Sorting out-to and što: Bulgarian and Macedonian Relative Markers*

Catherine Rudin Wayne State College

1. Introduction

Despite their very close relationship, Bulgarian and Macedonian are often surprisingly different in the details of their grammars. In this paper I examine one area of divergence: the status and behavior of their relative markers. Both languages have an invariant suffixed element which occurs only on relative *wh*words, not interrogative ones: -to in Bulgarian and *što* in Macedonian. Compare the corresponding questions and relative clauses in (1a-b) and (2a-b):

(1a)	Koj	govori?		Bulgarian	
	who	talks		_	
	'Who is ta	alking?'			
(1b)	Tozi	koj <u>to</u>	govori		
	that	who-to	talks		
(2a)	Koj	zboruva?		Macedonian	
	who	talks			
'Who is talking?'					
(2b)	Onoj	koj <u>što</u>	zboruva		
	that	who-što	talks		
'The one who is talking'					

At first glance the two languages appear completely parallel: questions contain a bare *wh*-word, while relative clauses contain a *wh*-word with the extension -*to* or *što*. However, -*to* and *što* actually behave quite differently, both in relative clauses and in other constructions, suggesting that they do not share the same syntactic status. Deeper theoretical questions of the syntactic identity of *što* and especially -*to* are the subject of ongoing work in progress, e.g. Franks and Rudin (to appear).

^{*} I would like to thank Ognen Vangelov for assistance with the Macedonian data, and the audiences of both SLS8 Szczecin and the 19th Biennial Conference on Balkan and South Slavic Linguistics, Literature and Folklore for helpful comments on earlier versions of this work.

In the present paper I simply describe the usage of *-to* and *što* and briefly discuss possible analyses of each.

I start by presenting the facts of *što* and *-to*in relative clauses in Macedonian (section 2) and Bulgarian (section 3). Section 4 expands the discussion to constructions other than relative clauses. In section 5 I consider two competing analyses of *što* and conclude that it is a complementizer. Section 6 concerns competing analyses of *-to*, whose correct analysis is much less clear than that of *što*: complementizer, definiteness marker, and other possible analyses are examined. Finally, section 7 is the conclusion.

2. Macedonian što in relative clauses

In Macedonian relative clauses, the *što* element is optional.In addition to *wh* word plus *što*, relative clauses can contain a *wh* word alone, with no suffix, or *što* alone:

(3a) čovekot koj<u>što</u> zboruva	'the man who is talking'
(3b) čovekot koj zboruva	'the man who is talking'
(3c) čovekot što zboruva	'the man that is talking'

All *wh* words can occur with *što*. Spelling conventions dictate that *što* be written as a single word with *koj* and *čij* and as a separate word with other *wh* words, but it is clearly the same element in all cases.

(4a) mestoto kadešto... 'the place where...'
(4b) čovek, kakvo što e tvojot tatko... 'a person, how your dad is' (like your dad)
(4c) Jovan, čijšto sin ti e student... 'Jovan, whose son is your student,...

Different types of relative clauses and different *wh* words vary in how readily they accept *wh+što* vs. plain *wh*. According to Tomić (2012), *što* is required with *kade*, *kako*, and *kolku* ('where, how, how-much') in headed relatives but not in free relatives and not with any other *wh* words; it is more common in relative clauses with pronominal than non-pronominal heads, and it is rare in free relatives. Use of *što* differs in restrictive vs. non-restrictive relatives and in colloquial vs. standard usage. Individual speakers' preferences vary, as do prescriptivists' prescriptions. The picture is quite complicated, but on the whole, *što* is optional: possible but not obligatory.

3. Bulgarian *-to* in relative clauses

Compared to the rather complex optionality patterns for *što*, the generalization for *-to* is simple: in Bulgarian, all *wh* relatives require *-to*. Non-*wh* relatives

formed with the relative complementizer *deto* exist, in colloquial Bulgarian, but relatives with unsuffixed *wh* are impossible; compare (5) to the Macedonian situation in (3).

(5a) čovekât kojto govori (the man who is talking'
(5b) *čovekât koj govori
(5c) čovekât deto govori (the man that is talking'

Like *što*, *-to* occurs with all *wh* words, but unlike *što*, it is always obligatory. This is true in free relatives as well as headed ones:

(6a) Vzemi kakvoto iskaš. 'Take what(ever) you like.'

(6b) *Vzemi kakvo iskaš.

4. -to and što in other constructions

In addition to relative clauses, several other constructions require *wh+to* in Bulgarian. Some of these also permit *što* in Macedonian, but others disallow it. In this section I briefly review the status of *-to/što* in five types of *wh*constructions. First, equative and comparative clauses in Bulgarian are formed with the *wh* word *kolko* 'how much', and always require *-to*:

(7a) brâmbar, goljam kolkoto beetle big how-much-to palm your..." 'a beetle as big as your palm' (7b) toi po-goljam, otkolkoto trjabva... than-how-much-to it is bigger to-us is-necessary 'it's bigger than we need'

In Macedonian equatives and comparatives the picture is mixed; clauses like (8a-b), with a noun phrase following the *wh* have bare *kolku*, but those with a compared verbal constituent like (8c-d) allow and even prefer *kolku što*.

- (8a) telefon, golem kolku tablet phone big how-much tablet 'a phone as big as a tablet'
- (8b) Našata kuća može da primi poveće gosti **otkolku** vašata. our house can to hold more guests than-how-much yours 'Our house can hold more guests than yours.'
- (8c) Trošam onolku pari **kolku <u>što</u>** zarabotuvam. spend_{ISG} so-much money how-much *što* earn_{ISG} 'I spend as much money as I make.'
- (8d) Poveke umraat **odkolku <u>što</u>** se raģaat. more die_{3PL} than how-much *što* are-born 'More are dying than are being born.'

Incorrelative clauses, Bulgarian always requires -to (9a-b), while in Macedonian *što* is optional and most often absent (10a-c).

	(9a)	Koj <u>to</u>	se uči,	toj	šte	spoluči.		
		who-to	study _{3SG}	he	will	succeed _{3SG}		
		'(The one) who studies, he will succeed.'						
	(9b)	Koga <u>to</u>	igraeš	za	udovolstvie,	togava	pečeliš.	
		when-to	play _{2SG}	for	satisfaction	then	win_{2SG}	
	'When you play for fun, (then) you win.'						200	
	(10a)	Koj	uči,	toj	ќе	nauči.		
		who	study _{3SG}	he	will	succeed _{3SG}		
	'(The one) who studies, he will learn.'							
	(10b)	Koga	igraš	za	zadovolstvo,	togaš	pobeduvaš.	
		when	play _{2SG}	for	satisfaction	then	win_{2SG}	
	'When you play for fun, (then) you win.'							
	(10c)	Koga	<u>što</u>	sakaš,	togaš	dojdi.		
		when	što	want _{2SG}	then	come _{2SG}		
'When you want, (then) come.'								

Non-interrogative multiple wh constructions in Bulgarian, including multiple correlatives and multiple free relatives, require -to on the second wh word and often allow it also on the other wh word(s), as seen in (11). In Macedonian neither wh allows $\dot{s}to$; the examples in (12) would be ungrammatical with $\dot{s}to$.

(11a)	Praštajte	[koj	kolko <u>to</u>	može].			
	send _{3PL}	who	how-much-to		can _{3SG}		
	'Everybody send as much as you can'						
(11b)	Koj(<u>to</u>)	kâde <u>to</u>	živee,	tam	se svikva.		
	who-to	where-to	live _{3SG}	there	gets-accustomed		
	'Everyone gets used to wherever they live.'						
(12a)	Koj	kolku	saka	neka	zboruva.		
	who	how-much	want _{3SG}	let	talk _{3SG}		
	'Let everyone talk as much as they like.'						

¹ Multiple free relatives with *što* can be found online; Elena Dimova (pc) provides the following examples:

It is hard to know what to make of examples like this, which are rejected by the speakers I have consulted; they may be grammatical for some speakers or simply typing mistakes. No examples with repeated *što* have been found; sequences like *kojšto kolku što* seem not to occur even in online social media.

i. a. Neka zbori koj kolku što saka. 'Let each speak as much as they like.'

b. Da se napieme koj kolku što može. Lets each get as drunk as we can.

c. ...ako ne gi održuvame nie samite žiteli sekoj **koj kade što** živee. '(Who will pay for mosque upkeep...) if we the inhabitants don't keep them up ourselves, each wherever s/he lives.'

(12b) **Koj kade** živee, tamu se naviknuva. who where live_{3SG} there gets-accustomed 'Everyone gets used to wherever they live.'

Universal Concessive Conditionals ("wh i da" clauses asserting the truth of the main clause, regardless of conditions) are another case in which-to is obligatory in Bulgarian, as in (13), but što does not (and uncontroversially cannot) occur in Macedonian; e.g. (14):

- (13) **Kâde<u>to</u>**i da otideš,... 'Wherever you go,...'
- (14) **Kade**i da odeš,... 'Wherever you go,...'

Finally, both Bulgarian and Macedonian have complex pronominal expressions based on a variety of *wh* words; these again require *-to* in Bulgarian (15) but disallow *što* in Macedonian (16).

- (15) **kojto** i da e /**kojto** i da bilo 'whoever it may be'/'someone or other'
- (16) **koj** da e / **koj** bilo / **koj**-gode 'whoever it may be'/'someone or other'

In short, -to is much more prevalent than što; it is obligatory in essentially all non-interrogative wh constructions.² In contrast, što is optional in relative clauses, severely limited in other non-interrogative wh clauses, and disallowed in non-clausal constructions (nominal comparatives, complex pronominals) as well as UCCs. With this brief summary of the data in mind, let us turn now to possible analyses of što and -to. The identity of Macedonian što is quite clear: it is a complementizer, the C head of a clause. The identity of Bulgarian -to is far less certain. I begin with the easier case, that of Macedonianšto.

5. Analysis: what is što?

Sto in Macedonian has several functions. In addition to the relative-marking use at issue here, it can be an interrogative *wh* pronoun (17a), a *wh* adjective (17b), or an indicative complementizer (17c) replacing the usual complementizer *deka* (or *oti*).³

² The one exception is the rather marginal use of *wh* words as indefinite pronouns with verbs of existence.

³ These examples are due to Tomić (2012).

- (17a) **Što** sakaš? **'What** do you want?'
- (17b) **Što** čovek e toj? 'What kind of person is he?'
- (17c) Se raduvam, **što** ve gledam. 'I am happy **that** I see you' (i.e 'to see you')

It seems likely that relative-marking *što* is the same as one of these; that is, it is either a *wh* word or a complementizer. Tomić (2012) assumesthe first of these options, that *što* in relative clauses is a relative pronoun, pointing out that all interrogative *wh* words can also be used as relative *wh*, and *što* is a *wh* word in questions. However, there are numerous reasons not to treat relative *što* as a *wh* word:

- A. *Što* occurs with *wh* words in relative clauses, as we have already seen: *wh*+*što*. This is very odd behavior if *što* is a relative pronoun, since in general two relative *wh* elements cannot cooccur: **wh*+*wh* for any "other" *wh* words.⁴
- B. *Što* occurs with all "other" *wh* words, but not with *što*. If *što* is a relative pronoun we might expect to find **štošto* alongside *kojšto*, *kade što*, *kakov što*, etc., but in fact it is, as noted by Kramer (1999), absolutely impossible.
- C. Other *wh* words cannot follow *što*. If *što* in combinations like *kojšto* and *kade što* is a *wh* word there seems no reason not to allow "double *wh*" combinations in the other order, like **što koj* or **što kade*.
- D. As seen in (17) above, *što* is an undoubted complementizer in other (non-relative) constructions, suggesting that *što* in relative clauses is the same complementizer.
- E. Finally, *što*cannot be object of a preposition.Both Tomić (2012) and Kramer (1999) point out the impossibility of examples like (18b); a *wh* word like *koja* is needed instead of (or in addition to) *što* after a preposition like *za*.
 - (18a) studentkata, **za koja**(**što**) zboruvame... 'the student about whom we speak'
 - (18b) *studentkata, **za što** zboruvame...
 *'the student about that we speak'

This closely parallels the facts of *that* as opposed to relative pronouns like 'who' in English, as seen in the equally ungrammatical gloss of (18b). Such data are widely recognized as indicating that English *that* is not a *wh* word, but a complementizer.

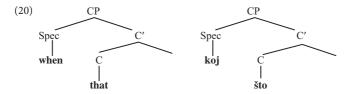
⁴ Except, of course, for multiple questions and other multiple wh constructions, which are irrelevant here.

Some of these arguments are weak; for instance, (B), the unacceptability of **štošto*, might well be due to a constraint on sequences of identical or homonymous words. However, taken all together they comprise overwhelming evidence that relative *što* is not a *wh* pronoun, but a complementizer.⁵ This is further confirmed by the fact that *što* occurs only in clausal environments (e.g. in clausal but not nominal comparatives). *Wh+što* is thus parallel to the Middle English [*Wh that*] relative construction, seen in the first line of Chaucer's Canterbury tales:

(19) **Whan that** Aprill, with his shoures soote, the droghte of March hath perced to the roote...

'When April with his sweet showers has pierced March's drought to the root...'

Here a *Wh*-word is followed by a complementizer, with structure (20). In both Middle English and Macedonian, the Spec (*wh* word) or the C head (complementizer *that* or *što*) or both can be overtly expressed, unlike in Bulgarian or Modern English, where one or the other must be null; the so-called 'doubly filled comp' filter.⁶



6. Analysis: what is -to?

Having concluded that *što* is a complementizer, I turn now to the more problematic issue of the identity of Bulgarian *-to*. Unlike *što*, *-to* never occurs as a

⁵ One reliable test of wh-word vs. complementizer status in Bulgarian does not apply in Macedonian. In Bulgarian, the relative complementizer *deto* is distinguished from *wh* words, which disallow clitic doubling, by the possibility of a resumptive pronoun. In Macedonian, the pervasive clitic doubling of definite DPs, including relative *wh* words, obscures this distinction. The doubling/resumptive clitic pronoun *go* is required in (i) regardless of whether the relativizer is *što* or a *wh* word.

⁽i) Čovekot što/kogo/kogošto go stretnavme... 'The man that/whom we met...'

⁶ An apparent problem for this view would be phrases like [kakovšto čovek] 'which person', claimed to exist by Tomić, with C after just part of the wh phrase. An analysis involving cliticization or split deletion could be possible, but the issue is moot since no Macedonian speaker I have consulted accepts such phrases.

complementizer in non-relative clauses⁷ and does not resemble a *wh* word. Ithas been treated as a definiteness marker (e.g. Izvorski (2000)), a modal element (Iliev (2011)), a specifically relative complementizer (Rudin (2009)), or simply a morphological mark of relative pronouns (Hauge (1999)). Each of these analyses raises troubling issues. There is general agreement that the *-tosuffix* on *wh* words is a relative marker in some sense, being used in all relative constructions. One could plausibly claim that all of the *wh-to* constructions surveyed in section 4 above are based on or parasitic on relative clauses. However, just what the formal content of the label "relative marker" is remains controversial. I consider three possible analyses here; that *-to* is a definiteness marker, a complementizer, or neither of the above:

6.1. The definiteness marker analysis

The view of -to as a definiteness suffix is often assumed uncritically. The idea is attractive because relative wh is generally said to be semantically definite, while interrogative wh is indefinite, and because-to looks looks identical to the neuter form of the postposed definite article: kafe/kafeto 'coffee/the coffee' cf. koj/kojto 'who_Q/who_{REL}. This resemblance is illusory, though, as I have argued at length in Rudin (2009); relative -to is NOT the definite article.

Relative *-to* is invariant, lacking the gender and number agreement of the article; compare (21a-b), with invariant relative *-to* even on gender-marked *wh* words, to the varying forms of the article in (22).

- (21a) koj<u>to</u> / koja<u>to</u> / koe<u>to</u> / koi<u>to</u> 'who (rel) m/f/n/p'
- (21b) kakâv<u>to</u> / kakva<u>to</u> / kakvo<u>to</u> / kakvi<u>to</u> 'which (rel) m/f/n/p'
- (22) xljab<u>ât</u>, / kola<u>ta</u>/ kafe<u>to</u> / knigi<u>te</u>

 'the bread (m)' /'the car (f)'/'the coffee (n)' /'the books (pl)'

It also differs from the definite article in not being replaceable by other determiners:

- (23a) dete<u>to</u> 'the child' **tova** dete 'that child' **edno** dete 'a child'
- (23b) kâde<u>to</u> 'where (rel)' *tova kâde 'that where' *edno kâde 'a where'

⁷ The normal indicative clausal complementizer isče, with *deto* occurring as a complementizer in relative clauses, though in some cases *deto* can also be synonymous with če (see e.g. Krapova 2010.)

Furthermore, pronouns other than relative *wh*-words never occur with either invariant *-to* or an agreeing article, as we might expect if e.g. *kojto* were actually a *wh* pronoun plus an article. For instance, personal pronouns like *toj* in (24b) cannot be suffixed.

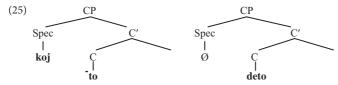
(24a) koj 'who' koj**to** (24b) toj 'he' *toj**to**/ *toj**ǔt**

Even though -to is not the definite article, it may still have a connection to definiteness, perhaps as an inflection instantiating definiteness without agreement features. Historically -to presumably comes from the same demonstrative element as the definite articles; this -t- element is still seen in the demonstratives tozi/tazi/tova/tezi,toja/taja/tija'this/that m/f/n/p'), pronouns toj/tja/to/te'he/she/it/they', and conjunctions like kato 'as' and nito 'neither'. But this does not mean definiteness is its current meaning or function. Analyses of wh constructions that depend on treating -to as definite face numerous challenges. Some of these involve contrast with the clear lack of a definite morpheme in Macedonian relative pronouns; for instance Izvorski's analysis of UCCs depends crucially on definite semantics supplied by -to, but Macedonian UCCs, in spite of lacking any suffix on the wh word, have the same properties as Bulgarian ones.

6.2. The Complementizer Analysis

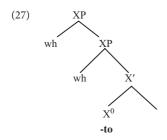
The idea that -to could be a relative complementizer is based partly on the parallelism of -to and Macedonian relative -štoseen in (1)-(2), partly on the similarity of -to to the undisputed relative complementizer deto in Bulgarian, and partly on its ability to account for otherwise mysterious facts about the interpretation of multiple free relatives and correlatives.

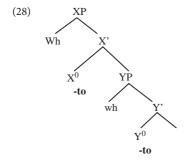
In Rudin (2009) I proposed that Bulgarian relative clauses have the same structure as those of Macedonian and Middle English, with Spec and C both filled. This analysis is represented in the trees in (25), identical to those in (20) but with one Bulgarian-specific quirk; the complementizer in relative clauses is realized as *-to* when preceded by a *wh* word but as *deto* otherwise. Thus, *-to* is essentially an allomorph of *deto*.



One nice consequence of this structure is that it can account for the correlation of single-pair readings with repeating -to. In multiple wh constructions like (26), -to can occur on all of the wh words or only the last one. This arguably indicates two different structures with one or more instances of the projection headed by -to. A single -to suggests a structure with all whs in Spec of a single projection, with just one head C (as in (27)); this structure results in pair-list interpretation. On the other hand two separate wh projections with two -to heads (as in (28)) gives single-pair interpretation; see Rudin (2009) for details.

(26a)	[Koj	kakvo <u>to</u>	iska],	da	vzeme.	
'Let everyone take what they want'						(pair-list)
(26b)	[Koj <u>to</u>	kakvo <u>to</u>	iska]	da	vzeme.	
	(single-nair)					





This account however has difficulty accounting for complex *wh* phrases of the type *kolkoto goljam* 'how big' or *kakvitopodarâci* 'which kind of presents', which appear to have the complementizer inserted within the *wh* phrase. These can even be stacked in some constructions, including multiple UCCs, making a cliticization or split deletion account of them particularly difficult. Unlike in Macedonian (see note 5), such phrases are fully normal in Bulgarian. The complementizer account also predicts more parallelism with *-što* than actually exists; for instance, if *što* is excluded in nominal comparatives because of its

complementizer status, a C analysis of *-to* incorrectly predicts that it also should not occur in nominal comparatives.

6.3. Other Analyses

It is, of course, possible that -to is neither a complementizer nor a definiteness marker; neither a C nor a D head. For instance, Iliev (2011) suggests, based largely on historical evidence, that -to is some type of modal element. This idea is attractive in some constructions (especially the Universal Concessive Conditional) which do have modal semantics, but most wh constructions do not involve any clear modalityand it is far from obvious why a modal element would suffix to the wh word. This analysis may well have some historical validity, but it does not seem to be a serious contender as a synchronic analysis of relative -to.8 Another possible approach to -to is the traditional morphological one: simply calling-to a relative suffix, a derivational morpheme creating relative from interrogative wh words. Under this approach kojto, kakvoto, and so onare labelled as relative wh wordswith no further analysis. This has the advantage of simplicity, and may turn out to be correct, but is not particularly satisfying in terms of explanatory force: it does not integrate -to into the syntax, nor clarify its relation to što, nor provide any reason why -to should be required on wh words in essentially all non-interrogative wh constructions. Vesela Simeonova suggests (p.c.) that -to is a demonstrative and that its demonstrative features account for the strong island properties of wh relatives. Again, this is an attractive idea, but it shares many of the problems of the definiteness head account.

7. Conclusion

Unfortunately, none of the proposed analyses looks fully satisfactory at this point. The relation between *-to* and *što* remains debatable, as does the syntactic identity of *-to* and the reason for its wider distribution and obligatory status. What is clear is, as often when comparing Bulgarian and Macedonian, that phenomena which superficially appear parallel or even identical diverge in unexpectedly complex ways when examined in detail. Relative*-to* and *što* exhibit several types of differences. Distributionally, *-to* is required in all relative clauses and in a number of related non-interrogative *wh* constructions, while *što* is optional in most relatives and disallowed in much of the broader range of *wh* constructions. In terms

⁸ Incidentally, Iliev alsoconsiders Macedonian and western Bulgarian štoamong the many modal-derived relative particles.

of form and historical source, *što* is homonymous with a complementizer and *wh* word, and derives from the *wh* word 'what', while -*to* resembles the Bulgarian definite article and derives from an old demonstrative root. Teasing apart the differences between these closely related languages will eventually shed enough light on their structures, one hopes, to finally resolve the puzzling identity of -*to*.

References

- Franks, S. and C. Rudin. "Invariant *-to* in Bulgarian". To appear in Festschrift for Ronald Feldstein,
- Hauge, K. R. 1999. A Short Grammar of Contemporary Bulgarian. Bloomington: Slavica.
- Iliev, I. G. 2011. "Otnositelno'-modalni razširiteli pri Bâlgaskite otnositelni mestoimenija" ['Relative' and modal elements in the structure of Bulgarian relative pronouns]. In Bojadžiev, T., Popova, V. and K. Aleksova (eds.): *Naučni Trudove v Pamet na Georgi Gerdžikov*. Sofia: Universitetsko izdatelstvo "Sv. Kliment Oxridski".
- Izvorski, R. 2000. "Free adjunct free relatives". In Billerey, R. and B. D. Lillehaugen (eds.): *WCCFL 19 Proceedings*. Somerville: Cascadilla Press, 232–245.
- Kramer, C. E. 1999. Macedonian. Madison: University of Wisconsin Press.
- Krapova, I. 2010. "Bulgarian relative and factive clauses with an invariant complementizer". *Lingua* 120(5), 1240-1272.
- Rudin, C. 2009. "The Bulgarian relative marker -to". In Franks, S., Chidambaram, V. and B. Joseph (eds.): *A Linguist's Linguist: Studies in South Slavic Linguistics in Honor of E. Wayles Browne*. Bloomington: Slavica, 403-422.
- Tomić, O. M. 2012. A Grammar of Macedonian. Bloomington: Slavica.