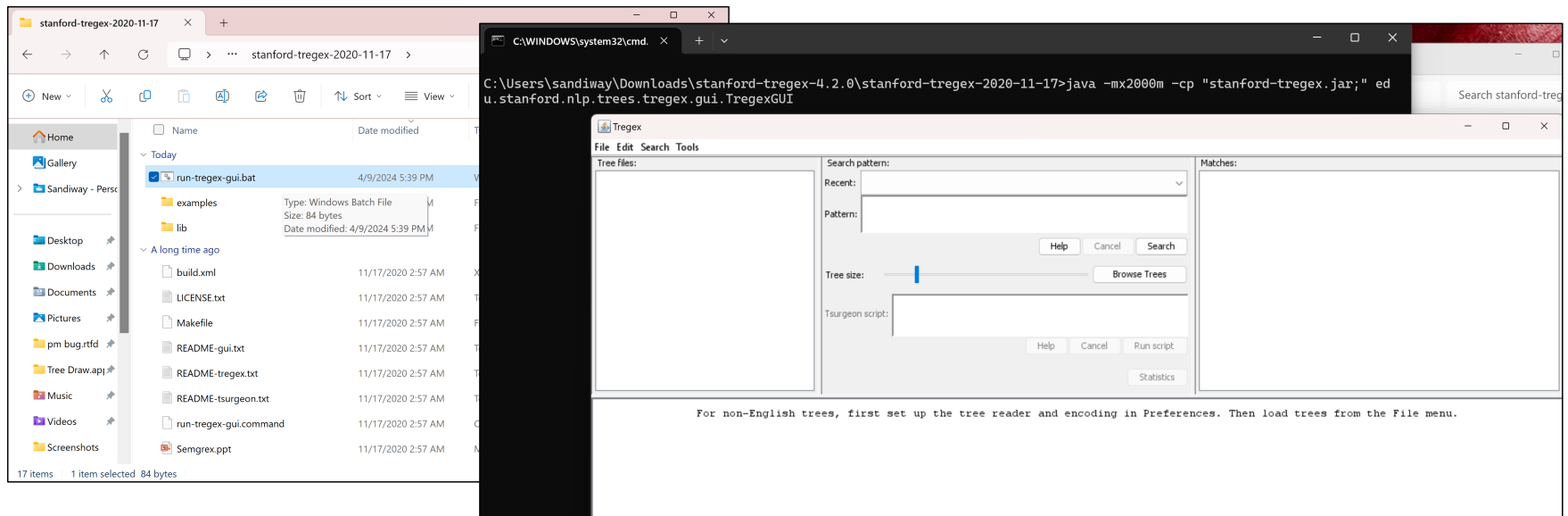


jre updated



run-tregex-gui.bat

- edit .bat file to increase memory: `-mx2000`



Ubuntu Install

- Terminal:

```
~/Downloads$ unzip stanford-tregex-4.2.0.zip
~/Downloads$ cd stanford-tregex-2020-11-17/
~/Downloads/stanford-tregex-2020-11-17$ ./run-tregex-gui.command
Gtk-Message: 20:44:58.957: Failed to load module "canberra-gtk-module"
Apr 09, 2024 8:44:59 PM java.util.prefs.FileSystemPreferences$1 run
INFO: Created user preferences directory.
```

- Install:

```
sudo apt install libcanberra-gtk-module
The following additional packages will be installed:
  libcanberra-gtk0
The following NEW packages will be installed:
  libcanberra-gtk-module libcanberra-gtk0
```

tregex on Ubuntu

Tregex

File Edit Search Tools

Tree files:
* brown
* wsj

Search pattern:
Recent:
Pattern:
Help Cancel Search

Tree size: 12 Browse Trees

Tsurgeon script:
Help Cancel Run script

Browse stats: 73451 trees found in the selected files

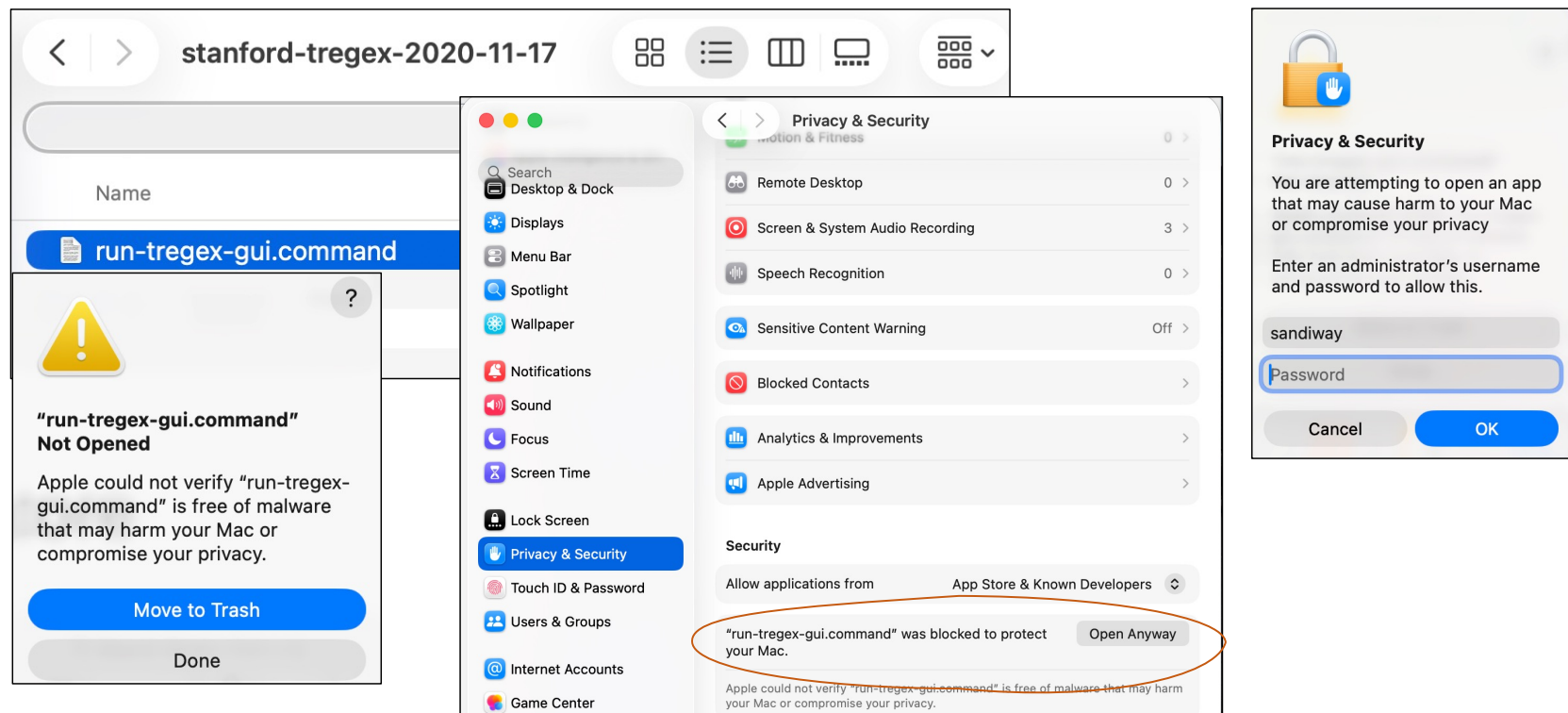
Matches:

- cr07.mrg-1 One day , the children had wanted *-1 to get up onto Gen
- cr07.mrg-2 They wanted *-1 to see what his back *ICH*-3 felt like *T*
- cr07.mrg-3 He looked so comfortable *-1 being straight .
- cr07.mrg-4 They wanted *-1 to touch the mystery .
- cr07.mrg-5 Arlene was boosting them up when the policeman came l
- cr07.mrg-6 He was very rude .
- cr07.mrg-7 Arlene had a hard voice , too , this time .
- cr07.mrg-8 The policeman 's eyes rather popped for a second ; ; but tl
- cr07.mrg-9 The policeman got a confused , funny look on his face , an
- cr07.mrg-10 But this is a public park and it *EXP*-1 's a city ordinance
- cr07.mrg-11 Arlene was so ashamed that she hung her head when sh
- cr07.mrg-12 The policeman walked on , but he looked back once .
- cr07.mrg-13 that had happened on the day when two other unusal

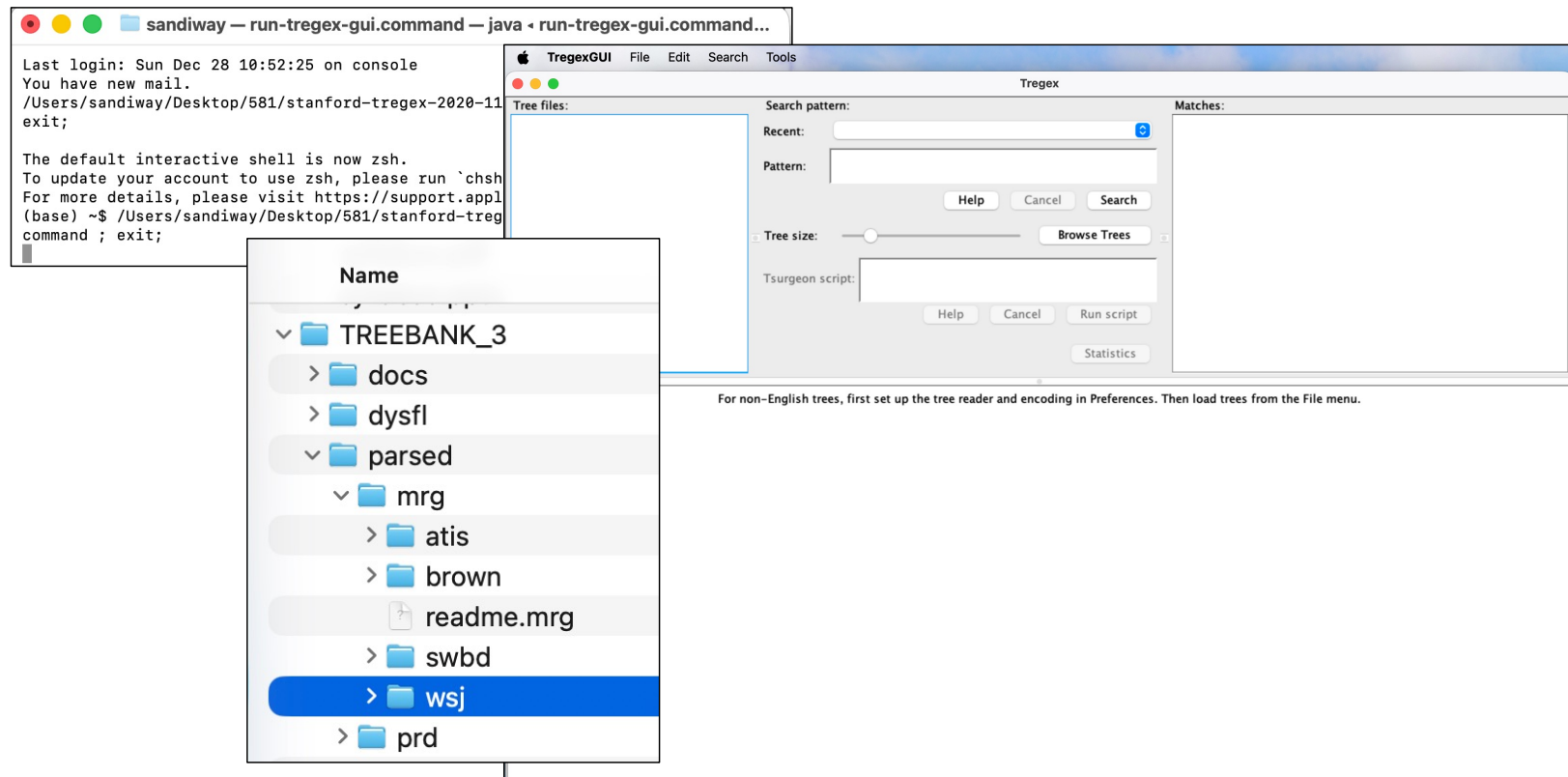
From file: /home/sandiway/Downloads/TREEBANK_3/parsed/mrg/brown/cr/cr07.mrg

One day , the children had wanted *-1 to get up onto General Burnside 's horse .

Possible macOS Problem 1



Possible macOS Problem 1



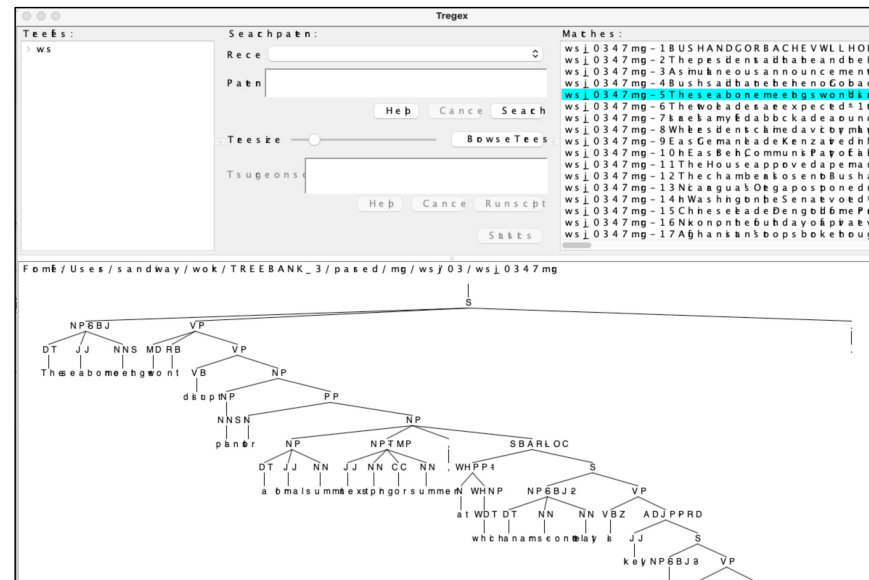
Possible macOS Problem 1

The screenshot shows the TregexGUI application interface. The 'Tree files' list on the left contains a file named 'wsj'. The 'Search pattern' section has a 'Recent' field and a 'Pattern' field, with buttons for 'Help', 'Cancel', and 'Search'. A 'Tree size' slider is set to a low value, and a 'Browse Trees' button is highlighted with an orange rectangle. The 'Matches' list on the right shows several results, with the first one being 'wsj_0347.mrg-1 BUSH AND GORBACHEV WILL HOLD two days of'. Below the search interface, the source file path is shown: 'From file: /Users/sandway/Desktop/581/TREEBANK_3/parsed/mrg/wsj/03/wsj_0347.mrg'. The main area displays a parse tree for the sentence 'BUSH AND GORBACHEV WILL HOLD two days of informal talks next month'. The tree structure is as follows:

```
graph TD
    S_HLN[S-HLN] --- NP_SBJ[NP-SBJ]
    S_HLN --- VP1[VP]
    NP_SBJ --- NNP1[NNP]
    NP_SBJ --- CC[CC]
    NP_SBJ --- NNP2[NNP]
    NNP1 --- BUSH[BUSH]
    CC --- AND[AND]
    NNP2 --- GORBACHEV[GORBACHEV]
    VP1 --- MD[MD]
    MD --- WILL[WILL]
    VP1 --- VP2[VP]
    VP2 --- VB[VB]
    VB --- HOLD[HOLD]
    VP2 --- NP[NP]
    NP --- CD[CD]
    CD --- two[two]
    NP --- NNS1[NNS]
    NNS1 --- days[days]
    NP --- PP[PP]
    PP --- IN[IN]
    IN --- of[of]
    PP --- NP2[NP]
    NP2 --- JJ[JJ]
    JJ --- informal[informal]
    NP2 --- NNS2[NNS]
    NNS2 --- talks[talks]
    VP2 --- NP_TMP[NP-TMP]
    NP_TMP --- JJ1[JJ]
    JJ1 --- next[next]
    NP_TMP --- NN[NN]
    NN --- month[month]
```

BUSH AND GORBACHEV WILL HOLD two days of informal talks next month .

- Disk image version, the Java runtime environment seems to pick wrong fonts. Display is hard to read.

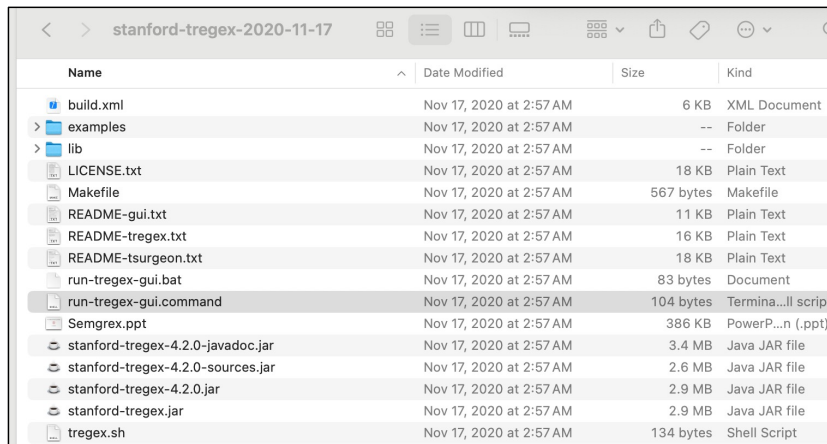


Possible macOS Problem 2

- If this happens, download the non disk image link

Download Tregex version 4.2.0 (source and executables for all platforms)

Download Tregex version 3.4 Mac OS X disk image (GUI packaged as Mac application; Java 1.7 runtime included)



Name	Date Modified	Size	Kind
build.xml	Nov 17, 2020 at 2:57 AM	6 KB	XML Document
examples	Nov 17, 2020 at 2:57 AM	--	Folder
lib	Nov 17, 2020 at 2:57 AM	--	Folder
LICENSE.txt	Nov 17, 2020 at 2:57 AM	18 KB	Plain Text
Makefile	Nov 17, 2020 at 2:57 AM	567 bytes	Makefile
README-gui.txt	Nov 17, 2020 at 2:57 AM	11 KB	Plain Text
README-tregex.txt	Nov 17, 2020 at 2:57 AM	16 KB	Plain Text
README-tsurgeon.txt	Nov 17, 2020 at 2:57 AM	18 KB	Plain Text
run-tregex-gui.bat	Nov 17, 2020 at 2:57 AM	83 bytes	Document
run-tregex-gui.command	Nov 17, 2020 at 2:57 AM	104 bytes	Terminal...ll script
Semgrex.ppt	Nov 17, 2020 at 2:57 AM	386 KB	PowerP...n (.ppt)
stanford-tregex-4.2.0-javadoc.jar	Nov 17, 2020 at 2:57 AM	3.4 MB	Java JAR file
stanford-tregex-4.2.0-sources.jar	Nov 17, 2020 at 2:57 AM	2.6 MB	Java JAR file
stanford-tregex-4.2.0.jar	Nov 17, 2020 at 2:57 AM	2.9 MB	Java JAR file
stanford-tregex.jar	Nov 17, 2020 at 2:57 AM	2.9 MB	Java JAR file
tregex.sh	Nov 17, 2020 at 2:57 AM	134 bytes	Shell Script

for macOS, run this
command

Possible macOS Problem 2

- Terminal

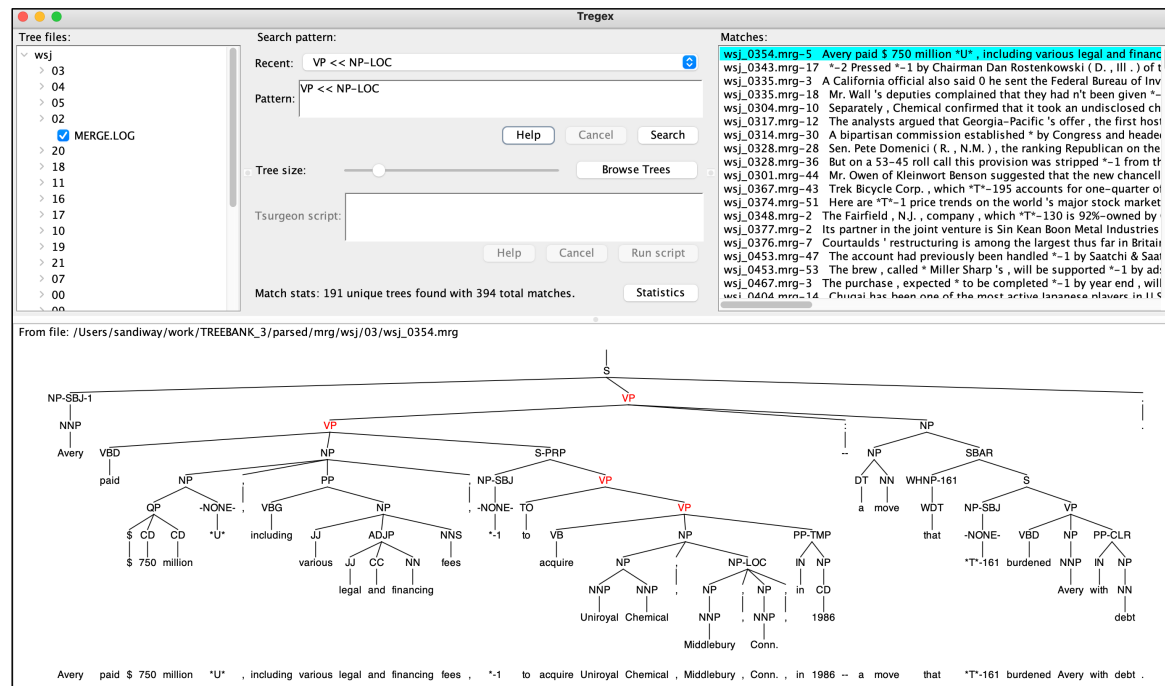
- (base) stanford-tregex-2020-11-17\$./run-tregex-gui.command
- Warning: the font "Times" is not available, so "Lucida Bright" has been substituted, but may have unexpected appearance or behavior. Re-enable the "Times" font to remove this warning.
- Warning: the font "Times" is not available, so "Lucida Bright" has been substituted, but may have unexpected

- Terminal

- *can be fixed, but maybe not worth doing ...*
- (base) stanford-tregex-2020-11-17\$./run-tregex-gui.command

Possible macOS Problem 2

Lucida Bright
is a
reasonable
substitute for
Times



with Times font restored to macOS using instructions here <https://stackoverflow.com/questions/68608157/how-can-i-fix-this-warning-the-fonts-times-and-times-are-not-available-fo>

Tregex

Tree files:
wsj

Search pattern:

Recent: VP << NP-LOC

Pattern: VP << NP-LOC

Help Cancel Search

Tree size: [Slider] Browse Trees

Tsurgeon script: [Text Area]

Help Cancel Run script

Match stats: 191 unique trees found with 394 total matches.

Matches:

- wsj_0354.mrg-5 Avery paid \$ 750 million *U*, including various legal and
- wsj_0343.mrg-17 *-2 Pressed *-1 by Chairman Dan Rostenkowski (D , III
- wsj_0335.mrg-3 A California official also said 0 he sent the Federal Bureau
- wsj_0335.mrg-18 Mr. Wall 's deputies complained that they had n't been giv
- wsj_0304.mrg-10 Separately , Chemical confirmed that it took an undisclos
- wsj_0317.mrg-12 The analysts argued that Georgia-Pacific 's offer , the fins
- wsj_0314.mrg-30 A bipartisan commission established * by Congress and I
- wsj_0328.mrg-28 Sen. Pete Domenici (R , N.M.) , the ranking Republican c
- wsj_0328.mrg-36 But on a 53-45 roll call this provision was stripped *-1 fr
- wsj_0301.mrg-44 Mr. Owen of Kleinwort Benson suggested that the new ch
- wsj_0367.mrg-43 Trek Bicycle Corp. , which *T*-195 accounts for one-qu
- wsj_0374.mrg-51 Here are "T*-1 price trends on the world 's major stock r
- wsj_0348.mrg-2 The Fairfield , N.J. , company , which "T*-130 is 92%-owne
- wsj_0377.mrg-2 Its partner in the joint venture is Sin Kean Boon Metal Indu
- wsj_0376.mrg-7 Courtaulds ' restructuring is among the largest thus far in
- wsj_0453.mrg-47 The account had previously been handled *-1 by Saatchi
- wsj_0453.mrg-53 The brew , called * Miller Sharp 's , will be supported *-1

From file: /Users/sandwich/work/TREEBANK_3/parsed/mrg/wsj/03/wsj_0354.mrg

Avery paid \$ 750 million *U*, including various legal and financing fees , *-1 to acquire Uniroyal Chemical , Middlebury , Conn. , in 1986 -- a move that *T*-161 burdened Avery with debt