

LING/C SC 581:

Advanced Computational Linguistics

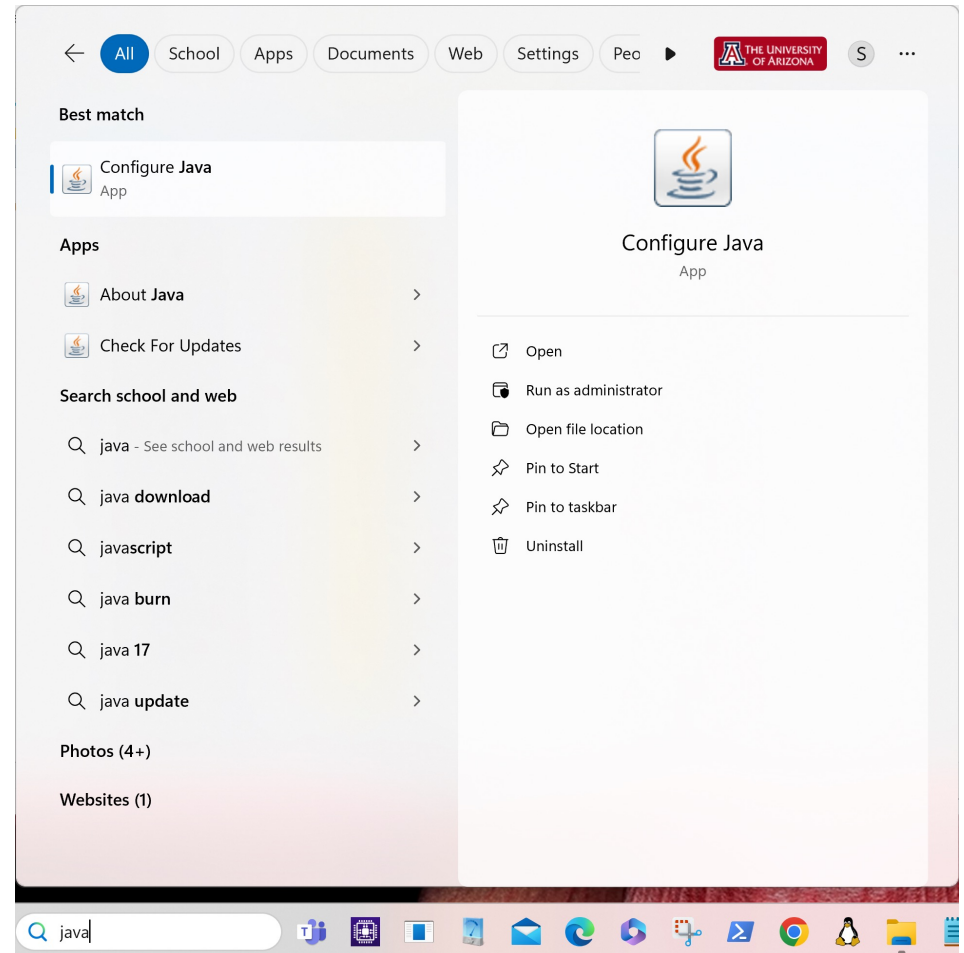
Lecture 24

Today's Topics

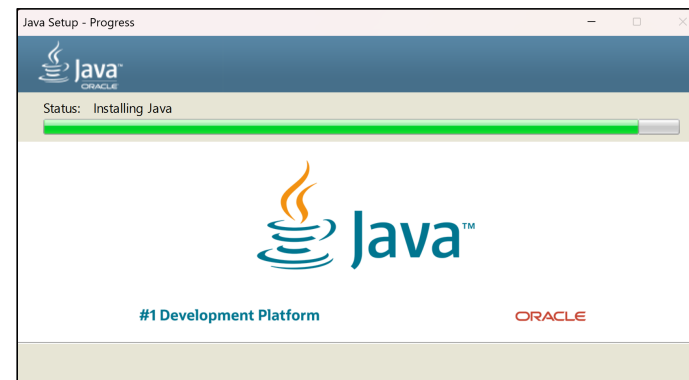
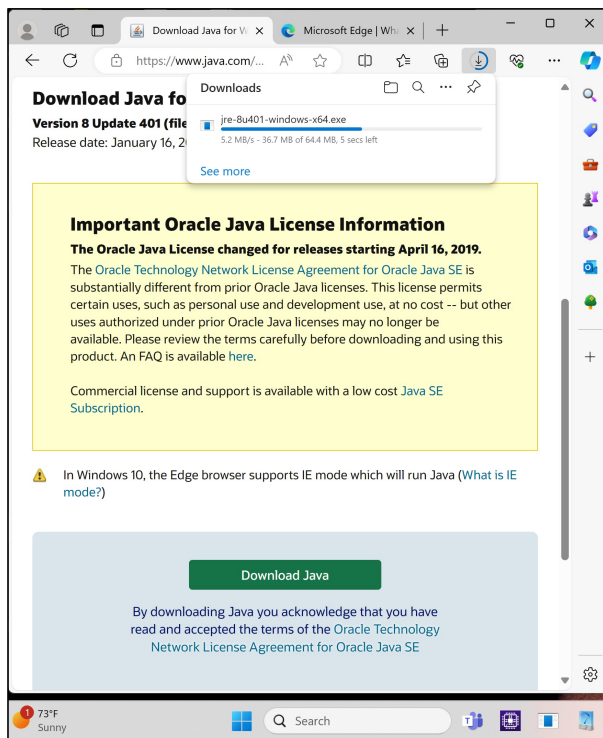
- tregex on Windows 11 and Ubuntu
 - *make sure jre is installed*
 - *I tested it by updating to jre 8 update 401*
- *Homework 10 questions?*
- TREEBANK_3 Theory
- Example: subject and object relative clauses
- Summary of tregex search

jre 8 update 401 January 2024

Make sure Java's run-time
environment is installed



jre-8u401-windows-x86.exe



jre updated

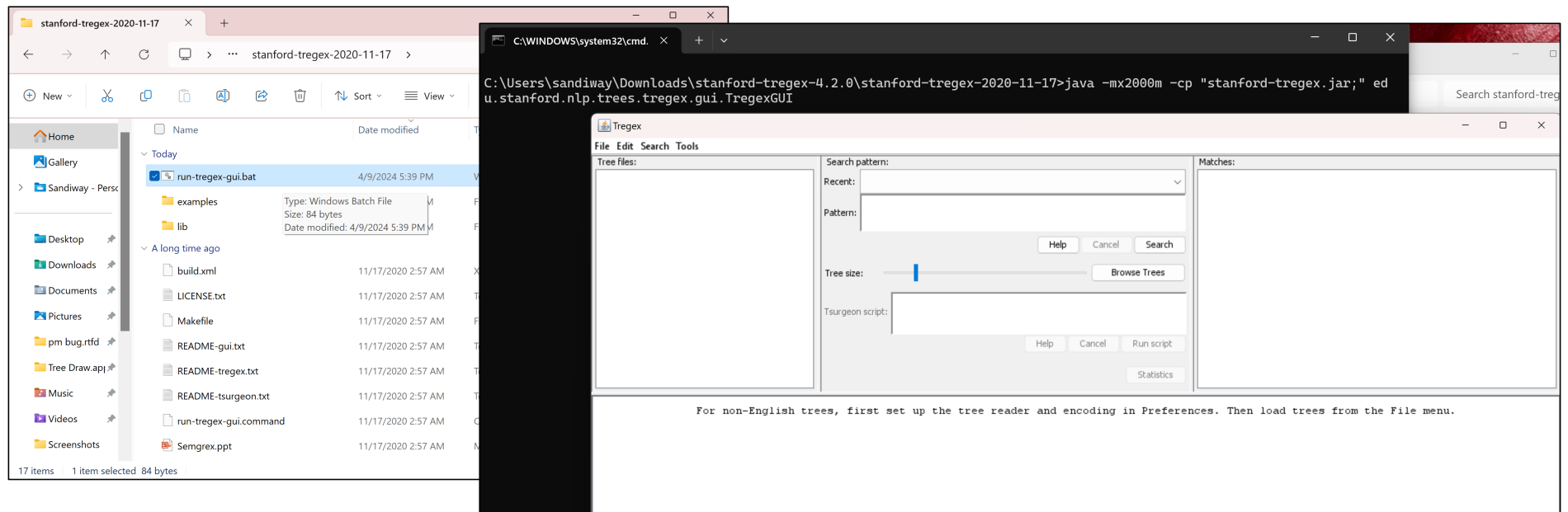
The image shows two overlapping windows. The background window is the Java Control Panel, displaying the 'General' tab with the 'Update Java' section. A 'Java Update' dialog box is overlaid on top, showing a message: 'You already have the latest Java Platform on this system.' with an 'OK' button. The foreground window is a Windows File Explorer showing the 'Downloads' folder. The file list includes:

Name	Date modified	Type	Size
Today			
TREEBANK_3.zip	4/9/2024 5:41 PM	Compressed (zipp...	63,714 KB
jre-8u401-windows-x64.exe	4/9/2024 5:29 PM	Application	65,976 KB
TREEBANK_3	4/9/2024 5:54 PM	File folder	
stanford-tregex-4.2.0	4/9/2024 5:20 PM	File folder	

```
PS C:\Users\sandiwai\Downloads> powershell -command "Expand-Archive" .\stanford-tregex-4.2.0.zip  
C:\Users\sandiwai\Desktop\
```

run-tregex-gui.bat

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



Load in brown and wsj mrg directories

The screenshot displays the Tregex application window. The top section contains a search interface with a 'Tree files' list on the left showing 'brown' and 'wsj'. The 'Search pattern' field is empty, and the 'Tree size' is set to 1. The 'Browse Trees' button is highlighted. Below the search interface, the 'Matches' list shows 18 results, with the first match highlighted: 'cf01.mrg-1 In American romance, almost nothing rates *-2 higher than what the mov'. The bottom section of the window shows two parse trees. The left tree is for the sentence 'In American romance, almost nothing rates higher than what the movie boy-meets-girl seems more adorable'. The right tree is for the sentence 'that is boy-meets-girl seems more adorable'. The parse trees use a hierarchical structure with nodes like S, NP, VP, PP, and ADJP, and terminal nodes like 'in', 'American romance', 'almost nothing rates', 'higher than', 'what the movie', 'boy-meets-girl', 'seems', 'more', and 'adorable'.

Example of search

The screenshot displays the Tregex application window. The interface is divided into several sections:

- Tree files:** A list of files including 'brown' and 'wsj'.
- Search pattern:** A text area containing the pattern `@S < (@VP < (@VP < (/^-NONE-$/)))`. Below it, a 'Recent' dropdown shows a similar pattern.
- Tree size:** A slider control and a 'Browse Trees' button.
- Tsurgeon script:** An empty text area with 'Help', 'Cancel', and 'Run script' buttons.
- Match stats:** A status bar indicating '7533 unique trees found with 8105 total matches' and a 'Statistics' button.
- Matches:** A list of search results, with the first match highlighted: 'cf01.mrg-11 This basic principle, the first in a richly knotted bundle, was conveyed *'.

Below the interface, the parse tree for the sentence is shown. The root node is 'S', which branches into 'NP-SBJ-1' and 'VP'. The 'NP-SBJ-1' node further branches into 'NP' (containing 'This basic principle') and another 'NP' (containing 'the first in a richly knotted bundle'). The 'VP' node branches into 'VBD' (containing 'was') and another 'VP'. This second 'VP' branches into 'VBN' (containing 'conveyed'), 'NP' (containing '*-1'), 'PP' (containing 'to'), 'NP' (containing 'me'), and 'PP' (containing 'by'). The 'PP' node under 'by' branches into 'NP-LGS' (containing 'Dr. Henry Lee Smith, Jr.') and 'PP-LOC' (containing 'at the University of Chicago'). The 'NP-LGS' node branches into 'NP' (containing 'Dr.', 'Henry', 'Lee', 'Smith', 'Jr.') and 'IN' (containing 'at'). The 'PP-LOC' node branches into 'IN' (containing 'at') and 'NP' (containing 'the University of Chicago'). The 'NP' node under 'at' branches into 'DT' (containing 'the'), 'NNP' (containing 'University'), 'IN' (containing 'of'), 'NNP' (containing 'Chicago'), and 'WHADVP-2' (containing 'in 1940').

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:

Tree size: 12 Browse Trees

Tsurgeon script: Help Cancel Run script

Browse stats: 73451 trees found in the selected files Statistics

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 unusual

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

```
graph TD
    S[S] --- NP_TMP[NP-TMP]
    S --- NP_SBJ[NP-SBJ]
    S --- VP1[VP]
    NP_TMP --- CD[CD]
    NP_TMP --- NN1[NN]
    CD --- One[One]
    NN1 --- day[day]
    NP_SBJ --- DT[DT]
    NP_SBJ --- NNS[NNS]
    DT --- the[the]
    NNS --- children[children]
    VP1 --- VBD[VBD]
    VBD --- had[had]
    VP1 --- VP2[VP]
    VP2 --- VEN[VEN]
    VEN --- wanted[wanted]
    VP2 --- NP_SBJ2[NP-SBJ]
    NP_SBJ2 --- NONE[NONE]
    NONE --- *-1[*-1]
    NP_SBJ2 --- TO[TO]
    TO --- to1[to]
    NP_SBJ2 --- VP3[VP]
    VP3 --- VB[V]
    VB --- get[get]
    VP3 --- PP_DIR[PP-DIR]
    PP_DIR --- RB[RB]
    RB --- up[up]
    PP_DIR --- NP3[NP]
    NP3 --- IN[IN]
    IN --- onto[onto]
    NP3 --- NP4[NP]
    NP4 --- NNP1[NNP]
    NNP1 --- General[General]
    NP4 --- NNP2[NNP]
    NNP2 --- Burnside[Burnside]
    NP4 --- POS[POS]
    POS --- 's['s]
    NP4 --- NN[NN]
    NN --- horse[horse]
```

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

tregex on Ubuntu

The screenshot shows the Tregex application window. The interface includes a menu bar (File, Edit, Search, Tools), a file list on the left, and a main workspace. The search pattern is set to `@S < (@VP < (@VP < (@NP < /^NONE-$/)))`. The search results list matches for various sentences, with the first match being: `cr07.mrg-18 I know something that *T*-1 is much more fun that we can do *T*-2 on our little lawn`. Below the results, a parse tree is displayed for this sentence, showing hierarchical relationships between phrases like NP, VP, S, and PP, with specific words and their grammatical functions labeled.

File Edit Search Tools

Tree files:
' brown
' wsj

Search pattern:
Recent: @S < (@VP < (@VP < (@NP < /^NONE-\$/)))
Pattern: @S < (@VP < (@VP < (@NP < /^NONE-\$/)))
Help Cancel Search

Tree size: 12 Browse Trees

Tsurgeon script:
Help Cancel Run script

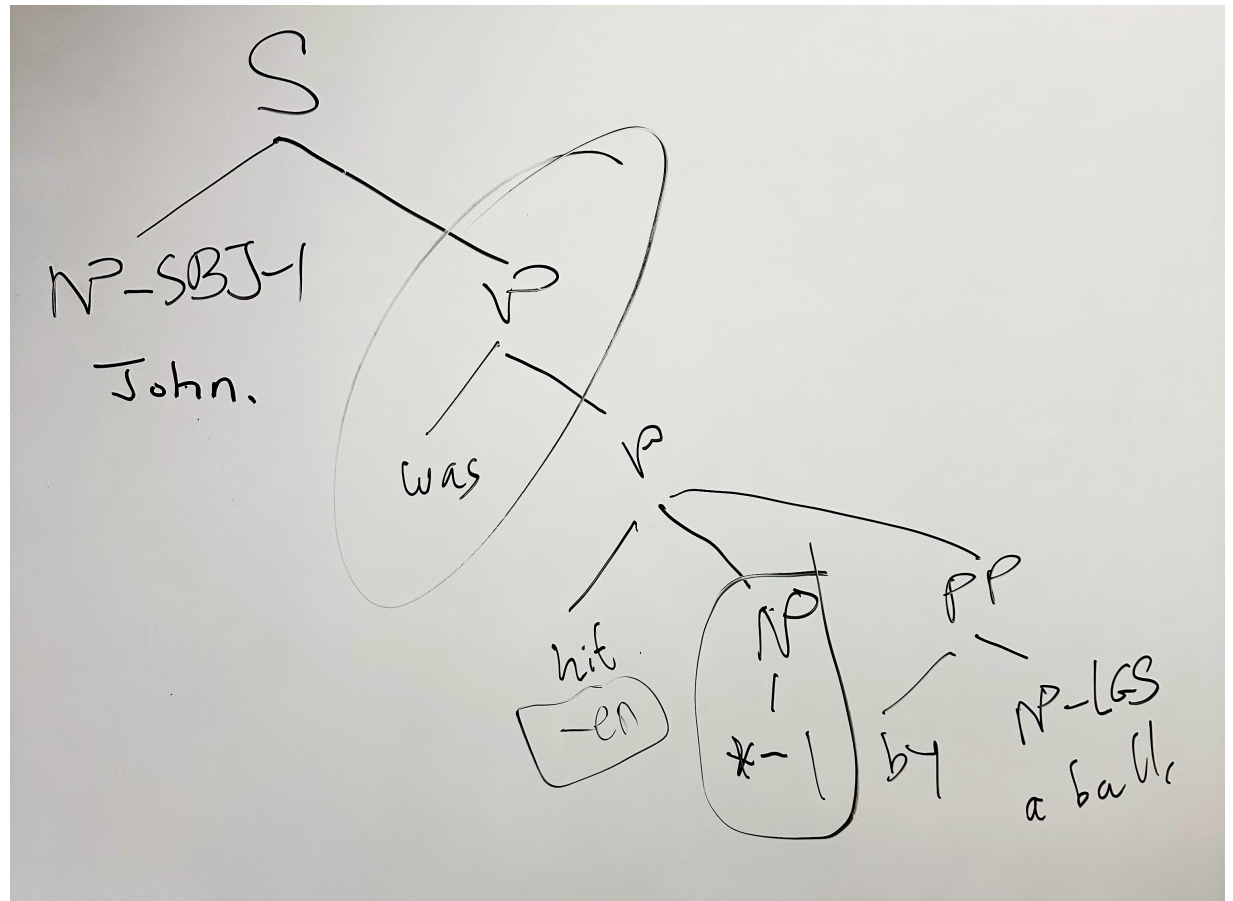
Match stats: 7533 unique trees found with 8105 total matches. Statistics

Matches:
cr07.mrg-18 I know something that *T*-1 is much more fun that we can do *T*-2 on our little lawn
cr07.mrg-48 `` This time `` , Arlene said *T*-5 , and she even kept on *3 wiggling
cr07.mrg-49 It all has something 0 * to do *T*-1 with General Burnside and his h
cr07.mrg-76 `` * Imagine 0 you wo n't get your allowance if you 're caught *2 *
cr07.mrg-115 As a result , he was persuaded *1 out to dinner .
cr07.mrg-119 But he was always persuaded *1 out .
cr07.mrg-123 This evening , they were pushed *1 in from the breakfast room ,
cr07.mrg-125 They were pushed *1 gently into the room by Arlene -- whose on
cr07.mrg-127 `` * Tell Mr. Gorboduc what you 're doing *T*-1 these days `` , my r
cr07.mrg-144 Usually , this was done *1 when attention was diverted *2 by so
cr07.mrg-148 This was delivered *1 in a forthright way , without coyne and o
cr07.mrg-157 She could see that Mr. Gorboduc was intrigued *1 ; ; the hostess i
cr02.mrg-1 I realized that Hamlet was faced *1 with an entirely different probl
cr02.mrg-2 The most that *T*-2 was accomplished *1 was * adding Mrs. Beige 'l
cr02.mrg-3 I was saved *1 from * making the decision as the phone rang , and t

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

Parse tree for: I know something that *T*-1 is much more fun that we can do *T*-2 on our little lawn

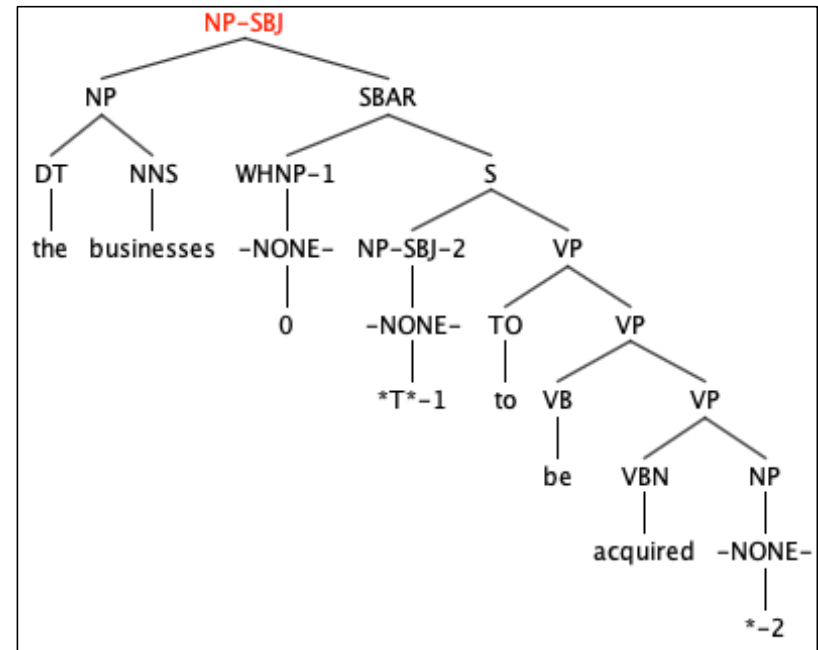
Homework 10: *English passivization*



TREEBANK_3 Theory

Theory includes empty categories: Example: *the businesses to be acquired*

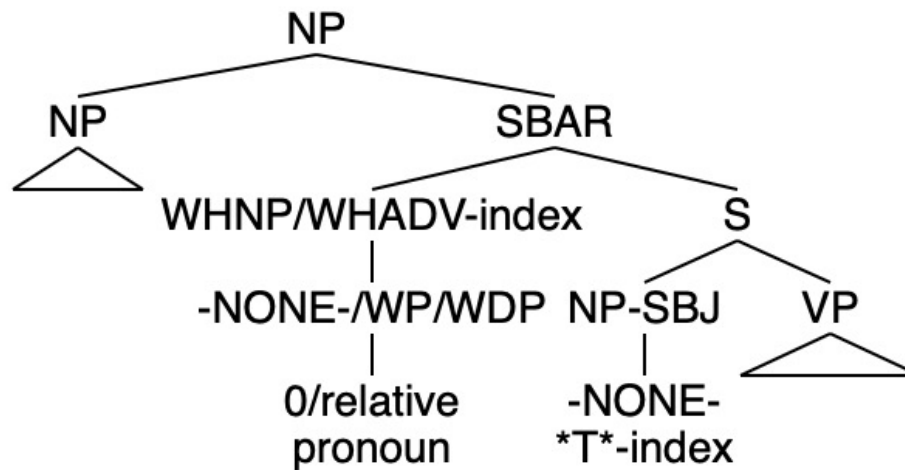
- –NONE– as a POS label is used in many circumstances, e.g.:
 - the zero complementizer under SBAR (\emptyset),
 - subject of nonfinite clauses (*),
 - object position in passives (*–*index*),
 - relativization of subjects, objects and adjuncts (*T*–*index*).



Template: subject relative clause

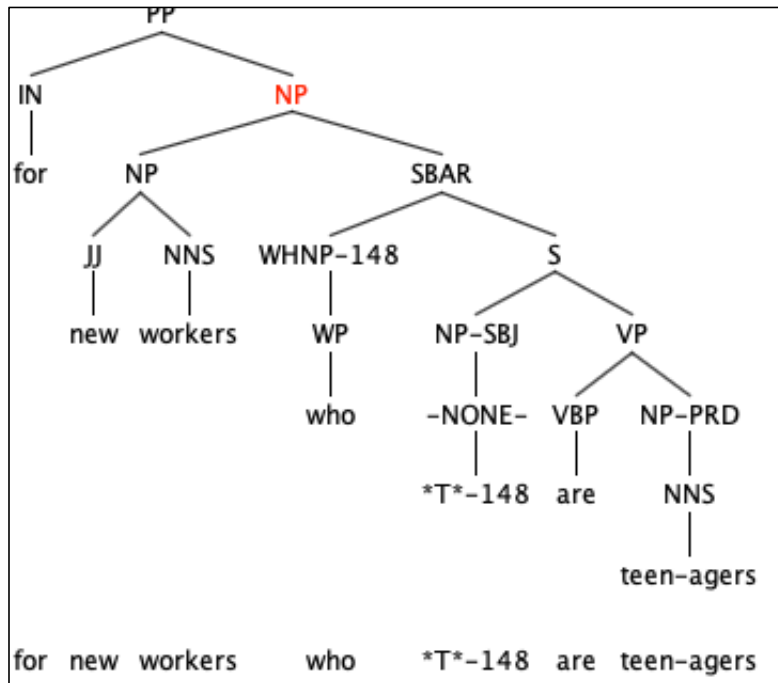
- Example:

- *the man who left the room at 6pm*
- *leave*: (subject, object)
- subject = *the man*

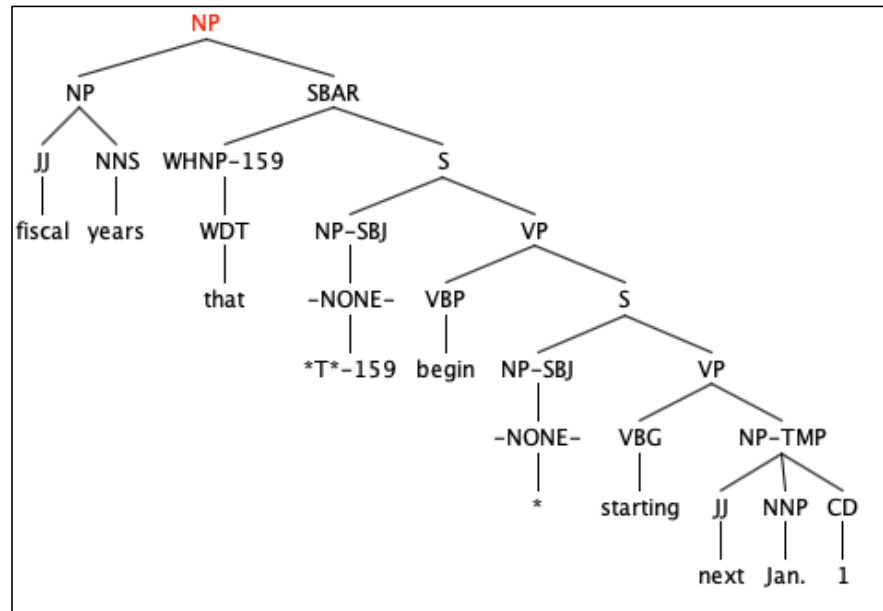


Example: subject relative clause

- relative pronoun *who*/WP (WHNP)

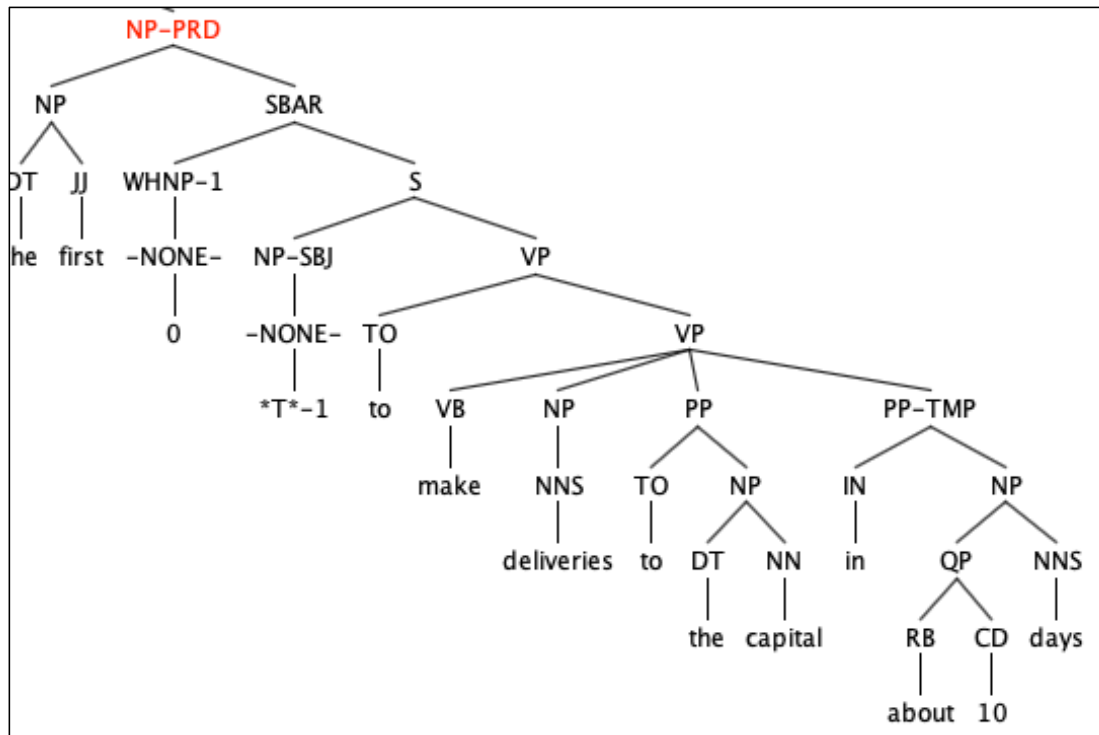


- relative pronoun *that*/WDT (WHNP)



Example: subject relative clause

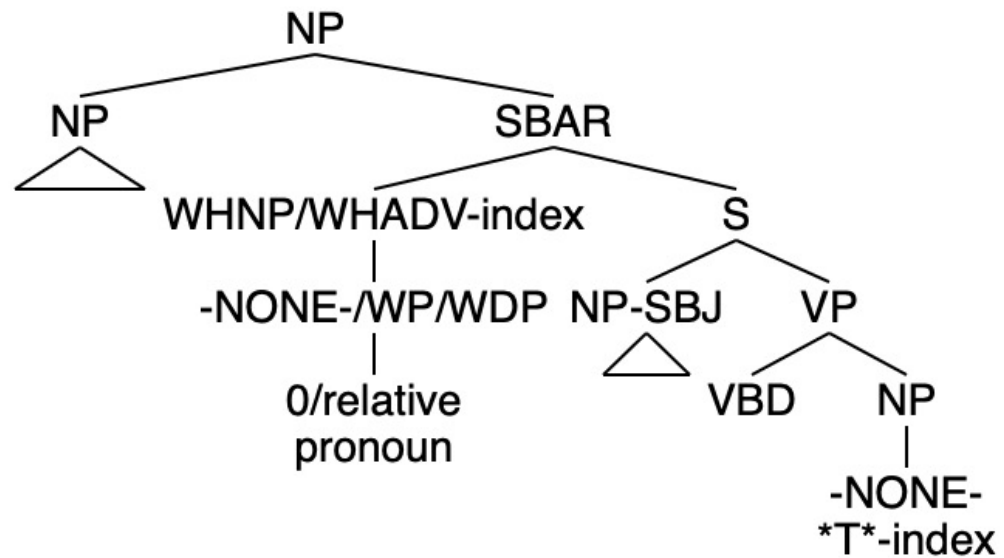
subject nonfinite relative clause (no relative pronoun, zero WHNP)



Template: object relative clause

- Example:

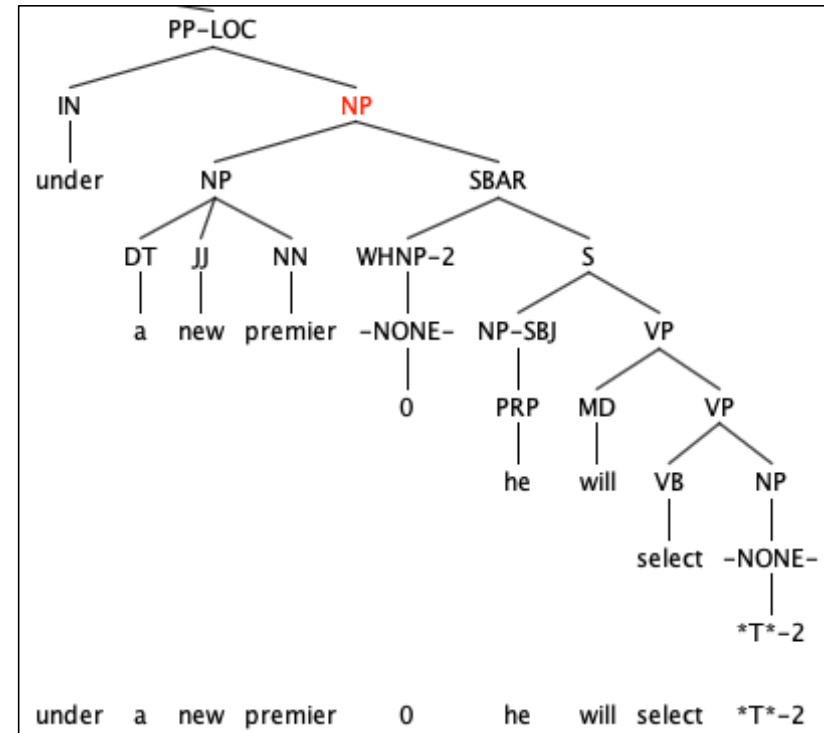
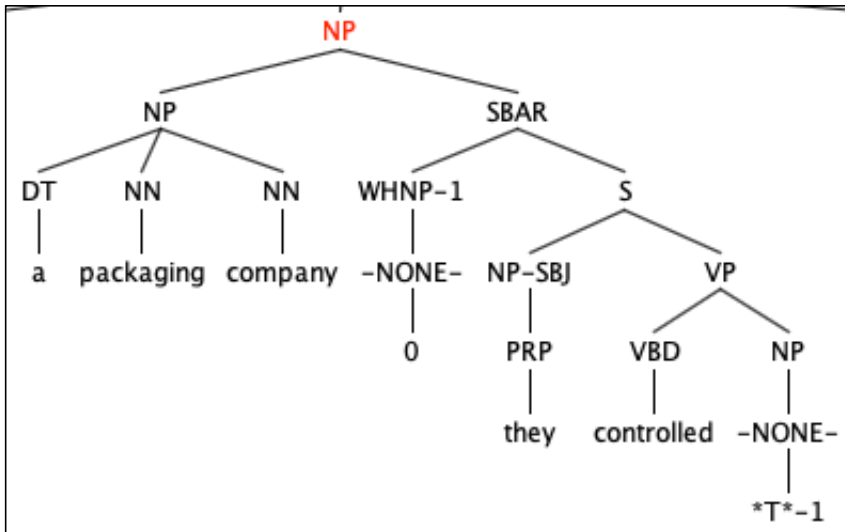
- *the room the man left at 6pm*
- *leave*: (subject, object)
- object = *the room*



Example: object relative clause

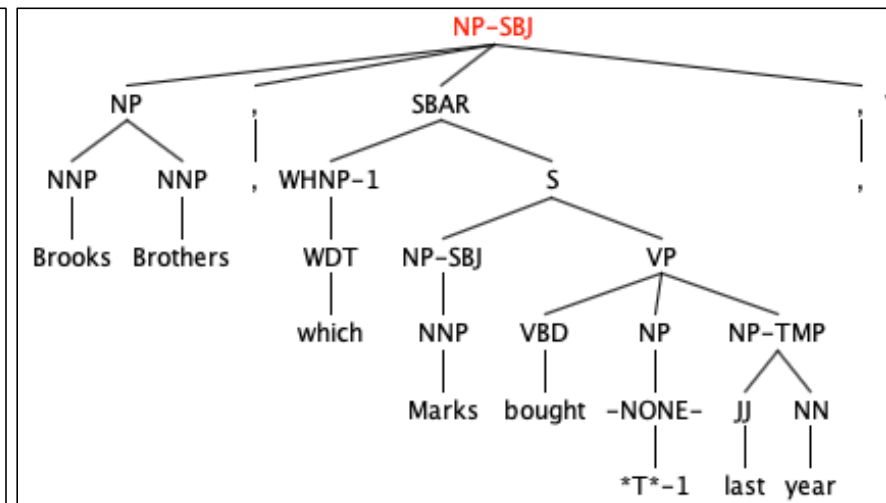
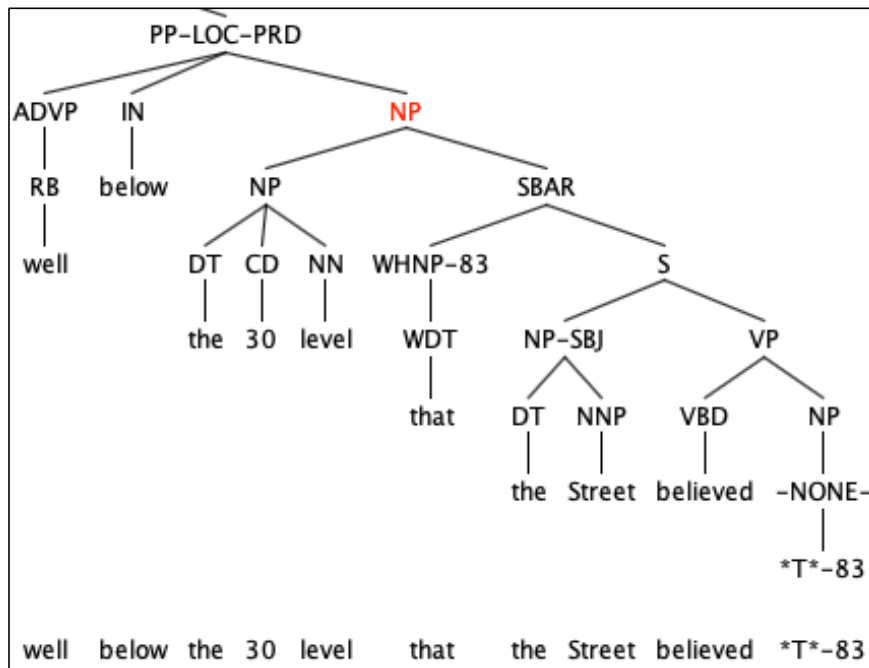
subject relatives generally need an overt WH word under SBAR

object finite relative clause **without** a relative pronoun (zero WHNP)



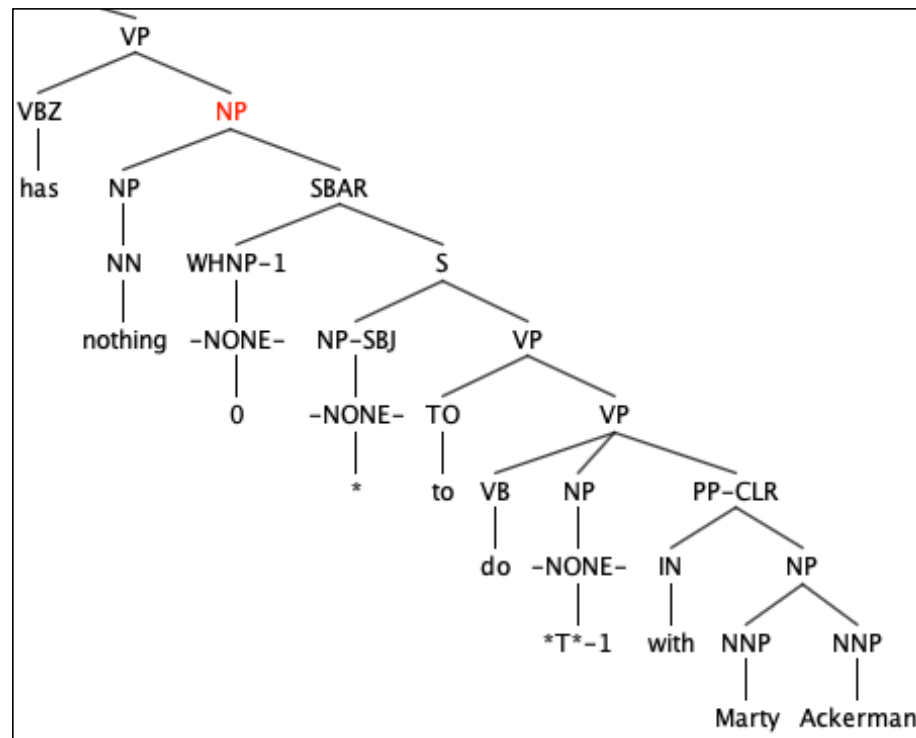
Example: object relative clause

object finite relative clause **with** relative pronoun *that*/WDT (WHNP)



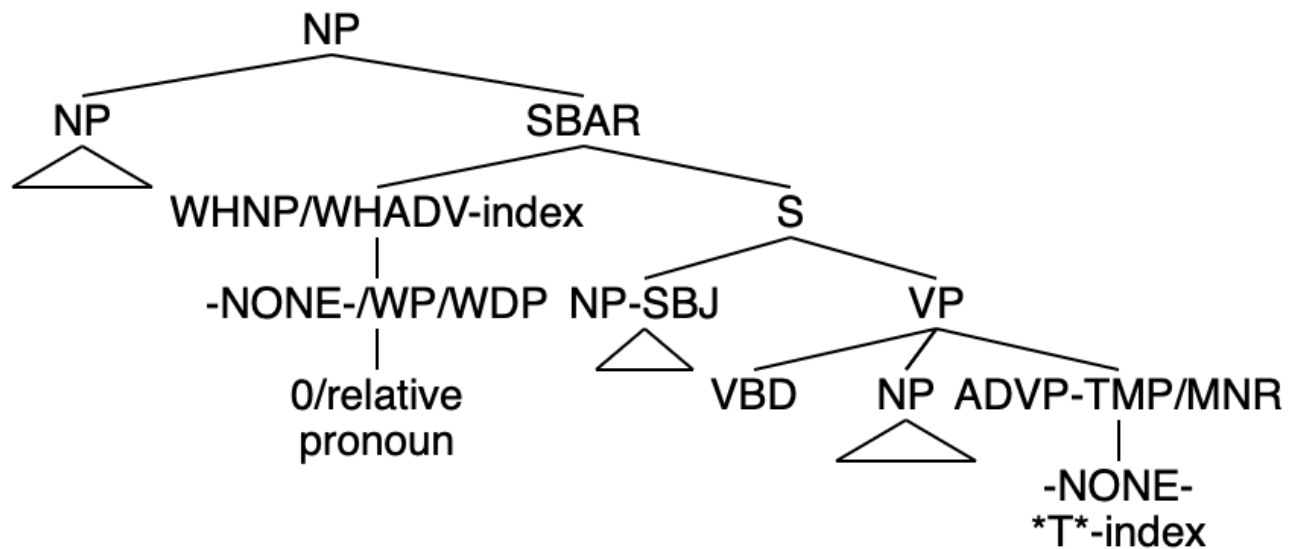
- relative pronoun *which*/WDT (WHNP)

Example: object relative clause



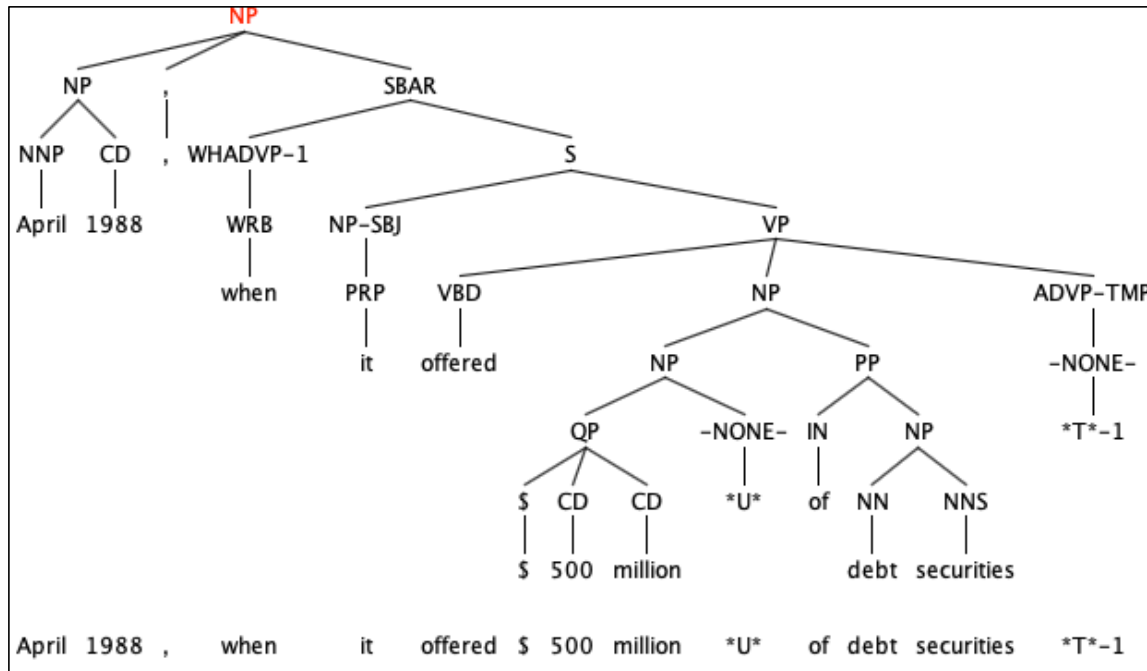
Relative Clause Examples

- from adjunct positions (not subject, not object)



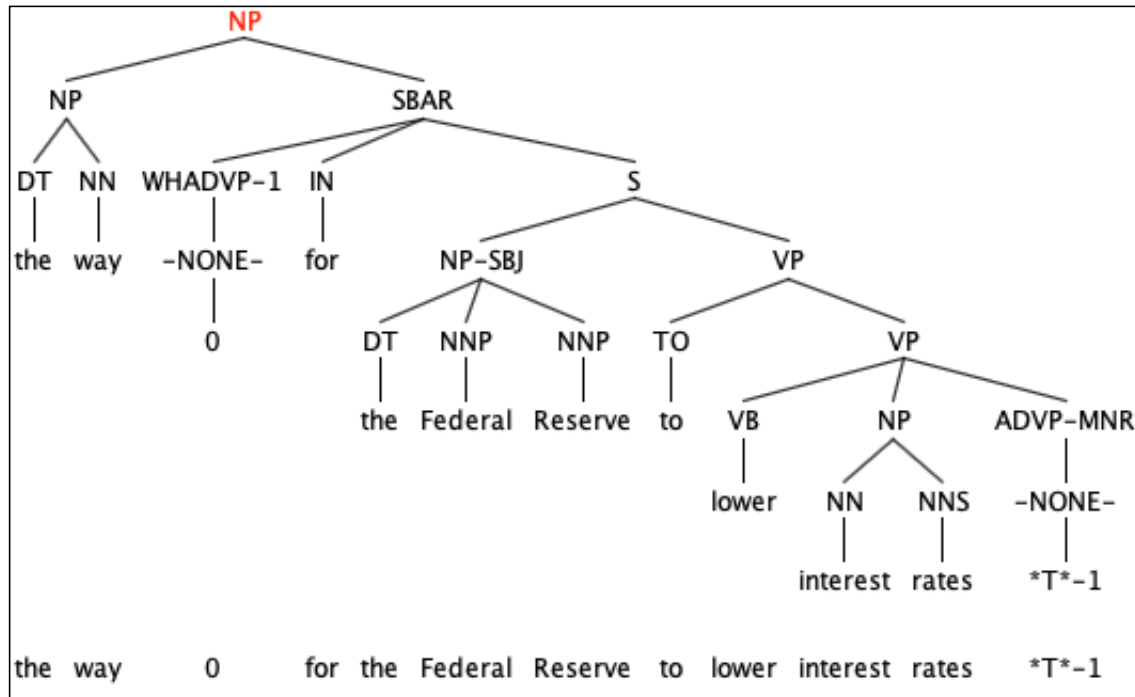
Example: adjunct relative clause

adjunct finite relative clause **with** a relative pronoun *when*/WRB (WHADVP)



Example: adjunct relative clause

adjunct nonfinite relative clause **without** a relative pronoun (zero WHADVP)



tregex = regex + structural relations

- combines regex with **structural relations** between tree node labels:
 - dominance (< or <<),
 - sisterhood (\$+ or \$),
 - precedence (. or . .) etc.
 - leftmost, nth, only, rightmost child of (<, or <*n* or <:*i* or <-)
- nodes can be matched:
 - literally (*exact match*),
 - @name, or
 - /.../ regex pattern
- groups from regex patterns can be linked via naming,
 - e.g. #1%index

@SBAR < /^WH.*-
([0-9]+)\$/#1#index
<< (__=empty < (/^-
NONE- / < ^*T*-([0-
9]+)\$/#1#index)

