"Who what to whom said?": An Argument from Bulgarian against Cyclic WH-Movement[1]

Catherine Rudin
Indiana University

In Bulgarian, all WH-words must move to COMP.[2] When a sentence contains two or more WH-words, multiple WH-movement occurs and results in COMP containing more than one WH-word:

1. Koj kŭde e otišŭl?   'Who went where?'
   who where went
2. Čudja se   koj kakvo na kogo e kazal.   'I wonder who said
   wonder-1sg who what to whom said   what to whom.'

If any of the WH-words fails to move to COMP, the result is ungrammatical (except, in some cases, as an "echo" question):

3. *Koj e otišŭl kŭde?
4. *Čudja se koj e kazal kakvo na kogo.

Similar multiple WH-movement occurs in Polish and Russian, among other languages; see Wachowicz and Sinicyn.

This data is interesting in itself, since it constitutes a superficial violation of Chomsky and Lasnik's "doubly filled COMP filter" (their filter 53) and other claims about the impossibility of doubly filled COMP, eg. Kayne's statement (p.122) that the structure [COMP WH-phrase WH-phrase] is universally excluded by the Empty Category Principle, and Chomsky's (1980, p.5) very similar statement that this structure is impossible "because of the c-command requirement for bound anaphora". In this paper, however, our major concern is the evidence which this data provides bearing on the competing analyses of WH-movement as successive cyclic (COMP-to-COMP) or as unbounded movement over a variable. These two analyses differ in their predictions with respect to syntactic islandhood in a language which allows multiple WH-movement to COMP, so Bulgarian provides a good test case for chosing between them.

If WH-movement is successive cyclic, as argued by Chomsky (1977 and elsewhere), one would expect WH-"islands" not to be islands in a language which permits two or more WH-words in COMP, as Bulgarian does. On this theory, the existence of WH-islands in languages like English is attributed to the fact that, in order to move "across" a WH-COMP, a WH-word must move into that COMP at an intermediate stage of the derivation:

5. a) *Whom do you wonder who saw?

   b) [$_\bar{S}$ [$_{COMP_1}$ whom$_i$] [$_S$ you wonder [$_\bar{S}$ [$_{COMP_2}$ who$_j$] t$_j$ saw t$_i$]]]

This derivation is blocked because of the prohibition on two WH-words in COMP; whom cannot move into (and subsequently out of) COMP$_2$, since COMP$_2$ is already occupied by a WH-phrase. Movement of whom directly to COMP$_1$ without passing through COMP$_2$ is ruled out by subjacency. The "island" status of embedded questions is thus

attributed to the same factors which make *Who whom saw or *Whom who saw ungrammatical, namely the inability of COMP to contain more than one WH-phrase.

Since Bulgarian does allow $[_{COMP}$WH-phrase WH-phrase$]$, it should freely permit extraction from clauses headed by a COMP containing WH. Contrary to this prediction, however, it is in most cases not possible to question out of an embedded question (introduced by a WH-word or the WH-complementizer dali 'whether'). WH-islands behave much as in English: compare (6) and (7) to (8), which has a non-WH complementizer, and note that the grammaticality judgements in most cases are the same as those of the English glosses.

6. *Kogo se čudiš koj e vidjal? '*Whom do you wonder who saw?'
   whom wonder-2sg who saw-3sg
7. *Koj kogo se čudiš dali e vidjal? {'*Whom do you wonder whether who saw?
                     whether   {'*Who do you wonder whether saw whom?}
8. Koj kogo misliš če e vidjal? {'Who do you think [that] saw whom?'
       think-2sg that   {'*Whom do you think [that] who saw?'}

The ungrammaticality of (7) is not due to the multiple movement. Movement of a single WH-word over dali is also bad:

    9. *Koj se čudiš dali e došůl? '*Who do you wonder whether came?'
    10. *Kogo se čudiš dali e vidjal? 'Whom do you wonder whether (he) saw?'
but:11. Koj misliš če e došůl? 'Who do you think [that] came?'
    12. Kogo misliš če e vidjal? 'Whom do you think [that] (he) saw?'

Conversely, it has been noted (eg. for Italian by Rizzi[3] and for Hebrew by Reinhart) that some languages which do not normally allow multiple WH-words in COMP nonetheless can violate at least some WH-islands. Rizzi explains this by a language-specific relaxation of subjacency to permit non-COMP-to-COMP WH-movement as long as the movement crosses no more than one S boundary. Reinhart argues for an analysis of both Hebrew and Italian utilizing an extra COMP position, which is never attested on the surface in Italian and only very rarely and marginally in Hebrew, and whose only function is to act as an "escape hatch" for the extraction of WH-words from WH-clauses. This double COMP, as Reinhart herself points out in section 4.3, allows only two WH-words to move out of a single clause. As we have seen (example (2)), in Bulgarian three WH-words can be fronted, and in fact four or perhaps even more can. Although it is hard to find a context in which a sentence like (13) is pragmatically normal, it probably is grammatical.

    13. Koj kakvo na kogo ot kogo kůde misliš če e predal?
        who what to whom from whom where think-2sg that delivered
        'Who do you think delivered what to whom from whom where?'

Reinhart suggests (section 4.1) that this sort of construction may be due to scrambling rules and not to WH-movement at all. I will return to this possibility below, but for the present I will assume all of

the fronted WH-words have been placed in COMP (a phrase-structure rule such as (14) having provided room for them) and will show that adherence to WH-islands can in fact still be accounted for.

14. $[_{COMP} +WH] \Longrightarrow +WH*$ (ie. I suggest the structure of COMP is:)

$$[_{COMP} +WH]$$
$$+WH \quad +WH \quad +WH$$

Both Rizzi and Reinhart assume that WH-movement is subject to subjacency and moves only COMP-to-COMP (although Rizzi bends the definition of subjacency a bit). It is this assumption which forces Reinhart to either postulate one COMP for every element moved or to fall back on "scrambling". An alternative worth considering is that of unbounded WH-movement of the type argued for by Bresnan, which makes no predictions concerning islandhood based on the number of WH-words allowed in COMP. The existence of WH-islands is entirely independent of the form of COMP in a particular language, on this theory, since WH-movement over a COMP does not involve movement into that COMP at any stage in the derivation. In place of (5b) the un-bounded movement analysis has (15), that is, movement of a WH-word directly to COMP across a variable, without intermediate stops:

15. $[_{\bar{S}} [_{COMP_1} whom_i] [_S \text{ you wonder } [_{\bar{S}} [_{COMP_2} who_j] t_j \text{ saw } t_i]]]$

This derivation is ruled out (and sentence (5a) thereby starred) by conditions on the form of possible variables, that is, on what can intervene between the constituent to be moved and the position it moves to. One such condition which seems to work quite well for the problem at hand is the Variable Interpretation Convention (VIC) proposed by Wilkins. Very informally stated, the VIC says that no rule may move a constituent over a variable containing an element of the same category as any of the "crucially affected terms" of that rule. A crucially affected term is anything which must be stated in the rule; variables are assumed between all other terms, but are not stated.

If WH-movement is formulated as in (16), that is, "WH moves into

16.   COMP - WH-phrase $\Longrightarrow 2 - \emptyset$

COMP over a variable", movement over either another WH-phrase or another (filled)[4] COMP should be ruled out by the VIC since both are categories crucially involved in the rule. This is, however, not true for Bulgarian, since the complementizer če (which, unlike English that, is obligatorily present) does not block movement. For examples with če, see sentences 8, 11, and 12. This can be remedied by stating WH-movement with COMP instead of simply COMP as the first term of the rule: +WH

17. COMP
    +WH   - +WH-phrase $\Longrightarrow 2 - \emptyset$

Če in the variable string will not now block movement, since it is neither a +WH-COMP(lementizer) nor a +WH-phrase. This formulation of the rule has the additional advantage of correctly predicting that embedded questions are islands only to questioning and not to relativization, which moves a non-interrogative ("-WH") phrase into a -WH COMP. The relative clause in (18) is perfectly grammatical, while the very similar question (19) is not.

18. Tova e knigata, kojato ne znam   koj e kupil.
    this is book-the which neg know-1sg who bought
    '?This is the book which I don't know who bought'
19. *Koja kniga ne znaeš   koj e kupil
    which book neg know-2sg who bought
    '*Which book don't you know who bought?'

Given (20) as the relativization rule, I propose (21) as the general WH-movement rule for both questions and relatives in Bulgarian.[5]

20. $\begin{matrix} \text{COMP} \\ -\text{WH} \end{matrix}$ – –WH-phrase $\Longrightarrow$ 2 – $\emptyset$

21. $\begin{matrix} \text{COMP} \\ \alpha\text{WH} \end{matrix}$ – αWH-phrase $\Longrightarrow$ 2 – $\emptyset$

Note that the same analysis will work for Italian as well, since in Italian it is only relativization that can violate the islandhood of embedded questions. Rizzi attributes the unacceptability of questioning out of a question to some undefined "independent property of the interrogative construction" in Italian. Under the VIC analysis it is accounted for automatically if the WH-movement rule in Italian is the same as in Bulgarian. The only difference between Italian and Bulgarian WH-movement is that Italian lacks multiple filling of COMP, i.e. it lacks anything like rule (14).

It is worth noting that relativization out of embedded questions is easier than questioning out of them, not only in Bulgarian and Italian, but also in Scandinavian (according to Reinhart, who cites Maling) and even English. I find (22) considerably better than (23) although admittedly not wonderful. (cf. also (18,19) glosses)

22. ?The only house which I don't know who lives in is that
    green one.
23. *Which house don't you know who lives in?

Thus the dichotomy often assumed between languages which absolutely do and absolutely do not obey WH-island constraints appears rather exaggerated, and differences between interrogative and non-interrogative WH-movement in this respect seem to be quite common.

Hebrew as described by Reinhart does seem genuinely different from the other languages considered here, since it allows not only relativization but also questioning out of WH-islands. I am not sure how this can or should be handled within the unbounded movement/VIC framework. Perhaps the statement of the rule of WH-movement is different in Hebrew, or perhaps, as indeed seems likely, Wilkins'

VIC in its present form is not entirely correct or universal. I am not arguing specifically for the VIC or any particular constraint on movement over a variable, rather, I am arguing more generally that, since the crucial connection between the form of COMP and the existence of WH-islands in a given language predicted by the successive cyclic WH-movement hypothesis does not hold, at least for Bulgarian, a more promising approach to explaining WH-islands is through unbounded WH-movement with universal and/or language specific constraints on how and where such movement can apply. Such an analysis can treat Bulgarian and English WH-islands (which are very similar in spite of the multiple movement in Bulgarian) in a unified way, while the successive cyclic analysis has to give two unrelated accounts for this phenomenon: English WH-islands are explained by subjacency and the single WH-position in COMP, while Bulgarian islands must receive some other type of explanation, what is not clear.

I will now return to a crucial problem which has been mentioned in passing but not dealt with seriously up to this point, namely the question of whether the clause-initial WH-words are actually in COMP. It is at least possible that some or all of them are in some other position, having either been generated in place in regular NP or PP positions or moved there by "scrambling". This kind of analysis of multiple fronted WH-words has been suggested by Reinhart in general (see above) and by Sinicyn for Russian.

In Russian, multiple WH-fronting is apparently possible only within a single clause, not out of embedded clauses; (24) and (25) are Sinicyn's (2) and (5):

24. Kto, komu, čto podaril?
    who to whom what presented
25. *Kakoj fil'm, kogda, gde Petr skazal, my uvidim
    what movie when where   said   we shall see

Sinicyn's solution is WH-placement by (non-configurational) base generation of WH-phrases clause-internally <u>and</u> successive cyclic WH-movement across clause boundaries. This will not work for Bulgarian, clearly, since multiple WH-movement out of a clause is possible: (26) is the Bulgarian equivalent of (25).[6]

26. Koj film koga kaza Petŭr če šte gledame?
    what film when said   that will watch-2pl
    'What film did Peter say we would see when?'

A Bulgarian sentence can have up to three preverbal "slots": TOPIC, COMP, and FOCUS, in that order:

27. $[_{TOPIC}$ Ivan$][_{COMP}$ dali $][_{FOCUS}$ na kino$]$ šte hodi?
        Ivan   whether   to the movies will go
       'Is Ivan going to go to the movies?'

The most plausible position for WH-words aside from COMP, and at first sight a rather attractive possibility, is that they are in the FOCUS

position. This is especially true since when a WH-word (or words) is present, no other focussed element is normally possible. Any preverbal NP must be TOPIC, not FOCUS:

    28. a) Ivan kakvo pravi?   'What is Ivan doing?
            what does-3sg
        b) *Kakvo Ivan pravi?[7]

However, this analysis has some problems. First, if non-cooccurrence is taken as evidence of occupying the same "slot", then exactly the same argument can be used to suggest that the fronted WH-words are in COMP: just as WH-words do not cooccur with a(nother) focussed constituent, they also do not cooccur with complementizers in true questions. Thus this line of argumentation is fruitless.

    Secondly, the complementizer _dali_ 'whether' (which I assume by definition to be in COMP) behaves in a somewhat similar manner with respect to FOCUS. (29a) is an entirely normal question. (29b) is not ungrammatical, but also is not a direct question; rather, it expresses musing or wondering with no expectation of an answer.

    29. a) Ivan dali e otišŭl?   'Did Ivan go?
            whether went-3sg
        b) Dali Ivan e otišŭl?   'I wonder whether Ivan went."

What fills the FOCUS "slot" is semantically controlled, as its name implies. Since WH-words and,to a lesser extent, _dali_ are the natural semantic focus of a sentence in which they appear, it is not surprising that they do not normally cooccur with focussed NPs - and it also proves nothing about the syntactic position of WH-words.

    Thirdly, if WH-words are always in argument positions (FOCUS or otherwise), it is unclear to me how the distinction between "real" and "echo" questions can be maintained. If the WH-word in an echo question such as (30b) is in an argument position, while the WH-word in a real question like (30a) is in COMP, the basis for the difference in interpretation is clear.

    30. a) Pitaha kogo si vidjal.   'They asked who you saw'
          asked-3pl whom saw-2sg
        b) Pitaha dali kogo si vidjal? 'They asked whether you saw
              whether   who?'

    The case for COMP as the surface position of WH-words in Bulgarian is thus at least as strong as the case against, although neither can be claimed to be conclusive at this point. Given such a weak argument for any other position, it seems best to analyze WH-words as being in COMP, if for no other reason than that COMP is so widely assumed to be the recipient of WH-words in other languages. The only real reason to suspect otherwise in Bulgarian is that it violates pronouncements about the "universal" impossibility of a multiply-filled COMP, which is a circular argument at best. I therefore continue to assume that all WH-words in true questions and

relatives in Bulgarian are moved to COMP, and that the resulting arguments against cyclic WH-movement are valid.

Finally, I would like to point out that even if it should turn out that some or all of these WH-words are not in COMP, it would not constitute an argument for successive cyclic movement. COMP-to-COMP movement obviously cannot place WH-words anywhere other than in COMP, so they would have to be either generated in place or fronted by some other type of rule, and the WH-island facts would still have to be accounted for in some way other than subjacency, probably by a constraint on the (variable) string over which a WH-word can be associated with a gap in a lower clause.

NOTES

[1] This paper is based on research done at the Bulgarian academy of sciences and supported by a grant from the International Research Exchange Board.

[2] Or, more neutrally, they all occur in some more or less clause-initial (but post-topic) position. I assume at least for the time being that their appearance in this position is a result of WH-movement placing them in COMP.

[3] As we shall see below, Rizzi's WH-island constraint violations do not really "count" as violations, since they involve only relativization and not movement of interrogative WH-phrases.

[4] Optional or empty nodes do not count for purposes of the VIC, according to Wilkins.

[5] This formulation would predict that relativization over the -WH complementizer če should be disallowed by VIC. i.e. that a sentence like (i) should be ungrammatical:

(i) Tova e šapkata, kojato kazvat če Ivan e otkradnal.
    this is hat-the which say-3pl that   stole
    'This is the hat which they say that Ivan stole.'

I have not been able to check this with a native speaker, and I am not sure if it is good or not.

[6] The second WH-adverb of the Russian model is impossible in Bulgarian unless it is in a conjoined structure with the first one: Koj film koga i kŭde ... 'Which film when and where...'. This is a very general restriction on WH-adverbs and has no bearing on the point under discussion.

[7] Some speakers will accept (28b) with very strong stress or for "talking to a child".

REFERENCES

Bresnan, Joan. 1977. "Variables in the theory of transformations" in Culicover, Wasow, and Akmajian, eds. Formal Syntax. New York: Academic Press. pp. 157-196.

Chomsky, Noam. 1977. "On Wh-movement" in Culicover, Wasow, and Akmajian, eds. Formal Syntax. New York: Academic Press. pp. 71-132.

_____. 1980. "On binding" Linguistic Inquiry 11.1:1-46.

_____ and Howard Lasnik. 1977. "Filters and control" Linguistic Inquiry 8.3:425-504.

Kayne, Richard S. 1981. "ECP extensions" Linguistic Inquiry 12.1: 93-134.

Reinhart, Tanja. "A second COMP position" (to appear in the proceedings of the Pisa Colloquium on Markedness, Anali della Scuola Normale Superiore - Pisa.)

Rizzi, Luigi. "Violations of the WH-island constraint in Italian and the subjacency condition" (to appear in Italian Linguistics.)

Sinicyn, Michael. "On one base parameter and some of its consequences for Russian grammar" (paper read at GLOW 1981 meeting; abstract in GLOW 1981 newsletter.)

Wachowicz, Kristina A. 1974. "Against the universality of a single WH-question movement" Foundations of Language 11:155-166.

Wilkins, Wendy. 1980. "Adjacency and variables in syntactic transformations" Linguistic Inquiry 11.4:709-758.