

# Wh-Questions Tgrep2 Corpus Annotation Guide

Morgan Moyer

October 17, 2020

---

## 1 Macros

### 1. Disfluencies @ DISFL:

/EDITED|UH|PRN|-UNF/

### 2. Wh-Node @ WH:

/WP|WRB|WDT/ (The word-level wh-node)

!< /<sup>t</sup>/ (that doesn't terminate in a that/That)

!>> @DISFL (isn't dominated by a disfluency)

>> /<sup>S</sup>/ (but is dominated by an S node)

!> /<sup>N</sup>|<sup>A</sup>/ (isn't parented by any NPs or ADV/ADJPs)

## 2 Categorical Variables

### 1. WhPhraseType:

Does the Wh-head have a sister? (e.g., which children); 'monomorphemic' otherwise (e.g., who).

(a) Complex: @WH \$ /<sup>N</sup>|<sup>D</sup>|<sup>J</sup>|<sup>RB</sup>|<sup>RP</sup>|<sup>PP</sup>/

(b) Monomorphemic: @WH !\$ /<sup>N</sup>|<sup>J</sup>|<sup>RB</sup>|<sup>PP</sup>/

### 2. ModalPresent: (yes/no): is there a modal auxiliary (can, could, shall, should, may, might, must) in the question?

(a) Yes: @WH . /MD/

(b) No: @WH !. /MD/

### 3. Finite (yes/no): Does the question have a tensed (non-modal) verb? If yes = Finite, if not= Infinitive.

(a) Finite: @WH [!. /MD/ | [!. (/^-NONE-/ . /TO/)]

(b) Infinitival: @WH . (/^-NONE-/ . /TO/)

### 4. QuestionType: what kind of syntactic construction is the question appearing in? The most important ones are "Embedded" and "Root"

(a) Embedded: @WH

>> (/^SBAR/ [\$ /<sup>VB</sup>|<sup>JJ</sup>/] | [> /<sup>PP</sup>/ \$ /<sup>JJ</sup>|<sup>VB</sup>/])

!>> (/SBAR-NOM/ [!\$ /<sup>V</sup> |<sup>JJ</sup>/] | [!> (/<sup>PP</sup>/ \$ /<sup>VB</sup>|<sup>JJ</sup>/)])

- (b) Root: @WH  
 >> (/^SQ|^SBARQ/ [!\$ / ^VB/ | !> / ^N/])  
 !>> /SBAR-NOM/
- (c) Relative: @WH  
 >> (/^S/ [> / ^N/ | > (/^P/ > / ^N/)])  
 !>> (/SBAR-NOM/ [\$ / ^VB|^JJ/] | [> (/^PP/ \$ / ^VB|^JJ/)])
- (d) embadjunct: @WH  
 >> (/^SBAR/ \$ / ^VP|^AD/ !\$ / ^VB|^JJ/)
- (e) adjunct: @WH  
 [>> (/^SBAR/ [> (/^PP/ !\$ / ^VB|^JJ/)] | [!\$ / ^VB|^JJ/])]
- (f) nominalized: @WH  
 >> (/SBAR-NOM/ [!\$ / ^VB|^JJ/] | [!> (/^PP/ \$ / ^VB|^JJ/)])
- (g) subject: @WH  
 [ >> /SBAR-SBJ/ | >> (/^S/ >> /TOP/)]
- (h) fragment: @WH  
 >> (/^SBAR/ > /FRAG/)
- (i) exclam: @WH  
 >> (/^S/ !<< / ^V/)  
 !>> / ^SBAR-NOM/

## 5. DegreeQ

- (a) Yes: @WH \$ / ^JJ|^RB/
- (b) No: @WH !\$ / ^JJ|^RB/

## 6. WhAll

- (a) Yes: @WH >> (/^WH/ << (/^NP/ < / ^DT/ !< / ^N/))

## 7. SubjectAux:

- (a) Yes: @WH >> (/^SBAR/ << / ^SQ/)

# 3 String Variables

1. **MatVerb:** Print all the matrix predicates who have the wh-clause as complement.  
 @WH >> (/^S/ \$ / ^VB|^JJ/=print)
2. **MatVerbPart:** Print the matrix predicates who have questions as complement to PP (e.g., surprised by, agree on, depend on).  
 @WH >> (/^S/ (/^PP/ \$ / ^VB|^JJ/=print))
3. **Wh:** Print the wh-word terminal node.  
 @WH < / ^w|^h/=print
4. **Modal:** Print any modals following the @WH node.  
 @WH . /MD/=print

5. **Verb1:** Print the first verb in the question clause.  
@WH >> (/^S/ [<< /<sup>VP</sup>/ << /<sup>VB</sup>/=print])
6. **Verb2:** Print the second verb in the question clause.  
@WH >> (/^S/ [<< /<sup>VP</sup>/ [< /<sup>VP</sup>/ < /<sup>VB</sup>/=print]])
7. **DeterminerSubject:** Print any determiners in the NP-SBJ.  
@WH >> (/^S/ << (/DT/=print >> /<sup>NP-SBJ</sup>/))
8. **DeterminerNonSubject:** Print any determiners in the VP.  
@WH >> (/^S/ << (/DT/=print >> /<sup>VP</sup>/))
9. **FullWhPhrase:** Print the full WH-Phrase.  
@WH > /<sup>W</sup>/=print \$ /<sup>N</sup>|<sup>J</sup>|<sup>RB</sup>/
10. **WhParse:** Print the Wh-phrase with POS.
11. **Sentence:** Print the full sentence in which the @WH node appears.  
@WH >> (\*=print !<sub>i</sub> \*)
12. **Question:** Print the full Question.  
@WH >> /<sup>S</sup>/=print
13. **QuantifiedPredicate:** Print any quantifiers in the question predicate.  
@WH  
>> (/<sup>WH</sup>/ > (/^S/ << (/<sup>VP</sup>/ << /all|every|some|one|any|most|least|more|most|much/=print)))
14. **QuantifiedSubject:** Print any quantifiers in the question subject.  
@WH  
>> (/<sup>WH</sup>/ > (/^S/ << (/<sup>NP-SBJ</sup>/ << /all|every|some|one|any|most|least|more|most|much/=print))))