Statistical approaches to hierarchical data in sociophonetics: The case of variable rhoticity in Scottish Standard English*

Ole Schützler

University of Bamberg

Hierarchical data structures in variationist linguistics are given when, for example, each of a number of speakers makes several utterances of interest and individual observations are therefore not independent. Structures of this kind are often not fully reflected in the analytic tools used to model language variation. Against this background, the paper investigates the presence or absence of coda-/r/ in a hierarchically structured dataset produced by 27 middle-class speakers of Scottish Standard English (SSE). Two regression-based statistical techniques are compared: multiple logistic regression and a hierarchical generalized linear model (HGLM). The latter emerges as a model that is not only more correct in theoretical terms because it takes the nested structure of the data into account, but also detects cross-level effects that go unnoticed in multiple logistic regression. The most striking finding is, for example, that in prepausal position coda-/r/ is less likely to be deleted, an effect, however, that is not general but depends critically upon the age of the speaker and the dialect contact to which he or she has been exposed. The paper thus provides one example of the potential of HGLM to explain complex patterns of variation in hierarchically structured data.

1. Introduction

This paper aims to contribute to variationist linguistics in two ways. First, it investigates the partial loss of rhoticity, i.e. the vocalisation or deletion of the consonant /r/ in syllable codas (e.g. in *car*, *bird*) in Scottish Standard English (hereafter SSE). Secondly, it makes methodology a focus of attention and introduces the Hierarchical Generalized Linear Model (hereafter HGLM) as a statistical tool for the explanation of sociophonetic variation. The value and special quality of HGLM is demonstrated by comparing its results to those of a multiple logistic regression model (hereafter LR), which does not take the multilevel structure of the dataset into account.[1]

The contribution as a whole will have to be a compromise: neither is it possible to present the full theoretical framework of HGLM, nor is it possible to include a review of the literature on rhoticity in Scottish English that goes beyond a cursory mentioning of a few relevant titles. Lastly, the analysis presented here is preliminary in the sense that it does not include all factors that are of theoretical interest or have been reported in the literature.

It is certainly a challenge to attempt to combine variationist insight with methodological advance. I would therefore refer readers who are interested in more detailed introductions to HGLM to the monographs by Raudenbush & Bryk (2002) or Snijders & Bosker (1999). For an application of these techniques in a full-scale linguistic study, see my more comprehensive work on selected features of the SSE accent currently in preparation.[2]

* The author is grateful to the Spanish Ministry of Science and Innovation and the European Regional Development Fund (under CONSOLIDER grant HUM2007-60706 for the research project *Variation, Linguistic Change and Grammaticalization*).

[1] The HGLM analysis was produced using the program HLM 6.08.

[2] This includes analyses using both HGLM, i.e. a hierarchical model predicting categorical outcome variables, and HLM, i.e. a hierarchical model predicting continuous outcome variables.

2. HIERARCHICAL MODELS IN LINGUISTICS: PRELIMINARY CONSIDERATIONS

According to Snijders & Bosker (1999:1), "multilevel analysis is a methodology for the analysis of data with complex patterns of variability, with a focus on nested sources of variability." *Nested* means that "[u]nits at one level are contained within units of another level" (Agresti & Finlay 2009:523). The basic statistical tool used to analyse nested data is the hierarchical linear model, which is a more complex relative of the (single-level) multiple linear regression model (Snijders & Bosker 1999:2), and its extension, the hierarchical generalized linear model, which is the multilevel relative of multiple logistic regression.[3]

The application of hierarchical models will provide analyses that (1) do not result in the loss of information, (2) do not result in statistical fallacies and violations of assumptions, and (3) are more accurate in the representation of predictable and unpredictable variation. I will briefly explain each of these points, using my dataset for illustration. Point (1), the loss of information, would result if, instead of looking at specific instances of the occurrence or non-occurrence of coda-/r/, speaker averages were calculated and analysed. It would then no longer be possible to take into account low-level information (e.g. the level of stress or the syntactic position of an observation). Point (2), statistical fallacies, would ensue if each observation was not only coded for lower-level characteristics (e.g. STRESS[4]), but also for characteristics of the next higher level (e.g. text style) or indeed for characteristics of the highest level, the speaker (e.g. AGE or GENDER). In this case, no information would be lost since all cases and all characteristics are accessible. However, this procedure would violate one of the basic assumptions of regression, namely the independence of observations, and would furthermore pool the entire un-modelled variation into a single error term (Luke 2004:6-7). Finally, point (3) refers to the option in multilevel models to model not only direct effects on the outcome variable at each level, but to include cross-level effects as well. For example, a question that could be answered with a multilevel model is "Does the predictor WORDLIST have a significant effect on the outcome, and if so, is it the same for men and women?" Indirect effects of this kind will not only be identified but quantified as well.

In agreement with Luke (2004:4), I would suggest that if data are multilevel in nature, their analytic techniques should also be multilevel in nature. While hierarchical data structures are just as common in the field of variationist linguistics as in the social sciences in general, I am under the impression that linguists' choices of statistical tools very often do not fully reflect these structures. Luke (2004:2) describes most systems investigated in the social sciences as open systems in which it is not easily possible to control for complex sources of error and variability. I would argue that this also fully applies to variationist linguistics, and that multilevel models are a big step forward since they are better able to account for that complexity. The present paper is intended as a case in point.

3. DERHOTICISATION IN GENERAL AND IN SCOTTISH STANDARD ENGLISH

A concise definition of rhoticity is that by Harris (2006:357): "In non-rhotic systems, *r* is […] said to be licensed in onsets but not in codas." Conversely, in rhotic accents /r/ would phonologically be licensed in *all* environments. It seems helpful to interpret *license* quite literally as this implies non-compulsoriness of coda-/r/ in rhotic accents: it may (and usually will) occur, but the rule is to some extent variable. For variationist approaches, rhoticity is

[3] For a general description of (multiple) linear regression see Bortz (2005:183-196) or Agresti & Finlay (2009:265-269, 321-355); for (multiple) logistic regression, see Agresti & Finlay 2009:483-512.

[4] I render concrete predictor variables in small capital letters if they are used in this form for the present paper, thus: GENDER, STRESS, CONTACT. Descriptions of general factors that do not immediately constitute predictor variables will appear in normal script, thus: level of stress, text style, syntactic position (cf. section 5).

viewed as potentially variable in rhotic accents, with /r/ in weak syllables (*surprise*, *forget*) being particularly prone to be reduced or deleted (Scobbie 2006:340). In non-rhotic accents, coda-/r/ is generally of less interest unless the research is concerned with juncture phenomena (e.g. Allerton 2000; Brown 1988; Bauer 1984), the (re-) introduction of rhoticity in non-rhotic accents (Labov 2006; Feagin 1990), or special discourse-structuring functions coda-/r/ may fulfil in otherwise non-rhotic accents (French 1988).

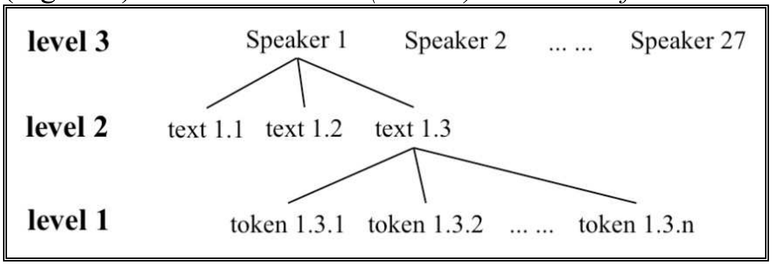
SSE is generally represented as a rhotic accent (e.g. Wells 1982:410; Jones 2002:26; Stuart-Smith 2008:57). These accounts clearly describe the general perception and classification of the accent as a whole and thus do not take away from the findings of empirical research showing that under certain conditions, rhoticity in SSE may be variable or prone to erosion. Suffice it to say that most of the predictors I include have in some way been found to be of relevance in Scottish accents of English in connection with this variability or partial derhoticisation. These include style (Johnston 1984; Schützler 2010b), PREPAUSAL POSITION (Romaine 1978; Stuart-Smith 2003; Lawson *et al.* 2008; Schützler 2010b), GENDER (Romaine 1978; Stuart-Smith 2003; Schützler 2010b), AGE (Johnston 1984; Stuart-Smith 2003; Schützler 2010b), and STRESS (Lawson *et al.* 2008, Schützler 2010b). However, comparing the findings especially of Schützler (2010b) for middle-class speakers with those of Romaine (1978), Stuart-Smith (2003) and Lawson *et al.* (2008) for working-class speakers suggests that for GENDER and PREPAUSAL, the effect on coda-/r/ may be rather different depending on the social class of speakers.

4. THE DATA AND THEIR HIERARCHICAL STRUCTURE

The data used for this study were collected in Edinburgh in 2008. Interviews were conducted with 27 speakers who were 17-62 years old at the time. The sample is middle-class throughout. Coda-/r/ was elicited in three speech styles: (1) a reading passage, (2) a careful speech component, and (3) a wordlist. The careful speech element consisted of a more spontaneous task based on the reading passage. The number of cases is fairly balanced between careful speech and reading passage with N = 1073 for the former and N = 1192 for the latter, but the number of cases in wordlist-style is considerably lower (N = 294). The wordlist had to be shorter as it was designed to target several different variables in a concise format (cf. Schützler 2009, 2010a, 2011).

Describing the data in a multilevel fashion, $N_1 = 2519$, i.e. there are 2519 level-1 units or individual cases. At level 2, the number of observations is $N_2 = 81$, i.e. there are 81 text units (3 per speaker) in which the individual observations at level 1 are nested. At the highest level, $N_3 = 27$ – these are the 27 speakers. This structure is shown in Figure 1.

(Figure 1) *Hierarchical (3-level) structure of the dataset*



The arrangement of level-2 units is perfectly regular, as the there are always three text-units per speaker: reading passage, careful speech, and wordlist. At level 1, the number of cases/tokens varies not only depending on the text unit but in the case of careful speech also on the general productivity of a given speaker. Therefore, the diagram represents the maximal number of units as fixed at levels 3 and 2, but variable at level 1.

5. OUTCOME VARIABLE AND PREDICTOR VARIABLES

It is for present purposes assumed that coda-/r/ will be either present or absent. The value of the outcome variable thus denotes the probability of the articulation of coda-/r/ in a specific instance. While the phoneme /r/ is "prone to vary in many and subtle ways" (Scobbie 2006:338) and is therefore a challenge for sociophoneticians, this paper cannot do justice to the subtleties of variation that go far beyond simple presence or absence of the consonant. The full-scale analysis of coda-/r/ in SSE, from which the present methodological discussion is derived, treats the outcome not as binary but as a multinomial variable that can take one of several categorical values, e.g. approximant, tap/trill, or zero.

The independent (predictor) variables fall into three groups corresponding to the three levels of the model shown in section 4:

A. Social parameters (level 3) [5]

1. **GENDER** (male = 0, female = 1)
2. **AGE Y** (17-22 years old)
3. **AGE M** (40-47 years old)
4. **CONTACT**
5. **REGION**

B. Text type related parameters (level 2) [6]

6. **WL**
7. **TX**

C. Low-level/contextual parameters (level 1) [7]

8. **STRESS**
9. **PREPAUSAL**

6. THE MULTIPLE LOGISTIC REGRESSION MODEL

Logistic regression is similar to linear regression in that it represents a probable outcome as the sum of an intercept (baseline value or constant) and one or several products of predictor coefficients with predictor values. However, as the outcome in this case is not continuous but binary (coda-/r/ present or absent), no direct linear relationships between predictors and outcome are possible. LR therefore employs a link-function, the so-called logit-function:

---

[5] The uncoded reference category for both age predictors is the group of older speakers, aged 52-62. CONTACT = 1 in the case of speakers who were (or are) exposed directly and extensively to speakers of Southern Standard British English (SSBE), either in their own families or during longer stays in England as students or in a professional capacity. CONTACT thus does not mean the more diffuse exposure to Southern English accents that is to be generally expected in Edinburgh. REGION 1 = 1 denotes speakers from "unmixed" Edinburgh backgrounds. Strictly speaking, the region of geographical provenance of a speaker would of course constitute a fourth level. Nevertheless, it was tested as a level-3 predictor, with very little effect (s. below).

[6] The uncoded reference category in this case is the careful speech component of the sample.

[7] STRESS was given categorical values of 0-3 which were treated as a scale in the analysis. STRESS = 3 corresponds to stressed syllables in focus stress position, STRESS = 0 is used for unstressed function words, and STRESS = 1-2 describes degrees of lexical stress weaker than focus stress. PREPAUSAL in my definition is at least partly synonymous with "utterance-final" or "before a speech pause" as used by other authors. It does not necessarily denote a following longer pause or a change of turn, (cf. French 1988) but simply a natural break in discourse or the end of a sentence.

$$logit_{HGLM}(\mathrm{r}\,|\,1) = \ln\left(\frac{p(r\,|\,1)}{1-p(r\,|\,1)}\right)$$

The logit is the natural logarithm of the odds in favour of the occurrence of the outcome (articulated coda-/r/). The log-odds have the advantage that they are continuous (can take any value) and theoretically unbounded (with possible values of $-\infty < \text{logit} < \infty$). They are therefore clearly eligible for prediction by linear regression. A (multiple) LR using the logit link-function can therefore also be described as an indirect (multiple) linear regression, in which a continuous, linearly predicted pseudo-outcome is related to a true, non-linear outcome via the link-function. A multiple LR equation thus takes the general form:

$$\text{logit}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + r$$

In this equation, β_0 is the intercept, i.e. the expected outcome if all relevant predictors take the value zero. β_1 is the slope of the first predictor, i.e. the change in the outcome if that predictor increases by the value one, X_1 is the value of the first predictor which is multiplied by that predictor's slope, and *r* is the residual, or error term, i.e. the difference between the predicted value and the actually observed value of the outcome variable. The coefficients are calculated so as to minimize the average value of *r* across all cases.

The LR for the current analysis progressed in a backward fashion by including all predictors at the outset and deleting them from the model if non-significant. The sequence of deletion was: AGE M (after step 1; p = .914), REGION 1 (after step 2; p = .629), and TX (after step 3; p = .586). Table 1 represents the final model, in which 6 of the original 9 predictors are retained.

(Table 1) *Predicting articulation of coda-/r/: Multiple LR*

Variable	B	SE	Wald	df	p	Exp(B)
constant	.176	.0895	3.88	1	.049	1.19
GENDER	- .498	.0944	27.80	1	.000	.61
AGE Y	.157	.0951	2.72	1	.099	1.17
CONTACT	- .736	.1132	42.26	1	.000	.48
WL	.867	.1962	19.53	1	.000	2.38
STRESS	.458	.0531	74.34	1	.000	1.58
PREPAUSAL	1.720	.1799	91.45	1	.000	5.58

The model can also be written as a regression equation. It must be remembered that this does not directly predict the probability of the outcome, but the log-odds:

$$\text{logit}_{LR}(\mathrm{r}|1) = .176 - .498\,(\textbf{GENDER}) + .157\,(\textbf{AGE Y}) - .736\,(\textbf{CONTACT}) + .867\,(\textbf{WL}) + .458\,(\textbf{STRESS}) + 1.720\,(\textbf{PREPAUSAL}) + \boldsymbol{r}$$

Thus the log-odds in favour of articulated coda-/r/ for a young male speaker producing a prepausal token carrying focus stress are .176 + .157 + 3*(.458) + 1.720 = 3.427. The coefficients associated with GENDER, CONTACT and WL do not feature in this equation, because these two predictors take values of zero.[8]

The logit of 3.427 is not immediately meaningful, it only tells us that under the given circumstances it is very likely that coda-/r/ should be articulated, as the logit has a high

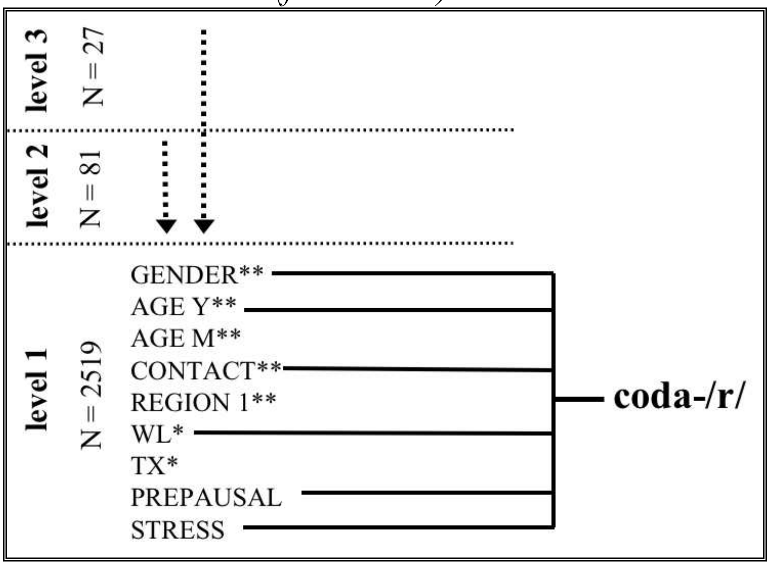
[8] Theoretically, WL could have been included as an additional positive predictor, but since in my data all wordlist tokens were embedded in a carrier phrase and WL and PREPAUSAL were thus mutually exclusive, I decided against this.

positive value. However, it can be transformed into a probability value using the inverse of the logit link-function. This is called the logistic function and takes the following form:

$$p_{LR}(\mathrm{r}\,|\,1) = \frac{\mathrm{e}^{\mathrm{logit}_{LR}(\mathrm{r}|1)}}{1+e^{\mathrm{logit}_{LR}(\mathrm{r}|1)}} \qquad \text{In the above example, this is} \qquad p_{LR}(\mathrm{r}\,|\,1) = \frac{\mathrm{e}^{3.427}}{1+e^{3.427}} = .969$$

Under the given circumstances, the probability of articulated coda-/r/ is .969, or 96.9%. The example with the lowest possible probability is a female speaker exposed to contact who produces an unstressed token in non-prepausal position. The logit for this scenario is .176 - .498 - .736 = -1.058 which results in a probability of .258 or 25.8%. Figure 2 shows the LR model in diagram form.

(Figure 2) *Predicting articulation of coda-/r/: Schematic representation of LR. To indicate level of origin, disaggregated predictors are marked ** (from level 3) and * (from level 2)*



The diagram shows that all higher-level predictors have been assigned to the individual cases. The standard errors (and consequently the p-values) of level-2 and level-3 coefficients will therefore be underestimated, because they are based on an inflated number of observations.

7. THE HIERARCHICAL GENERALIZED LINEAR MODEL

Essentially, level-1 outcomes in HGLM are predicted using LR, but these predictions are conditional upon estimates made for higher-level units. Thus HGLM avoids both the loss of information characteristic of aggregated data and the violation of statistical assumptions characteristic of erroneously applied LR that assumes the independence of observations where it is not given (see sections 2 and 6). Not only the level-1 intercept can be modified depending on the higher-level unit to which an observation belongs, but the slopes can vary as well: HGLM is able to assess, for example, if the effect of PREPAUSAL is equal in the wordlist and the reading passage, and if its effect is the same between different speakers. Further, it can model the differences in slope or intercept at higher levels, using higher-level variables as predictors, before inserting them into the level-1 equation. Thus, in the present case of a 3-level analysis, the level-1 model is in fact a nested 'model-within-a-model-within-a-model' that reflects the nested structure of the data.

Table 2 shows the final model with seven of the original nine predictors. However, some of them have more than one function. TX and REGION 1 were found to have no effect that would have improved the model. Note that the degrees of freedom (*df*) take different values for different predictors, depending on the number of observations made on the level to which a predictor belongs. Also note that there are now three error terms (*E*, *R0*, *U00*) instead of one: unlike LR, this model acknowledges that predictions at all three levels will to some extent be incomplete or imperfect.[9]

(Table 2) *Predicting articulation of coda-/r/: HGLM*

Fixed effects	Coeff.	SE	T	df	p
intercept	.234	.1444	1.62	24	.118
—— GENDER	- .474	.1845	- 2.57	24	.017
—— CONTACT	- .646	.2358	- 2.74	24	.012
— WL	.875	.2081	4.21	79	.000
STRESS	.465	.0550	8.45	2510	.000
PREPAUSAL	1.495	.3477	4.30	2510	.000
—— AGE Y	.850	.4327	1.96	2510	.049
—— AGE M	1.643	.6371	2.58	2510	.010
—— CONTACT	- 1.312	.4026	- 3.26	2510	.002
Random effects	σ	σ^2	df	χ^2	p
E	1.8138	3.2899			
R0	.2604	.0678	53	77.95	.014
U00	.3725	.1388	24	70.21	.000
σ/σ^2 (fitted values)	1.1242	1.2638			

GENDER and CONTACT have a direct negative impact on the logit, i.e. female speakers and those extensively exposed to the SSBE accent tend to delete coda-/r/ more often than others. In wordlist style, coda-/r/ is less likely to be deleted as the coefficient of WL is positive, and the same is true for STRESS. Clearly, PREPAUSAL is the most interesting level-1 predictor: it depends on several level-3 predictors, namely AGE Y, AGE M, and CONTACT. Its intercept or baseline value is 1.495 which denotes its effect on the logit if the three variables affecting PREPAUSAL itself take value 0 (i.e. for for an older, no-contact speaker). However, if a case is nested in a young speaker or a middle aged speaker, the positive effect of PREPAUSAL will be magnified as the respective coefficient of AGE Y or AGE M will be added to that of PREPAUSAL. On the other hand, if CONTACT = 1, the effect of PREPAUSAL will be dramatically reduced.

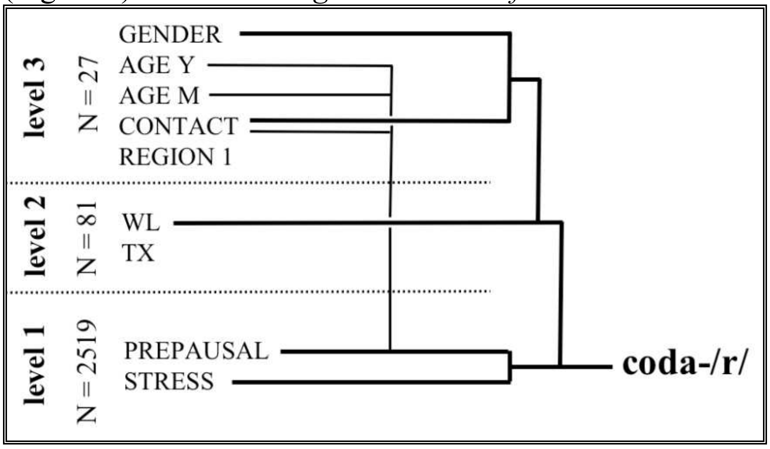
The logit of any individual case can be predicted using the following equation which contains both the direct intercept-effect and the three cross-level effects, the latter given in square brackets:

$$\begin{aligned} \text{logit}_{HGLM}(r|1) = {} & .234 - .474\,(\textbf{GENDER}) - .646\,(\textbf{CONTACT}) + .875\,(\textbf{WL}) + .465\,(\textbf{STRESS}) \\ & + [1.495 + .850\,(\textbf{AGE Y}) + 1.643\,(\textbf{AGE M}) - 1.312\,(\textbf{CONTACT})]\,(\textbf{PREPAUSAL}) \\ & + \boldsymbol{e} + \boldsymbol{r}_0 + \boldsymbol{u}_{00} \end{aligned}$$

For example, in the case of a young male speaker producing a prepausal token carrying focus stress (s. also section 6), the logit is .234 - .474*0 - .646*0 + .875*0 + .465*3 + [1.495 + .850 + 1.643*0 – 1.312*0] = 3.974, which is equivalent to a probability of .982 or 98.2%. Figure 3 shows this hierarchical generalized linear model in diagram form. It can be seen how the direct level-1 effects depend on level-2 effects, which in turn depend on level-3 effects. Also, the effect of PREPAUSAL depends on the age of the speaker and the level-3 predictor CONTACT.

9 Model fit, e.g. using R^2, will be discussed below.

(Figure 3) *Predicting articulation of coda-/r/: Schematic representation of HGLM*



8. COMPARISON OF MODELS

In the way it treats the impact of higher-level predictors, HGLM is theoretically the more conservative of the two procedures compared in sections 6 and 7: it is more reluctant to recognise a significant contribution of higher-level variables because it does not disaggregate them onto a lower level and thus does not create an artificially enlarged number of independent observations. Nevertheless, there are several predictors that are not only identified by both LR and HGLM in the present analysis, but also have similar coefficients in both analyses, as Table 3 shows. This is especially true for GENDER, CONTACT, WL, and STRESS. The p-values for both GENDER and CONTACT are lower in LR, while WL is robust enough to produce very low p-values in both models, and the level-1 predictor STRESS is equally robust in both.

(Table 3) *Direct comparison of coefficients in LR and HGLM*

Variable	HGLM Coeff.	HGLM p	Variable	LR Coeff.	LR p
intercept	.234	.118	constant	.176	.049
—— GENDER	- .474	.017	GENDER	- .498	.000
			AGE Y	.157	.099
—— CONTACT	- .646	.012	CONTACT	- .736	.000
WL	.875	.000	WL	.867	.000
STRESS	.465	.000	STRESS	.458	.000
PREPAUSAL	1.495	.000	PREPAUSAL	1.720	.000
—— AGE Y	.850	.049			
—— AGE M	1.643	.010			
—— CONTACT	- 1.312	.002			

Beyond the direct predictors GENDER, CONTACT, WL and STRESS, there are some profound differences between the two models. For example, AGE Y is retained in LR as a direct predictor, albeit not a very valuable one: the coefficient is low, not significant at $p < .05$, and would normally be excluded. In HGLM, AGE Y does not feature as a direct (intercept) effect. However, this difference is partly explained when the predictor PREPAUSAL is inspected. In LR, this simply is a strong predictor, in HGLM it is not as strong but potentially enhanced by other level-3 predictors that have an effect on it. Here, AGE Y features as a slope predictor in combination with CONTACT and AGE M – the latter is a predictor not at all included in LR.

Statements about the amount of explained variation in both models based on values equivalent to the R^2 used in linear models have to be treated with caution, since in HGLM there are error terms at three levels and the calculation (and the theoretical concept) of explained variance in these models is sometimes described as incompatible with the corresponding practices in single-level regressions (e.g. Luke 2004:58). However, using the R^2_{HGLM} proposed by Snijders & Bosker (1999:225), which is based on the three random effects and the variance of fitted values shown in Table 2, and comparing it to Nagelkerke's pseudo R^2 for LR, suggests that the hierarchical model is somewhat better in this respect: $R^2_{HGLM} = .266$ and $R^2_{LR} = .185$. HGLM appears to be more successful at the overall explanation of variation, but one should be careful with this comparison since both values are necessarily calculated differently.

What can be directly compared, however, is the number of correct predictions made by each model. For this purpose, a p-value describing the probability of a positive outcome (articulated coda-/r/) is calculated from the logit of each observation. The result is compared to a threshold value of $p = .500$ (equivalent to logit = 0). A value of $p \geq .500$ is interpreted as a predicted value of 1, a value of $p < .500$ is interpreted as a predicted value of 0. If the predicted value matches the binary outcome, the prediction is true. Table 4 compares the results of this test for the empty model (which simply predicts the majority outcome Y = 1 in each case), LR, and HGLM.

(Table 4) *Correctly predicted outcomes: empty model, LR, and HGLM (cut value at p = .500)*

	Empty model			**LR**			**HGLM**		
	Predicted			Predicted			Predicted		
Observed	/r/	ø	correct	/r/	ø	correct	/r/	ø	correct
/r/	1655	0	100.0%	1457	198	88.0%	1479	176	89.4%
ø	864	0	0.0%	618	246	28.5%	633	231	26.7%
Total %			65.7%			67.6%			67.9%

The proportion of correct predictions is only a rough indicator: even in the empty model the number of correct predictions is not strikingly lower than in the other two models. But the trend is nevertheless clear. If nothing but the majority outcome is predicted, observations match predictions in 1655 cases (65.7%), if predictions are based on LR there are 1703 matches (67.6%), and if HGLM is applied there are 1710 matches (67.9%).

11. Conclusions

The comparison of analyses using LR and HGLM shows that, given the structure of the dataset under investigation, the latter is not only more adequate and legitimate for theoretical reasons, but also produces rather different results. For example, the coefficients of the two predictors GENDER and CONTACT have considerably lower p-values when LR is applied. This is the result of disaggregation: GENDER and CONTACT are characteristics of speakers, and estimating them based on the true number of speakers (N = 27), as HGLM does, results in a more conservative and realistic assessment of their levels of significance. While these predictors are significant in both analyses, this finding nevertheless points to the risk of committing a statistical type I error in LR: this procedure will more readily (but fallaciously) find coefficients to be significant than HGLM if they belong to a higher level.

In HGLM, AGE Y does not features as a direct intercept predictor (as in LR), but as a slope predictor affecting PREPAUSAL. This partly explains why LR was not able to fully capture its effect: in this model it has a low coefficient and is statistically not significant. Putting it differently, LR identifies AGE Y as a weak (in fact untenable) *general* predictor while HGLM identifies it as a strong *specific* predictor. This can serve as an illustration of one main limitation of LR: the impact of all predictors is assumed to be directed at the outcome in the same way (although differing in magnitude). HGLM on the other hand is much better able to model complexity, not only in terms of different levels, but also in terms of slope-effects, i.e. indirectness of effects.

Compared to LR, HGLM seems to some extent to be simply *different*. However, it also seems to explain rather more of the variability present in the data, and in this respect it is also *better*. Both the value of R^2 is – with all due caution – higher for the final model using HGLM, and the same is true for the number of correctly predicted outcomes, using a threshold value of p = .500 (or 50%).

As several predictors (GENDER, CONTACT, WL, STRESS) were identified as significant by both models and had coefficients whose estimated values were rather similar in both models, it seems that LR is able to capture intercept effects quite successfully, even if fallaciously applied to hierarchical data. However, some effects – and arguably the most interesting ones – elude the single-level approach, as has been shown for the effect of the level-1 predictor PREPAUSAL which was itself affected by three level-3 predictors. The new model introduced in this paper thus appears to be not only more adequate for theoretical statistical reasons. It also explains variability more successfully and brings to light highly interesting effects that would otherwise go unnoticed.

REFERENCES

Agresti, Alan and Barbara Finlay (2009) *Statistical Methods for the Social Sciences.* Upper Saddle River, NJ: Pearson Education.

Allerton, David J. (2000) Articulatory Inertia vs. Systemzwang: Changes in liaison phenomena in recent British English. *English Studies* 6: 574-581.

Bauer, Laurie (1984) Linking /r/ in RP: Some Facts. *Journal of the International Phonetic Association* 14.2: 74-79.

Bortz, Jürgen (2005) *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.

Brown, Adam (1988) Linking, intrusive, and rhotic /r/ in pronunciation models. *Journal of the International Phonetic Association* 18.2: 144-151.

Feagin, Crawford (1990) Dynamics of sound change in Southern States English: From r-less to r-ful in three generations. In Jerold A. Edmondson, Crawford Feagin and Peter Mühlhäusler (eds.) *Development and Diversity: Language Variation across Time and Space.* Arlington: SIL/University of Texas, pp. 129-146.

French, J. Peter (1988) Word Final /r/ In A Northern English Accent: An Interactional Account Of Variable Pronunciation. In Peter Auer and Aldo di Luzio (eds.) *Variation and Convergence*. Berlin: Mouton De Gruyter, pp. 125-133.

Harris, John (2006) Wide-domain r-effects in English. *UCL Working Papers in Linguistics* 18: 357-379.

Johnston, Paul A. (1984) Variation in the Standard Scottish English of Morningside. *English World-Wide* 4.2: 133-185.

Jones, Charles (2002) *The English Language in Scotland: An Introduction to Scots.* East Linton: Tuckwell Press.

Labov, William (2006) *The Social Stratification of English in New York City.* Cambridge: Cambridge University Press.

Lawson, Eleanor, Jane Stuart-Smith and James Scobbie (2008) Articulatory Insights into Language Variation and Change: Preliminary Findings from an Ultrasound Study of Derhoticisation in Scottish English. *University of Pennsylvania Working Papers in Linguistics* 14.2: 99-110.

Luke, Douglas A. (2004) *Multilevel Modeling.* Thousand Oaks: Sage Publications.

Raudenbush, Stephen W. and Anthony S. Bryk (2002) *Hierarchical Linear Models. Applications and Data Analysis Methods.* Thousand Oaks: Sage Publications.

Romaine, Suzanne (1978) Postvocalic /r/ in Scottish English: sound change in progress? In Peter Trudgill (ed.) *Sociolinguistic Patterns in British English.* London: Arnold, pp. 144-157.

Schützler, Ole (2009) Unstable Close-mid vowels in Modern Scottish English. In Carlos Prado-Alonso, Lidia Gómez-García, Iria Pastor-Gómez and David Tizón-Couto (eds.) *New Trends and Methodologies in Applied English Language Research. Diachronic, Diatopic and Contrastive Studies*. Bern: Peter Lang, pp. 153-182.

Schützler, Ole (2010a) Which WAY do Scottish monophthongs GO? Charting vowel variation in Scottish Standard English. In Ewa Waniek-Klimczak (ed.) *Issues in Accents of English 2: Variability and Norm.* Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 133-145.

Schützler, Ole (2010b) Variable Scottish English consonants: The cases of /ʍ/ and non-prevocalic /r/. *Research in Language* 8. [DOI: 10.2478/v10015-010-0010-9] http://versita.com

Schützler, Ole (2011) Charting vowel spaces in Edinburgh middle-class speech. *English World-Wide* 32.1: 24-45.

Scobbie, James M (2006) (R) as a Variable. In Keith Brown (ed.) *Encyclopedia of Language and Linguistics.* Vol. 10. Amsterdam: Elsevier, pp. 337-344.

Snijders, Tom and Roel Bosker (1999) *Multilevel Analysis.* London: Sage Publications.

Stuart-Smith, Jane (2003) The Phonology of Modern Urban Scots. In John Corbett, J. Derrick McClure and Jane Stuart-Smith (eds.) *The Edinburgh Companion to Scots.* Edinburgh: Edinburgh University Press, pp. 110-137.

Stuart-Smith, Jane (2008) Scottish English: Phonology. In Bernd Kortmann and Clive Upton (eds.) *A Handbook of Varieties of English. Vol.1: The British Isles.* Berlin: Mouton de Gruyter, pp. 48-70.

Wells, John C. (1982) *Accents of English*. 3 vls. Cambridge: Cambridge University Press.

Ole Schützler

English Linguistics
Universität 9
University of Bamberg
D-96045 Bamberg
Germany

ole.schuetzler@uni-bamberg.de