APPENDIX 1: CELL VALUES AND SUMMARIES OF PREVIOUS STUDIES

TABLE 1A. *Rhotic tokens out of total tokens per speaker / vowel combination*

speaker	lettER	NEAR	NORTH/FORCE	NURSE	SQUARE	START
1	2/20 (10%)	2/15 (13.33%)	1/20 (5%)	16/20 (80%)	4/20 (20%)	6/20 (30%)
10	10/20 (50%)	2/10 (20%)	6/20 (30%)	16/20 (80%)	2/20 (10%)	9/20 (45%)
19	15/20 (75%)	7/15 (46.67%)	9/20 (45%)	20/20 (100%)	13/20 (65%)	18/20 (90%)
2	10/20 (50%)	14/17 (82.35%)	11/20 (55%)	16/20 (80%)	9/20 (45%)	12/20 (60%)
20	2/20 (10%)	2/20 (10%)	3/20 (15%)	6/20 (30%)	2/20 (10%)	7/20 (35%)
21	20/20 (100%)	12/13 (92.31%)	15/20 (75%)	20/20 (100%)	19/20 (95%)	19/20 (95%)
22	2/20 (10%)	1/20 (5%)	0/20 (0%)	14/20 (70%)	4/20 (20%)	2/20 (10%)
23	8/20 (40%)	8/12 (66.67%)	15/20 (75%)	17/20 (85%)	8/20 (40%)	13/16 (81.25%)
24	10/20 (50%)	14/14 (100%)	8/20 (40%)	19/20 (95%)	8/11 (72.73%)	5/16 (31.25%)
25	9/20 (45%)	8/10 (80%)	8/20 (40%)	17/19 (89.47%)	9/14 (64.29%)	4/8 (50%)
26	13/20 (65%)	7/20 (35%)	8/20 (40%)	19/20 (95%)	5/20 (25%)	9/20 (45%)
27	16/20 (80%)	15/17 (88.24%)	12/20 (60%)	20/20 (100%)	14/20 (70%)	13/20 (65%)
28	0/20 (0%)	6/20 (30%)	1/20 (5%)	1/20 (5%)	4/20 (20%)	0/20 (0%)
3	14/20 (70%)	15/20 (75%)	13/20 (65%)	16/20 (80%)	6/13 (46.15%)	11/12 (91.67%)
4	1/19 (5.26%)	5/20 (25%)	1/20 (5%)	4/20 (20%)	0/20 (0%)	3/20 (15%)
5	11/20 (55%)	11/16 (68.75%)	6/20 (30%)	18/20 (90%)	8/20 (40%)	18/20 (90%)
6	1/20 (5%)	0/9 (0%)	0/20 (0%)	7/20 (35%)	3/20 (15%)	0/12 (0%)
7	1/20 (5%)	3/6 (50%)	0/20 (0%)	11/20 (55%)	0/17 (0%)	0/12 (0%)
8	7/20 (35%)	3/15 (20%)	2/20 (10%)	14/20 (70%)	2/20 (10%)	9/20 (45%)
b12	91/118 (77.12%)	23/26 (88.46%)	108/128 (84.38%)	135/141 (95.74%)	76/92 (82.61%)	50/61 (81.97%)
b13	77/143 (53.85%)	41/42 (97.62%)	51/170 (30%)	112/135 (82.96%)	70/91 (76.92%)	34/53 (64.15%)
b3	98/118 (83.05%)	33/37 (89.19%)	49/65 (75.38%)	72/74 (97.3%)	33/51 (64.71%)	27/27 (100%)
b5	184/213 (86.38%)	43/43 (100%)	166/171 (97.08%)	133/133 (100%)	102/108 (94.44%)	90/90 (100%)
b6	71/87 (81.61%)	29/29 (100%)	69/77 (89.61%)	50/54 (92.59%)	60/68 (88.24%)	24/25 (96%)
b7	108/115 (93.91%)	18/18 (100%)	94/101 (93.07%)	68/69 (98.55%)	95/98 (96.94%)	54/54 (100%)
b8	141/149 (94.63%)	67/67 (100%)	128/138 (92.75%)	72/72 (100%)	105/108 (97.22%)	40/40 (100%)

TABLE 2A. *Rhotic tokens out of total tokens per speaker and other context*

speaker	morpheme final	morpheme internal	word final	non-prepausal	prepausal	content word	function word
1	8/24 (33.33%)	16/45 (35.56%)	7/46 (15.22%)	26/103 (25.24%)	5/12 (41.67%)	28/84 (33.33%)	3/31 (9.68%)
10	7/23 (30.43%)	23/44 (52.27%)	15/43 (34.88%)	35/94 (37.23%)	10/16 (62.5%)	38/80 (47.5%)	7/30 (23.33%)
19	10/21 (47.62%)	49/54 (90.74%)	23/40 (57.5%)	72/99 (72.73%)	10/16 (62.5%)	67/87 (77.01%)	15/28 (53.57%)
2	10/21 (47.62%)	30/46 (65.22%)	32/50 (64%)	59/98 (60.2%)	13/19 (68.42%)	56/83 (67.47%)	16/34 (47.06%)
20	3/24 (12.5%)	12/45 (26.67%)	7/51 (13.73%)	18/99 (18.18%)	4/21 (19.05%)	16/85 (18.82%)	6/35 (17.14%)
21	18/21 (85.71%)	38/39 (97.44%)	49/53 (92.45%)	90/98 (91.84%)	15/15 (100%)	74/79 (93.67%)	31/34 (91.18%)
22	2/25 (8%)	14/37 (37.84%)	7/58 (12.07%)	21/102 (20.59%)	2/18 (11.11%)	16/85 (18.82%)	7/35 (20%)
23	14/23 (60.87%)	31/37 (83.78%)	24/48 (50%)	45/75 (60%)	24/33 (72.73%)	49/71 (69.01%)	20/37 (54.05%)
24	19/23 (82.61%)	25/44 (56.82%)	20/34 (58.82%)	53/83 (63.86%)	11/18 (61.11%)	51/77 (66.23%)	13/24 (54.17%)
25	14/20 (70%)	22/30 (73.33%)	19/41 (46.34%)	41/72 (56.94%)	14/19 (73.68%)	45/66 (68.18%)	10/25 (40%)
26	11/31 (35.48%)	21/32 (65.63%)	29/57 (50.88%)	58/116 (50%)	3/4 (75%)	43/72 (59.72%)	18/48 (37.5%)
27	22/30 (73.33%)	29/38 (76.32%)	39/49 (79.59%)	87/114 (76.32%)	3/3 (100%)	62/78 (79.49%)	28/39 (71.79%)
28	4/20 (20%)	1/45 (2.22%)	7/55 (12.73%)	9/103 (8.74%)	3/17 (17.65%)	7/84 (8.33%)	5/36 (13.89%)
3	9/16 (56.25%)	36/44 (81.82%)	30/45 (66.67%)	63/84 (75%)	12/21 (57.14%)	59/76 (77.63%)	16/29 (55.17%)
4	3/19 (15.79%)	6/42 (14.29%)	5/58 (8.62%)	11/110 (10%)	3/9 (33.33%)	13/87 (14.94%)	1/32 (3.13%)
5	13/22 (59.09%)	35/43 (81.4%)	24/51 (47.06%)	64/103 (62.14%)	8/13 (61.54%)	58/80 (72.5%)	14/36 (38.89%)
6	2/15 (13.33%)	7/41 (17.07%)	2/45 (4.44%)	7/72 (9.72%)	4/29 (13.79%)	7/73 (9.59%)	4/28 (14.29%)
7	1/18 (5.56%)	10/33 (30.3%)	4/44 (9.09%)	11/76 (14.47%)	4/19 (21.05%)	13/67 (19.4%)	2/28 (7.14%)
8	6/19 (31.58%)	18/40 (45%)	13/56 (23.21%)	31/95 (32.63%)	6/20 (30%)	28/78 (35.9%)	9/37 (24.32%)
b12	46/57 (80.7%)	186/191 (97.38%)	251/318 (78.93%)	403/484 (83.26%)	80/82 (97.56%)	297/323 (91.95%)	186/243 (76.54%)
b13	58/88 (65.91%)	125/208 (60.1%)	202/338 (59.76%)	266/482 (55.19%)	119/152 (78.29%)	247/377 (65.52%)	138/257 (53.7%)
b3	51/69 (73.91%)	80/83 (96.39%)	181/220 (82.27%)	200/251 (79.68%)	112/121 (92.56%)	200/218 (91.74%)	112/154 (72.73%)
b5	94/95 (98.95%)	229/229 (100%)	395/434 (91.01%)	496/535 (92.71%)	222/223 (99.55%)	471/476 (98.95%)	247/282 (87.59%)
b6	41/46 (89.13%)	82/87 (94.25%)	180/207 (86.96%)	185/220 (84.09%)	118/120 (98.33%)	186/203 (91.63%)	117/137 (85.4%)
b7	62/67 (92.54%)	127/128 (99.22%)	248/260 (95.38%)	325/342 (95.03%)	112/113 (99.12%)	275/280 (98.21%)	162/175 (92.57%)
b8	95/100 (95%)	144/149 (96.64%)	314/325 (96.62%)	426/447 (95.3%)	127/127 (100%)	325/338 (96.15%)	228/236 (96.61%)

TABLE 3A. *Internal effects on rhoticity reported in previous studies*

Study	variety/ies	direction	preceding vowel	tautosyllabic C	other /r/	prepausal	morpheme-final	word-final	stress	emphasis	functionword	word length	word frequency
Asprey, 2007	Black Country	nonrhoticity	*	+									
Barras, 2010	Lancashire	nonrhoticity	back vowels > front vowels*			+	-	-	+				
Baxter, 2008	Quebec	rhoticity	*					+					
Becker, 2014	New York City	rhoticity	*	+		+	-	-	+		-		
Dudman, 2000	Cornwall	nonrhoticity	*					-	+				
Elliott, 2000	American films	rhoticity											
Ellis, Groff, & Mead, 2006	Philadelphia	rhoticity			-								
Feagin, 1990	Alabama	rhoticity	*	+									
French, 1988	Yorkshire	nonrhoticity				+							
Hartmann & Zerbian, 2010	South Africa	rhoticity		0						+			
Hinton & Pollock, 2000	Iowa	*	0						+				
Hollitzer, 2013	Berkshire, Wiltshire, Somerset	nonrhoticity	*	+									
Irwin & Nagy, 2007	Boston	rhoticity	back vowels > front vowels*	+			0	+			-	-	0
Jones, 1998	Devon; West Somerset	nonrhoticity	*										
Labov, 1966 [1972]	New York City	rhoticity	*					+					
Miller, 1998	Philadelphia	rhoticity	*		-								

Continued

TABLE 3A. *Continued*

Study	variety/ies	direction	preceding vowel	tautosyllabic C	other /r/	prepausal	morpheme-final	word-final	stress	emphasis	functionword	word length	word frequency
Myhill, 1988	Philadelphia	nonrhoticity	*	0	-	-		0	+				
Nagy & Irwin, 2010	Boston; New Hampshire	rhoticity		+	0	+	-	-					-
Parslow, 1967, 1971	Boston	rhoticity	NURSE > other vowels										
Piercy, 2006, 2007, 2012	Dorset	nonrhoticity	*	+		+	*	-	+		0		-
Pollock & Berni, 1997	Tennessee	*	*										
Schützler, 2010	Edinburgh	nonrhoticity				+			+				
Sharbawi & Deterding, 2010	Brunei; Singapore	rhoticity	0										
Simpson, 1996	Shropshire	nonrhoticity						-	+				
Sudbury & Hay, 2002[1]	New Zealand	nonrhoticity	back vowels > front vowels						+				-
Sullivan, 1992	Devon	nonrhoticity						-					
Trudgill & Gordon, 2006	Australia	nonhoticity	*										
Villard, 2009	New Hampshire; Vermont	rhoticity	*										
Vivian, 2000	Lancashire	nonrhoticity	*					-	+				
Watt, Llamas, & Johnson, 2014	Scottish-English Border	nonrhoticity		*									
Williams, 1991	Isle of Wight	nonrhoticity											

Note: Key: + = favors rhoticity, - = disfavors rhoticity, 0 = no effect, * = mixed or multiple effects

TABLE 4A. *Effects of preceding vowel on rhoticity reported in previous studies*

Study	Variety	Effect of preceding vowel
Asprey, 2007:96–98	Black Country	NURSE > lettER > SQUARE > NEAR > NORTH > START
Barras, 2010:115, 175	Lancashire	back vowels > front vowels FORCE > NURSE > START > NORTH > SQUARE > NEAR > lettER
Baxter, 2008	Stanstead (Quebec)	NURSE > back vowels > front vowels > lettER
Becker, 2014:155–156	New York City	NURSE > NEAR > START > SQUARE > NORTH/FORCE[2]
Dudman, 2000:36	Cornwall	CURE > START (f) > NURSE > NEAR > SQUARE > NORTH/FORCE > START (b) > lettER (?)
Feagin, 1990:132	Alabama	NURSE > NEAR > SQUARE > START > NORTH > FORCE > lettER
Hinton & Pollock, 2000	Davenport (Iowa)	no effect[3]
Hollitzer, 2013	Berkshire, Wiltshire, Somerset	NURSE > lettER > other vowels (?NURSE > NEAR > lettER > START > SQUARE > NORTH/ FORCE)[4]
Irwin & Nagy, 2007:140–142; Nagy & Irwin, 2010:256–257	Boston & New Hampshire	NURSE > START > SQUARE > CURE > NEAR > NORTH/ FORCE > lettER
Jones, 1998	Devon, West Somerset	START > FUR > ‘farmer, darning’, NORTH/ FORCE > FIR
Labov, 1972	New York City	NURSE > lettER back vowels > front vowels
Miller, 1998	Philadelphia	NURSE > all other vowels > lettER
Myhill, 1988	Philadelphia	NURSE > all other vowels > lettER (more integrated into white community) NURSE > START > all other vowels (less integrated into white community)
Nagy & Irwin, 2010:258–259, 277	Boston	NURSE > START > CURE > FUR > NORTH/ FORCE > NEAR > lettER > SQUARE (older speakers) CURE > START > NURSE > SQUARE > NEAR > NORTH/ FORCE > lettER (younger speakers)
Nagy & Irwin, 2010:260, 277–278	New Hampshire	NURSE > SQUARE > NEAR > START > NORTH/ FORCE > lettER (older speakers) START > SQUARE > NORTH/ FORCE > NURSE > NEAR > lettER (younger speakers)[5]
Parslow, 1967; 1971	Boston	NURSE > other vowels
Piercy, 2012:81–82[6]	Dorset	NURSE > NEAR > START > lettER > CURE > SQUARE > NORTH/ FORCE
Pollock & Bernie, 1997	Memphis (Tennessee)	NURSE > front vowels > back vowels > lettER
Sharbawi & Deterding, 2010	Brunei, Singapore	no effect[7]
Sudbury & Hay, 2002:289–290	New Zealand	back vowels > front vowels[8]
Sullivan, 1992:82–83	Exeter	(NEAR) > NURSE > START > SQUARE > FORCE > lettER > NORTH
Trudgill & Gordon, 2006:240	Australian English	NORTH/ FORCE, lettER > others[9]
Villard, 2009	Upper Valley (New Hampshire, Vermont)	NURSE > lettER

TABLE 5A. *External effects on rhoticity reported in previous studies (cells report the social group found to favor rhoticity)*

Study	Variety	gender	class	ethnicity	locality	style	exposure
Elliott, 2000	American films	female					
Becker, 2014	New York	female	middle-class	change only for white & Jewish		formal	
Feagin, 1990	Alabama	female	working-class				
Irwin & Nagy, 2007	Boston	*	middle-class				
Nagy & Irwin, 2010	Boston, New Hampshire	*	*	0			
Ellis, Groff, & Mead, 2006	Philadelphia	*		disfavored by African Americans			
Villard, 2009	Upper Valley (New Hampshire, Vermont)	female	middle-class				
Baxter, 2008	Stanstead (Quebec)	female	middle-class				
Parslow, 1967, 1971	Boston (Massachusetts)						
Labov, 1966 [1972]	New York	female		favored by whites		formal	
Myhill, 1988	Philadelphia			0 (but favored by speakers more integrated into the white community)			
Miller, 1998	Philadelphia			favored by African Americans			
Hinton & Pollock, 2000	Davenport (Iowa)					0	
Pollock & Berni, 1997	Memphis					0	
Cychosz & Johnson, 2017	American English (Buckeye Corpus)	female					
Hartmann & Zerbian, 2010	South Africa	female	affluent				
Sharbawi & Deterding, 2010	Brunei; Singapore						
Asprey, 2007	Black Country				rural		
Barras, 2010	Lancashire				rural		
Vivian, 2000	Lancashire	male			*		
Jones, 1998	Devon; West Somerset						
Piercy, 2007	Dorset	male			rural		

Continued

TABLE 5A. *Continued*

Study	Variety	gender	class	ethnicity	locality	style	exposure
Williams, 1991	Isle of Wight					minimal pairs wordlist > casual speech > wordlist	
Sudbury & Hay, 2002	New Zealand				*		
Trudgill & Gordon, 2006	Australia						
Watt, Llamas, & Johnson, 2014	Scottish-English Border				*		
Schützler, 2010	Edinburgh	male				wordlist	more exposed to SSBE
Sullivan, 1992	Devon	male	working-class			casual speech	
Simpson, 1996	Shropshire						
Dudman, 2000	Cornwall					casual speech	
French, 1988	Yorkshire						
Hollitzer, 2013	Newbury, Swindon, Taunton				western		

Elastic net regression is not widely used in linguistics. Since it is best understood through the matrix approach to regression, we start by describing ordinary least squares regression for context before describing the different forms of penalized regression: ridge, lasso, and elastic net regression.[10]

Least squares regression

In normal linear regression, we have a set of p predictor variables x_1, $x_2 \ldots x_p$ and a response variable y. We aim to estimate the values of coefficients β_1, $\beta_2 \ldots \beta_p$ such that:

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \epsilon$$

where ϵ is Gaussian white noise. We have t observations of our predictor and response variables, so that we actually have a vector y responses of length t, a t by p matrix of predictors X called the design matrix, a vector of random noise ϵ, and a vector of coefficients β. We can then express our model as:

$$y = X\beta + \epsilon$$

We estimate the best possible values of β using a method called least squares, which minimizes the sum of squared residuals:

$$\sum |y - X\beta|^2$$

The solution to this is the matrix equation:

$$\beta = (X^T X)^{-1} X^T y$$

where X^T is the transpose of the design matrix X and $(X^T X)^{-1}$ is the inverse of $X^T X$. This gives us an estimation of β which is unbiased and as precise as possible.

Ridge regression

This procedure fails when some of the predictors are highly correlated. From a conceptual standpoint, this is easy to understand. If two predictors rise and fall in tandem, and these rises and falls are linearly related to changes in the response variable, it is difficult or impossible to determine which of the two predictors is responsible for the changes in the response. The 'best' result we can achieve will be coefficient estimates with very large errors, representing the fact that either predictor might actually be irrelevant if all of the observed effect is assigned to the other predictor. From the point of view of our least squares method, if two predictors (two columns of our design matrix) are perfectly correlated then $X^T X$ has a determinant equal to zero and so has no inverse. If two predictors are nearly perfectly correlated, the determinant of $X^T X$ is close to zero and so it is difficult to find the inversion precisely.

In ridge regression, we solve this problem by using a different method of estimating our coefficients. Instead of minimizing the sum of squared residuals, we minimize the following:

$$\sum ||y - X\beta||^2 + ||\lambda\beta||^2$$

As a result, in addition to minimizing the residuals, we are also minimizing the coefficients. The solution to this is the following equation:

$$\beta = (X^T X + \lambda^2)^{-1} X^T y$$

Because we have added λ^2 to X^TX, we can now find an inverse, even if our design matrix contains columns which are perfectly correlated. The result is no longer a truly unbiased estimate and will tend to underestimate the coefficients, but does give better results in cases of collinearity. The difference between the results of this method and the ordinary least squares method depends on the λ parameter. As we increase λ, we increase the penalty for large coefficients and so increase the degree to which coefficients are minimized: if $\lambda = 0$ then the result is identical to the least squares method; if $\lambda = \infty$ then all of our coefficient estimates will be zero. Because we are minimizing the sum of the *square* of the coefficients, this penalty is stronger for larger coefficients. The result is that large coefficients are shrunk to a 'reasonable' size while small coefficients are affected relatively little.

Lasso regression

A different approach is Lasso regression (standing for Least Absolute Shrinkage and Selection Operator). Here, instead of minimizing the following:

$$\sum ||y - X\beta||^2 + ||\lambda\beta||^2$$

we minimize:

$$\sum ||y - X\beta||^2 + ||\lambda\beta||$$

Again, we are penalizing large coefficients: if $\lambda = 0$ then the result is identical to the least squares method, and if $\lambda = \infty$ then all of our coefficient estimates will be zero.

However, penalizing the coefficient estimates themselves rather than the squared coefficient estimates gives lasso regression some quite different behaviors to ridge regression. Unlike ridge regression, shrinkage is not greater for larger coefficients, so lasso regression does not offer us a tool to deal with individual inflated coefficient estimates. However, at reasonable values of λ, lasso regression tends to reduce all small coefficients to zero, leaving only the larger coefficients in the model. Thus, lasso regression builds in a form of feature selection: because only the larger coefficients are retained, it tends to give us as simple a model as possible.

Elastic net regression

In cases with a very large number of predictors, neither of these options may suffice. With a sufficiently large number of possible predictors, some are likely to be highly correlated, making ridge regression an attractive option. However, variable selection is difficult with ridge regression: with normal regression we might use a stepwise procedure where we use significance tests to progressively add or remove predictors to the model; since there is no straightforward significance test for ridge regression, we cannot follow this approach and are left with a maximally complex model with all the potential predictors.

To solve this problem, we can use elastic net regression, combining the advantages of ridge regression (robust with highly correlated predictors) with lasso regression (automated variable selection). In elastic net regression, we include both the ℓ_1- and ℓ_2- penalty, minimizing the following:

$$\sum ||y - X\beta||^2 + ||\lambda_2\beta||^2 + ||\lambda_1\beta||$$

Elastic net regression has some of the properties of ridge and lasso regression: increasing either λ_1 or λ_2 sufficiently high will shrink all coefficient estimates to zero; if both λ_1 and λ_2 are equal to zero, the model is the same as the ordinary least squares; the model performs well with highly correlated predictors; very high coefficient estimates are shrunk towards reasonable values; small coefficients are reduced to zero, leaving us with a relatively simple model.

Parameter setting

We then have to determine the values of λ_1 and λ_2. There are two broad approaches to this. One is cross-validation. The idea here is to use the existing data to find the model that best predicts some new dataset. Because we generally cannot acquire a whole new dataset easily, we instead split our existing dataset into a training set and a test set. To avoid some accidental properties of the data we assign to the test set having disproportionate influence over the final model, we can use *k*-fold crossvalidation: we divide the dataset into *k* equally sized subsets each of which is treated as the test set in turn; we then select the values λ_1 and λ_2 that perform best on average across all the test sets.

An alternative approach is to select an ‘information criterion’ measure, a statistic that measures model goodness-of-fit offset by model complexity, such as the Aikaike Information Criterion (AIC). By setting λ_1 and λ_2 so as to minimize the AIC of the model, we find the model with the optimal trade-off between complexity and fit.

APPENDIX NOTES

1. Internal factors only investigated for linking r.

2. However, Becker states that, when the data is broken down into ethnic groups, only the effect of NURSE is consistent and that “no overall pattern for preceding full vowels is evident” (2014:158–159).

3. As with Trudgill and Gordon’s (2006) study of Australian English and Nagy and Irwin’s (2010) study of New Hampshire English, we might hypothesize that the lack of effect here is due to the fact that the change had almost gone to completion: either because conditioning systems tend to disappear in the final stages of change, or because the very low frequency of one variant inevitably makes it hard to detect significant effects without an extremely large sample.

4. Hollitzer's analysis divides the data up into three towns: Newbury, Swindon, and Tauton; although rates of rhoticity per vowel are calculated for each town (2013:34–35), several categorically nonrhotic speakers are included in these calculations for Newbury and Swindon, making the hierarchies suspect. Hollitzer's only strong conclusion is that NURSE and lettER favor rhoticity, since this is consistent across the three towns (2013:35).

5. Nagy and Irwin point out that disagreements in constraint rankings between the younger New Hampshire speakers and all other groups might be the result of the fact that the change is almost gone to completion in this group and that constraints must necessarily fade as the conservative variant becomes vanishingly rare (2010:259–260).

6. The analysis of Piercy (2012) is used rather than the less statistically sophisticated analysis of the same data in Piercy (2006:55).

7. Sharbawi and Deterding examine only START, NORTH, and NURSE. Comparison of their data for these vowels shows no significant difference in rates of rhoticity for either variety studied: for Brunei English, 10/18 START, 24/54 NURSE, and 25/54 NORTH tokens were rhotic ($\chi^2 = 0.68$, $p = 0.7118$); for Singapore English, 1/12 START, 4/36 NURSE, and 2/36 NORTH tokens were rhotic ($\chi^2 = 0.727$, $p = 0.6952$). However, as the sample size is tiny, no strong conclusions should be drawn from this.

8. Sudbury and Hay's (2002) finding applies only to linking r and not coda r.

9. No statistical evidence of the relative effect of the different contexts is offered, and the sample is relatively small; the authors suggest that the mismatch with other studies is the result of the fact that this "must represent the last surviving traces of earlier, fuller rhoticity" (2006:240).

10. Note that, while the specific model actually used in the paper is logistic elastic net regression, here, for reasons of space, we describe linear elastic net regression.