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)
   !< /ˆt|ˆT/ (that doesn't termininate in a that/That)
   !≫ @DISFL (isn't dominated by a disfluency)
   ≫ /ˆS/ (but is dominated by an S node)
   !> /ˆN|ˆA/ (isn't parented by any NPs or ADV/ADJPs)
   ≫ /ˆWH/

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 [$ /ˆN|ˆD|ˆJ|ˆRB|ˆRP|ˆPP/ | ≫ (/ˆWH/ $ /ˆAD|ˆJJ|ˆNP|ˆPP/)]

   (b) Monomorphemic: @WH [!$ /ˆN|ˆD|ˆJ|ˆRB|ˆRP|ˆPP/ | !≫ (/ˆWH/ $ /ˆAD|ˆJJ|ˆNP|ˆPP/)]

2. **ModalPresent**:
   Returns "yes" if there is a modal auxiliary (can, could, shall, should, may, might, must) in the clause that is sister to the WH-phrase node that dominates the @WH (where we can find coffee); Returns "no" if there isn't any.

   (a) Yes: @WH ≫ (/ˆWH/ ≫ /ˆS/ $ (/S|SQ/ ≪ /MD/))

   (b) No: @WH !≫ (/ˆWH/ ≫ /ˆS/ $ (/S|SQ/ ≪ /MD/))

3. **HaveNeed**:
   Returns 'yes' if there's a have to or need to construction. Returns 'no' otherwise.

   (a) Yes: @WH ≫ (/ˆWH/ $ (/ˆS/ ≪ (/ˆVP/ ≪ /need|needs|needed|have|has|had|'ve|'d/) ≪ /TO/))

   (b) No: @WH !≫ (/ˆWH/ $ (/ˆS/ ≪ (/ˆVP/ ≪ /need|needs|needed|have|has|had|'ve|'d/) ≪ /TO/))

4. **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 ≫ (/ˆWH/ !$ (/ˆS/ !$ /ˆVB/ < (/ˆNP/ < /-NONE-/ $. (/ˆVP/ < /TO/))))

(b) **Infinitival**:
@WH ≫ (/ˆWH/ (/ˆ*S*/! /ˆVB/ < (/ˆNP/ < /-NONE-/ $. (/ˆVP/ < /TO/))))

5. **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 QUESTIONTYPES

(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
≫ (/ˆSQ|ˆSBARQ/ [!$ /ˆVB|ˆJJ|ˆBE/ | !> /ˆN/])
!≫ /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
≫ (/ˆSBAR/ [$ /ˆVB|ˆJJ|ˆBE/] | [> /ˆPP/ $ /ˆJJ|ˆVB|ˆBE/])
!≫ (/SBAR-NOM/ [!$ /ˆV |ˆJJ|ˆBE/] | [!> (/ˆPP/ $ /ˆVB|ˆJJ|ˆBE/)])

(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
≫ (/ˆSBAR/ $ /ˆVP|ˆAD/ !$ /ˆVB|ˆJJ|ˆBE/)

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 Crunsh). @WH
≫ (/ˆS/ [> /ˆN/ | > (/ˆP/ > /ˆN/)])
!≫ (/SBAR-NOM/ [$ /ˆVB|ˆJJ|ˆBE/] | [> (/ˆPP/ $ /ˆVB|ˆJJ|ˆBE/)])

(b) **Adjunct**
Wh-clauses that are neither complements of verbs/adjectives, nor the children of PPs that are complements of verbs/adjectives @WH
[≫ (/ˆSBAR/ [> (/ˆPP/ !$ /ˆVB|ˆJJ|ˆBE/)] | [!$ /ˆVB|ˆJJ|ˆBE/])]

(c) **Nominalized**
(this is now redundant with the above) Nominalized wh-clauses that are not complements.
@WH
≫ (/SBAR-NOM/ [!$ /ˆVB|ˆJJ|ˆBE/] | [!> (/ˆPP/ $ /ˆVB|ˆJJ|ˆBE/)])

(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/

6. **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/

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

   (a) Yes: @WH ≫ (/^WH/ ≪ (/^NP/ < (/^DT/ < /all/) !< /^N/))

8. **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|^BE/=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|^BE/=print))

3. **Wh**: Print the wh-word terminal node.
@WH < /^w|^h/=print

4. **Modal**: Print any modals following the @WH node.
@WH ≫ (/^WH/ ≫ /^S/ $ (/S|SQ/ ≪ /MD/=print))

5. HaveNeed: Print any have to / need to modal constructions @WH ≫ (/^WH/ $ (/^S/ ≪ (/^VP/ ≪ /need|needs|needed|have|has|had|'ve|'d/=print) ≪ /TO/))

6. **Verb1**: Print the first verb in the wh-clause.
@WH ≫ (/^S/ [≪ /^VP/ ≪ /^VB|^JJ|^BE/=print])

7. **Verb2**: Print the second verb/predicate in the wh-clause.
   @WH $\gg$ (/^S/ [$\ll$ /^VP/ [< /^VP/ < /^VB|^JJ|^BE/=print]])

8. **Verb3**: Print the third verb/predicate in the wh-clause.
   @WH $\gg$ (/^S/ [$\ll$ /^VP/ [$\ll$ /^VP/ [< /^VP/ < /^VB|^JJ|^BE/=print]]])

9. **DeterminerSubject**: Print any determiners in the wh-clause NP-SBJ.
   @WH $\gg$ (/^S/$\ll$ (/DT/=print $\gg$ /^NP-SBJ/))

10. **DeterminerNonSubject**: Print any determiners in the wh-clause VP.
    @WH $\gg$ (/^S/ $\ll$ (/DT/=print $\gg$ /^VP/))

11. **FullWhPhrase**: Print the full WH-Phrase.
    @WH $\gg$ (/^WH/=print $ /S|SQ/ $\gg$ /^S/ $\gg$ /^N|^J|^RB|^PP|^A/)

12. **WhParse**: Print the Wh-phrase with POS.

13. **Sentence**: Print the full sentence in which the @WH node appears.
    @WH $\gg$ ($*$=print !> $*$)

14. **Question**: Print the full wh-clause.
    @WH $\gg$ /^S/=print

15. **QuantifiedPredicate**: Print any quantifiers in the wh-clause predicate.
    @WH
    $\gg$ (/^WH/ > (/^S/ $\ll$ (/^VP/
    $\ll$ /all|every|some|one|any|most|least|more|most|much/=print)))

16. **QuantifiedSubject**: Print any quantifiers in the wh-clause subject.
    @WH
    $\gg$ (/^WH/ > (/^S/ $\ll$ (/^NP-SBJ/
    $\ll$ /all|every|some|one|any|most|least|more|most|much/=print)))