

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)

!< /[^]t|[^]T/ (that doesn't terminate in a that/That)

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

>> /[^]S/ (but is dominated by an S node)

!> /[^]N|[^]A/ (isn't parented by any NPs or ADV/ADJPs)

>> /[^]WH/

2 Categorical Variables

1. WhPhraseType:

Returns "complex" if @WH has a sister (e.g., which children); returns "monomorphemic" otherwise (e.g., who).

(a) Complex: @WH \$ /[^]N|[^]D|[^]J|[^]RB|[^]RP|[^]PP/

(b) Monomorphemic: @WH !\$ /[^]N|[^]J|[^]RB|[^]PP/

2. ModalPresent:

Returns "yes" if there is a modal auxiliary (can, could, shall, should, may, might, must) following @WH (where we can find coffee); Returns "no" if there isn't any.

(a) Yes: @WH >> (/[^]WH/ \$ (/[^]S/ << /MD/))

(b) No: @WH !>> (/[^]WH/ \$ (/[^]S/ << /MD/))

3. Finite

Returns "finite" if the question contains a tensed (=non-modal) verb (where they found coffee); returns "infinite" if the question contains an infinitival clause (where to find coffee)

(a) Finite:

@WH >> (/[^]WH/ !\$ (/[^]S/ !\$ /[^]VB/ < (/[^]NP/ < /-NONE-/ \$. (/[^]VP/ < /TO/))))

(b) Infinitival:

@WH >> (/[^]WH/ (/[^]S/! /[^]VB/ < (/[^]NP/ < /-NONE-/ \$. (/[^]VP/ < /TO/))))

4. QuestionType

Categorizes the kind of wh-question. We are mostly interested in Root (Where can we find coffee?) and Embedded (Dana knows where we can find coffee), and possibly Embedded Adjuncts. Embedded Adjuncts may look on the surface like Embedded Clauses, the difference only being whether the wh-clause is a complement to a verb or an adjunct.

CRITICAL QUESTION TYPES

(a) Root

Wh-clauses that have interrogative illocutionary force (e.g., Where can I find coffee?).
Are not complements to adjectives/verbs or children of Ns

@WH

>> (/^SQ|^SBARQ/ [!\$ / ^VB|^JJ|^BE/ | !> / ^N/])

!>> /SBAR-NOM/

(b) Embedded

Wh-clauses that are complements (sisters) to a verb (VB) or an adjective (JJ), OR that are embedded in PP which are complements of verbs or adjectives.

@WH

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

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

(c) EmbAdjunct

Wh-clauses that are not complements to a verb/adjective, but linearly follow the verb/adjective (i.e., that look like embedded clauses from a linear perspective).

@WH

>> (/^SBAR/ \$ / ^VP|^AD/ !\$ / ^VB|^JJ|^BE/)

NON-CRITICAL QUESTION TYPES

These are tags meant to weed out spurious root/embedded questions.

(a) Relative

The Wh-clause modifies a noun. (Dana saw the boy who ate Captain Crunch). @WH

>> (/^S/ [> / ^N/ | > (/ ^P/ > / ^N/)])

!>> (/SBAR-NOM/ [\$ / ^VB|^JJ|^BE/] | [> (/ ^PP/ \$ / ^VB|^JJ|^BE/)])

(b) Adjunct

Wh-clauses that are neither complements of verbs/adjectives, nor the children of PPs that are complements of verbs/adjectives @WH

[>> (/^SBAR/ [> (/ ^PP/ !\$ / ^VB|^JJ|^BE/)] | [!\$ / ^VB|^JJ|^BE/])]

(c) Nominalized

(this is now redundant with the above) Nominalized wh-clauses that are not complements.

@WH

>> (/SBAR-NOM/ [!\$ / ^VB|^JJ|^BE/] | [!> (/ ^PP/ \$ / ^VB|^JJ|^BE/)])

(d) Subject

Wh-clauses that are subjects.

@WH

[>> /SBAR-SBJ/ | >> (/^S/ >> /TOP/)]

(e) Fragment

Wh-clauses that are fragments. @WH

>> (/^SBAR/ > /FRAG/)

(f) Exclam

Wh-clauses that are exclamatives (don't have a verb), e.g., what a meal! @WH

>> (/^S/ !<< /^V/)

!>> /^SBAR-NOM/

5. DegreeQ

Returns 'yes' if the wh-clause is a degree-question (how many cups of salt do we need?); returns 'no' otherwise.

(a) Yes: @WH \$ / ^JJ| ^RB/

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

6. WhAll

Returns 'yes' if the wh-phrase contains an all; returns 'no' otherwise.

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

7. SubjectAux

Returns 'yes' if the wh-clause contains subject-aux-inversion.

(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 wh-clause 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 wh-clause.

@WH >> (/ ^S/ [<< / ^VP/ << / ^VB/=print])

6. Verb2: Print the second verb in the wh-clause.

@WH >> (/ ^S/ [<< / ^VP/ [< / ^VP/ < / ^VB/=print]])

7. DeterminerSubject: Print any determiners in the wh-clause NP-SBJ.

@WH >> (/ ^S/ << (/DT/=print >> / ^NP-SBJ/))

8. DeterminerNonSubject: Print any determiners in the wh-clause VP.

@WH >> (/ ^S/ << (/DT/=print >> / ^VP/))

9. FullWhPhrase: Print the full WH-Phrase.

@WH > / ^W/=print \$ / ^N| ^J| ^RB/

10. WhParse: Print the Wh-phrase with POS.

11. Sentence: Print the full sentence in which the @WH node appears.

@WH >> (*=print !; *)

12. **Question:** Print the full wh-clause.

```
@WH >> /^S/=print
```

13. **QuantifiedPredicate:** Print any quantifiers in the wh-clause predicate.

```
@WH
```

```
>> (/^WH/ > (/^S/ << (/^VP/ << /all|every|some|one|any|most|least|more|most|much/=print)))
```

14. **QuantifiedSubject:** Print any quantifiers in the wh-clause subject.

```
@WH
```

```
>> (/^WH/ > (/^S/ << (/^NP-SBJ/ << /all|every|some|one|any|most|least|more|most|much/=print)))
```