

Revisiting transatlantic relatives: Evidence from British and Canadian English

Stephen Levey & Heike Pichler

1. Introduction

Restrictive relative clauses, which delimit the reference of an antecedent head NP, are reported to exhibit alternation between WH-relativizers (e.g. *who*, *which*), *that* and *zero* (\emptyset) in contemporary spoken varieties of English, as illustrated in (1)-(3):

(1) The lady ***who*** we bought this house off was a friend of my mam's. (Salford/RoSE)

(2) You only get the nose ***that*** suits your face. (Berwick/BwE)

(3) It was easily the craziest thing \emptyset I've ever done. (Ottawa/OEC)

The literature addressing this variation reveals a preoccupation with two fundamental but unresolved issues: (i) To what extent have modern vernaculars been infiltrated by WH-relativizers, widely construed as extraneous, 'learned' variants (Romaine 1982: 213)? (ii) Is there a vernacular norm with respect to the distribution of relativizer variants (Ball 1996)?

In this study, we re-visit these issues, drawing on speech data from five contemporary varieties of English. Using the framework of comparative sociolinguistics (Tagliamonte 2013), we go beyond comparing relativizer *rates*, and closely examine the underlying *conditioning* of the alternation between restrictive WH-relativizers, *that* and *zero*, as revealed by multivariate analysis. This approach allows us to: (i) assess the infiltration of WH-relativizers into the selected varieties; and (ii) determine the extent to which the social and linguistic conditioning of relativizer choice exhibits cross-varietal parallels.

2. Data

Table 1 lists the corpora of sociolinguistic interviews targeted in our analysis. They include data from the geographically peripheral variety spoken in Berwick-upon-Tweed, a small town in the far north-east of England. Two additional northern varieties represent major urban centres: Tyneside English, spoken in the north-eastern Tyneside conurbation, which includes the city of Newcastle; and Salford English, spoken in the third-largest metropolitan area of England, Greater Manchester. In the south of England, we examine London English, a major metropolitan variety known to be rapidly innovating (Cheshire et al. 2011). Beyond the British Isles, we target Ottawa English, a mainstream urban variety spoken in Ontario, Canada. The majority of speakers in the five corpora share a white Anglo background. Those in the corpora from England are primarily of working-class origin, whereas the Ottawa corpus is balanced between middle- and working-class speakers. In each corpus, speakers are categorized into two age groups: <40 and >40 in order to detect any ongoing change in relativizer preferences in apparent time.

Table 1 Corpora¹

Variety/corpus	Location	N _{speakers}	Ages	N _{tokens}
Berwick (<i>BwE</i>)	north-east England	36	17-81	937
Tyneside (<i>DECTE</i>)	north-east England	53	16-84	630
Salford (<i>RoSE</i>)	north-west England	30	17-63	312
London (<i>LIC</i>)	south-east England	44	16-91	986
Ottawa (<i>OEC</i>)	Ontario, Canada	37	20-79	947
TOTAL		200		3812

¹ We thank Karen Corrigan, Jenny Cheshire, Paul Kerswill and their research teams for access to DECTE (Corrigan et al. 2012) and LIC (Kerswill et al. 2004-07).

By comparing these five geographically disparate varieties, we assess claims that WH-relativizers are less prevalent in northern British varieties compared to their southern British counterparts from which they reportedly spread (Buchstaller & Corrigan 2015: 88), as well as in mainstream urban Canadian English (D’Arcy & Tagliamonte 2010: 389). An innovative component of our analytical approach resides in its capacity to determine whether any cross-variety differences in variant rates mask structural similarities in the underlying conditioning of relativizer choice.

3. Method

From each corpus, we extracted every construction comprising a matrix clause (or a lone NP) and a restrictive relative clause. Following previous studies, we excluded near-invariant contexts (e.g. *all*-clefts) as well as other types of relative clause (e.g. non-restrictive and adverbial relative clauses). We coded the remaining 3812 restrictive relativizers for the extra-linguistic predictors variety, speaker sex and age, in addition to including linguistic predictors widely reported to condition relativizer choice, such as syntactic position of the relativizer, type and animacy of the antecedent head NP, adjacency of the relative clause to the head NP and length of the relative clause (see Tagliamonte et al. 2005). We also coded for less widely investigated predictors: the lexical category of the head NP and the grammatical subject of non-subject relatives (see Temperley 2003; Wiechmann 2015).

4. Variant distributions

Table 2 shows that although variants are not identically distributed across the datasets, all five varieties share the same major variant repertoire comprising *that*, *zero* and *who* in the same descending order of frequency. These three variants together account for between 94% (Salford) and 99.8% (Ottawa) of all restrictive

relativizer tokens. *Who* is the only productive WH-relativizer in these varieties, with other WH-variants such as *which*, *whose* and non-standard *what* occurring at negligible rates. The frequency of *who* fluctuates across varieties, ranging from a high of 24% in Tyneside to a low of 12% in Berwick, with intermediate rates found in the other varieties. Contra Buchstaller & Corrigan (2015: 88), there is no compelling evidence indicating that these rate differentials correlate with a north-south regional distinction. Admittedly, Berwick in the far north-east has the lowest rate of *who*, but a low rate is also found in London in the south. By comparison, the two northern conurbations exhibit slightly higher rates of *who*, suggesting that the reduced rate in Berwick reflects its relative peripherality to the northern urban varieties. In Ottawa, the relatively low rate of *who* (18%) is congruent with what has been documented in other varieties of urban Canadian English (see D’Arcy & Tagliamonte 2010: 389).

Table 2 Variant distributions²

Variety/corpus	<i>THAT</i>		<i>ZERO</i>		<i>WHO</i>		<i>OTHER</i>	
	N	%	N	%	N	%	N	%
Berwick (<i>BwE</i>)	514	55%	300	32%	108	12%	15	2%
Tyneside (<i>DECTE</i>)	266	42%	191	30%	151	24%	22	3%
Salford (<i>RoSE</i>)	159	51%	74	24%	60	19%	19	6%
London (<i>LIC</i>)	453	46%	330	34%	161	16%	42	4%
Ottawa (<i>OEC</i>)	570	60%	203	21%	172	18%	2	0%
TOTAL	1962	51%	1098	29%	652	17%	100	3%

5. Multivariate analyses

We next assess the data using multivariate analysis. In Tables 3-5, the input value provides a measure of the overall frequency of the relativizer variant in question.

² The ‘other’ category includes these variants: *which*, *whose*, *what*, *as*.

Factor weights (FW) vary between 0 and 1: the higher the figure, the greater the probability that a given relativizer will be chosen in the context listed in the left-most column. In our interpretation of the results, we draw on *statistical significance* at the .05 level (square brackets indicate non-significant predictors), and the *constraint hierarchy* within each predictor. Comparison of the latter will identify cross-variety structural (dis)similarities in the ‘grammar’ underlying relativizer choice (see Poplack & Tagliamonte 2001: 92-93). As relativizer choice in subject and non-subject relatives patterns differently, we analyze these environments separately.

Table 3 examines extra-linguistic patterns of variation in subject relative clauses.³ In the varieties where non-standard subject *zero* is sufficiently instantiated to permit multivariate analysis (Berwick, Tyneside, London), the constraint hierarchies for age indicate that *zero* is waning in apparent time. A reverse age effect for relativizer *that* in Salford and London suggests that this overt variant is gaining ground in subject position in these locations, where a concomitant sex effect implicates females as possible agents of change. In Ottawa, by contrast, *that* is on the decrease. Turning to *who*, the apparent-time effects point to its increasing use in Berwick and Ottawa, but a decrease in Salford.

³ Non-subject relativizer variation is not greatly affected by extra-linguistic predictors.

Table 3 Multivariate analysis of the contribution of social predictors to relativizer choice in subject relative clauses

	BERWICK (BwE)			TYNESIDE (DECTE)			SALFORD (RoSE)			LONDON (LIC)			OTTAWA (OEC)		
	THAT	WHO	ZERO	THAT	WHO	ZERO	THAT	WHO	ZERO	THAT	WHO	ZERO	THAT	WHO	ZERO
INPUT	.639	.175	.156	.454	.439	.056	.616	.315	1%	.619	.276	.044	.643	.336	2%
	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW
AGE															
Old (>40)	[]	.42	.62	[]	[]	.78	.34	.62	N.A.	.36	[]	.79	.59	.41	N.A.
Young (<40)	[]	.60	.35	[]	[]	.40	.58	.44		.59	[]	.30	.44	.56	
SEX															
Male	[]	[]	[]	[]	[]	[]	.42	[]	N.A.	.41	.60	[]	[]	[]	N.A.
Female	[]	[]	[]	[]	[]	[]	.58	[]		.62	.37	[]	[]	[]	

Summarizing, the results indicate that subject *zero* is either already severely restricted, or in the process of receding in the varieties shown in Table 3. However, varieties differ in terms of which co-variant – *that* or *who* – may be influencing *zero*'s demise. This awaits further exploration of the data.

Table 4 Multivariate analysis of the contribution of linguistic predictors to relativizer choice in subject relative clauses

SUBJECT RELS.	BERWICK (BWE)			TYNESIDE (DECTE)			SALFORD (RoSE)			LONDON (LIC)			OTTAWA (OEC)		
	That	Who	Zero	That	Who	Zero	That	Who	Zero	That	Who	Zero	That	Who	Zero
INPUT	.639	.175	.156	.454	.439	7%	.616	.315	1%	.619	.276	7%	.643	.336	2%
	FW	FW	FW	FW	FW	%	FW	FW	FW	FW	FW	%	FW	FW	FW
MATRIX CL.															
Existential	.24	.34	.89	[]	.34	32%	.42	.54		.48	.31	25%	[]	[]	
Stative Possessive	.36	.76	.54	[]	.63	10%	.59	.29		.33	.63	15%	[]	[]	
<i>It/that</i> -cleft	.67	.19	.55	[]	.34	7%	.89	.15		.72	.36	0%	[]	[]	
Centre-embedded	.44	.61	.56	[]	.76	0%	.22	.83		.46	.60	7%	[]	[]	
Other copula	.70	.50	.24	[]	.56	0%	.74	.26		.68	.39	4%	[]	[]	
Loan NP	.48	.66	.38	[]	.51	0%	.36	.65		.43	.67	0%	[]	[]	
Other	.56	.57	.35	[]	.47	0%	.40	.60		.51	.51	0%	[]	[]	
TYPE OF ANTECEDENT NP															
Unique	.81	K.O.	.39	.75	[]	0%	[]	[]		[]	[]	6%	.75	.19	
Definite	.55	[]	.40	.62	[]	1%	[]	[]		[]	[]	1%	.55	.47	
Indefinite	.43	[]	.58	.46	[]	10%	[]	[]		[]	[]	10%	.45	.55	
Pronoun	.56	[]	.42	.41	[]	0%	[]	[]		[]	[]	5%	.58	.45	
ANIMACY															
Inanimate	.72	K.O.	[]	.86	K.O.	3%	.79	K.O.		.70	K.O.	5%	.94	K.O.	
Human	.41	25%	[]	.33	61%	9%	.36	47%		.44	37%	7%	.17	54%	
Other significant effects: <i>Who</i> is favoured with longer relative clauses in Tyneside, and with generic antecedents in Berwick and Tyneside															

We next consider the linguistic conditioning of variation in subject relativizers, displayed in Table 4. In those varieties where data are sufficient to judge, subject *zero* correlates with existential and stative possessive matrix clauses, as illustrated in (4)-(5). These correlations can be linked to the structural characteristics and presentational functions of existential and stative possessive relative constructions. They consist of a semantically light matrix clause verb (e.g. *BE*, *HAVE (got)*) and a new referent in post-verbal position (e.g. *two houses*, *a family next door*), which is subsequently characterized in the embedded relative clause. They routinely function as semi-conventionalized structural units for introducing and elaborating on discourse-new referents (Lambrecht 1988). Unsurprisingly, then, they favour a

relativizer variant that signals a high degree of integration of the matrix and relative clause: *zero* (see Hopper & Traugott 2003: 202).

(4) There **was two houses** \emptyset became vacant. (Berwick/BwE)

(5) They **had a family next door** \emptyset was driving her mad. (London/LIC)

In most varieties we examined, matrix clause construction type also makes a significant contribution to the selection of *who* and *that*. All of our British varieties favour *who* in centre-embedded relative clauses, illustrated in (6), which post-modify the subject of the matrix clause. Given that grammatical subjects tend to encode human referents and that *who* has specialized as a marker of human antecedents (see the results for animacy at the bottom of Table 4), *who* would be expected to surface in precisely this environment (see Guy & Bayley 1995; Kikai et al. 1987).

(6) **The people** *who* are actually in that school fight each other as well. (London/LIC)

Another effect common to Berwick, Salford and London concerns the propensity of *it/that*-clefts to co-occur with *that*. The presence of an overt relativizer with clefts may reinforce their archetypical focal function (see Wiechmann 2015: 194).

Though not identical across varieties, the constraint hierarchy for type of antecedent NP exhibits a consistent definiteness effect for *that*, particularly with unique antecedents in Berwick, Tyneside and Ottawa, as in (7). And across all varieties, *that* is also highly favoured with inanimate antecedents.

(7) That's probably **the most interesting thing** *that* could happen. (Ottawa/OEC)

Turning to non-subject relativizers, where the variable context is restricted to *that* and *zero*, Table 5 reveals additional evidence of structured affinities shared by the comparison varieties. Despite cross-variety inconsistencies in constraint hierarchies for matrix clause construction type, the following effects are reproduced across the board: stative possessive matrix clauses tend to surface with *that*, and *it/that*-clefts favour the *zero* variant.

Table 5 Multivariate analysis of the contribution of linguistic predictors to relativizer choice in non-subject relative clauses

NON-SUBJECT RELS.	BERWICK (BwE)		TYNESIDE (DECTE)		SALFORD (RoSE)		LONDON (LIC)		OTTAWA (OEC)	
	That	Zero	That	Zero	That	Zero	That	Zero	That	Zero
INPUT	.407	.554	.385	.558	.366	.519	.296	.626	.558	.431
	FW	FW	FW	FW	FW	FW	FW	FW	FW	FW
MATRIX CL.										
Stative Possessive	.69	.25	.63	.23	.68	.33	.77	.23	.64	.33
Existential	.62	.39	.77	.21	.29	.61	.50	.49	.51	.44
Other copula	.62	.42	.42	.62	.63	.51	.40	.61	.39	.62
Other	.60	.33	.57	.42	.68	.29	.66	.35	.58	.42
Lone NP	.49	.50	.64	.39	.44	.56	.47	.57	.61	.41
Centre-embedded	.46	.58	.19	.82	.12	.82	.21	.71	.42	.60
<i>It/That</i> -cleft	.36	.67	.38	.64	0%	100%	.40	.58	.36	.64
REL. CLAUSE SUBJECT										
Other	.73	.24	.82	.22	[]	[]	.79	.15	.85	.12
Pronoun	.48	.53	.47	.52	[]	[]	.49	.52	.47	.53
ADJACENCY										
Non-adjacent	.74	.08	[]	[]	.94	0%	.70	.17	.75	.24
Adjacent	.47	.56	[]	[]	.44	57%	.49	.53	.48	.52
LENGTH OF REL. CLAUSE										
Long (6+ words)	.64	.29	[]	[]	[]	[]	[]	[]	.68	.36
Medium (4-5 words)	.51	.50	[]	[]	[]	[]	[]	[]	.47	.52
Short (≤ 3 words)	.45	.57	[]	[]	[]	[]	[]	[]	.45	.54
Other significant effects: <i>Zero</i> is favoured with unique NP antecedents in all varieties, and with generic antecedent NPs in Tyneside and London. <i>That</i> is favoured with indefinite NPs in 4/5 varieties and non-generic NPs in London.										

More striking parallels emerge from the results for relative clause subject type. With the exception of Salford, all varieties exhibit a significant preference for *zero* with pronominal subjects and for *that* with non-pronominal subjects. Similarly, with the exception of Tyneside, all varieties favour *zero* when the head NP and its embedded relative clause are adjacent, and *that* when they are non-adjacent. Surface-

level processing constraints offer a likely explanation for these effects. If a non-subject relative clause containing a non-pronominal subject is directly preceded by a lexical NP, there is a risk of local syntactic ambiguity when no overt relativizer is present (Temperley 2003). Similarly, processing complexity is increased when a *zero*-marked relative clause is separated from its antecedent head NP (Wiechmann 2015). Use of an overt relativizer attenuates these processing difficulties.

A parsing-based explanation also applies to the results for relative clause length in Berwick and Ottawa. Non-subject relative clauses comprising 6+ words favour *that*, as would be expected of the real-time processing demands imposed by long-distance dependencies.

6. Discussion

What is the pertinence of these results to elucidating our major research objectives: (i) the infiltration of WH-relativizers into our comparison varieties; and (ii) the existence of a vernacular norm in relativizer distributions? We have shown that relativizer *who* has permeated our comparison varieties to varying extents, although it is overwhelmingly restricted to subject relative clauses where it has staked out its preferred niche with human antecedent NPs. Crucially, though, *who* is the *only* WH-relativizer to have affected these varieties; no other WH-relativizer plays an appreciable role in the restrictive relativization system in the vernaculars investigated here.

Divergent trends in the *social conditioning* of variant choice, even within the same geographical region (e.g. the north of England), lend superficial credence to the claim that the restrictive relativization system in British and North American varieties lacks a vernacular norm (Ball 1996). However, detailed inspection of the *linguistic conditioning* of variant selection proved fundamental in demonstrating

such a claim to be insufficiently accountable to the full range of evidence. Notwithstanding cross-variety dissimilarities shown in Tables 3-5 (not discussed due to space constraints), comparison of a range of linguistic constraints on variant distributions revealed several highly structured parallels in the system conditioning relativizer selection. Certain parallels are attributable to the discourse-pragmatic and information-structural functions of specific relative constructions whereas others appear to be epiphenomenal of on-line processing constraints.

In conclusion, our analytical approach demonstrates that the fine conditioning of relativizer choice serves as a pivotal check on earlier claims about the absence of a vernacular norm in this sector of the grammar, many of which have been impervious to insights from the structure of variability expressed in terms of hierarchies of effects. The results we have presented here suggest that closer consideration of these effects, particularly those associated with discourse-pragmatic and processing constraints, is an indispensable prerequisite to achieving a more accurate understanding of patterns of restrictive relativizer variation and change in contemporary varieties of English.

References

- Ball, Catherine 1996. A diachronic study of relative markers in spoken and written English. *Language Variation and Change* 8/2, 227-258.
- Buchstaller, Isabelle and Corrigan, Karen. 2015. Morphosyntactic features of Northern English. In Raymond Hickey (ed.) *Researching Northern English*. Amsterdam: John Benjamins, 71-98.
- Cheshire, Jenny, Kerswill, Paul, Fox, Sue and Torgersen, Eivind. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15/2, 151-196.

- Corrigan, Karen P., Buchstaller, Isabelle, Mearns, Adam J. and Moisl, Hermann L. 2012. *The Diachronic Electronic Corpus of Tyneside English*. <http://research.ncl.ac.uk/decte/>
- D'Arcy, Alexandra and Tagliamonte, Sali. 2010. Prestige, accommodation and the legacy of relative *who*. *Language in Society* 39/3, 383-410.
- Guy, Gregory R. and Bayley, Robert. 1995. On the choice of relative pronouns in English. *American Speech* 70/2, 148-162.
- Hopper, Paul and Traugott, Elizabeth C. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.
- Kikai, Akio, Schleppegrell, Mary and Tagliamonte, Sali. 1987. The influence of syntactic position on relativization strategies. In Keith Denning, Sharon Inkelas, Faye McNair-Knox & John Rickford (eds.) *Variation in language - NWAV-XV at Stanford: Proceedings of the Fifteenth Annual Conference on New Ways of Analyzing Variation*. Stanford, CA: Stanford University, Department of Linguistics, 266-277.
- Kerswill, Paul, Cheshire, Jenny, Fox, Sue and Torgersen, Eivind. 2004-07. Linguistic innovators: The English of adolescents in London. Economic and Social Research Council Project RES-000-23-0680.
- Lambrecht, Knud. 1988. There was a farmer had a dog: Syntactic amalgams revisited. *Berkeley Linguistics Society* 14, 319-339.
- Poplack, Shana and Tagliamonte, Sali. 2001. *African American English in the Diaspora*. Oxford: Blackwell.
- Romaine, Suzanne. 1982. *Socio-Historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge University Press.

- Tagliamonte, Sali. 2013. Comparative sociolinguistics. In J.K. Chambers & Natalie Schilling (eds.) *The Handbook of Language Variation and Change*. 2nd ed. Oxford: Blackwell, 128-156.
- Tagliamonte, Sali, Smith, Jennifer and Lawrence, Helen. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17/1, 75-112.
- Temperley, David. 2003. Ambiguity avoidance in English relative clauses. *Language* 79/3, 464-484.
- Wiechmann, Daniel. 2015. *Understanding Relative Clauses: A Usage-Based View on the Processing of Complex Constructions*. Berlin: Mouton de Gruyter.