



---

On the Choice of Relative Pronouns in English

Author(s): Gregory R. Guy and Robert Bayley

Source: *American Speech*, Vol. 70, No. 2 (Summer, 1995), pp. 148-162

Published by: Duke University Press

Stable URL: <http://www.jstor.org/stable/455813>

Accessed: 26-09-2017 21:55 UTC

## REFERENCES

Linked references are available on JSTOR for this article:

[http://www.jstor.org/stable/455813?seq=1&cid=pdf-reference#references\\_tab\\_contents](http://www.jstor.org/stable/455813?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

*Duke University Press* is collaborating with JSTOR to digitize, preserve and extend access to *American Speech*

# ON THE CHOICE OF RELATIVE PRONOUNS IN ENGLISH

GREGORY R. GUY

*York University*

ROBERT BAYLEY

*The University of Texas at San Antonio*

THE RELATIVE CLAUSE IN ENGLISH is the site of a clear case of syntactic variation, namely the alternation among the three relative pronoun options: a *wh*-word, *that*, or zero, for example:

1. a. This is the house *which* I told you about.
- b. This is the house *that* I told you about.
- c. This is the house  $\emptyset$  I told you about.

The fact that English speakers must choose among these three options whenever they produce a relative construction presents the linguist with some interesting questions. How do English speakers choose a relative pronoun? What factors influence this choice, and why? How can we model this option within the productive grammar of English? Relative pronouns are a particularly convenient site for investigating questions of this type for several reasons. First, they are fairly frequent in speech as well as writing. Thus, compared to such other syntactic variables as resumptive pronouns and passives, they are considerably easier to collect data on using standard sociolinguistic methodology. Second, it is fairly clear from intuitive reflection that they are subject to strong constraints. For example, nonrestrictive relatives seem to demand a *wh*-relative pronoun. And third, since this variation involves three clearly distinct variants (rather than the two that standard variable rules accommodate), it offers a richer range of theoretical/methodological issues that can be addressed.

In this study, we examine relative pronoun choice in 943 relative clauses from speech and writing. Results of VARBRUL analyses indicate that the animacy of the antecedent, the channel of communication, the syntactic position of the relativized element in the embedded clause, and the adjacency of the antecedent and the relativized element all have significant effects on the choice of relativizer.

We will also present the basic parameters of the variation, examining aspects of both the linguistic and the social/stylistic context. Where possible, we will explore explanations for the constraints we find. In addition, we will examine several analytical issues that arise at several points in our analysis, questions which depend crucially on the theoretical assumptions being made. Without an examination of these issues, some of the quantitative results would be obscure or incomparable.

## PREVIOUS STUDIES

In recent years, variation in the choice of relativizers has attracted considerable attention. Romaine (1982), for example, studied relative-pronoun use in Middle Scots from a diachronic perspective, and Dekeyser (1984) and Rissanen (1984) examined relativizers in Early Modern English and seventeenth-century American English, respectively. From a synchronic perspective, Prideaux and Baker (1986) examined processing of English relative clauses. In addition, Biesenbach-Lucas (1987) compared relative-pronoun use in speech and journalistic writing, and Kikai, Schleppegrell, and Tagliamonte (1987) examined the effect of syntactic position on relative-pronoun choice; both studies found that zero relatives were more common in speech than in writing. Recently, Adamson (1992) and Tottie (1993) have also examined relative-pronoun choice using variable rule analysis. Adamson found that zero was strongly disfavored in subject position in the relative clause and that the choice of zero as a relative pronoun was socially stratified, with working-class speakers favoring and upper class speakers disfavoring zero relative pronouns; Tottie found that personal pronouns as relative-clause subjects favored zero relatives.

## THE PRESENT STUDY

DATA. The source for speech data is *The White House Transcripts* (Nixon 1974), the published version of the surreptitious recordings made by Richard Nixon during the Watergate crisis. While clearly limited to a very narrow segment of the population, namely white male high-level government officials, the transcripts of Nixon's conversations with his staff are a convenient source of unmonitored standard American English. No one on Nixon's staff was aware of being recorded, and the late president himself was not particularly conscious of recording devices when he engaged in conversations whose later disclosure led to his forced resignation.

Written data are drawn from a variety of articles on language variation (Carden 1973; Guy 1978; Horvath 1987; Labov 1973, 1986; Naro 1981; Schiffrin 1986). Academic articles, which exemplify formal style, were chosen to represent the opposite end of the standard English stylistic continuum from the informal conversations that took place in the Oval Office during the Watergate years.

The nature of the sources of spoken and written data admittedly limits our interpretation of the results. Our sources provide a very clear stylistic contrast, but the use of academic articles makes it difficult to disentangle the effects of different levels of formality from the effect of the channel of

communication. Thus, although we refer to the effect of speech and writing in our discussion, we recognize that further research comparing formal and informal speech and formal and informal writing is necessary to fully assess the effect of this factor.

**CODING.** Our data confirmed our intuition that relatives are categorically realized as *wh*- words in nonrestrictive clauses. We therefore confined our statistical analysis to the remaining 827 restrictive clauses. Separate analyses were performed for each of the three variants (*wh*- words, *that*, and zero) using MacVarb (Guy 1989), a Macintosh version of VARBRUL 2S (Rousseau and Sankoff 1978). Our analysis examined the following factors:

1. animacy of the antecedent ([± human])
2. channel of communication (speech or writing)
3. syntactic function of the relativized element in the lower clause; here we distinguished subjects, direct objects, objects of pied-piped prepositions (e.g., *The man to whom I gave the money. . .*), objects of stranded prepositions (e.g., *The man I gave the money to. . .*), and the adjunct elements of locatives, manner adverbials, and temporals
4. adjacency of the antecedent and the relative pronoun; this group included three factors: relative pronouns adjacent to their antecedents, relative pronouns separated from their antecedents by another relative clause, and relative pronouns separated from their antecedents by a phrase other than a relative clause
5. syntactic function of the antecedent in the matrix clause (subject, direct object, other)

The last factor group proved not to be significant and will not be discussed further.

**RESULTS.** The results indicate that relative pronoun choice shows typical patterns of sociolinguistic variation, jointly constrained by syntactic and social/stylistic factors. In general, our overall results agree with those of Kikai, Schleppegrell, and Tagliamonte (1987; see table 1); however, our analysis and interpretation differ from theirs in several respects. For example, they did not investigate the effect of adjacency of the relative pronoun and its antecedent. Our analysis indicates that adjacency strongly

TABLE 1  
Totals of *wh*-, *that*, and  $\emptyset$  in Relative Clauses

	<i>Kikai et al. 1987</i>		<i>Current Study</i>	
	<i>N</i>	%	<i>N</i>	%
<i>wh</i> -	598	34	288	35
<i>that</i>	782	45	362	44
$\emptyset$	360	21	177	21
TOTAL	1740		827	

favors a zero relative, while the intervening occurrence of a phrase or clause other than another relative clause strongly inhibits relative pronoun deletion. Varbrul analysis also indicates that all other factor groups have significant effects ( $p < .05$ ) on use of at least one of the three variants.

Table 2 details our statistical results.

ANIMACY. Whether the antecedent is human is clearly the principal constraint affecting the choice between *who* and *which* as a relative pronoun. (Our data contained no instances of animate nonhuman antecedents.) Does it also affect the higher level choice between a *wh-* word, *that*, and zero? Table 3 gives our results for this factor group.

TABLE 2  
Numbers and Percentages of *wh-*, *that*, and  $\emptyset$  by Factor  
(restrictive clauses only)

Factor	<i>wh-</i>		<i>that</i>		$\emptyset$		Total
	N	%	N	%	N	%	
1. $\pm$ Human Antecedent							
+ Human	104	65	35	22	21	13	160
- Human	184	28	327	59	156	23	667
2. Channel							
Speech	159	26	286	47	159	26	604
Writing	129	58	76	34	18	8	223
3. Grammatical Function in the Relative Clause							
Subject	176	47	199	53	1 <sup>a</sup>	0	376
Direct object	19	7	124	46	125	47	268
Object of pied-piped preposition	36	100	0	0	0	0	36
Object of stranded preposition	3	5	28	47	29	48	60
Genitive	10	100	0	0	0	0	10
Complement	0	0	3	38	5	53	8
Locative	41	87	5	11	1	2	47
Manner	0	0	1	8	12	92	13
Temporal	3	33	2	22	4	44	9
4. Adjacency of the antecedent							
Adjacent to antecedent	231	33	303	43	172	24	706
Separated by another relative clause	14	44	15	47	3	9	32
Separated by another element	43	48	44	49	2	2	89
TOTAL	288	35	362	44	177	21	827

a. One spurious token was introduced to avoid a knockout in Varbrul analysis.

TABLE 3  
Human and Inanimate Antecedents: VARBRUL Weights for *wh-*, *that*, and  $\emptyset$

	<i>wh-</i>	<i>that</i>	$\emptyset$
+ Human	.80	.23	.59
- Animate	.20	.77	.42

These results indicate that human antecedents strongly favor the use of *wh-* forms and strongly disfavor the use of *that*. For zero forms, the effect of this factor group is statistically significant, but much less dramatic: zero is moderately favored for human antecedents and disfavored for nonhuman antecedents. The overall picture could be interpreted as reflecting a linguistic norm which inhibits *that* in reference to humans.

SPEECH AND WRITING. Writing, particularly the formal writing included in our sample, clearly provides more opportunity for monitoring than conversation. Kroch and Small's (1978) theory of "grammatical ideology" would predict that an increase in monitoring would lead to a higher incidence of the use of *wh-* forms as relative pronouns. These forms are generally perceived as more formal, and in some instances, as more "correct." Further, the greater explicitness attributed to writing by literacy theorists such as Olson (1977) would predict that writing would disfavor deletion as a relativization strategy. Our results, summarized in table 4, provide qualified support for both predictions.

*Wh-* forms are favored in formal written discourse and disfavored in conversation, while *that* and zero are favored in informal conversation and disfavored in formal writing. Individual differences in the written sources, however, suggest that the apparent robustness of the results may be partially a consequence of the particular sample we chose. The use of *wh-* forms in writing is greatly influenced by stylistic preferences, which vary considerably even between different articles by the same author. In Labov (1973), for example, *wh-* forms are favored (81% of those coded), while in Labov (1986) *that* is favored (76% of those coded). (It is, of course, possible that these differences reflect the influence of different editors rather than changes in the author's practice.) The choice of pied-piping is also a matter of individual preference. Carden (1973) and Guy (1978), for example, account for the majority of pied-piped prepositions in the written data.

TABLE 4  
Speech and Writing: VARBRUL Weights for *wh-*, *that*, and  $\emptyset$

	<i>wh-</i>	<i>that</i>	$\emptyset$
Speech	.24	.68	.63
Writing	.76	.32	.37

Despite the high level of significance obtained for this factor group, our suggestion that the difference is due more to the channel of communication than to the register is offered tentatively. More robust conclusions await studies that compare relative pronoun choice in writing and speech of similar levels of formality and across different types of written texts.

SYNTACTIC POSITION OF THE RELATIVIZED ELEMENT IN THE EMBEDDED CLAUSE. The syntactic position of the relative pronoun in the embedded clause clearly affects the choice of relativization strategy. For example, in standard English, relative pronouns are rarely deleted in subject position. As (2) indicates, genitives categorically require a *wh*-form.

2. a. The friend whose book I borrowed. . . .
- b. \*The friend that book I borrowed. . . .
- c. \*The friend ( $\emptyset$ ) book I borrowed. . . .

Our results for this factor group, given in table 5, indicate that subject position also has an effect on the choice between a *wh*-word or *that*. Other syntactic positions also affect the choice of relativizer.

Embedded clause subject position strongly favors the choice of *that* ( $p_i = .89$ ). In contrast, the value for use of a *wh*-form as the subject of the embedded clause is only .33. In our sample, *wh*-forms are used primarily with human antecedents in this position. No cases of zero relative pronoun in subject position were found in our sample, although such forms are known to occur in casual speech, in constructions like those in (3) (cf. reports in Adamson 1992; Biesenbach-Lucas 1989; Kikai, Schleppegrell, and Tagliamonte 1989; and especially Shnukal 1981):

TABLE 5  
Syntactic Position in the Embedded Clause: *wh*-, *that*, and  $\emptyset$

	<i>wh</i> -			<i>that</i>			$\emptyset$		
	<i>N</i>	%	$p_i$	<i>N</i>	%	$p_i$	<i>N</i>	%	$p_i$
Subj	176	47	.33	199	53	.89	1 <sup>a</sup>	0	.01
DirObj <sup>b</sup>	19	7	.08	127	46	.71	130	47	.77
PrpObj	39	41	.37	28	29	.63	29	30	.68
Gen. <sup>c</sup>	10	100	—	0	0	—	0	0	—
Loc.	41	87	.95	5	11	.21	1	2	.09
Man. <sup>d</sup>	0	0	—	1	8	.16	12	92	.99
Temp.	3	33	.67	2	22	.37	4	44	.69
Total	288	35	—	362	44	—	177	21	—

- a. One spurious token was introduced to avoid a knockout in VARBRUL analysis.
- b. Includes Complements (cf. table 2).
- c. Tokens excluded from VARBRUL analysis.
- d. Factor included in analysis of *that* and  $\emptyset$  only.

3. a. We want a guy  $\emptyset$  can do the job.
- b. There's this guy  $\emptyset$  lives across the street from me.

In order to preserve comparability of results across the three analyses and to avoid having to drop the subject data because of a knockout, a single spurious zero token in subject position was added for the analysis of the zero relative pronouns. This accounts for the  $p_i$  of .01 reported for this category in table 5 despite the fact that  $N = 0$ .

A choice of *that* or zero is favored in direct and indirect object positions in the embedded clause, while *wh-* forms are strongly disfavored in direct object position ( $p_i = .08$ ) and moderately disfavored as prepositional objects ( $p_i = .37$ ). Two considerations help to explain these results. First, the low values for *wh-* words may reflect a general avoidance of objective case *whom*. Speakers avoid a form they consider affected but preserve grammatical correctness by choosing *that* or zero. (Our written data did not contain any human antecedents for relative pronouns in this position.) Second, the relatively higher probability of use of a *wh-* form as a prepositional object is a result of our decision to include pied-piped prepositional objects (which are categorically *wh-*) along with stranded prepositional objects under this factor.

Results for other embedded-clause factors indicate that locatives strongly favor the choice of a *wh-* form ( $p_i = .95$ ). Zero forms are almost categorically favored for manner clauses ( $p_i = .99$ ), and both *wh-* and zero forms are favored for temporal clauses ( $p_i = .67$  and  $.69$ , respectively). The results for manner and temporal clauses, however, must be qualified. Our data contained only 13 tokens of manner clauses and nine of temporal clauses.

ADJACENCY OF THE ANTECEDENT AND RELATIVIZED ELEMENT. The effects of the adjacency of the relative pronoun and its antecedent on relativization strategies have not been examined in previous studies. Results are summarized in table 6.

These results indicate that zero is favored when the relativized element is adjacent to its antecedent or separated by another relative clause. The occurrence of some other constituent between the antecedent and the

TABLE 6  
Adjacency of Relative Pronoun and Antecedent: *wh-*, *that*, and  $\emptyset$

	<i>wh-</i>	<i>that</i>	$\emptyset$
Relpro adjacent to antecedent	.42	.48	.76
Relpro separated by another relative clause	.59	.44	.62
Relpro separated by another element	.49	.58	.16

relativized element strongly inhibits use of zero form. Usually an explicit relative pronoun is required to allow for parsing, as in (4):

4. a. Things may turn up that we don't see now. (Nixon 1974, 182)
- b. We put a number of people into this that we had at work on other things. (Nixon 1974, 237).

If, however, we rewrite (4a) and (4b) so that the relative pronoun is adjacent to its antecedent, zero becomes an option. Compare the following possible revisions:

5. a. Things ( $\emptyset$ ) we don't see now may turn up.
- b. We put a number of people ( $\emptyset$ ) we had at work on other things into this.

The choice between a *wh*- form or *that*, however, is not significantly affected by the adjacency of the antecedent and the relativized element.

**SUMMARY OF RESULTS.** To summarize, our analysis has identified several categorical environments and a rich patterning of variable constraints. Nonrestrictive clauses and genitives are always realized as *wh*-. Moreover, *wh*- forms are favored in formal writing and for human antecedents in embedded-clause subject position. Choice of *that* is favored in informal speech and for nonhuman antecedents. Zero is moderately favored for human antecedents, especially in embedded-clause direct-object position, and in informal speech. Zero is very strongly disfavored in embedded-clause subject position in both speech and writing. Zero is also strongly disfavored when the relative clause is separated from the antecedent by a phrase other than another relative clause.

**ANALYTICAL ISSUES.** We now turn to some analytical issues that have arisen in connection with this study and can potentially surface in any study of variation. These arise when the quantitative analyst is confronted with a choice among several alternative classifications of elements in the data, and the alternatives imply or derive from different theoretical positions on the variable in question. Often such a choice will have a substantial effect on the quantitative results. We will illustrate this problem with some examples from the study of English relatives.

The first example concerns the treatment of pied-piped prepositions. When a preposition is fronted along with the relative pronoun in English, the only grammatical relative pronoun choice is a *wh*- word. Compare the stranded prepositions in (6) with their pied-piped counterparts in (7).

6. a. . . . the bed which he slept in last night. . . .
- b. . . . the bed that he slept in last night. . . .
- c. . . . the bed  $\emptyset$  he slept in last night. . . .

7. a. . . . the bed in which he slept last night. . . .
- b. \* . . . the bed in that he slept last night. . . .
- c. \* . . . the bed in  $\emptyset$  he slept last night. . . .

What does this imply for our analysis? If we treat this as a distinct environment in our quantitative study, it is a categorical context. The normal procedure is to exclude such contexts from the analysis; in fact, this is a mathematical requirement of the VARBRUL algorithm. But such a procedure implies a particular theoretical position about the derivation of these examples. Note that preposition-fronting is itself optional in English; it can apply as in (7a), or fail to apply as in (6). The question is, how is the choice speakers must make about pied-piping related to the choice they make about relative pronouns? If we model these choices as ordered rules, the question is one of rule ordering: does pied-piping precede or follow relative-pronoun selection? The analytical procedure of excluding these tokens from our quantitative study implies that pied-piping precedes pronoun choice, so that the pied-piped context is then available as a conditioning factor to the relative pronoun rule. (We are assuming that it is impossible in an ordered rule-set to have a variable rule be conditioned by an event which follows in the derivation.)

Now such a theoretical model of events is not unreasonable. It is essentially the kind of model one would associate with transformational theories based on the *Aspects* model (Chomsky 1965). But the point we wish to make here is that there are other possible theoretical treatments of these facts that would imply different analytical judgments about this situation. First, one might wish to argue for the reverse ordering of these two rules: relative-pronoun choice precedes pied-piping. This would be a very reasonable proposal in, for example, a Hallidayan systemic grammar, where syntagmatic decisions about the arrangement of items are usually postponed until the end of a derivation (Halliday 1985). This theoretical position would imply an entirely different treatment of these data: all prepositional objects should be classified together, regardless of their syntagmatic position with respect to the preposition. Under this arrangement, the categorical nonoccurrence in examples (7b) and (7c) would be due to a constraint on pied-piping: given that you have selected a relative pronoun other than *wh*-, pied-piping is rendered impossible; it only becomes an option when *wh*-forms are selected.

Another possible theoretical position on this issue is that of a unification-based grammar (e.g. GPSG, etc.; cf. Sheiber 1986), which takes a nonderivational “declarative” approach, stating a set of unordered constraints on cooccurrences. This would make the whole issue of ordering irrelevant. The problem would reduce to some statement to the effect that

only *wh-* forms successfully unify with the structure in which the preposition and relative pronoun are both at the front of the relative clause. In this case our quantitative analysis would depend on the kinds of structures we wished to define as possible representations of relative clauses.

What effects would these contrasting theoretical positions have on our quantitative study? The results are quite different. Our corpus includes 36 tokens of relative clauses followed by pied-piped prepositions plus a *wh-* form and 60 tokens of other prepositional-object relatives. Table 7 shows the effects of including and excluding the pied-piped cases. If we adopt a theory that takes the fact of pied-piping as an available conditioning factor which turns out to be a knockout, we exclude them from the analysis and find that other prepositional objects have only 5% *wh-* forms; in fact *wh-* becomes the least-favored choice in this context. But a unified treatment of prepositional objects that includes the pied-piped cases gives a figure of 41% *wh-* forms: *wh-* becomes the most frequent choice in this context.

The effects of this decision on the VARBRUL probabilities is also nontrivial. As we would expect, the values for the syntactic function factor group are the most affected. Prepositional objects, with a value of .07, are least-favored environments for *wh-* realizations when fronted cases are excluded. But they climb to a middle position, with a value of .37, greater than those for subjects and objects, when pied-piped cases are included. (Compare figures in table 5 for subject and object.)

This decision also has a substantial effect on the significance of other factor groups in the analysis of the deletion strategy. Excluding the pied-piped cases causes the factor groups of animacy and channel of communication to lose significance. Why should this be the case? Evidently it results from the skewed distribution of the pied-piped cases across these two dimensions. Many more pied-piped cases were found in writing than in speech (which is not surprising, given the highly formal character of this construction), and almost all of them occurred with inanimate antecedents. Therefore animacy and channel constraints lose their statistical significance when these extreme values are removed from the corpus.

TABLE 7  
Syntactic Function: Comparison of Results of Alternative Treatments  
of Fronted Prepositions

	<i>Fronted Prepositions Included</i>			<i>Fronted Prepositions excluded</i>		
	<i>N</i>	<i>%</i>	<i>p<sub>i</sub></i>	<i>N</i>	<i>%</i>	<i>p<sub>i</sub></i>
<i>wh-</i>	39	41	.37	3	5	.03
<i>that</i>	28	29	.63	28	47	.72
∅	29	30	.68	29	48	.77

Which result is the more accurate or correct, and which of the competing theories do we find to be superior? One is tempted to try to give a quantitative answer, in line with a tradition dating back to Labov's (1969) copula study. But we wish to claim that in fact THERE IS NO CORRECT QUANTITATIVE ANSWER. What measure would we use to try to answer this question? Log-likelihood, chi-square per cell, iterations to convergence? All of these are perturbed by the different numbers of cells and tokens in the competing analyses, and even if these could be controlled for, it is not clear that a small difference in such a global measure would justify the selection of one theory over another. Other kinds of evidence would be required as well.

We suggest that the better question to ask about quantitative findings is: Which of these analyses is more revealing for particular purposes? If our interest is focused primarily on the syntactic function of elements, then we will find the figures with pied-piped cases included more revealing: the high rates of *wh*-realization here illustrate the unremarkable fact that prepositions prefer bona fide NPs as their objects, rather than COMPs, traces, or gaps. But if our attention is focused more on the syntagmatic flexibility of these elements, then the results of excluding the pied-piped cases are probably more interesting. They show very low rates of *wh*-occurrence and high rates of zero. This reflects that fact that, aside from some highly constrained contexts like pied-piping, prepositions in English actually have strongly particle-like characteristics, among which is the ability to dangle without objects. And finally, if we are examining the spoken/written distinction, it would seem that throwing away all tokens of a construction which is far more prevalent in writing (the pied-piped cases) would obscure rather than illuminate our quest. Thus each analysis illuminates some angle on the truth, but they do not by themselves allow us to choose among conflicting theoretical treatments.

A second example of this class of analytical problems is the ternary character of the variable under study. With binary variable rules consisting of an input and an output, how do we get three surface alternants? Many scenarios are possible, depending on a series of analytical decisions. First there is the question of what form is taken as basic or underlying: one could postulate any one of the three surface alternants, or some abstract alternative, giving at least four possible starting points for our variable rules. Next, one must decide whether the rules are to be ordered, so that some forms can potentially be intermediate steps in the derivation of others; alternatively, one could use unordered rules without intermediate steps (cf. Sankoff and Rousseau 1989). Some possible models for relatives given by different decisions on these factors are given in figure 1.

FIGURE 1  
Models of Rule-Ordering for Relative Pronoun Choice

Ordered Rules	Unordered Rules
a. Underlying form: <i>wh</i> - 1. <i>wh</i> - → <i>that</i> 2. <i>that</i> → $\emptyset$	c. Underlying form: <i>wh</i> - 1. <i>wh</i> - → <i>that</i> 2. <i>wh</i> - → $\emptyset$
b. Underlying form: $\emptyset$ 1. $\emptyset$ → <i>that</i> 2. <i>that</i> → <i>wh</i> -	d. Underlying form: RELPRO 1. [RELPRO] → <i>wh</i> - 2. [RELPRO] → <i>that</i> 3. [RELPRO] → $\emptyset$

Still other approaches are possible. For example, one could generate the three forms by three different phrase-structure rules (one starting the relative clause with a *wh*- form, one with *that* as a COMP structure, and one with just the basic clause structure and a gap). The choice would be the selection of one of these optional PS rules to realize the relative clause role, which we could characterize as a kind of metarule about which rules to use.

Given all these options, how do we conduct our analysis? Once again the quantitative results will be affected, at least somewhat, by our decision. Consider for example the categorical use of *wh*- in nonrestrictive relative clauses. A model like (b) which postulates zero as the underlying form, and then orders rules of  $\emptyset \rightarrow \textit{that}$  and *that* → *wh*-, would report a 100% rate of application of the rule which generates *that* (rule b1), in order to feed the second rule (*that* → *wh*-) and get the right outcome for this environment. By contrast, a model which postulated underlying *wh*- would report a zero probability of *that* realization in this environment. Once again, we cannot ask the quantitative model to tell us which analysis is “right.” Probabilities can be assigned to any of the rule systems we have described, and it is likely that at least some of them will do essentially equivalent jobs in fitting the data. So other kinds of evidence must be adduced to decide questions of this sort. In this case, for example, we could note that there are several environments that appear to demand a *wh*- form, and at least one that virtually excludes a zero form, but none that require *that* or zero. Therefore *wh*- would seem a better choice as underlying form on grounds of its wider distribution.

CONCLUSION. We wish to highlight here two aspects of the foregoing discussion: the problem of explanation and the problem of methods. The rich and diverse patterns of constraints on variation in the choice of relative pronouns illustrate two explanatory principles. First, we suspect

that facilitation of parsing is strongly implicated as an explanation for several of the principal constraints: low rate of zero relatives for subject NPs, low rate of zero in nonadjacent relative clauses, and categorical *wh*-occurrence in nonrestrictives and genitives. Second, several of the constraints seem strongly influenced by normative attitudes: the high rate of *wh*-usage in writing can be explained in terms of Kroch and Small's theory of grammatical ideology, and the low rate of use of *wh*-forms in objective positions with human antecedents strongly implicates avoidance of the highly marked pronoun *whom*, possibly motivated by fear of being prescriptively wrong.

On the question of methods, we have been pessimistic as to the capacity of quantitative analysis to resolve purely theoretical debates. We should add that we believe this to be true only insofar as the theories make no substantively different empirical predictions. Furthermore, this is a question of sufficiency of evidence rather than of necessity; quantitative evidence may still be necessary to support arguments derived from other sources. But the converse of this issue should also receive our attention: that is that there is rarely any theory-neutral way to approach an analytical problem. Any analysis of this sort requires us to make theoretical assumptions and decisions. It is essential for purposes of reliability and validity to spell out those decisions for our scholarly audience. Without knowing what treatment has been adopted of, say, fronted prepositions or nonrestrictive relatives, the numbers that another study of this phenomenon might produce would be difficult to interpret and not comparable with our results. Variationists should be mindful of this general problem in the task of exploring the relations between theoretical constructs and observable language use.

#### NOTE

Data for this paper were coded by student linguists at Cornell and Stanford Universities (M. Blocker, E. Hume, A. M. Hutchison, A. Hylden, R. Mark, B. McElhinney, P. O'Neill, D. H. Rau, D. Silva), whose assistance is gratefully acknowledged. An earlier version of the paper was presented in 1989 at NWAV-18 at Duke University.

#### REFERENCES

- Adamson, H. D. 1992. "Social and Processing Constraints on Relative Clauses." *American Speech* 67: 123-133.
- Biesenbach-Lucas, Sigrun. 1987. "The Use of Relative Markers in Modern American English." *Variation in Language: NWAV-XV at Stanford*. Ed. Keith Denning, Sharon Inkelas, Faye C. McNair-Knox, and John Rickford. Stanford, CA: Stanford U, Dept. of Linguistics, 13-21.

- Carden, Guy. 1973. "Dialect Variation and Abstract Syntax." *Some New Directions in Linguistics*. Ed. Roger Shuy. Washington: Georgetown UP, 1–34.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT.
- Dekeyser, Xavier. 1984. "Relativizers in Early Modern English: A Dynamic Quantitative Study." *Historical Syntax*. Ed. Jacek Fisiak. Berlin: Mouton, 61–87.
- Guy, Gregory R. 1978. "Locating Variation in the Grammar." *Penn Review of Linguistics* 3: 17–23.
- . 1989. *MacVarb: Variable Rule Analysis on the Macintosh*. Software and user documentation. Stanford, CA: Stanford U, Dept. of Linguistics.
- Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. London: Arnold.
- Horvath, Barbara. 1987. "Text in Conversation: Variability in Story-Telling Texts." *Variation in Language: NWAV-XV at Stanford*. Ed. Keith Denning, Sharon Inkelas, Faye C. McNair-Knox, and John Rickford. Stanford, CA: Stanford U, Dept. of Linguistics, 212–23.
- Kikai, Akio, Mary Schleppegrell, and Sali Tagliamonte. 1987. "The Influence of Syntactic Position on Relativization Strategies." *Variation in Language: NWAV-XV at Stanford*. Ed. Keith Denning, Sharon Inkelas, Faye C. McNair-Knox, and John Rickford. Stanford, CA: Stanford U, Dept. of Linguistics, 266–78.
- Kroch, Anthony, and Catherine Small. 1978. "Grammatical Ideology and Its Effect on Speech." *Linguistic Variation: Models and Methods*. Ed. David Sankoff. New York: Academic, 45–55.
- Labov, William. 1969. "Contraction, Deletion, and Inherent Variability of the English Copula." *Language* 45: 715–62.
- . 1973. "The Boundaries of Words and Their Meanings." *New Ways of Analyzing Variation in English*. Ed. Charles-James N. Bailey and Roger W. Shuy. Washington: Georgetown UP, 340–73.
- . 1986. "Sources of Inherent Variation in the Speech Process." *Invariance and Variability in Speech Processes*. Ed. Joseph S. Perkell and Dennis H. Klatt. Hillsdale, NJ: Erlbaum, 402–25.
- Naro, Anthony. 1981. "The Social and Structural Dimensions of a Syntactic Change." *Language* 57: 63–98.
- Nixon, Richard. 1974. *The White House Transcripts: Submission of Recorded Conversations to the Committee on the Judiciary of the House of Representatives*. Ed. Gerald Gold. Toronto: Bantam.
- Olson, David R. 1977. "From Utterance to Text: The Bias of Language in Speech and Writing." *Harvard Educational Review* 47: 257–81.
- Prideaux, Gary D., and William J. Baker. 1986. *Strategies and Structure: The Processing of Relative Clauses*. Philadelphia: Benjamins.
- Rissanen, Matti. 1984. "The Choice of Relative Pronouns in 17th Century American English." *Historical Syntax*. Ed. Jacek Fisiak. Berlin: Mouton, 417–35.
- Romaine, Suzanne. 1982. *Socio-Historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge UP.
- Rousseau, Pascale, and David Sankoff. 1978. "Advances in Variable Rule Methodology." *Linguistic Variation: Models and Methods*. Ed. David Sankoff. New York: Academic, 57–69.
- Sankoff, David, and Pascale Rousseau. 1989. "Statistical Evidence for Rule Ordering." *Language Variation and Change* 1: 1–18.
- Schiffrin, Deborah. 1986. "Turn-Initial Variation: Structure and Function in Conversation." *Diversity and Diachrony*. Ed. David Sankoff. Amsterdam: Benjamins, 367–80.

- Sheiber, Stuart. 1986. *An Introduction to Unification-Based Approaches to Grammar*. Stanford: CSLI.
- Shnukal, A. 1981. "There's a Lot Mightn't Believe This . . . : Variable Subject Relative Pronoun Absence in Australian English." *Variation Omnibus*. Ed. David Sankoff and Henrietta R. Cedergren. Edmonton: Linguistic Research.
- Tottie, Gunnel. 1993. "The Man Ø I Love: An Analysis of Factors Favouring Zero Relatives." 22nd Conference on New Ways of Analyzing Variation in English and Other Languages. University of Ottawa, October.

#### 1995 NEH OVERVIEW OF ENDOWMENT PROGRAMS

The 1995 issue of the *National Endowment for the Humanities Overview of Endowment Programs* is now available. The *Overview* describes the 35 funding opportunities available from the NEH as well as application deadlines through 1995, how to get application forms, addresses and phone numbers of state humanities councils, and whom to contact for assistance or more information. To obtain your free copy of this year's *Overview*, write, phone, or e-mail:

NEH Overview, Room 402  
1100 Pennsylvania Avenue, NW  
Washington DC 20506

Phone: 202/606-8400  
NEH Bulletin Board: 202/606-8688  
E-mail: [nehopa@gwuvm.gwu.edu](mailto:nehopa@gwuvm.gwu.edu)