The Bulgarian Relative Marker *-to*

Catherine Rudin

1. Introduction

Relative pronouns and other relative *wh*-words in Bulgarian are distinguished from interrogative *wh*-words by a suffix, *-to*, as can be seen in the chart in (1).

(1) **Bulgarian *wh*-words**

<table>
<thead>
<tr>
<th>who (m)</th>
<th>koj</th>
<th>koj*to</th>
</tr>
</thead>
<tbody>
<tr>
<td>who (f)</td>
<td>koja</td>
<td>koja*to</td>
</tr>
<tr>
<td>who (n)</td>
<td>koe</td>
<td>koe*to</td>
</tr>
<tr>
<td>who (p)</td>
<td>koi</td>
<td>koi*to</td>
</tr>
<tr>
<td>whom</td>
<td>kogo</td>
<td>kogo*to</td>
</tr>
<tr>
<td>what/which/what kind (n)</td>
<td>kakvo</td>
<td>kakvoto</td>
</tr>
<tr>
<td>which/what kind (m)</td>
<td>kakâv</td>
<td>kakûv*to</td>
</tr>
<tr>
<td>which/what kind (f)</td>
<td>kakva</td>
<td>kakva*to</td>
</tr>
<tr>
<td>which/what kind (p)</td>
<td>kakvi</td>
<td>kakvi*to</td>
</tr>
<tr>
<td>whose (m)</td>
<td>čij</td>
<td>čij*to</td>
</tr>
<tr>
<td>whose (f)</td>
<td>čija</td>
<td>čija*to</td>
</tr>
<tr>
<td>whose (n)</td>
<td>čie</td>
<td>čie*to</td>
</tr>
<tr>
<td>whose (p)</td>
<td>čii</td>
<td>čii*to</td>
</tr>
<tr>
<td>where</td>
<td>küde</td>
<td>küde*to</td>
</tr>
<tr>
<td>when</td>
<td>koga</td>
<td>koga*to</td>
</tr>
<tr>
<td>how</td>
<td>kak</td>
<td>kak*to</td>
</tr>
<tr>
<td>how much/how many</td>
<td>kolko</td>
<td>kolko*to</td>
</tr>
</tbody>
</table>

*A version of this paper was presented at the 2008 Slavic Linguistics Society meeting in Columbus, Ohio. I would like to thank Wayles Browne, not only for his comments at that meeting, but for all the wise suggestions and eye-opening questions he has given me on other work over the decades.*

Steven Franks, Vrinda Chidambaram, and Brian Joseph, eds. *A Linguist's Linguist: Studies in South Slavic Linguistics in Honor of E. Wayles Browne*. Bloomington, IN: Slavica, ##–##.
This suffix occurs obligatorily on all and only relative \textit{wh}-words. \textit{Wh}-words without the -\textit{to} suffix are used in questions, both main-clause and embedded, while those with -\textit{to} form headed and headless relative clause constructions. The two are never interchangeable:

(2) a. \textbf{Koj imam vreme?} \hspace{1cm} (*\textit{kojto})
   \begin{quote}
   who has time
   ‘Who has time?’
   \end{quote}

b. \textbf{Ne znam koj ima vreme.} \hspace{1cm} (*\textit{kojto})
   \begin{quote}
   NEG know_{1PL} who has time
   ‘We don’t know who has time.’
   \end{quote}

c. Čovek, \textbf{kojto ima vreme...} \hspace{1cm} (*\textit{koj})
   \begin{quote}
   person who-\textit{to} has time
   ‘A person who has time...’
   \end{quote}

d. \textbf{Kojto ima vreme (može da dojde)...} \hspace{1cm} (*\textit{koj})
   \begin{quote}
   who-\textit{to} has time can to come
   ‘Whoever/he who has time...’
   \end{quote}

These facts are straightforward and well known, stated as a simple descriptive generalization in virtually every grammar of Bulgarian. However, even complete reference grammars generally give no attention to -\textit{to} beyond simply noting its existence,\footnote{Scatton 1984, for instance, simply states “Relative pro-forms are formed from interrogative pro-forms with the addition of the postponed particle -\textit{to}’ (154). Hauge 1999 is similarly brief: his entire comment on relative pronouns is “Most of them are formed by adding -\textit{to} to the interrogative pronouns” (59) plus a short discussion of the one exception, the stylistically marked (old-fashioned, bookish, and/or dialectal) alternate word for ‘what,’ \textit{što}, which among other idiosyncracies has no suffixed form.} and to the best of my knowledge the -\textit{to} suffix has never been discussed from a generative perspective. The goal of this paper is to investigate the status of -\textit{to} in more depth. In section 3 I propose an analysis of relative -\textit{to} as a marker of syntactic status, namely as head of a projection that attracts [-interrogative] \textit{wh}-words to its Spec position. In essence I propose to treat -\textit{to} as a relative complementizer. Section 4 introduces supporting and complicating evidence from two multiple-\textit{wh}- relative constructions, and section 5 discusses some remaining problems. In section 2,
before introducing the complementizer analysis, I briefly dispense with an appealing but incorrect alternate analysis of -to as a marker of definiteness.

2. Rejected Analysis: -to Is Not an Article or Definiteness Marker

The -to suffix superficially resembles the postpositive definite article; compare (1) and (3):

(3) selo ‘village’ vs. selo to ‘the village’

It is tempting to identify relative -to as a definiteness marker, closely related if not identical to the definite article. I have made this assumption myself in earlier work (Rudin 1986), and the idea has a long history in Bulgarian studies, going back at least to Ivanov 1894.2 Cyxun 1981 divides Balkan Slavic dialects into those whose relative pronouns are “articled” (členie), that is, suffixed with -to, versus “un-articled” (nečlenie), with no -to. The assumption is reasonable on general linguistic grounds, since question words are presumably indefinite, while relative pronouns can be considered definite; see eg., Bianchi 1999, Bhat 2004, Šimík 2007. Relative pronouns are historically derived from definite determiners or demonstratives in many languages worldwide, indicating a clear crosslinguistic link between relative pronouns and definiteness; see e.g., Andrews 1975. A few examples are given in (4) of languages in which some relative pronouns incorporate at least an etymological definite article or in which the relative pronouns are actually identical to definite articles:

(4) a. Relative Incorporates Definite Article

<table>
<thead>
<tr>
<th>Language</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romance</td>
<td>lequel, il quale, el cual</td>
</tr>
<tr>
<td>Modern Greek</td>
<td>o opios</td>
</tr>
<tr>
<td>Archaic English</td>
<td>the which</td>
</tr>
<tr>
<td>Albanian</td>
<td>i cili</td>
</tr>
</tbody>
</table>

2 Ivanov is quoted in Cyxun 1981 (108) as saying “I believe that this is the very same -to [as the article] albeit invariable in all numbers and genders” (Translation mine).
(4) b. *Relative = Definite Article
   German:  **der**, etc.
   Arabic:  **al-** ?? (see Belkacemi 1999 for critical discussion)

Bulgarian-internal evidence for a distinction in definiteness between bare and suffixed *wh*-words can be found in the use of bare *wh*-words as indefinite pronouns, as in (5a). This usage is quite restricted even for bare *wh-* but completely impossible for *wh*-to: see (5b).

(5) a. Njamam **koj** da mi pomaga.
   "not-have\textsubscript{1SG} who to me help"
   ‘I don’t have anyone (=lit. who) to help me.’

b. *Njamam **kojto** da mi pomaga.

However, appealing though the idea may be, *-to* is not the definite article. First, *-to* lacks the gender and number agreement of the article, appearing always in what looks like the neuter singular article form, even when accompanying non-neuter *wh*-words. The *wh*-words which are marked for gender or number are repeated in (6).

(6) koj\textit{to} / kojat\textit{to} / ko\textit{eto} / ko\textit{ito}  \quad \text{‘who’ (relative)}
   \begin{array}{cccc}
   & M / & F / & N / & PL \\
   kakvi\textit{to} / kak\textit{vato} / kakov\textit{to} / kak\textit{vito} & \text{ ‘which’ (relative)}
   \end{array}
   \begin{array}{cccc}
   & M / & F / & N / & PL \\
   \end{array}

The invariant ending of these *wh*-words is quite different from the true definite article, which has four distinct forms:

(7) masculine  **-\textalpha** or **a**:  \begin{array}{ccc}
   & xlab ‘bread’ / & xlab\textalpha, xlab\textbeta ‘the bread’
   \end{array}
   \begin{array}{ccc}
   & kola ‘car’ / & \text{kolata ‘the car’}
   \end{array}
   \begin{array}{ccc}
   & kafe ‘coffee’ / & \text{kafeto ‘the coffee’}
   \end{array}
   \begin{array}{ccc}
   & knigi ‘books’ / & \text{knigite ‘the books’}
   \end{array}

\textsuperscript{3} The two masculine forms are nominative vs. non-nominative case in literary Bulgarian; in colloquial Bulgarian this distinction is lost.
If the relative -to suffix were the definite article we would expect unattested forms like those in (8) instead of the actual forms in (6).

(8) *kojāt / *kojata / *koite  
    M / F / PL
*kakāvāt / *kakvata / *kakvite  
    M / F / PL

‘who’ (relative)

‘which’ (relative)

Secondly, if wh-words did in fact combine with a definite article, we could expect them to occur also with other determiners such as possessives, demonstratives, or the indefinite article edno. None of these determiners are actually found with wh-words, however. Compare the range of determiners found with a noun in (9a) with the impossibility of any of them in (9b).

(9) a. deteto mi  ‘my child’
    tova dete  ‘that child’
    edno dete  ‘a child’

b. *koj(to) mi  ‘my who’
*tova kūde  ‘that where’
*edno kakvo  ‘a what’

Finally, no pro-forms other than relative wh-words occur either with invariant -to or with an agreeing definite article, so it would be at least unexpected for relative pronouns to occur with an article. The starred forms in (10) show that personal pronouns, demonstrative pronouns, and adverbial pro-forms cannot take an article or the invariant -to suffix; this suffix is found only on relative wh-words.

(10) toj ‘he’  *tojto / *tojāt
    tja ‘she’  *tjato / *tjata
    onezi ‘those’  *onezito / *onezite
    tam ‘there’  *tamto
    togava ‘then’  *togavato
It is, in short, abundantly clear that -to is not the definite article, regardless of their etymological connection.\(^4\) One might be tempted to argue that it is still some type of morphological definiteness marking; after all, definiteness can certainly be shown in other ways than through a definite article. However, this also does not hold water: -to does not always or reliably indicate definiteness. Even indefinite relative clauses take -to; in (11b), for instance, we have kojto in a situation where it is not even specific, much less definite; the speaker is referring to any chair, not a certain specific chair.\(^5\) The wh-word must have the -to suffix in this example, just as in the definite relative in (11a).

(11) a. stolât na kojto sedna
chair\(_{DEF}\) on which sat
‘the chair on which s/he sat’

b. stol na kojto da sedna
chair on which to sit
‘a chair on which to sit’

3. Proposal: -to Heads a Relative Clause Projection

I take it as established that -to is not a marker of definiteness. In fact, it appears to be, as the traditional grammars imply, simply a relative clause marker. But what exactly does this mean? I suggest that, rather than being just a bit of morphology attached to wh-words, the -to suffix has independent syntactic status as the head of a non-interrogative wh-clause; that is, it is essentially a relative complementizer.

One hint that this analysis might be right comes from the very closely related Macedonian, whose relative marker, optionally added to most interrogative wh-words, actually has the form of a complementizer. The main wh-words, interrogative and relative, are shown in (12).

\(^4\) Relative -to and the definite article do derive from the same source, a Common Slavic demonstrative pronoun. See Mladenova 2007 (241–43) for discussion and references on this historical development.

\(^5\) A reviewer points out that the same is true for French lequel, Italian il quale, etc. I would argue that this indicates these forms also are not necessarily definite, their etymological connection to the definite article notwithstanding.
(12) Macedonian *wh*-words.\(^6\)

<table>
<thead>
<tr>
<th></th>
<th>interrogative</th>
<th>relative (što is optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>who, which</td>
<td>koj</td>
<td>kojšto</td>
</tr>
<tr>
<td>whom</td>
<td>kogo</td>
<td>kogo što</td>
</tr>
<tr>
<td>what kind</td>
<td>kakov</td>
<td>kakov što</td>
</tr>
<tr>
<td>whose</td>
<td>čij</td>
<td>čijšto</td>
</tr>
<tr>
<td>where</td>
<td>kade/kaj</td>
<td>kade što, kaj što</td>
</tr>
<tr>
<td>how</td>
<td>kako</td>
<td>kako što</td>
</tr>
<tr>
<td>how much, how many</td>
<td>kolko</td>
<td>kolku što</td>
</tr>
<tr>
<td>how big</td>
<td>kolkav</td>
<td>kolkav što</td>
</tr>
</tbody>
</table>

Relative *wh*-words are formed “by adding što, written without any space after koj and čij (hence, kojšto), and written as a separate word after all others” (Friedman 2001). This što element appears to be the same word as the relative complementizer word što, so that a Macedonian relative clause like (13a–b) has the same structure as a Middle English one like (13c). This structure is presumably that shown in (14), with the *wh*-phrases in SpecCP, immediately followed by što or *that* in C.

(13) a. čovekot [so kogo što se šetaše včera]
    person\(_{DEF}\) with whom that REFL walked yesterday
    ‘the person with whom he walked yesterday’
    (slightly modified from Friedman 2001)

    b. čovekot [kojšto mislam deka te videl tebe]
    person\(_{DEF}\) whom-that think\(_{1SG}\) that you saw you
    ‘the person that I think saw you’
    (slightly modified from Friedman 2001)

    c. I am she [which that saved hath youre lyf].
    (Middle English (14th century) from Allen 1980)

\(^6\) *Wh*-words shown in the table are those which can occur with što. Koga ‘when’ and što ‘what’ do not take the relative ending što. Gender and number marked forms do take što: kojašto ‘whoj’ etc. The conditions under which suffixed što occurs are not clear to me; grammars of Macedonian, including Friedman 2001, state simply that it is optional.
Let us suppose that Bulgarian also has a structure like this; that is, let us hypothesize that -to is merged in the head of the projection which attracts fronted wh-words. Given that Bulgarian is a true wh-movement language, this projection is CP. We could see -to as spelling out the features which distinguish relative from interrogative wh-clauses, features which are located in C. The structure of a Bulgarian relative clause will then be (15), identical to (14) except that the material in C ends up as a suffix and not an independent word.

(15)  
```
(14)  
CP
  /\      /\                  /
 so kogo koj which C'
     C0 \   /                     /
       TP -to                     
       that
```

The claim is thus essentially that -to is a complementizer. Not surprisingly, it is in complementary distribution with other complementizers, especially the relative complementizer deto, as seen in (16). One possibility would be to see -to in fact as an allomorph of deto; the relative complementizer appears as deto when alone and as -to when accompanying a wh-word.7 Under this analysis, the resumptive pronoun which regularly appears in deto relatives (gi in (16a)) could be seen as corresponding to the wh- portion of the wh+to word in (16b) and other wh- relatives.

---

7 This is not the source of -to historically; in fact, deto itself derives from the wh-word (with -to suffix) meaning ‘where’: kūdeto > gdeto > deto.
(16) a. filmite, deto gi gledahme movies_{DEF} that them watched_{1PL}
   ‘the movies that we watched’

b. filmite, koito gledahme movies_{DEF} which watched_{1PL}
   ‘the movies which we watched’

c. *filmite, koito deto (gi) gledahme movies_{DEF} which that them watched_{1PL}

d. *filmite, koito deto (gi) gledahme movies_{DEF} which that them watched_{1PL}

Other complementizers, declarative če and interrogative dali, also
do not cooccur with -to. Not only would they be in competition for a
single C slot, they also conflict in clause type.

4. Single and Repeating -to in Multiple Relatives

The analysis of -to as a clausal head is supported by evidence from a
perhaps unexpected quarter: an otherwise odd fact about multiple wh-
constructions. In spite of the clear pattern we saw in (2), that bare wh-
words in Bulgarian are normally only interrogative, there is one situ-
tion in which bare wh-words, without -to, can unexpectedly function as
relative pronouns. This one anomalous case is multiple wh- free relatives
and correlatives. In the examples of multiple correlatives in (17)
and (18); notice that both of the wh-words can have the -to suffix, or the
first wh-word can be unsuffixed.

(17) a. [Koj kakvoto iska], da vzeme.
   who what-to wants to take
   ‘Let everyone take what they want’ (pair-list)

b. [Kojto kakvoto iska] da vzeme.
   who-to what-to wants to take
   ‘Whoever wants whatever, let them take it’ (single-pair)
(18) a. [Kogo kakvoto go boli], za nego prikazva.\textsuperscript{8}
whom what-to him hurts about it talks
‘Everyone talks about whatever is hurting them.’ (proverb)
(pair-list)

b. [Kogoto kakvoto go boli], za nego prikazva.
whom-to what-to him hurts about it talks
‘The person who has something hurting, talks about it.’
(single-pair)

These two options are not quite synonymous. The difference between the (a) examples with initial bare \textit{wh}- and (b) with -to on both \textit{wh}-words is one of pair-list vs. single-pair interpretation.\textsuperscript{9} So, (17a) refers to a set of people, each of whom wants something, possibly a different thing in each case, while (17b) refers to a single unknown person who wants a certain thing. Similarly (18a) implies multiple people, one with a headache, another with a broken heart, and so on, while (18b) refers to a single individual. These judgments are subtle, but fairly robust.

How does this fact help us come to grips with the syntax of the -to suffix? An obligatory pair list reading in multiple questions is well known to correlate with movement of \textit{wh}- to SpecCP. Thus, multiple questions in English (with \textit{wh}-movement to SpecCP) normally elicit a pair-list answer. Single-pair answers are normal only in languages with no overt \textit{wh}-movement, for example, Japanese. Among Slavic languages, single-pair answers are possible in, e.g., Serbo-Croatian, which arguably fronts \textit{wh}-words to a focus position instead of SpecCP in most contexts (see Grebenyova 2006, Bošković 2003, and sources cited there, especially Hagstrom 1998).\textsuperscript{10} This suggests that the (a) and

\textsuperscript{8}Some speakers prefer \textit{koi}(to) instead of \textit{kogo}(to) in (18a, b), perhaps for reasons connected with the ongoing replacement of \textit{kogo} with \textit{koi} in colloquial Bulgarian; the remnants of the case distinction are collapsing much as for English who/whom. The clitic \textit{go} is obligatory with the experiencer verb \textit{boli} and clitic-doubles the accusative \textit{kogo}.

\textsuperscript{9}A reviewer suggests the distinction could also be seen in terms of indefinite vs. specific interpretation, an idea I leave for further research.

\textsuperscript{10}Hagstrom proposes that multiple questions are interpreted via a Q morpheme which is an existential quantifier moved to C; Bošković adopts this
(b) examples in (17–18) differ in the position of the wh-words; specifically that the (a) versions have the wh-words in SpecCP, while the (b) versions do not. In other words, single and repeating -to are in different positions.

This line of thought is immediately reminiscent of the well-known typology of multiple wh-movement first sketched out by Rudin 1988 and refined by large numbers of researchers over the last two decades. Some multiple wh-fronting languages place their wh-words all in SpecCP, while others place only one or even none in SpecCP, fronting the rest to some other left-periphery position such as SpecFocusP or Spec of some other functional projection. There is some evidence that different constructions within a language can have different wh-positions. Bulgarian normally (which is to say, in multiple questions) fronts all wh-words to SpecCP. Perhaps the construction in (17b) and (18b) shows that Bulgarian is not monolithic in this regard: in this one construction Bulgarian places its wh-words in a non-SpecCP position, a configuration more typical of languages like Polish or Serbian/Croatian.\footnote{Just as an aside, the ability to place some or all wh-words in a position other than SpecCP may account for the ability for at least some Bulgarian speakers to place adverbs and parentheticals between the wh-words of a multiple question, albeit not as freely as in some other languages.}

There are two multiple relative constructions in Bulgarian, multiple free relatives and multiple correlative. These constructions differ in their external syntax, that is, their relation to the matrix clause. Correlatives are always sentence-initial, external to the main clause and linked to it by a coreferential pronominal or demonstrative, the “correlate.” Free relatives are inside the matrix clause and function as DP or AdvP within it. These characteristics are sketched in (19). (See Rudin 2007a for more detailed discussion.)
(19) **Two types of (multiple) headless relative constructions**

<table>
<thead>
<tr>
<th>Correlative</th>
<th>Free Relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ [\text{CP wh wh ...}], [\text{CP ... CORRELATE}, ...] ]</td>
<td>[ [\text{CP ... {DP/AdvP [\text{CP wh wh ... }} ... ] } ]</td>
</tr>
<tr>
<td>• adjoined to CP: left peripheral position</td>
<td>• occupies argument or ad-verbal position</td>
</tr>
<tr>
<td>• not dominated by any XP</td>
<td>• dominated by DP or AdvP</td>
</tr>
<tr>
<td>• coreferential pro-form in main clause</td>
<td>• no coreferential pro-form</td>
</tr>
</tbody>
</table>

Correlatives and free relatives also differ in their ability to have `-to` on one or both *wh*-words. Both patterns (*wh-to wh-to* and *wh wh-to*) are found in multiple correlatives like those in (17–18), but only the second pattern is possible in free relatives. An example of a multiple *wh*- free relative is given in (20), showing that this construction cannot have repeated `-to`.\(^\text{12}\) As we might expect, this construction has only the pair-list reading; i.e., the reading which normally goes with a single `-to` at the end of the *wh*-word string.

(20) a. Vzemajte **koj** kakvoto može. (pair-list)
    take who what can
    ‘Let everyone take whatever they can’

b. *Vzemajte **kojto** kakvoto može.

The chart in (21) summarizes possibilities for both kinds of multiple relatives.

---

\(^{12}\) One Bulgarian speaker reports that (20b) is good for her, but most speakers I have consulted do not accept it, and in internet searches I have found the repeating `-to` pattern only in correlatives.
(21) **Possible -to configurations, readings, and wh-positions in Bulgarian multiple relatives**

<table>
<thead>
<tr>
<th>Multiple Correlative</th>
<th>Multiple Free Relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CP wh wh-to ... ]</td>
<td>[CP wh wh-to ... ]</td>
</tr>
<tr>
<td>pair list</td>
<td>pair list</td>
</tr>
<tr>
<td>wh’s in SpecCP</td>
<td>wh’s in SpecCP</td>
</tr>
<tr>
<td>[CP wh-to wh-to ... ]</td>
<td>single pair</td>
</tr>
<tr>
<td>wh’s lower than C</td>
<td></td>
</tr>
</tbody>
</table>

I have argued elsewhere (Rudin 2007a; Rudin 2008) that the multiple free relative construction (unlike multiple correlative) is possible only in languages which place at least one wh-word in SpecCP, while multiple correlatives also occur in languages like Polish which have no true wh-movement to SpecCP. This dovetails nicely with the suggestion that single -to is possible only with wh- in SpecCP.

Returning to the idea that -to is a spelling-out of features associated with relative wh-, consider how this meshes with the pattern in (21). Suppose that the features spelled out by -to are located in whatever functional head(s) attract wh-words in a given language and construction, whether C or another projection such as FocusP. When this projection has multiple specifiers, as is the case for CP in languages of the Bulgarian type (the so-called +MFS languages of Rudin 1988), a single instance of -to will occur at the end of the string of wh-phases occupying those specifiers, as in (22a) or perhaps (22b). Both configurations have been proposed as the structure of multiple fronted wh-constructions in Bulgarian (by Richards 2001 and Rudin 1988, respectively). The choice between the two structures in (22), though obviously important for other purposes, is immaterial for the analysis I am proposing here; in either case, both wh-words are in a Spec-head relation with a single C0, resulting in a surface appearance of -to suffixed only to the last wh-word. This, I suggest, is the configuration of multiple free relatives and also of multiple correlatives with pair-list reading.
When two instances of the -to suffix appear, that is, when each wh-word has its own -to, the wh-words are in Spec of separate projections, and the head of each projection is spelled out as -to, as in (23). This is the structure of multiple correlatives with single-pair reading.

Notice that this analysis provides an explanation for the impossibility of -to distribution other than on the last of a group of wh-words or on each wh-. In particular, a multiple relative cannot have -to only on the first wh- or no -to at all; compare (24) to (16).
(24) a. *[Kojto kakvo iska], da go vzeme.
   who what wants to it take
b. *[Koj kakvo iska], da go vzeme
   who what want to it take

Assuming that C (and the “X” and “Y” heads in (22), on which more shortly) cannot be null, i.e., that the relative wh-attracting heads are obligatory spelled out as -to, and given that specifiers are to the left of the head, there will always at least be a -to at the end of the wh-string.

5. Remaining Problems

Although the analysis of -to as head of a wh-attracting projection accounts nicely for several aspects of its syntax, including the single vs. repeating -to of multiple relative constructions, it also raises some issues which remain unsolved at this point. The first such problem is the identity of the projections labelled XP and YP in (23). One of these could be a focus projection; wh-fronting is often said to be attraction to Focus. Perhaps the other is TP or perhaps another discourse-related category. I leave this as an open question for now; what matters for this analysis is that they must be two separate projections with separate heads. One other point which is clear is that neither XP nor YP can be CP for Bulgarian sentences like (17b) and (18b), since movement of even one wh- to CP should force the pair-list reading.\(^\text{13}\)

A second unresolved issue raised by this analysis of -to is how to deal with wh-phrases like those in (25). In such multi-word wh-phrases -to is suffixed to the wh-word, not to the whole phrase, as we might expect if the wh-phrase occupies the Spec position of the head that is spelled out as -to.

(25) a. kojato kniga
    (*koja knigato)
    ‘which book’

\(^\text{13}\) However, it is possible that languages other than Bulgarian might have a multiple wh correlative structure with one wh in SpecCP (and pair-list reading). I have not investigated interpretation of multiple correlatives in other multiple fronting languages.
(25) b. kakvoto kamâče (kakvo kamûče) 
   ‘what kind of little stone’

This position of -to is reminiscent of the position of other DP clitics in Bulgarian, namely, the definite article and possessive clitics, which also suffix to a modifier rather than the entire phrase. This is essentially a type of second-position cliticization within the nominal phrase. (The precise rule for placement of these clitics is more complex, but need not concern us here.)

(26) a. novata kniga (nova knigata) 
   ‘the new book’

b. mladoto ti dete (mlado deteto ti) 

Apparently a similar second-position placement applies in relative wh-phrases, though the exact mechanism for raising -to into the wh-phrase remains to be explored. Phrase-internal -to occurs also in multiple relatives, in sentences like those in (27). Under the present analysis these must have a structure something like (28), with -to somehow cliticized into the 2nd position of the preceding wh-phrase.

(27) a. Sâsedite bjaha si otmûkvali komu kakvoto 
   neighborsDEF had REFIt carried-off whomDAT which 
   kamûče potrjabvalo.
   stoneDIM was.necessary
   ‘The neighbors had carried off whichever little stone each one needed’ (Daskalov)

b. Vzemi otkâde kojato kniga iskaš.
   take from-where which book want2SG
   ‘Take whatever book you want from wherever.’
(28)  
```
      CP
     /\  /
    komu CP
   /  \ /  
kakvo kamâče C'
   \  |
     |
    C'  -to  TP
```

A cliticization analysis of some kind is independently supported by the fact that -to behaves like a clitic in terms of morphology in any case. As noted by Pashov 1999, it attaches to the end of the fully inflected word, outside of all gender, number, and case inflections:

(29)  
```
kj+a+to   which
kav+i+to   what.kind
ko+go+to   whom
```

This morpheme order is highly unexpected if -to were a derivational suffix; in that case it should attach to the root with inflectional affixes further out, as in the starred forms in (29). But attaching to the fully inflected word is entirely normal for a clitic. It is thus not surprising that it has the same type of syntactic mobility as other phrase-internal clitics, i.e., roughly 2nd place within its phrase, even if the apparent movement required is somewhat problematic.

It is worth noting that the same problem applies in Macedonian, where što, if present, follows the \textit{wh}-word in complex \textit{wh}-phrases just as -to does in Bulgarian:

(30)  
```
a. ...kakov što čovek ...
    ...what.kind that man
    ‘which(ever) kind of man’

b. ...katolici vo čij što crkvi služali ...
    ...catholics in whose that churches serve
    ‘catholics in whose churches they serve’
```

[archives.vmacedonia.com/15285.htm]
If Macedonian što is actually the complementizer što, examples of this type pose the same conundrum as the Bulgarian phrases in (25): apparent movement of an element in C into a wh-phrase. One possibility is that in both languages the wh-phrase moves first to a position below C, perhaps SpecFocusP, and that the wh-word alone then moves further to SpecCP; this idea has problems of its own, however, and I leave the issue open for further research.

6. Conclusion

In this paper I have suggested an analysis of relative -to as a marker of syntactic status. Specifically I suggest that -to spells out features which identify a relative clause, perhaps something like [–interrogative, +wh]. These features are located in the head of projections which can attract the relative wh-word or phrase. In Bulgarian this projection is normally C: wh-words, including all the wh-words in a multiple-wh-construction, normally move to SpecCP. Thus -to is essentially a relative complementizer. Fronted wh-words in one construction, the multiple correlative, do not necessarily move to SpecCP, and those cases where they front to a non-CP projection are marked by the occurrence of multiple -to markers, one per projection (i.e., each head that attracts a relative wh-word is spelled out separately as -to).

Questions remain, obviously. One problem is the identity of the heads/projections involved in the multiple -to construction. Another is the mechanism for placing -to correctly in multi-word wh-phrases. But in spite of these issues I find the analysis appealing enough to be worth pursuing.

Some of the results of the paper hold regardless of whether the relative complementizer analysis turns out to be correct. In the first part of the paper I demonstrated that -to is clearly not an article, or even a definiteness marker. It is also clear that it is a clitic, and not a derivational suffix. Finally, whether we actually call it a complementizer or not, it clearly has connections to the complementizer što in Macedonian.

To end with a broader view, I close by noting that overlap of determiners and complementizers is known from languages as far afield as Siouan and other Native American language families; clause-typing morphology tends to merge with, develop from or into, and generally be hard to distinguish from nominalizing morphology and especially
definite determiners. So it is not surprising typologically that a particle which began life as a demonstrative and may at one stage have been an article seems to have turned into a complementizer or at least some kind of clause typing marker.

Wayne State College
carudin1@wsu.edu

References


——. (2007b) “Multiple vs. single -to in Bulgarian multiple wh-relatives”. Presented at SEESA meeting, Columbus, OH.

