

# 1 Comparing Syntactic Variables

---

*Sali A. Tagliamonte*

## 1.1 Introduction

Studies of language variation and change in the variationist tradition (Labov 1969) are based on the assumption that both linguistic and social factors are implicated in language variation and change. Indeed, the embedding of linguistic phenomena in the speech community is one of the five founding problems for the study of variation (Weinreich, Labov and Herzog 1968, 185–6):

- i. *Constraints* What are the constraints on change?
- ii. *Transition* How does language change?
- iii. *Embedding* How is a given language change embedded in social and linguistic systems?
- iv. *Evaluation* How do members of a speech community evaluate a given change and what is the effect of this evaluation on the change?
- v. *Actuation* Why did a given linguistic change occur at a particular time and place that it did?

This chapter grapples specifically with the embedding problem and the evaluation problem, which involve both social and linguistic systems. On one hand, language-internal mechanisms are involved, including analogy, reanalysis, metaphorical extension and others (Joseph 2004, 61). On the other hand, social influences can impact variation as well, from broad categorizations such as (biological) sex, level of education, social class and other externally defined factors (Labov 1963, 1966) to style, attention to speech, audience and stance (Bell 1984, 2002).

Over the past forty years or more, variationist work has consistently demonstrated these cross-cutting influences of the language/society interface (e.g. Labov 1963; Sankoff 1980; Tagliamonte 1998). Studies using quantitative

I am grateful to the Economic and Social Research Council of the United Kingdom (ESRC) for research grants from 1996 to 2003, and the Social Science and Humanities Research Council of Canada (SSHRC) for research grants from 2001 to the present. I thank my co-authors in the original studies of these variables – Alexandra D’Arcy, Helen Lawrence and Jennifer Smith – and the many research assistants, from 1997 to the present, who extracted and coded the data.

methods typically test social categories such as age, sex, education and job type along with a broad range of linguistic factors. More recently, additional predictors have led to novel insights such as considerations of processing (Grondelaers et al. 2009), psycholinguistic influences (Grondelaers and Speelman 2007), prescriptivism-related predictors (Hinrichs, Szmrecsanyi and Bohmann 2014) and stance (Kiesling 2009). However, simply testing and reporting the results of a myriad of variegated predictors is not sufficient to understand and explain syntactic variation; it is also necessary to understand what the function of the variation is in the grammar and what it means in the history and current state of the community. As I demonstrate in this chapter, deconstructing two syntactic variables and comparing the patterns of variation across them offers fresh insights into the relationship between linguistic and social predictors in the analysis of variation and its explanatory adequacy.

To elucidate these ideas I consider two syntactic linguistic variables: the alternation between *that* and zero complementizers, henceforth variable (that); and between *that*, *who* and zero relative pronouns, henceforth variable (who). The foundation of my observations and discussion comes from previous analyses conducted on these variables in large spoken language corpora from two communities: York, England (YRK) and Toronto, Canada (TOR) (Tagliamonte and Smith 2005; D'Arcy and Tagliamonte 2010; Tagliamonte 2012).<sup>1</sup>

## 1.2 The Variables – Complementizers and Relative Pronouns

Variable (that) and variable (who) involve syntactic structure. Both focus on the linguistic form that links a subordinate clause with a matrix clause. The first variable is the choice of complementizer, as in (1). The second case is the choice of subject relative pronoun, as in (2). Note the alternation in closely proximate utterances.

- (1) I always said *that* I wouldn't leave it a five year gap . . . and so I always said  $\emptyset$  I wanted them very close together. (YRK, female, 31)
- (2) There's one lady *that* lives in my building *who* had been in a concentration camp. (TOR, female, 83)

The mechanisms that underlie the frequency and patterning of these variants are essentially linguistic, involving the nature of the syntactic constituents and grammatical categories involved, for example subject versus object, lexical

<sup>1</sup> Due to the nature of these data sets, one constructed in the early 2000s on materials collected in 1997 and the other constructed in 2010 on materials collected in 2003–4, in some cases current statistical practices cannot be implemented.

verb, syntactic construction (e.g. existential) and others. Issues involved with syntactic reanalysis come to the forefront with regard to the complementizer (Elsness 1984; Thompson and Mulac 1991a; Cheshire 1996; Jaeger 2005; Tagliamonte and Smith 2005; Torres Cacoullos and Walker 2009). Social, interactional and register-based factors are prominent in discussions of the choice of relative pronouns (Shnukal 1981; Guy and Bayley 1995; Ball 1996; Sigley 1997; Nevalainen and Raumolin-Brunberg 2002; Tagliamonte 2002b; D'Arcy and Tagliamonte 2010). The question is: What do the different variable profiles of these syntactic variables reveal about the synergy of social and linguistic factors more generally?

At the outset, there are important distinguishing characteristics that set these two variables apart. One variable simply involves presence or absence of *that* in its function as a complementizer. The other involves a similar overt versus covert alternation, *that* and zero, but with the added dimension of an overt *wh-* form, mostly *who*. While both linguistic variables involve the same overt form *that*, the internal structure of each variable is unique. The different variants of each variable appear with varying degrees of productivity. For variable (that), the zero variant dominates and there is no attested social nuance attached to the use of zero. The internal origin, that is, change from below (see Labov 1972), of the zero variant of the complementizer may be the explanation. For variable (who) the *that* variant dominates, but *who* is prescribed as standard. In this case, the origin and history of *who* is key. The *wh-* forms entered the English relative pronoun system as an exogenous change, instigated by contact with another system (i.e. French) (see D'Arcy and Tagliamonte 2015). This external origin of the *wh-* variants as a change from above has a major impact on the way that linguistic and social factors play out in variation.

### 1.3 The Data

The data under consideration comprise an uncommonly large compendium of vernacular spoken language. These materials were collected in the UK and Canada between 1997 and 2010 according to standard sociolinguistic procedures, using ethnological fieldwork, conversational interviewing (i.e. the 'sociolinguistic interview'; and judgement sampling (Labov 1972; Tagliamonte 2006; Schilling 2013). In the UK, the data come from York, a city in the north-east of England (Tagliamonte 1998), and small towns and villages all over the UK (Tagliamonte 2013). In Canada, the data come from Toronto, the largest city in Canada (Tagliamonte 2003–6). The corpora comprise speakers born and raised in the communities, and in most cases from pre-adolescents to senior citizens. For all intents and purposes, these data provide a comprehensive body

of materials for analysing variable (that) and variable (who) in two major varieties of English.

#### 1.4 Method

In order to study linguistic variation so as to provide a useful characterization of the grammatical mechanism(s) giving rise to variability, it is necessary to use careful methodological practice and appropriate statistical tools. Each of the ensuing analyses was founded on the exacting procedures developed in language variationist and change research. First, all contexts of the variable were circumscribed, extracted and coded according to existing protocols (Tagliamonte and Smith 2005; Tagliamonte, Smith and Lawrence 2005). Second, the main constraints tested in contemporary studies in the extant literature were operationalized. Third, each of the variables was probed using distributional analyses and cross-tabulations (e.g. Guy 1993; Wolfram 1993; Tagliamonte 2006). Finally, statistical tools were used to model the simultaneous application of multiple predictors (Labov 1994a, 3) while at the same time taking into account their possible interactions. In these investigations fixed effects logistic regression using Goldvarb (Sankoff, Tagliamonte and Smith 2005) was employed in the original analyses, and some newer techniques using R, a language and environment for statistical computing (R Core Team 2007), were implemented in this updated comparison. The results expose regularities and tendencies from the data, namely the predictors that predispose the occurrence of the variants and the strength of the influence of each predictor. The choice of linguistic form may be probabilistically conditioned by specific characteristics of the internal linguistic environments in which it occurs, providing decisive insights into the inner mechanisms of grammatical organization.

The evidence for interpreting and understanding the results from the analysis comes from (1) frequency, (2) patterns, that is, the constraint hierarchy of the relevant predictors, and (3) the relative strength of the predictors (Tagliamonte 2002a, 2006). If the variable is conditioned by the same factors across communities, which in turn are ranked in the same order, this will be evidence of shared grammatical patterns (Poplack and Tagliamonte 2001; Labov 2007). If the patterns of the variants are found to be systematically different between the UK and Canada, then this will be evidence of locally situated usage patterns. Synthesizing all this information will lead to a greater understanding of the underlying processes that have led to contemporary patterns. This will establish a more accurate perspective of the synchronic variability across diverse populations and offer a map of the trajectory of linguistic change.

The evidence from variant frequency provides an indication of the appropriation and diffusion of forms as well as a baseline for comparison. However, frequency alone is not definitive because it can fluctuate considerably from one

individual to the next, or one situation to the next, under the influence of topic, style or another external force (Tagliamonte 2002a). Patterns (i.e. constraints) are known to remain stable across diverse circumstances. They provide a measure of the variable grammar of the new form and offer insight into its phase of development (Poplack and Tagliamonte 2001, chapter 5). The method of comparing frequency, constraints and the relative weight of factors is often referred to as ‘comparative sociolinguistics’ (e.g. Tagliamonte 2002a). This technique was specifically developed for assessing correspondences across corpora and so is particularly appropriate for making comparisons across the UK and Canadian communities.

### 1.5 Variable (that)

Variation in the presence versus absence of the English complementizer *that* versus zero is widely studied and considered ubiquitous in English (Pesetsky 1982; Warner 1982; Elsness 1984; Rissanen 1991; Thompson and Mulac 1991a; Rohdenburg 1998; Tagliamonte and Smith 2005). Jespersen (1954, 38) suggested that the alternation is simply the result of ‘momentary fancy’. Since then, two theories regarding this variation have been proposed. The first claims that the zero variant is the result of grammaticalization of certain collocations into epistemic parentheticals, particularly *I think* (Thompson and Mulac 1991a, 1991b). These constructions are thought to have developed out of the structure in which complementizers are found, that is, *I think that*, but instead of functioning as a matrix clause these collocations have become reanalysed as discourse-pragmatic features indicating the speaker’s degree of ‘commitment to a proposition’ or to his or her beliefs about it (Denis 2015, 152, fn. 1). The second theory suggests that the alternation between the overt form and the zero variant is the result of processing effects whereby the complementizer, *that*, only occurs under conditions of structural complexity (Rohdenburg 1998). Which explanation is correct? The next step is to subject these hypotheses to empirical testing.

A valuable starting point is to situate variation. Where does it fit in time, space and with respect to society? Consider the use of complementizer variation in the history of English, as in Figure 1.1.

This trajectory shows that the zero complementizer increases incrementally from Wycliffe’s sermons (1300s) (Warner 1982) to Early Modern English (1400–1700) (Rissanen 1991). These points in the evolution of this system come from written materials. The last three points on the trajectory come from contemporary spoken English: two Canadian locations (Quebec and Toronto) and one British (York) (Tagliamonte and Smith 2005; Torres Cacoullos and Walker 2009; Tagliamonte 2012). While written and spoken data are undoubtedly very different

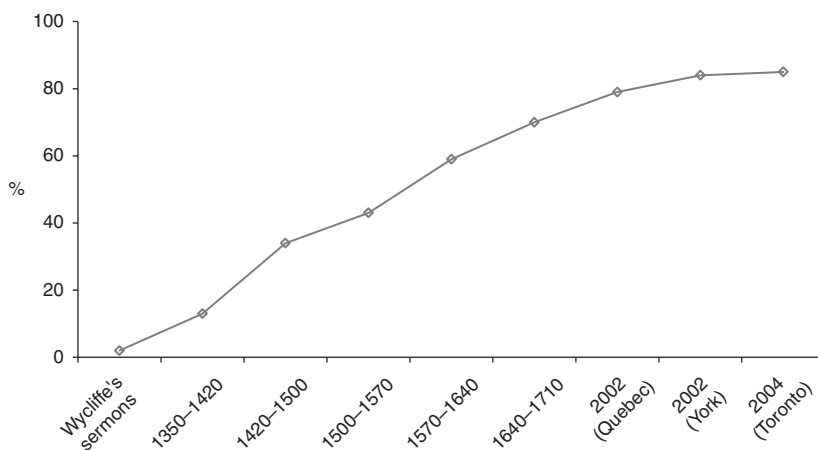


Figure 1.1 Frequency of zero complementizers in the history of English

types of data, Figure 1.1 shows a regular pattern of development towards more and more zero forms.

Despite the longitudinal trajectory of change visible in Figure 1.1, the zero variant is not a change that has gone to completion. In contemporary spoken language, variation between overt and zero forms can be found within most speakers, and in the same speaker in the same stretch of discourse, as in (3a–b) from Walter Edwards,<sup>2</sup> an elderly man aged seventy-two born and raised in York, England.

- (3) a. Uh my mother decided *that* uh she'd have a- a new house built. (YRK, male, 72)  
 b. My mother, at the end of the meal, suddenly decided  $\emptyset$  she'd go to- in to town. (YRK, male, 72)

The question is, what is influencing the choice of one variant over the other? The extensive body of research on this variable has uncovered a set of significant constraints operating on the choice of form. These include the matrix verb (e.g. *think*), the grammatical person of the matrix, tense and intervening material (e.g. *I really think*). These constraints can be related to the two theories about the variation. If there is an ongoing process of grammaticalization in which particular collocations such as *I think*, *I mean*, *I guess* are gradually becoming epistemic parentheticals, then certain features of the contexts will become more prominent, such as the verb *think*, first-person and present tense. Indeed, there will also be an intervening period of ambiguity during which time

<sup>2</sup> All names are pseudonyms that have been specifically selected to reflect the intrinsic nature of the original.

some of the contexts that have no *that* but appear in constructions that are consistent with matrix + complement clauses may be interpreted as either complements or epistemic parentheticals, as in (4a–b). However, because the grammatical development takes place over a period of time, different constructions can remain layered in the language as well as in individuals: *I think* can function as an epistemic parenthetical (4a–b) or it can function as the matrix clause of a complement, as in (5a–b).

- (4) a. *I think* they mostly went into service in those days. (YRK, female, 63)  
 b. *I think* we pretty well all sound the same, you know. (TOR, male, 72)
- (5) a. *I think that* if you start sitting about vegetating you've had it haven't you? (YRK, female, 63)  
 b. *I think that* the government is doing it on purpose. (TOR, male, 72)

Through the transition period tendencies can be observed in the linguistic data. Epistemic parentheticals tend to occur with first- and second-person subjects over other grammatical persons (Thompson and Mulac 1991a, 242), and the expression of the speaker's beliefs are typically constructed with present tense (Tagliamonte and Smith 2005). This is observed earlier in (4) as well as in (6).

- (6) a. *I guess* we're not doing that this year. (TOR, female, 19)  
 b. *You know* they didn't know what you were saying. (TOR, female, 83)  
 c. *I mean* I used to go down to the Kensington Market. (TOR, male, 60)

Table 1.1 shows what happens when all the constructions in the data that comprise *I think*, *you know* and *I mean* are examined separately. The frequency of zero complementizers is near categorical, suggesting that these constructions have already undergone reanalysis to epistemic parentheticals and should be removed from consideration when the variation between *that* and zero is under the microscope.

Once these collocations are removed, the remaining data set still exhibits robust variation, including the possibility of overt *that* when these same matrix verbs (i.e. *think*, *know*, *mean*) occur in linguistic environments other than their collocation, that is, with grammatical subjects, for example *she* or *we* with *think* and *mean*, as in (7a–b), or first-person singular, *I*, with *know*, as in (7c). This

Table 1.1 *Frequency of zero in I think, you know, I mean*

Matrix collocation	%	N	Total N
<i>I think</i>	98.5	974	989
<i>you know</i>	99	535	541
<i>I mean</i>	100	428	428

suggests that the foundation for the emergence of epistemic parentheticals was a variable system that was already hospitable to this development.

- (7) a. She *thinks that* fish can get in your pool. (TOR, female, 13)  
 b. We didn't know *that* I'd actually go there. (TOR, female, 12)  
 c. I *know that* he is going to sell this in a week. (TOR, female, 54)

Another influence that operates on this variation is the nature of the subject of the complement clause. Pronominal subjects are said to be more likely to encode the topic of the discourse (Thompson and Mulac 1991a, 248). Thompson and Mulac's claim is that this makes the preceding material – if it is a matrix clause – more likely to be epistemic, and therefore the zero option more favourable, as in (8). Of course, the preceding clause could also be a main clause, as in (7b–c). While these differences cannot easily be determined on a case-by-case basis, they can be discovered by quantitative analysis, from which the relevant patterns emerge as trends in variable data.

- (8) a. I *think*  $\emptyset$  it's really funny. (YRK, male, 20)  
 b. You *know*  $\emptyset$  they didn't think it was worthwhile. (TOR, male, 72)

In contrast, when the complement subject is a noun phrase, as in (9), Thompson and Mulac (1991a, 248) claim that the matrix subject is more likely to function as the topic, making it more prone to be non-epistemic, producing an overt complementizer.

- (9) a. I *know that* Kennedy won the election. (TOR, male, 66)  
 b. I *think that* a bit of that must have rubbed off on me. (YRK, male, 58)

Another explanation for variation between *that* and zero is that there are psycholinguistic influences underlying the realization of the complementizer. In this view, anything that increases the processing load of the matrix + complement construction will lead to more use of an overt complementizer. Matrix/complement constructions are considered more complex when they involve negation, past tense, complex tenses and modals, leading to more use of *that*, as in (10). Similarly, if any linguistic material intervenes between the matrix clause and the complement clause, this leads to more overt forms as well. This can be observed in the examples in (11).

- (10) a. We [*weren't*] aware that there was such devastation. (TOR, female, 75)  
 b. I [*can't*] see really that it would change a great deal. (TOR, female, 81)  
 c. I [*'m not saying*] that we would have gotten as far like. (TOR, female, 22)
- (11) a. I must say that [as much as I miss the way things used to be] I'm having the time of my life. (TOR, male, 61)  
 b. The man knows [it was um for economic reasons] that [um they wanted to um] give us what they could. (TOR, female, 34)



Table 1.2 *Fixed effects logistic regression of predictors conditioning zero complementizer – Toronto (Canada), excluding tokens of I think, I mean and you know*

Input probability		0.87		
Total N		2,148		
		FW	%	N
Lexical verb in matrix clause				
<i>think</i>		.70	93	461
<i>say</i>		.54	85	647
<i>know</i>		.46	85	196
<i>other</i>		.38	80	680
<i>tell</i>		.30	64	164
	<i>Range</i>	40		
Matrix subject				
1st-person singular		.60	88	985
Other pronoun		.45	83	894
Other		.30	67	267
	<i>Range</i>	30		
Additional elements in matrix verb phrase				
Nothing		.56	87	1,512
Something		.37	76	636
	<i>Range</i>	19		
Complement clause subject				
Pronoun		.54	86	1,703
Other		.37	74	443
	<i>Range</i>	17		
Intervening material				
None		.52	85	1,963
Some		.35	71	185
	<i>Range</i>	17		
Matrix verb tense				
Present		.56	87	917
Past		.44	84	861
	<i>Range</i>	12		

All these influences are multiplex and variegated. Which of them exert a statistically significant influence on the spoken language data? Tables 1.2 and 1.3 display fixed effects logistic regressions of the simultaneous contribution of these factors when all of them are included in a statistical model. This method permits the combined contribution of all the contextual factors to be modelled simultaneously and determine which of them contribute statistically significant effects to the variation, the nature of the constraints and their relative strength (see, e.g., Tagliamonte 2006). Note that these models exclude the (near) categorical epistemic parenthetical cases (see Table 1.1).

Table 1.3 *Fixed effects logistic regression of predictors conditioning zero complementizer – York, England, excluding tokens of I think, I mean and you know*

Input probability	0.89		
Total N	1,810		
	FW	%	N
Lexical verb in matrix clause			
<i>think</i>	.75	95.5	829
<i>say</i>	.57	76.8	228
<i>other</i>	.34	65.4	619
<i>know</i>	.33	68.7	134
<i>Range</i>	41		
Matrix subject			
1st-person singular	.72	91.8	1,167
NP	.39	61.0	246
Other pronoun	.38	61.2	397
<i>Range</i>	34		
Verb tense			
Present	.59	87.4	1,285
Past	.42	65.0	525
<i>Range</i>	17		
Intervening material			
None	.58	85.0	1,425
Some	.42	65.7	385
<i>Range</i>	16		
Complement clause subject			
Pronoun	.58	82.9	1,356
Other	.42	74.9	464
<i>Range</i>	16		
Additional elements in matrix verb phrase			
Nothing	.57	85.1	1,356
Something	.43	68.3	464
<i>Range</i>	14		

Table 1.2 shows that verb type trumps every other predictor, with a range value of forty: the matrix verbs *think* especially and *say* strongly favour the zero complementizer. The nature of the complement subject also exerts a strong influence. Other predictors are significant, but less so.

The same model can be tested in another community, in this case a different majority variety of English, namely British English as spoken in York. The regression returns the same result. Verbs such as *think* and *say* favour zero. Simple present tense favours zero. First-person singular favours zero and pronominal subjects in the complement clause favour zero. The model is virtually identical to the one in Table 1.2 for Toronto.

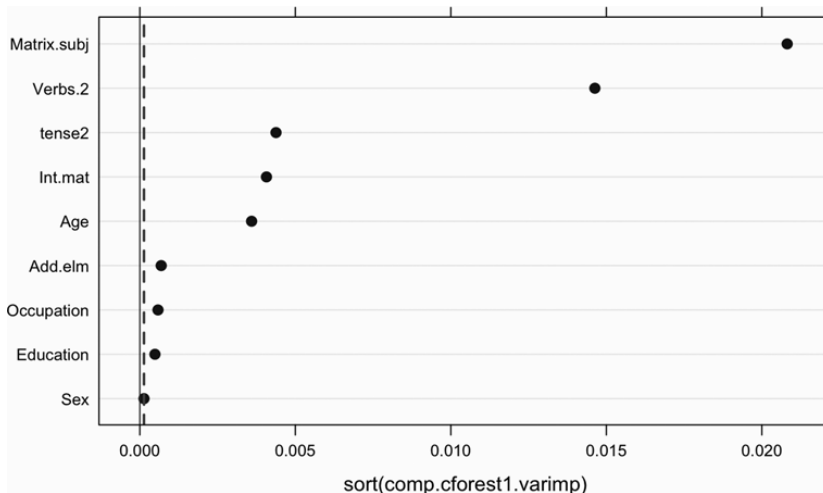


Figure 1.2 Random forest analysis of internal and external predictors for complementizer variation in York

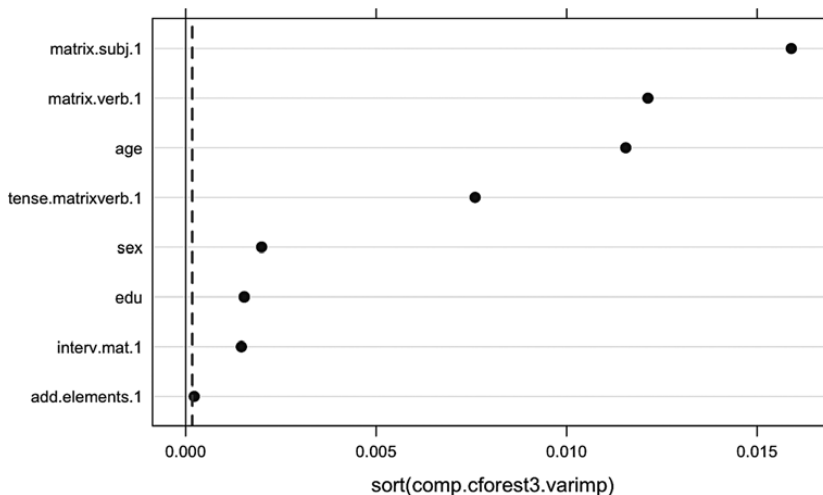


Figure 1.3 Random forest analysis of internal and external predictors for complementizer variation in Toronto

An updated perspective on the York and Toronto data can be achieved using a random forest analysis (Strobl, Malley and Tutz 2009; Tagliamonte and Baayen 2012), which can expose the relative importance of the social and

linguistic predictors involved in complementizer variation. This is shown in Figure 1.2 (York) and Figure 1.3 (Toronto). Note that these models include all the data in the analysis, including the near categorical cases in Table 1.1, but exclude the subject of the complement clause to facilitate comparison.

In this type of analysis, the farther to the right of the dot, the greater the importance of the predictor. Predictors to the right of the vertical dashed line are significant. The solid line shows zero on the x axis. Figure 1.2 (York) shows that although both linguistic and social factors are significant, the matrix subject and matrix verb are the most important predictors. Of much less importance are the tense of the matrix clause, intervening material and individual age. Social factors such as occupation, education and sex are even less important. Figure 1.3 (Toronto) shows a similar profile in that the variation is again almost entirely explained by the matrix subject and matrix verb, whereas social factors are less important. While the relative strength of the social factors, particularly age, differs across varieties in both locales, the variation is overwhelmingly governed by the same two linguistic constraints. Note that the York and Toronto data structures were built separately at different points in time and are not internally consistent with each other, so the effect of variety cannot be tested across them.

Taken together, these results demonstrate that contemporary English speakers use complementizer *that* variably determined by a suite of the following strong linguistic contexts:

- with verbs other than *think*, *say* or *know*
- with tense and aspects other than the simple present
- with matrix subjects other than first-person singular ‘I’, and
- with NP subjects in the complement clause.

### 1.5.1 Summary

Variable (that) in contemporary English is a stable variable that has been part of the English language since Wycliffe’s sermons in the 1400s (Figure 1.1) and is present in both written and spoken registers in the early twenty-first century. In the contemporary literature it is widely studied and consistently exhibits intricate linguistic conditioning. Statistical modelling of the variable constraints on its use in two varieties of English demonstrates that the matrix verb and matrix subject are the most important influences, followed by other internal factors and social influences. The significance and direction of these patterns in the data do not support the old idea of ‘momentary fancy’ (Jespersen 1954, 38), but instead align with contemporary hypotheses of the grammatical development of certain constructions into epistemic parentheticals, for example, *I think*. In the canonical structure of matrix + subordinate sequence, for example, *I think that’s it*, these constructions may appear to be matrix clauses, but they are not.

Over and above these frequent collocations, a range of constraints – matrix verb, complement subject, tense, intervening material – maintain a strong effect. However, the idea that *that* surfaces only in contexts where the syntactic structure is complex and/or interrupted by false starts and disfluency is also very strong. In essence, the syntactic strings themselves are not monolithic. In some cases, the construction has already grammaticalized into a different form and function. In other cases, a matrix + complement construction in which use of an overt complementizer emerges when there is complexity in the syntactic structure is still structurally sound. While additional social influences are present (see Figures 1.2 and 1.3), these operate well below the linguistic constraints in the system.<sup>3</sup>

## 1.6 Variable Relative Pronouns

Variation in the forms used to mark English relative clauses is also widely studied, and variation in form appears to be present in every variety of English that has been studied to date (e.g. Quirk 1957; Shnukal 1981; Rissanen 1984; Montgomery 1989; Guy and Bayley 1995; Tottie 1995; Ball 1996; Tottie and Harvie 2000; Beal and Corrigan 2002; Nevalainen and Raumolin-Brunberg 2002; Tagliamonte 2002b; D’Arcy and Tagliamonte 2010; Cheshire, Adger and Fox 2013).

However, the relative pronoun system is critically partitioned by type in terms of its preferred variants. First, there are two types of relative clauses. Non-restrictive relatives, as in (12), present add-on information that is supplemental to what is expressed in the rest of the sentence. These types are near categorically marked with *wh-* forms, either *who* or *which* (Quirk et al. 1985, 1239; Huddleston and Pullum 2002, 1035). The nature of these relative clauses as additional commentary is made clear in (12c), where *which* does not refer to the ‘doorman’.

- (12) a. In those days he built a log house, *which* is still sitting here. (TOR, male, 83)  
 b. I worked with a guy named Robin B, *who*’s pretty famous. (TOR, male, 40)  
 c. Now we have a doorman, *which* I like. (TOR, female, 22)

The disproportionate, in fact mostly categorical, use of *wh-* forms in non-restrictive relatives is why most studies only include restrictive relative clauses in the analysis of variation. If non-restrictive clauses were included, they would raise the incidence of *wh-* forms and mask the variation within the restrictive relative cohort.

<sup>3</sup> Recent research suggests that stance, a pragmatic factor, is also significantly implicated in complementizer variation, but consistent with the results here, less important than purely linguistic factors (Gadanidis et al., 2021).

Restrictive relative clauses can be identified semantically by the fact that they ‘serve to identify their antecedent’ (Denison 1998, 278). This is where the relative clause system is variable, since the relative clause can be marked by either *that*, *who* or zero, as in (13). However, the varying linguistic characteristics of restrictive relative clauses by antecedent type, grammatical role and other factors distinguish relative clauses near categorically by form to the point where they have been described as ‘different populations’ (Ball 1996, 233). Therefore, at the outset of variation analysis it is critical to separate subject relative clauses, as in (13a–c), from all other types, as in (13d–e).

- (13) a. He’s a person *that* has qualities *that* tick me off. (TOR, female, 19)  
 b. There was a huge trestle  $\emptyset$  went across the Etobicoke Creek to carry the trolley on. (TOR, male, 82)  
 c. I’ve got a friend *that* lives across the street. (TOR, male, 11)<sup>4</sup>  
 d. Of course any samples *that* you got in the candy line, you ate. (TOR, female, 81)  
 e. Well, nudes is all  $\emptyset$  I’ve ever sold. (TOR, female, 22)

Moreover, as we shall see, the nature of the antecedent is also of great importance, namely the contrast between human antecedents, as in (14), or non-human antecedents, as in (15). Note too that the examples in both (14) and (15) come from the same individual, a female aged 81 in (14) and a male aged 82 in (15), demonstrating intra-speaker variability.

- (14) a. The chap  $\emptyset$  I was going with went over.  
 b. I have a sister *who* is a nun.  
 c. The boys *that* played rugby with my brother . . . (TOR, female, 81)
- (15) a. Again that’s all a tradition *that’s* gone by the boards.  
 b. There was a huge trestle  $\emptyset$  went across the Etobicoke Creek to carry the trolley on. (TOR, male, 82)

With these characteristics of the variability in mind, the next step is to probe the historical trajectory of the forms vying for marking relative clauses in the history of English in order to understand how this system evolved. Figure 1.4 shows the results of a quantitative investigation of the relative pronoun system in the history of English (Ball 1996).

This trajectory shows that the *wh*- forms increased dramatically from the seventeenth to the eighteenth century to the point of virtual saturation of the system by the twentieth century, at least for subject relatives with human subjects. Non-human relative clauses follow the same trajectory, but remain

<sup>4</sup> I include existential constructions, cleft sentences and possessives in the restrictive relative clause data. Although the syntactic status of these constructions is controversial (e.g. Ball 1996, 235), they are included here in order to view their different distributional patterns in the data compared to other constructions (see also Tagliamonte et al. 2005, 96–7).

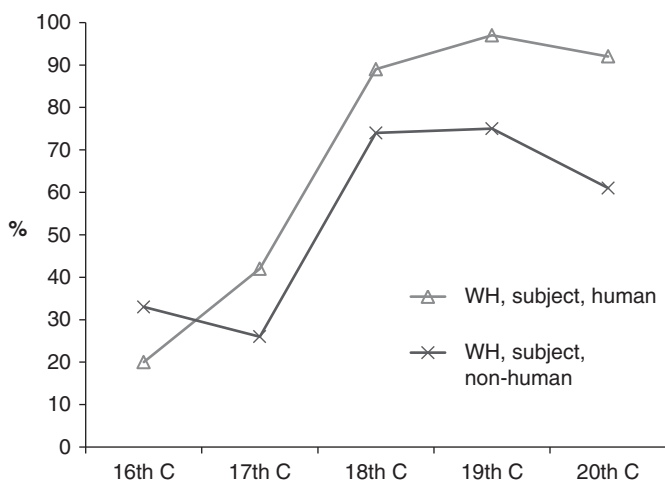


Figure 1.4 Frequency of restrictive *wh*-relatives in the history of English

more robustly variable. As with previous studies of the complementizer system, these results come from written materials. Here too the question arises as to what is influencing the choice of forms and, further, what led to the dramatic rise in *wh*-relatives in the eighteenth century?

Although Figure 1.4 makes it appear that the *wh*-relatives are moving towards completion, several studies have questioned this conclusion. In a widely cited statement from Romaine (1982, 212), she claims that ‘the infiltration of WH into the relative system can be seen as completed in the modern written language. ... but it has not really affected the spoken language’.

Let us now turn to an analysis of the contemporary spoken language. As observed in (14–15), two overt forms and zero can be found within most speakers, and in the same speaker in the same stretch of discourse.

Tables 1.4 and 1.5 display the distribution of relative pronouns in subject relative clauses and in non-subject relative clauses respectively in the city of York in northern England (Tagliamonte 2002b).

With this perspective, it now becomes apparent that the *wh*-form *who* is actually a fairly minor part of the system – occurring at a frequency of only 21 per cent in its most favoured context: human antecedents in subject function. Instead, *that* is a dominant form in both subject and non-subject relatives, and in the latter,  $\emptyset$  actually dominates.

Figure 1.5 exposes yet another dimension to this variability. The form *who* is dramatically more frequent in York, an urban centre in northern England

Table 1.4 *Variation of relative markers (subject only)*

	%	N
<i>that</i>	62	850
Ø	12	170
<i>who</i>	21	294
<i>which</i>	3	46
<i>what</i>	1	16
<i>as</i>	0.07	1
Total		1,377

Table 1.5 *Distribution of relative markers (non-subject)*

	%	N
<i>that</i>	41	358
Ø	54	465
<i>who</i>	1	8
<i>which</i>	3	23
<i>what</i>	2	15
Total		869

(Yorkshire), than in the outlying small towns and villages of Cumnock (Ayrshire), Maryport, Wheatley Hill, Tiverton (Devon) or Wincanton (Somerset).

Given the results exhibited in Figure 1.4, it is clear that *who* has not diffused very far in spoken varieties. Moreover, it has penetrated these spoken vernaculars to different degrees. In Tagliamonte (2002b, 103) I argued that the geographic split was due to ‘the relative proximity of the dialects to mainstream norms’. As an urban centre, York was further ahead in the encroachment of *who* into the English relative system, while the small, peripheral communities lagged behind.

While this geographic perspective is informative with respect to the frequency of *who* for subject relatives, it is now important to understand what governs its choice. Table 1.6 shows a fixed effects regression of the constraints underlying these overall frequencies.

Consistent with Figure 1.5, *who* is most likely in York. In certain locales – the northern towns of Ayrshire (southwest Scotland) and Maryport (northwest England) – *that* and zero predominate. Zero predominates in the south, in Somerset and Devon. As expected, *who* is favoured with human subjects. Sentence type is most relevant for the choice of zero. Zero is highly favoured



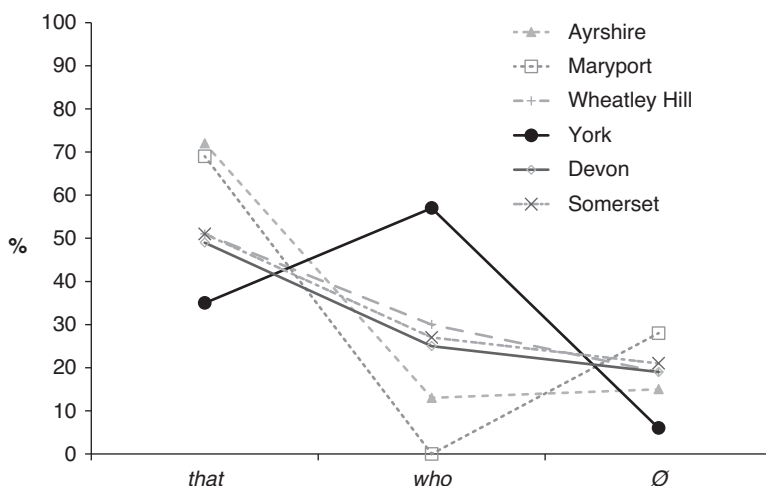


Figure 1.5 Frequency of relative *who* across communities, UK

for clefts, possessive constructions and particularly existentials. In an earlier study, separate analyses of the different places also revealed cross-dialectal consistency of the internal constraints on relative pronoun choice (Tagliamonte et al. 2005). However, at least in the UK, *who* is still undergoing diffusion into the relative pronoun system.

The next step is to corroborate these results with a study of another major variety of English, in this case Canadian English as spoken in Toronto. In the interests of brevity, an analysis of subject relative clauses only is presented. In this context of maximal variation amongst all the forms – *who*, *that* and zero – let us first assess whether the linguistic constraints on subject relative pronouns are parallel in Toronto.

D’Arcy and Tagliamonte (2010, 392) demonstrated that use of *who* is strongly influenced by antecedent type, consistent with the findings for the UK shown in Table 1.6, as can be seen in Table 1.7.

Human subjects (humans plus the lexical item *people*) are the main locus for the use of *who*. The socially stratified Toronto data permits quantitative assessment of the social conditions on the variation by including categories such as sex, education and job type in the model that were not possible in the UK corpora.<sup>5</sup> This also enables me to probe the potential pathway *who* may be

<sup>5</sup> The Roots Corpora are internally consistent in that the individuals are all fairly old and mostly less educated. Where scrutiny of these factors is possible, however, there is an indication that higher education and community leadership lead to higher use of *who*. For further discussion, see Tagliamonte et al. 2005, 92).

Table 1.6 *Three fixed effects logistic regression analyses of the contribution of factors to the probability of subject restrictive relative clauses in the UK*

	<i>that</i>	zero	<i>who</i>	Total N
Community				Ns/cell
Ayrshire	.68	.59	.26	355
Maryport	.63	.78	0	65
Wheatley Hill	.44	.51	.47	113
York	.38	.30	.74	470
Somerset	.45	.69	.44	208
Devon	.48	.53	.43	166
Range	23	48	48	
Antecedent type				
Human	.74	.55	.86	927
Non-human	.38	.41	.02	450
Range	36	14	84	
Sentence type				
Other	.58	.29	.55	819
Cleft, possessive	.52	.60	.49	332
Existential	.21	.94	.34	226
Range	37	65	21	

Table 1.7 *Distribution of relative markers by animacy of the antecedent, subject function only in Toronto, Canada*

	<i>that</i>		<i>who</i>		Ø		Total N
	%	N	%	N	%	N	
Things	96.2	583	0.0	0	3.5	21	606
Humans	45.2	306	50.8	344	3.7	25	677
<i>people</i>	41.9	122	54.6	159	3.1	9	291
Collectives	71.4	50	24.3	17	5.7	4	71
Animals	87.1	27	6.5	2	6.5	2	31
Total N		1,088		522		61	1,675

Not shown: *whose* and *which* (N = 5)

taking across time by using the apparent time construct as a proxy for change in progress (Labov 1994b). Table 1.8 displays a fixed effects logistic regression analysis of social predictors.

Table 1.8 shows that even in a place where subject relative *who* represents nearly half of the system (input = 0.488), social constraints are highly important. We can now see that it is highly favoured amongst middle-aged speakers (30–

Table 1.8 *Fixed effects logistic regression analyses of the contribution of factors to the probability of subject relative who – Toronto, Canada*

Input	0.488		
Total N	968		
	FW	%	N
Age			
10–16	.41	50.7	140
17–29	.45	55.3	219
30–59	.71	72.2	259
60–92	.40	35.4	350
<i>Range</i>	<i>31</i>		
Education			
+ post-secondary	.59	62.1	605
– post-secondary	.27	24.1	212
<i>Range</i>	<i>32</i>		
Occupation			
Professional	0.55	57.4	484
Non-professional	0.40	36.0	253
<i>Range</i>	<i>15</i>		
Sex			
Female	[0.53]	58.6	490
Male	[0.47]	45.2	478

59), who use it significantly more than anyone else in the community. Moreover, post-secondary education and professional-level jobs significantly favour its use.

It is curious that despite this, the variation does not exhibit an effect of speaker sex as per Labov's Principles 3 and 4 (Labov 2001), in which women are widely held to lead linguistic change. Here, sex is not selected as significant despite the fact that females show a higher frequency of use (58.6% > 45.2%). It may be that this is not a change in progress, but the result of long-term stability. Probing the data further, we conducted a cross-tabulation of sex and occupation, as in Figure 1.6.

Figure 1.6, depicting the proportions, suggests that neither effect is significant. Chi square tests of the differences reveal non-significance for both:  $p = 0.0247$  for professionals, and  $p = 0.7802$  for non-professionals. However, when we coded the data for the nature of the conversational dyad, as in Figure 1.7, a new insight emerged.

Figure 1.7 shows that the depressed use of *that* and therefore heightened use of *who* have to do with the nature of the dyad amongst professional-level speakers. When both the interviewer and the interviewee are women (F+F), the rate of *who* in subject relative pronouns rises, distinguished from all other

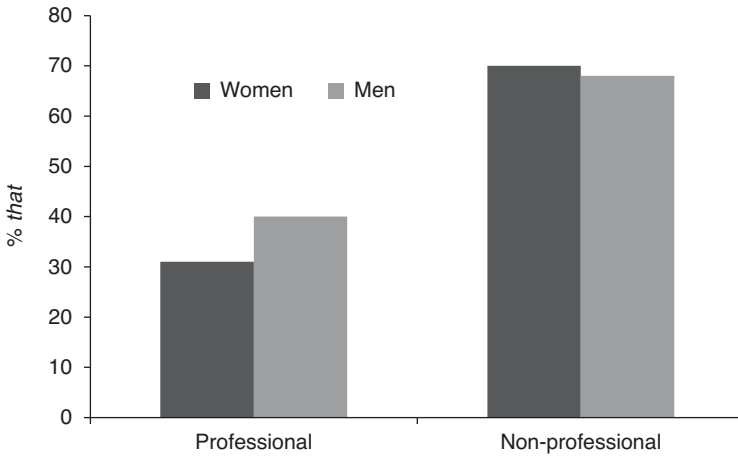


Figure 1.6 Distribution of *that* in subject relative clauses by job type

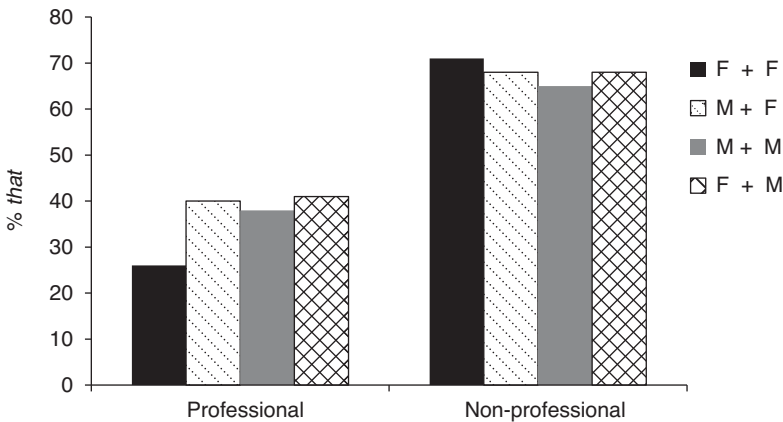


Figure 1.7 Distribution of *that* in subject relative clauses by job type and interlocutors

dyads of professionals as a whole, with a chi square of  $p = 0.0098$  and no significant difference amongst the other dyads.

The conclusion is that contemporary English speakers use relative *who* according to a suite of predictors as follows:

- variably with human antecedents in subject relatives
- amongst middle-aged speakers

- who are educated professionals
- and especially when two women are talking together.

### 1.6.1 *Summary of Relative Pronouns*

The findings up to this point confirm that the use of *who* is a partitioned variable in both contemporary varieties of English: the syntactic function of the antecedent and humanness explain most of the variation (Tagliamonte 2002b, Tagliamonte et al. 2005, D'Arcy and Tagliamonte 2015). Within the highly circumscribed locations of variability relative pronoun shows signs of being stable and age-graded in contemporary English. Relative *who* is much less frequently used than previously hypothesized, and highly socially circumscribed. Consistent with its original entry into the relative system (Nevalainen and Raumolin-Brunberg 2002), *who* is used most often by middle-aged people who are educated and hold a professional-level job or, as in the case of northern British communities (Tagliamonte and Smith 2005), hold local leadership positions. The discovery in the Toronto English data of a new 'interactional' effect, women talking to women, adds an additional status-based nuance to this suite of predictors. Women, widely known to favour prestige forms (Principle 2; Labov 2001) have even more enhanced uses. Relative *who* was a change from above, but it appears to be maintained in English as a stable linguistic variable that marks prestigious associations and social alignment, a fact that offers a possible test of the famed 'sociolinguistic monitor' (see Smith and Holmes-Elliott, this volume). This finding offers a test of the sociolinguistic monitor (Gadanidis et al., 2021) and is consistent with the results from Smith and Holmes-Elliott (Chapter 2 in this volume) where interviews with a local interviewer differed substantially from those of a non-local.

## 1.7 **Discussion**

I have now provided an overview of findings arising from the analysis of two syntactic variables in two major varieties of English. Variable (that) and variable (who) are both ubiquitous, and have been studied quantitatively from the perspective of reanalysis, complexity and language variation and change. It is evident that exhaustive probing of patterns from all sources of potential influence – social, geographic, linguistic, cognitive – is necessary in order to 'get to the bottom' of the variable system and to understand what is going on. Of particular importance is to first identify the distributional characteristics of the data set and to distinguish between categorical, near-categorical and variable sections of the system under investigation (see an alternative procedure in Chapter 3 of this volume), where both categorical and variable uses are included in the meaning hypotheses tested).

It is also critical to study syntactic variables in the context of their social and historical evolution in order to understand why they operate as they do 'on the ground' in the existing sociolinguistic situation. On one hand, based on the extensive study of complementizer variation in written materials, the trajectory into the modern period using spoken data shows stable variation based in large part on grammaticalization, processing and complexity. On the other hand, wide-ranging studies of relative pronouns in written materials led to the expectation of increasing frequency of *who*; however, the spoken vernacular shows resistance to this development. In these conditions of diachronic trajectories of change and ongoing variation, it is not sufficient simply to test internal factors influencing syntactic variables. Their structural importance is only relevant in the context of the internal character of the variable system. Indeed, both syntactic variables seem to function meaningfully for different externally motivated situations. While social factors are of lesser importance to variation between the overt and zero complementizer, grammatical change is key, and once epistemic parentheticals have split away, processing and pragmatic factors can continue to influence variation. For the relative pronoun system, internal factors are strong and important, but the impact of age, sex, education and job type are crucial for understanding the current situation. With these considerations in mind, let me return to an explicit discussion of what the approach I have taken here offers for understanding the functions of complementizer *that* and relative *who* in contemporary English. Taking the overarching patterns for each variable as a focal point, what does making the choice of *that* and *who* encode for language users?

### 1.7.1 Function of 'That'

In matrix + complement clause constructions, complementizers are typically thought to mark the relationship between a matrix and complement clause (Brittain 1778). However, many researchers have argued that the use of *that* also signals register. It is associated with written language, particularly formal and institutional genres (Biber and Finegan 1994, 1997). As a consequence, it is considered less personal, friendly and emotive (Storms 1966; Quirk et al. 1972; Leech and Svartvik 1975; Huddleston and Pullum 2002). In some cases, researchers have said that *that* is simply the result of 'momentary fancy' (Jespersen (1954, 38). However, what *that* actually seems to be doing, at least in contexts where it is still functioning as a complementizer, is ensuring intelligibility. If you want to make yourself absolutely clear, you use it.

### 1.7.2 Function of 'Who'

Relative pronouns are typically thought to mark the type of relative clause subject, *who* for human beings (e.g. Denison 1998, 278), *which* for things (e.g.

Curme 1947) and *that* for either people or things (Curme 1947, 166; Swan 1995). However, there are inconsistencies in the literature as to whether these claims are true, and, indeed, just how far *who* has infiltrated the contemporary English system (Romaine 1982; Ball 1996). In fact, the story of *who* in contemporary English – at least in the spoken languages – presents a decidedly social story. It is reported to be used in high registers, and is considered a learned variant with formal connotations (Dekeyser 1986; Nevalainen and Raumolin-Brunberg 2003). It is used by certain individuals, with a profile of advanced education, involvement in community affairs and with class aspirations (Romaine 1982; Beal and Corrigan 2002; Tagliamonte et al. 2005). Moreover, it is used most often in female-to-female speech (D'Arcy and Tagliamonte 2010). If you want to sound smart, you use it.

All this serves to emphasize that it is useful and important for the explanatory adequacy of our interpretations to assess syntactic variables in context, not simply to assess syntactic configurations or provide complex statistical models, nor even to elucidate single tokens or socially imbued interpretations. I suggest it is the dialectic of linguistic and social interpretations that are key to understanding variation. Linguistic, social, stylistic, cognitive, prescriptive and possibly other factors impact syntactic variation. However, the details are inevitably different, depending on the nature of the variable, whether it has evolved as change from above or change from below, how it is situated in time and place, and in the nuances one variant or the other holds in discourse. The frequency of forms, the details of the predictors, in patterning and strength combine to inform explanatory insights. Synthesizing across all these influences leads to informed explanations.

## 1.8 Conclusion

The results of distributional analysis and statistical modelling with a comparative perspective grounded in social and historical context provide insights into the mechanics of variation and offer discernment to the embedding problem and the evaluation problem. First, I can now categorize the two syntactic variables according to type. In the case of complementizers, the variation comprises an overt and unrealized form. In contrast, the choice of subject relative pronouns is almost always between competing overt forms (i.e. *that* and *who*), which have contrasting historical origins and a legacy of social evaluation. Whether this is a systematic difference between linguistic variables that predominately involve information load and clarity and therefore implicate processing, that is, cognitive factors, and those that predominately involve a choice amongst distinct forms with differing social evaluations and therefore implicate external factors, remains for future comparative study. Second, I can evaluate the application or not of different types of predictors. While variation

in the choice of complementizer is relatively indifferent to sex, education or job type, the choice of relative pronoun is highly predisposed to these same factors as well as interactional factors. Third, the relative contributions of predictors add another dimension. In the case of complementizers, the overwhelming influence of verb and matrix subject demonstrates how particular collocations may have begun to grammaticalize away from matrix + complement constructions into epistemic parenthetical (e.g. *I think*), while the preponderance of *who* for subject, animate antecedents reflects a well-known typological pattern favouring the marking of human subjects that is overlaid with social evaluation from the speech community. These interpretations of the dialectic between linguistic and social embedding are key to understanding how variation functions in the speech community and brings us closer to addressing the elusive actuation problem, all part of the oeuvre that Labov set his sights on in the early 1960s ‘to gather data from the secular world’ (Labov 1972, xvi–xvii).

### References

- Ball, Catherine. 1996. ‘A Diachronic Study of Relative Markers in Spoken and Written English’. *Language Variation and Change* 8 (2): 227–58.
- Beal, Joan C. and Karen P. Corrigan. 2002. ‘Relatives in Tyneside and Northumbrian English’. In *Relativization on the North Sea Littoral* (LINCOM Studies in Language Typology), edited by Patricia Poussa, 125–34. Munich: Lincom Europa.
- Bell, Allan. 1984. ‘Language Style as Audience Design’. *Language in Society* 13 (2): 145–204.
- Bell, Allan. 2002. ‘Back in Style: Reworking Audience Design’. In *Style and Sociolinguistic Variation*, edited by Penny Eckert and John R. Rickford, 139–69. Cambridge: Cambridge University Press.
- Biber, Douglas and Edward Finegan. 1994. *Sociolinguistic Perspectives on Register*. New York: Oxford University Press.
- Biber, Douglas and Edward Finegan. 1997. ‘Diachronic Relations among Speech-Based and Written Registers in English’. In *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, edited by Matti Rissanen, Terttu Nevalainen and Leena Kahlas-Tarkka, 253–75. Helsinki: Société Néophilologique.
- Brittain, Lewis. 1778. *Rudiments of English Grammar*. Louvain.
- Cheshire, Jenny. 1996. ‘That Jacksprat: An Interactional Perspective on English *That*’. *Journal of Pragmatics* 25: 369–93.
- Cheshire, Jenny, David Adger and Sue Fox. 2013. ‘Relative *Who* and the Actuation Problem’. *Lingua* 126 (1): 51–77.
- Curme, George O. 1947. *English Grammar*. New York: Barnes and Noble.
- D’Arcy, Alexandra and Sali A. Tagliamonte. 2010. ‘Prestige, Accommodation and the Legacy of Relative *who*’. *Language in Society* 39 (3): 383–410.
- D’Arcy, Alexandra and Sali A. Tagliamonte. 2015. ‘Not Always Variable: Probing beneath the Vernacular Grammar’. *Language Variation and Change* 27 (3): 255–85.



- Dekeyser, Xavier. 1986. 'English Contact Clauses Revisited: A Diachronic Approach'. *Folia Linguistica Historica* 7: 107–20.
- Denis, Derek. 2015. 'The Development of Pragmatic Markers in Canadian English'. PhD Dissertation, Department of Linguistics, University of Toronto.
- Denison, David. 1998. 'Syntax'. In *The Cambridge History of the English Language, 1776–Present Day*, edited by Suzanne Romaine, 92–329. Cambridge: Cambridge University Press.
- Elsness, Johan. 1984. 'That or Zero? A Look at the Choice of Object Clause Connective in a Corpus of American English'. *English Studies* 65: 519–33.
- Gadanidis, Timothy, Nicole Hildebrand-Edgar, Angelika Kiss, Lex Konnelly, Pabst, Katharina, Schlegl, Lisa, Umbal, Pocholo and Tagliamonte, Sali A. 2021. 'Integrating Qualitative and Quantitative Analyses of Stance: A Case Study of English That/Zero Variation'. *Language in Society*: 1–24. DOI: <https://doi.org/10.1017/S0047404521000671>.
- Grondelaers, Stefan and Dirk Speelman. 2007. 'A Variationist Account of Constituent Ordering in Presentative Sentences in Belgian Dutch'. *Corpus Linguistics and Linguistic Theory* 3: 161–93.
- Grondelaers, Stefan, Dirk Speelman, Denis Drieghe, Marc Brysbaert and Dirk Geeraerts. 2009. 'Introducing a New Entity into Discourse; Comprehension and Production Evidence for the Status of Dutch Er "There" as a Higher-Level Expectancy Monitor'. *Acta Psychologica* 130: 153–60.
- Guy, Gregory R. 1993. 'The Quantitative Analysis of Linguistic Variation'. In *American Dialect Research*, edited by Dennis Preston, 223–49. Amsterdam: John Benjamins.
- Guy, Gregory R. and Robert Bayley. 1995. 'On the Choice of Relative Pronouns in English'. *American Speech* 70: 148–62.
- Hinrichs, Lars, Benedikt Szmrecsanyi and Axel Bohmann. 2014. 'Which-Hunting and the Standard English Relative Clause'. International Society for the Linguistics of English 3 [ISLE], Zurich, 24–7 August 2014.
- Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Jaeger, Florian. 2005. 'Optional That Indicates Production Difficulty: Evidence from Disfluencies'. In *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop, 10–12 September 2005*, 103–8. Aix-en-Provence: DELIC, Université de Provence.
- Jespersen, Otto H. 1954. *A Modern English Grammar on Historical Principles. Part VI: Morphology*. London: George Allen and Unwin.
- Joseph, Brian. 2004. 'Rescuing Traditional (Historical) Linguistics from Grammaticalization Theory'. In *Up and Down the Cline: The Nature of Grammaticalization*, edited by Olga Fischer, Muriel Norde and Harry Perridon, 45–69. Amsterdam: John Benjamins.
- Kiesling, Scott F. 2009. 'Style as Stance: Can Stance Be the Primary Explanation for Patterns of Sociolinguistic Variation?' In *Sociolinguistic Perspectives on Stance*, edited by Alexandra Jae, 171–94. Oxford: Oxford University Press.
- Labov, William. 1963. 'The Social Motivation of a Sound Change'. *Word* 19: 273–309.
- Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

- Labov, William. 1969. 'Contraction, Deletion, and Inherent Variability of the English Copula'. *Language* 45 (4): 715–62.
- Labov, William. 1972. 'The Study of Language in Its Social Context'. In *Sociolinguistic Patterns*, edited by William Labov, 183–259. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1994a. *Principles of Linguistic Change, vol. 1: Internal Factors*. Cambridge: Blackwell Publishers.
- Labov, William. 1994b. 'The Study of Change in Progress: Observations in Apparent Time'. In *Principles of Linguistic Change: Social Factors*, edited by William Labov, 43–72. Cambridge: Blackwell Publishers.
- Labov, William. 2007. 'Transmission and Diffusion'. *Language* 83 (2): 344–87.
- Leech, Geoffrey N. and Jan Svartvik. 1975. *A Communicative Grammar of English*. London: Longman.
- Montgomery, Michael B. 1989. 'The Standardization of English Relative Clauses'. In *Standardizing English: Essays in the History of Language Change, in Honor of John Hurt Fisher*, edited by Joseph B. Trahern, 111–38. Knoxville: University of Tennessee Press.
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 2002. 'The Rise of Relative *Who* in Early Modern English'. In *Relativisation on the North Sea Littoral* (LINCOM Studies in Language Typology), edited by Patricia Poussa, 109–21. Munich: Lincom Europa.
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Pearson Education.
- Pesetsky, David. 1982. 'Complementizer-Trace Phenomena and the Nominative Island Constraint'. *Linguistic Review* 1: 297–343.
- Poplack, Shana and Sali A. Tagliamonte. 2001. *African American English in the Diaspora: Tense and Aspect*. Malden, MA: Blackwell Publishers.
- Quirk, Randolph. 1957. 'Relative Clauses in Educated Spoken English'. *English Studies* 38: 97–109.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1972. *A Grammar of Contemporary English*. New York: Harcourt Brace Jovanovich.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.
- R Core Team. 2007. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. [www.r-project.org](http://www.r-project.org).
- Rissanen, Matti. 1984. 'The Choice of Relative Pronouns in 17th Century American English'. In *Historical Syntax* (Trends in Linguistics. Studies and Monographs, 23), edited by Jacek Fisiak, 417–35. Berlin: Mouton de Gruyter.
- Rissanen, Matti. 1991. 'On the History of *That/Zero* as Object Clause Links in English'. In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, edited by Karin Aijmer and Bengt Altenberg, 272–89. London: Longman.
- Rohdenburg, Gunter. 1998. 'Clausal Complementation and Cognitive Complexity in English'. In *Anglistentag Erfurt*, edited by Fritz-Wilhelm Neumann and Sabine Schülting, 101–12. Trier: Wissenschaftlicher Verlag.
- Romaine, Suzanne. 1982. *Socio-Historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge University Press.

- Sankoff, David, Sali A. Tagliamonte and Eric Smith. 2005. 'Goldvarb X'. <http://individual.utoronto.ca/tagliamonte/goldvarb.html>.
- Sankoff, Gillian, ed. 1980. *The Social Life of Language*. Philadelphia: University of Pennsylvania Press.
- Schilling, Natalie. 2013. *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.
- Shnukal, Anna. 1981. 'There's a Lot Mightn't Believe This ... Variable Subject Relative Pronoun Absence in Australian English'. In *Variation Omnibus*, edited by David Sankoff and Henrietta R. Cedergren, 321–8. Edmonton, Alberta: Linguistic Research, Inc.
- Sigley, Robert. 1997. 'The Influence of Formality and Informality on Relative Pronoun Choice in New Zealand English'. *English Language and Linguistics* 1: 207–32.
- Storms, G. 1966. 'That-Clauses in Modern English'. *English Studies* 47: 249–70.
- Strobl, Carolin, James Malley and Gerhard Tutz. 2009. 'An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests'. *Psychological Methods* 14 (4): 323–48.
- Swan, Michael. 1995. *Practical English Usage*, 2nd ed. Oxford: Oxford University Press.
- Tagliamonte, Sali A. 1998. 'Was/Were Variation across the Generations: View from the City of York'. *Language Variation and Change* 10 (2): 153–91.
- Tagliamonte, Sali A. 2002a. 'Comparative Sociolinguistics'. In *Handbook of Language Variation and Change*, edited by Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes, 729–63. Malden, MA: Blackwell Publishers.
- Tagliamonte, Sali A. 2002b. 'Variation and Change in the British Relative Marker System'. In *Relativisation on the North Sea Littoral* (LINCOM Studies in Language Typology), edited by Patricia Poussa, 147–65. Munich: Lincom Europa.
- Tagliamonte, Sali A. 2003–6. Linguistic Changes in Canada Entering the 21st Century. Research Grant. Social Sciences and Humanities Research Council of Canada (SSHRC). #410–2003–0005. <http://individual.utoronto.ca/tagliamonte/research.html>.
- Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation* (Language in Society, 40). Malden, MA: Wiley-Blackwell.
- Tagliamonte, Sali A. 2013. *Roots of English: Exploring the History of Dialects*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. and R. Harald Baayen. 2012. 'Models, Forests and Trees of York English: Was/Were Variation as a Case Study for Statistical Practice'. *Language Variation and Change* 24 (2): 135–78.
- Tagliamonte, Sali A. and Jennifer Smith. 2005. 'No Momentary Fancy! The Zero 'Complementizer' in English Dialects'. *English Language and Linguistics* 9 (2): 1–21.
- Tagliamonte, Sali A., Jennifer Smith and Helen Lawrence. 2005. 'No Taming the Vernacular! Insights from the Relatives in Northern Britain'. *Language Variation and Change* 17 (2): 75–112.

- Thompson, Sandra and Anthony Mulac. 1991a. 'The Discourse Conditions for the Use of the Complementizer *That* in Conversational English'. *Journal of Pragmatics* 15: 237–51.
- Thompson, Sandra and Anthony Mulac. 1991b. 'A Quantitative Perspective on the Grammaticization of Epistemic Parentheticals in English'. In *Approaches to Grammaticalization*, edited by Elizabeth Closs Traugott and Bernd Heine, 313–29. Amsterdam: John Benjamins.
- Torres Cacoullos, Rena and James A. Walker. 2009. 'On the Persistence of Grammar in Discourse Formulas: A Variationist Study of *That*'. *Linguistics* 47 (1): 1–43.
- Tottie, Gunnel. 1995. 'The Man Ø I Love: An Analysis of Factors Favouring Zero Relatives in Written British and American English'. In *Studies in Anglistics* (Stockholm Studies in English), edited by Gunnel Melchers and Beatrice Warren, 201–15. Stockholm: Almqvist and Wiksell.
- Tottie, Gunnel and Dawn Harvie. 2000. 'It's All Relative: Relativization Strategies in Early African American English'. In *The English History of African American English*, edited by Shana Poplack, 198–230. Oxford: Blackwell.
- Warner, Anthony. 1982. *Complementation in Middle English and the Methodology of Historical Syntax*. London: Croom Helm.
- Weinreich, Uriel, William Labov and Marvin Herzog. 1968. 'Empirical Foundations for a Theory of Language Change'. In *Directions for Historical Linguistics*, edited by Winfred P. Lehmann and Yakov Malkiel, 95–188. Austin: University of Texas Press.
- Wolfram, Walt. 1993. 'Identifying and Interpreting Variables'. In *American Dialect Research*, edited by Dennis Preston, 193–221. Amsterdam: John Benjamins.