Non-stranded preposition + relative *who(m)*:
Syntactic discussion and corpus-related problems[1]
Carsten Breul

1　　　　Introduction

F. Aarts (1994a) summarises the pronouncements on the 'correct' use of the relative pronouns *who* and *whom* by three influential 18th and early 19th century grammarians[2] as follows:

1. *Who* should be used as subject only;
2. *Whom* should be used as direct object and as complement of a preposition, both in clauses where the preposition precedes the pronoun (*to whom*) and in clauses in which the preposition is stranded (*whom...to*).

As is well known, actual linguistic behaviour does not conform to these prescriptions. This fact has been taken into account in modern grammars of the 'descriptive' tradition such as e.g. Quirk et al. 1985 (= *CGEL*). Their authors, Aarts (1994a: 73) writes, "point out that it is necessary to distinguish between formal and informal style, and they regard *who* as grammatically correct in all contexts except when immediately preceded by a preposition."[3] In the present paper, I concentrate on the 'odd' case, i.e. where the relative pronoun is the complement of a non-stranded preposition such as in *the person to whom he spoke*. I will discuss the generalisation that in this case "the choice of *whom* is obligatory" (*CGEL*: 1249) from the perspective of generative syntax and relate it to corpus linguistic findings which cast some doubt on its validity. As the question to be addressed concerns the inner structure of the relative clause rather than its relation to its syntactic environment, the distinction between restrictive and non-restrictive relative clauses can be and is ignored here.

---

[1] My thanks go to the members of the Internet Corpora newsgroup who have discussed the corpus-related aspects of this paper with me. Thanks also to my local discussants Elizabeth von Schoff and Alan Webb.

[2] R. Lowth, *A Short Introduction to English Grammar* (1762), L. Murray, *English Grammar* (1795), W. Cobbett, *A Grammar of the English Language* (1818). On the last mentioned author and his grammar, which are less well-known than the others, see Aarts 1994b.

[3] More fully:

1. *Who* is used as the form for the subject, both in formal and in informal English;
2. In informal English *who* may also be used as the form for the direct object or the prepositional complement (if the preposition is stranded);
3. *Whom* is largely restricted to formal style, where it functions as object and prepositional complement. (Aarts 1994a: 73.)

2   Syntactic discussion of relative *whom/who* as complements of prepositions

The rule that relative *whom* is obligatory if governed and preceded by a preposition seems to be supported by a questionnaire based investigation and a corpus study mentioned by Aarts (1994a). In the questionnaire presented to Louisiana State University students by Walsh & Walsh (1989), no student substituted *who* and all students substituted *whom* for the gap in

(1)   That man to _____ you were speaking is my math teacher

Indeed, Aarts (1994a: 74) correctly points out that the questionnaire technique does not test actual usage. But Aarts (1994a: 76f.) reports that Quirk's (1957/1968) corpus of educated spoken English yields the same result as regards the aspect in focus here: There is no instance of *who* as complement of a non-stranded preposition in that corpus.[4]

From the point of view of a generative syntactician working in a more recent framework (as reflected by Culicover 1997, Roberts 1997, Radford 1997, Haegeman & Guéron 1999), these results may seem disturbing. There appears to be no syntactic reason to account for the judgement that (2a) is grammatical and (2b) ungrammatical given that (2c) and (2d) are grammatical, about which modern descriptive and didactic grammars as well as dictionaries seem to be unanimous:[5]

(2a)   That man to whom you were speaking is my math teacher
(2b)   That man to who you were speaking is my math teacher
(2c)   That man whom you were speaking to is my math teacher[6]
(2d)   That man who you were speaking to is my math teacher

Indeed, there is the classic account in the framework of early transformational grammar by Klima (1964). He proposed two 'styles' of English, the first allowing for Prep + *whom* and *whom* ... Prep, the second allowing for Prep + *whom* and *who* ... Prep. The two styles differ with respect

---

[4] I have not been able to verify that report by reading Quirk 1957/1968, which does not reproduce the the corpus itself. I assume that Aarts had access to the corpus.

[5] Besides the *CGEL*, see among the more recent grammars e.g. the descriptive one by Bache & Davidsen-Nielsen (1997: 430), the didactic grammars by Rutherford (1998: 139f.) and Swan & Walter (1997: 270ff., 290), among the dictionaries e.g. the *OALD* (5th edn., sv. *whom*), *LDCE* (3rd edn., sv. *who*), *CIDE* (sv. *whom*), *CCED* (sv. *whom*).

[6] Such constructions are judged grammatical by descriptive and pedagogic grammarians, but are rare; see Aarts 1994a: 77, *CGEL*: 1249. In the didactic grammar by Swan & Walter (1997), the phrase *the girl (whom) I was talking about* is given as a model (the brackets around *whom* indicate optionality). But note also that in generative syntactic accounts of relative clauses such constructions seem to be avoided (see e.g. Culicover 1997: 182, Radford 1997: 305f.). I will come back to this point in the course of the paper.

to the sequence of application of two transformational rules, Case marking and *Wh*-attachment, which operate in that order in the first style and in the reversed order in the second.

In G_2 [i.e. the grammar of style 1], the rule of case marking depends on the position of the elements of the sentences as they occur after *Wh*-attachment. The difference in order of rules reflects the fact that while in L_1 [i.e. style 1] case marking is dependent on function (namely on whether or not the element is a grammatical object), in L_2, *CASE* is a concomitant of position, reckoned only after certain rearrangements of the basic order of elements. In L_2, *CASE* is not associated with functional elements, if, at a particular point in the sequence of ordered transformational rules, those elements happen not to be situated after (that is, to the right of) the factors motivating CASE-attachment. (Klima 1964: 13.)

However, this line of argumentation is incompatible with more recent developments in Generative Grammar for several reasons. Most relevantly for our purposes here, the notion of there being a number of construction specific 'transformational rules' has been abolished (in favour of a general constituent movement rule, Move α), so that there is no Case marking rule and hence no way of transforming a constituent [P *who*] into a constituent [P *whom*] in the course of the derivation. The basic line of analysis in terms of recent ideas in Generative Grammar is as follows.
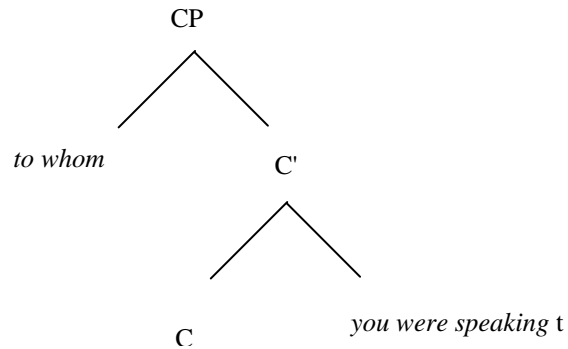
In all of (2a-d), *who/whom* is the complement of *to* (i.e. its sister constituent) at some earlier stage in the derivation of the sentence.

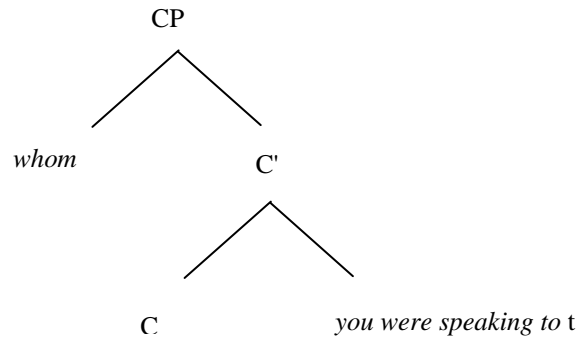(3)   that man [you were speaking *to who/whom*] is my math teacher

The *wh*-expression is assumed to move to the front of the relative clause (being bracketed in (3)), which yields the surface constituent order (see e.g. Haegeman & Guéron 1999: 185ff.). It is assumed that the relative clause is dominated by a complementiser phrase (CP), with its head C containing a morphologically "null relative clause affix **R**" (Radford 1997: 307; Radford's bold print). In terms of the checking theory developed within the Minimalist Program (seminally Chomsky 1995), this affix is said to carry a relative clause feature, which needs a corresponding feature in the specifier position of CP (Spec-CP) for the purpose of 'feature checking'.[7] Such a feature is provided by the *wh*-expression, which has to move to Spec-CP in order to fulfill the requirement of feature checking (the '*t*' in the following partial structural descriptions is the trace left behind by the movement of the *wh*-expression):

---

[7] Thus, movement is triggered by the requirement to check features (see Chomsky 1995, Radford 1997).
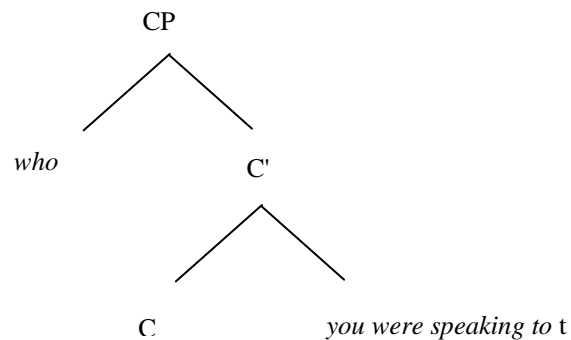
(4a)   that man [CP to whom [C' you were speaking *t*]] is my math teacher


```
                    CP
                   /  \
                  /    \
            to whom    C'
                      /  \
                     /    \
                    C    you were speaking t
```

(4b)   that man [CP whom [C' you were speaking to *t*]] is my math teacher


```
                    CP
                   /  \
                  /    \
            whom       C'
                      /  \
                     /    \
                    C    you were speaking to t
```

(4c)   that man [CP who [C' you were speaking to *t*]] is my math teacher


```
                    CP
                   /  \
                  /    \
            who        C'
                      /  \
                     /    \
                    C    you were speaking to t
```

The *wh*-expression to be moved may be either a PP carrying the relative clause feature by virtue of containing a *wh*-pronoun, as in (4a), or it may be the *wh*-pronoun itself, as in (4b, c). This analysis leaves the ungrammaticality supposed for (4d) mysterious.

(4d)   that man [CP to who [C' you were speaking *t*]] is my math teacher

Although I said above that this problem "may seem disturbing" to generative syntacticians, a discussion of it is hard to find in more recent generative syntactic analyses. Among the sources I have looked at, only Radford (1997) addresses it, although not with respect to relative but with respect to interrogative *who/whom* (I will come to his analysis below). It is astonishing that for example McCawley (1988) mentions the phrases *the cook from whom I learned this recipe* and *the cook who I learned this recipe from* within the range of half a page (ib.: 436) in order to illustrate an older version of the movement account just sketched, whithout, however, saying anything about the morphological difference between the relative pronouns in both phrases.

Note that since prepositions are heads which require objective case on their complements, we must maintain that not only *whom* but also *who* has objective case in (4). This is not problematic, for it is obviously true that word-forms may be morphologically ambiguous with respect to case.[8] While the case of *whom* is overtly objective, that of *who* is covertly objective in (4c) for example. Thus *who/whom* are alternatives as objective case forms in present day English (while *who*, but not *whom*, may also have subjective case covertly[9]). The existence of these alternatives is probably due to the fact that the development of English towards the analytic type of expressing grammatical relations has not come to an end yet. The preference of prescriptive grammars and present day manuals and dictionaries of usage for *whom* as the form of the relative pronoun functioning as complement of a preposition follows immediately from the thesis that linguistic prescriptivism aims at the suppression of emerging alternatives in the lexicon and in grammar (see Milroy & Milroy 1985/1991).

The generative syntactic analysis sketched above fails to address the question of how to account for the fact that either the whole *wh*-PP or only the *wh*-pronoun (*who/whom*) may move to Spec-CP; see e.g. (4a) versus (4b, c). Actually, the existence of this alternative seems to pose a problem for the economy principle underlying the Minimalist Program. For one of its

---

[8] This also holds for the system of pronouns in Modern English (which has retained the greatest amount of morphological case distinctions) as the pronoun *you* shows immediately.

[9] Speakers/writers occasionally use *whom* as a subject relative pronoun. I came across the following example in a novel by Ian Rankin (1999): *I tell you to keep an eye on a man whom experts say may murder someone* spoken by Chief Superintendent Watson to Detective Inspector Rebus (see also Quirk 1957/1968: 101f., Walsh & Walsh 1989). After giving two examples of this type, Aarts (1994a: 75) writes: "This construction, which is still very frequent today, involves what generative grammarians call LONG MOVEMENT (Haegeman 1991, 370), since the relative pronoun has been moved out of a clause that depends on the verb to its right." This reference to generative syntactic analyses can be misleading in that context. For it might be interpreted so as to suggest that the 'long movement' analysis has to do with the choice of *whom* in such sentences. This is not the case. There are no sentences with *whom* as subject relative pronouns in Haegeman 1991 and 1991/1994. Indeed, there are analyses of sentences of the type *This is the man whom John claims that he will invite*, but these have *whom* as object pronouns (having undergone long movement). Most grammarians, including generative syntacticians, will probably agree

requirements is "to move the smallest constituent possible the smallest distance possible" (Radford 1997: 277). How, then, can it be that a complete PP moves if movement of a *wh*-pronoun suffices to check the relative clause feature? As the following discussion indicates, an attempt at an answer to this question is provided by Radford (1997). But it rests on assumptions which may seem problematic.

I have already pointed out (fn. 6, above) that in generative syntactic analyses of relative clauses examples such as (4b), where we have *whom* with a stranded preposition, are avoided. We might perhaps conclude that some generative syntacticians tend to assume that such sentences are actually ungrammatical. If one makes this assumption, one may adduce an analysis by Radford (1997) to solve the problem of the then preposition + *whom* versus *who* + stranded preposition alternative with relative clauses. Radford (ib.: sect. 7.4) does not analyse relative clauses with respect to this problem but *wh*-interrogative clauses with stranded versus non-stranded prepositions, such as in (see ib.: 278f.)

(5a)   Who were you talking to?

(5b)   To whom were you talking?

Such sentences are assumed to be structurally similar – and with respect to our problem identical – to relative clauses in that they too show movement of a *wh*-expression to Spec-CP. Radford's account appears to satisfy the economy principle about the alternatives presented by (5) as well as the corresponding cases of relative clauses. But note that this account crucially involves the assumption that a sentence like

(5c)   Whom were you talking to?

is ungrammatical "in formal English" (see ib.: 279).

Radford's analysis can be summarised as follows: Merging the overtly case-marked *whom* with a governing preposition produces percolation of the case feature to the governing preposition. The percolating case feature pied-pipes other grammatical features of *whom* (the interrogative feature in interrogative clauses like (5) or the relative feature in relative clauses). Hence it is the PP which carries the interrogative or the relative feature after the merger of the preposition and the overtly case-marked *whom*. And this means that it is the PP which has to move to Spec-CP to check the interrogative or relative feature. In contrast, merging the covertly case-marked *who* with the governing preposition does *not* produce this kind of feature percolation. Hence, after the merger of a preposition and *who*, it is still the latter which carries

---

with the assessment made e.g. by Quirk et al. (1985: 368, note) and Walsh & Walsh (1989: 285) that *whom* as subject relative pronoun is hypercorrect, i.e. ungrammatical.

the interrogative or relative feature. And this means that it is *who* alone which can and must move to Spec-CP.[10]

However, as already pointed out, the crucial assumption underlying this account, namely that *whom* + stranded preposition constructions are ungrammatical, seems problematic in view of the judgements by descriptive grammarians and the facts of usage mentioned above. Once again in the words of the *CGEL*: "both *who* and *whom* can take initial position, as in [*Who/Whom is she working for?*]" (ib.: 370). Four examples of this kind from the British National Corpus (BNC) are the following (but note that my discussion in the next section of the reliability of corpus findings applies to (6)-(9) as well):

(6)   Whom did you dance with? (F9R: 703,[11] a written source)

(7)   Whom did you talk to? (G0B: 1208, a written source)

(8)   Whom does it come from? (HUC: 458, a spoken source)

(9)   Whom could she talk to? (HWE: 496, a written source)[12]

Thus I think that Radford's (1997) approach to the problem is not fully satisfying.

Apart from trying to restore the 'symmetry' with respect to grammaticality of (4a-d) by pronouncing (4b) ungrammatical, there is the alternative to opt for the grammaticality of (4d). This is not altogether eccentric as the discussion in the following section shows (but, it must be admitted, it leaves the economy problem just discussed completely unaccounted for).

---

[10] Of course, generative syntacticians have to provide a more general principle or set of principles from which it follows that such a feature percolation takes place with overtly case-marked, but not with covertly case-marked forms. One may also ask why French *qui*, which also seems to be not overtly case-marked, does not behave like *who* in this respect? I.e. why do we not have **Qui pensez-vous à?* but *À qui pensez-vous?* and not **Qui l'avez-vous reçu de?* but *De qui l'avez-vous reçu?*? See Roberts (1997: 212f.) for a suggestion as to why preposition stranding is *generally* not possible in French.

[11] References of this kind identify the BNC-document and the sentence number of the respective sentence in the BNC-document.

[12] On the other hand, Quirk et al's (1985: 370) claim that *wh*-interrogative clauses with non-stranded preposition plus *who*, as in *For who is she working?*, are ungrammatical does not seem to stand on perfectly firm ground. I have found a number of instances of preposition + interrogative *who* constructions – although, interestingly, no full clauses – in the BNC. E.g. (GW2: 681):

'Did you fall in love with him?' Walterkin asked.
'With who?'

For similar cases see: A7A: 970, AEB: 866, F7F: 1240, FUK: 2069, G1B: 2657, G48: 8, HWN: 2334, HYH: 60, HYH: 64, KBB: 6813, KBE: 9056, KC1: 2776, KCT: 9752, KCT: 12646, KCV: 4524, KCW: 679, KCW: 1102, KD5: 1515, KD8: 4202, KDA: 6573, KDM: 7570, KDP: 168, KNY: 918, KNY: 1864, KP0: 484, KP1: 8544, KP3: 2831, KSV: 1712. This usage of preposition + interrogative *who* "as a complete utterance" is also acknowledged by Bache & Davidsen-Nielsen (1997: 426), who additionally mention the existence of this construction "in clause-final position (*Di shacked up with who?*)" (ib.).

3   Corpus findings

I searched for sentences of the type exemplified by (4b), i.e. those which do have relative *who* as complement of a non-stranded preposition, in the British National Corpus (BNC). The BNC is a corpus of ca. 100 million words of British English texts and text extracts originating from 1960 till 1993. Ca. 90% of the material in the 4125 texts and text extracts are from written sources, the rest are orthographic transcripts from spoken sources.[13]

My search yielded the following sentences (bold print added):

(10)   A passionate lover of the Savoy Operas, she was a founder member of the Bradford Gilbert and Sullivan Society, **with who** she had a long association. (C8G: 418)

(11)   It's almost impossible to put him down in the tackle, and there are few players **about who** you an [sic; probably 'can' is intended] say that. (CB2: 396)

(12)   They need to keep their bankers happy by reducing a £3 million-plus debt in the next few months, and Robson, **in who** Leeds are already showing an interest, will become their most disposable asset. (CBG: 305)

(13)   It has one senior bishop, two suffragans, one dean, four archdeacons, and 21 rural deans **of who** one is a woman. (ED9: 3029)

(14)   They must be treated as adequate because they reflect the statutory provisions in regard to appeals by persons **upon who** intervention notices are served by S.I.B. in the exercise of the intervention powers delegated to it by the Secretary of State (as to which see section 97 of the Act). (FD1: 413)

(15)   Instead, the sites are used by local residents, **of who** some 25 per or so come by foot and use the sites like an urban park, primarily to take a walk (Harrison, 1981). (FR2: 453)

(16)   Through its founder Molly Braithwaite, **for who** we held a heartfelt admiration, we feel an affectionate living bond with all those in positions of responsibility within the MEDAU SOCIETY. (HU8: 57)

(17)   I would also I think put in a word for the work of the joint [...] policy panel [...] which is shared between this committee and the social services committee because it seems to me that it is not [unclear] for us to be thinking that there is a group of children **for who** nursery education is necessary [...] or desirable and a different group of children for

---

[13] For more information on the BNC and its construction see Aston & Burnard 1998: part I, ch. 2 and Burnard (ed.) 1995.

whom something else [...] is necessary and desirable, largely because of their parents'
position. (JWA: 149)[14]

All examples except for (17) are from written sources. The list is exhaustive for the BNC in so
far as it gives those sentences of the complete BNC which contain the linear sequence

noun + preposition + *who* + anything but verb

and in which *who* actually introduces a relative clause.[15] The list may fail to comprise instances
of relevant sentences in which the relative clause is, sequentially, not directly attached to a noun.
For example, if (11) had the adverb *around* after the noun to which the relative clause attaches,
as in

(18)   ... and there are few players around about who you can say that

it would not have been detected by the search procedure.

However, even if there are a few more relevant instances in the corpus, the crucial
methodological question regarding these findings would not have changed: Does the
orthographic representation faithfully reflect what the text producer really expressed or intended
to express. In other words, can we be sure that there are no errors of any kind involved – either in
the production of the source texts or in their conversion to BNC-files – whereby *whom* has
inadvertently and mistakenly been replaced by *who*?[16] Exemplifying it with sentences (10)-(17),
I have posed this question to the Internet Corpora newsgroup. Besides an invitation for the
discussion of this problem, I asked for reports on actual usage with respect to the preposition +
relative *who* pattern.[17] The responses I have got seem to confirm that there is indeed such a
reliability problem inherent in corpora so that absolute authenticity with respect to text-

---

[14] The search yielded one further sentence which, however, might involve a typing error at the crucial
position (see the unmotivated comma after *who*):

Other people, **for who,** [sic; typing error? ',' and 'm' being next to each other on the keyboard] anything which
isn't modern, is old English, claimed that it was old English. (JT2: 81)

[15] a) Punctuation is ignored. b) Obviously, the automatic extraction depends crucially on the the word-
class tagging of the BNC being correct and is only reliable to the degree that it is so. I will explain the
search procedure in more detail in the Appendix.

[16] Also, an originally stranded preposition could have been moved in front of the relative pronoun by an
editor or proof reader who aims at a more formal style but forgets to change the form of *who*. This idea
emerges from T. Mann's contribution to the Internet Corpora newsgroup discussion to be mentioned
below.

[17] The text of my question was this:

Do native-speakers of English actually use Prep + relative *who*? In other words: Do you think that
the sentences may very well be correct representations of the text-producers' usage? Or would you
consider it likely that there are printing/typing (or transcription) errors involved here (omission of
'm' in 'who'), which distort the picture?

production cannot be guaranteed in principle. For example, A. Renouf says (personal communication): "There is no means by which we can be sure that a corpus is a true representation of original production unless we have access to information about the original circumstances of the event, of course."[18] The informants' reports on usage range from "I have never noticed such usage in written or spoken texts" to "I have heard many similar sentences" or "the forms quoted are highly possible today". In addition, some informants do not find each of the sentences equally good or bad; i.e. the reactions vary from sentence to sentence. Occasionally, speakers report vacillating judgements on re-reading the sentences. Some informants point out that they would try to avoid both 'preposition + relative *whom*' and 'preposition + relative *who*' patterns in favour of the pattern 'relative *who* + stranded preposition'.[19]

4   Conclusion

In this paper I argue that if we consider

(19)   a) preposition + relative *whom* (as in (20a) below)

         b) relative *who* ... + stranded preposition (as in (20b) below)

         c) relative *whom* ... + stranded preposition (as in (20c) below)

grammatical – as is unanimously done in descriptive grammars and dictionaries – then there is no syntactic reason to judge

         d) preposition + relative *who* (as in (20d) below)

ungrammatical.

(20a)   That man to whom you were speaking is my math teacher
(20b)   That man who you were speaking to is my math teacher
(20c)   That man whom you were speaking to is my math teacher
(20d)   That man to who you were speaking is my math teacher

It appears that we actually find instances of the problematic d-construction in performance data (the BNC). But their frequency is extremely low, and we are confronted with a reliability

---

[18] She also points out (personal communication) that "[i]t depends on one's research focus what is important to be retained in the electronic version."

problem as regards the authenticity of the respective instances. Native speaker reports on the use of the problematic construction are equally inconclusive. However, the fact that some informants do report its use seems to indicate that it is *not* non-existent in performance. Obviously, there is a very strong normative pressure against the use of that construction. The following statement by G. Sampson (personal communication) emphasises this aspect:

> Surely the point is that the pattern "preposition + relative pronoun" is nonexistent in natural speech, it is something that people learn to produce for the purposes of writing; and because it is in that sense somewhat artificial, some people get it wrong. To me, the examples you quote are all barbarous and unacceptable, but then I am the type of professional who is well-versed in the norms of written English. No competent editor would allow these examples, but it does not seem surprising that some less-skilled writers produce them.

I would maintain that this normative pressure accounts for the low frequency of occurrence of the pattern and for very strong reactions against its acceptability by a number of informants. Hence, Quirk et al. (1985: 368) are absolutely right in writing that "there is a stylistic incompatibility between the preposition + relative pronoun construction (*to whom*), which is rather formal, and the use of *who* rather than *whom* as prepositional complement (*who ... to*), which is informal";[20] but it is doubtful if they are justified in writing (ib.) that a construction such as in

(21)   This is the person to who you spoke

"does not occur".[21] And we certainly have to distinguish between what may create a "stylistic incompatibility" or "clash of genres" (see fn. 20) and what is ungrammatical.

The only alternative to the conclusion presented so far is an approach that seems to be preferred by generative syntacticians, exemplified here by an account following Radford 1997. In this view, despite what descriptive grammars say, neither (19d) (and hence (20d)) nor (19c) (and hence (20c)) are grammatical. The symmetry in the patterns of (19) and (20) thus achieved has the theoretical advantage – at least with respect to recent ideas in generativism – that it appears to be compatible with assumptions about economy principles operative in grammar. Additionally, this grammatical rejection of the c-pattern may be used in an explanation of its rarity in performance. However, this view would seem to require a modification in the evaluation

---

[19] P. de Haan (in his contribution to the discussion in the Corpora newsgroup) points out that this avoidance strategy is not always possible. He mentions my (15); (13) is also a case in point.

[20] In the words of A. Kilgarriff (personal communication): "'prep + who' is a clash of genres as much as a grammatical issue". See also W.J. Crawford's message on the Internet LINGUIST Network vol. 10, issue 1049 (1999), which makes the same point. With respect to preposition + interrogative *who*, this point is also made by Bache & Davidsen-Nielsen (1997: 426).

[21] By saying 'doubtful' I do *not* mean the possibility of slips of the tongue or that we are confronted with non-standard features here – unless 'Standard English' is identified with 'English that conforms to normative grammar(s)'.

of the *whom* + stranded preposition construction from the perspective of descriptive grammar as represented by the *CGEL*.

Appendix: The search procedure

The search tool standardly accompanying the BNC, SARA (= SGML-Aware Retrieval Application; on SARA see Aston & Burnard 1998: part II), is not flexible enough to conveniently carry out searches of the kind needed for the present study. The reason is that SARA does not allow the search for word-class (or POS = part-of-speech) tags without specifying a word-form (see Aston & Burnard 1998: 141). I developed a tool (a program written in the programming language Perl) which extracts sentences from the BNC that contain any word-form or sequence of word-forms – with or without their corresponding POS-tags – to be determined by the user. Punctuation is ignored in the version of the tool which I used for the present investigation. In the BNC, the coding of word-forms and their POS-tags takes the form

'<w XYZ>word-form '

with 'XYZ' standing for one of a number of three-character POS-tags as, for example, 'NN1' for a singular common noun.[22] Word-forms and tags are nothing but character strings to the programming language, and it is possible to use the pattern matching options for character strings available in Perl. For example, we may use the '.'-symbol to match an arbitrary character, or we may use the '^'-symbol to match any character which is not identical to the character or characters following that symbol (with square brackets enclosing the '^' and the character(s) not to be matched). Thus the tool is not restricted to exact pattern matching, but is very flexible.

In terms of this tool, searching for the sequence

noun + preposition + *who* + anything but verb

was realised as searching for

–   'w N' in a first tag + word-form sequence, matching tags beginning with 'N', i.e. nouns;

–   'w PR' in the following tag + word-form sequence, matching tags beginning with PR, i.e. prepositions;

–   '>who ' in the next tag + word-form sequence, matching the word-form *who*;

---

[22] There are a few tags of the form 'UVW-XYZ'. These are the cases where the tagging program was unable to decide between two tags.

– 'w [^V]' in the last tag + word-form sequence, matching tags beginning with other characters than 'V', i.e. non-verb word-forms.

This last item in the determination of the pattern serves to filter out sentences like

(22)   I would especially wish to record my thanks to who retired in September 1990 from the chair of the Highland Area Committee after eleven successful years. (J1R: 41)

(23)   The corporation and the rural district council were in dispute about who should rehouse them. (APP: 886)

Running the search tool with these parameters over the complete BNC yielded 266 sentences, most of which are not of the type required, as e.g.

(24)   You adapt your personality depending on who you're dealing with. (A5Y: 1046)

(25)   He could build up a knowledge of who he was piece by piece. (AC4: 1659)

This list of 266 sentence had to be sieved 'manually' to filter out the relevant cases. The result is the list given in the preceding section, i.e. examples (10)-(17).

References

Aarts, Flor. 1994a. "Relative *who* and *whom*: Prescriptive rules and linguistic reality". *American Speech* 69: 71-79.

Aarts, Flor. 1994b. "William Cobbett's *Grammar of the English Language*". *Neuphilologische Mitteilungen* 95: 319-332.

Aston, Guy & Burnard, Lou. 1998. *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Bache, Carl & Davidsen-Nielsen, Niels. 1997. *Mastering English: An Advanced Grammar for Non-native and Native Speakers*. Berlin & New York: Mouton de Gruyter.

Burnard, Lou (ed.). 1995. *Users' Reference Guide to the British National Corpus.* Oxford: Oxford University Computing Centre.

[*CCED*] *Collins Cobuild English Dictionary*. Eds. John Sinclair et al. London: HarperCollins, 1995.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge (MA) & London: MIT Press.

[*CIDE*] *Cambridge International Dictionary of English*. Eds. Paul Procter et al. Cambridge: Cambridge University Press, 1995.

Culicover, Peter W. 1997. *Principles and Parameters: An Introduction to Syntactic Theory.* Oxford: Oxford University Press.

Haegeman, Liliane. 1991/1994. *Introduction to Government and Binding Theory*. 2nd edn. Oxford & Cambridge (MA): Blackwell. [Haegeman 1991 = 1st edn. of Haegeman 1991/1994]

Haegeman, Liliane & Guéron, Jaqueline. 1999. *English Grammar: A Generative Perspective.* Oxford: Blackwell.

Klima, Edward S. 1964. "Relatedness between grammatical systems". *Language* 40: 1-20.

[*LDCE*] *Longman Dictionary of Contemporary English*. 3rd edn. Eds. Adam Gadsby et al. Burnt Mill: Longman, 1995.

McCawley, James D. 1988. *The Syntactic Phenomena of English*. 2 vols. Chicago & London: University of Chicago Press.

Milroy, James & Milroy, Lesley. 1985/1991. *Authority in language: Investigating language prescription and standardisation.* London & New York: Routledge.

[*OALD*] Hornby, A.S. *Oxford Advanced Learner's Dictionary*. 5th edn. Eds. Jonathan Crowther et al. Oxford: Oxford University Press, 1995.

Quirk, Randolph. 1957/1968. "Relative Clauses in Educated Spoken English". In: Quirk, Randolph. 1968. *Essays on the English Language: Medieval and Modern.* London: Longman. 94-108.

Quirk, Randolph & Greenbaum, Sidney & Leech, Geoffrey & Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman. [= *CGEL*]

Radford, Andrew. 1997. *Syntactic Theory and the Structure of English: A Minimalist Approach.* Cambridge: Cambridge University Press.

Rankin, Ian. 1999. *Dead Souls: An Inspector Rebus Novel.* London: Orion.

Roberts, Ian. 1997. *Comparative Syntax.* London et al.: Arnold.

Rutherford, William. 1998. *A Workbook in the Structure of English: Linguistic Principles and Language Acquisition.* Malden (MA): Blackwell.

Swan, Michael & Walter, Catherine. 1997. *How English Works: A grammar practice book.* Oxford: Oxford University Press.

Walsh, Thomas & Walsh, Natasha. 1989. "Patterns of *who/whom* Usage". *American Speech* 64: 284-286.