LVC guidelines for reporting quantitative results

Gregory R. Guy
New York University

1. Basics of good practice.

The point of reporting quantitative results in a research paper is two-fold: to communicate objective information about the research question in an efficient manner, and to provide evidence for or against hypotheses or arguments. These objectives are best achieved when the presentation satisfies four criteria:

i. Clarity. Any quantitative data presented should aim to be easily interpretable. Tables and figures must be designed so that the informed reader can readily figure out whatever points they illustrate. Wherever possible, a table or figure should contain all the information necessary to interpret it without the reader having to go read something in the text of the paper.

ii. Efficiency. Authors should strive to report the right amount of information: include what is necessary to give a clear picture of the data and sustain the conclusions, without a flood of unnecessary material, or omissions that require extra work from the reader.

iii. Transparency. Everything about the quantitative analysis must be fully explained, including the data source, criteria for including or excluding data, analytical methodology, the quantitative or statistical model, the assumptions of the model, and the software that was used.

iv. Recoverability. Enough information should be given for readers to replicate the procedures used and evaluate the analysis. Maximum recoverability is achieved with open access corpora; consideration of this option is desirable.

Data quantity. Basic descriptive information about the data is the essential point of departure for all statistical analyses. Without adequate descriptive and summary statistics, the credibility of subsequent inferential and test statistics will be questionable. The most elementary requirement is N – the total number of observations or tokens of the dependent variable. This value should also be broken down by all the data sources and predictors that are relevant to the analysis. Show the N for all independent variables investigated: for each phonological, syntactic or discourse context, each morphological or lexical category, each social context or category, etc. When the data come from multiple speakers, the N for each speaker should be reported, and the distribution of speakers across any social categories investigated, "so that the reader can appreciate the balance (or imbalance) of the research design" (Paolillo 2017).

Summary statistics. Two types of dependent variables account for most studies of language variation: continuous variables – like vowel formants, and nominal variables – like deletion vs. retention of a segment, or pre- or post-verbal position of a subject. The basic summary statistic for a continuous variable is the mean (which may be accompanied by further statistics characterizing the distribution of results, such as the standard deviation). For nominal variables, the Ns or percentages of the various realizations are the appropriate summary statistics. In the case of binary dependent variables, the efficiency criterion indicates that the N or percentage should only be reported for one of the two variants; while for variables with more than two variants, N or percentage must be given for each one.

These summary statistics should also be given for all the independent variables and data sources used in the analysis: give percentages or means for each level of a predictor, and for each

speaker. These predictors will, in many studies, be used in multivariate analyses (e.g. using Goldvarb or R) to calculate factor weights, effect estimates, or odds ratios, but these values are derived from specific models with particular assumptions. The credibility of such results can only be assessed in light of the basic descriptive facts. A value based on a small N is always less reliable than one based on a large N, and an effect estimate that is greatly at odds with the basic percentage or mean value for that predictor should alert the analyst to the likelihood of a skewed distribution or an interaction with some other predictor.

Precision. The precision, or level of detail, of all statistical values depends directly on N. More data yields more precise statistics, more accurate estimates, and higher levels of significance. Therefore, the order of magnitude of all reported statistics should not exceed the order of magnitude of the N. Otherwise, you are implicitly claiming greater precision than the data support. A study that found two instances of outcome X in seven observations (N=7), is only accurate to within one part in seven; reporting this result as showing that X occurs at a rate of 42.857% implicitly claims an accuracy to within one part in 100,000. As a rule of thumb do not give more significant digits in your statistics than there are in the N: with tens of tokens, give two significant digits; with hundreds, give three digits, and with thousands of tokens, give four significant digits.

2. Hypotheses.

All choices of models imply hypotheses about what affects what, or at the very least, what things are correlated. Every independent variable included in a quantitative analysis entails a hypothesis that it has some effect on the dependent variable. Usually this is the whole point of the analysis – to test such hypotheses. A paper submitted for publication must necessarily be clear and explicit on what those hypotheses are, and how the quantitative results bear on them.

This applies to all the details of the analytical model, including the distinctions made in the independent variables: the different factors in a factor group, the levels of a predictor. The paper should state the rationale for making whatever distinctions are made in the analysis. Often this involves decisions about how refined or general an analysis is needed. For example, an independent variable for a phonological context might have only the levels {consonant, vowel}, or a more refined analysis contrasting, say, {stop, fricative, sonorant, vowel}. The latter analysis tests the hypothesis that different levels of obstruency have different effects on the dependent variable. The analyst should acknowledge these implicit hypotheses, and indicate whether the quantitative results support the given level of refinement, or demonstrate that some such distinctions do not have a significant effect on the dependent variable.

3. Modelling.

Most linguistic variation is affected by multiple independent variables, whether language internal or external. Furthermore, the distribution of linguistic items across the independent variables of interest is always lumpy. Certain contexts are highly frequent, others are rare; forms belonging to 'regular' paradigms will usually greatly outnumber the 'irregular' forms, etc. In this respect most data on variation and change in natural language will not occur in a balanced distribution, with all the cells in a cross-tabulation having equal numbers of tokens. Balanced distributions can be obtained by design, as in experimental studies, but in naturalistic data this can only be

achieved by deliberately selecting tokens in equal numbers, which introduces gross distributional distortions and is statistically dubious.

As a consequence of this lumpy distribution, variationist studies almost always require a multivariate analysis to tease out the separate effects of the independent variables. A series of separate univariate analyses, each testing one predictor at a time, is rarely adequate, because these will fail to control for the uneven distribution of the tokens across the other predictors. For example, consider a case in which some syntactic process is favored by preverbal subjects, and the study also seeks to test the effect of declarative vs. interrogative clauses. If most declarative clauses have preverbal subjects, then a univariate analysis for clause type (i.e. one which does not control for subject position) will report that declaratives also favor the process, being unable to recognize the skewed distribution across the subject position variable.

Multivariate analyses are therefore essential to the study of language variation and change. But there are a number of different multivariate models being used in contemporary research, and they have different virtues and limitations, and make different assumptions. Every research paper submitted for publication must explicitly describe and justify the quantitative model that is used. The analyst must show awareness of the assumptions behind the model, and be alert for any aspect of the results that might challenge those assumptions.

4. Reporting main effects.

The results of a quantitative analysis must be reported completely; this means that all the effects of every independent variable that is investigated must be presented. This includes every level of every predictor, the factor weights for every factor, *whether they are significant or not*. It is a common shortcoming to report only significant effects, but it is a basic scientific principle that non-significance of a potential predictor is itself an important finding. It means that the hypothesis that such an independent variable impacts on the realization of your dependent variable is rejected, which informs future research, helps other researchers to avoid blind alleys, and contributes to the development of theories about possible and impossible constraints on linguistic processes, and consequently, about the nature of grammar.

It should be emphasized that this requirement applies equally to reporting the effects of all the levels of a predictor. Goldvarb provides estimates (factor weights) for every factor in a group, centered on a notionally neutral value of .50, but models in common use in R (e.g. lmer) choose a reference level for every predictor and give estimates for the other levels of that predictor in terms of how far and in what direction their effect differs from the reference level. Consequently no estimate is given for the reference level, but this does not mean that the reference level can be neglected in reporting the results. On the contrary, it must be identified, and its associated N and percentage (or mean, in the case of continuous dependent variables) must be reported. Such models also return measures of the significance of each predictor level tested, meaning whether it is significantly different in estimated effect from the reference level. Again, reporting only the significant effects is invalid. In fact, for predictors with more than two levels, it is almost useless to report only the significant levels: knowing that level A is significantly different from reference level B tells us nothing about whether A is significantly different from C or D. All of these values must be presented regardless of what the significance statistic indicates, so that readers can see for themselves what the relationships among the various levels/factors are.

When reporting effects, researchers should take care that all the numbers they report are readily interpretable as to direction as well as magnitude of effect estimate. The sign (positive or negative) of each estimate should be specified as to what result it favors or disfavors. Similarly, in Goldvarb analyses, the 'application value' should be clearly indicated.

Reference levels. Selection of reference levels should not be done arbitrarily; the default in many R analyses is by alphabetical order of variant labels. This is absurd. The most obvious choice for a reference level is the variant with the largest N, as larger Ns are almost always associated with greater confidence and stability, and are more likely to represent the center of a normal distribution. Alternatively, some meaningful baseline should be chosen: e.g., one end of an age range to show a time trend, the highest or lowest social class, or one linguistic unit that can be taken as default or unmarked (e.g. if you are investigating a predictor variable for consonant manner, you might want to use stops as the reference level; if you are looking at subject position relative to verb in a Romance language, you might take SV as the default reference level.).

The analyst should also be aware that the choice of reference level may have a major impact on significance results for predictors with more than two levels. For example, when the levels form a cline, the middle point in the cline (say level B in a cline of A, B, C) may not be significantly different from the endpoints, and selecting this as the reference level might return no significant effects – meaning none of the other predictors have an effect that is significantly different from the central reference level. But using one end of the cline as a reference level (say A) could well show a major significant difference from the level at the other end (C). Care should be taken not to extrapolate the significance figures reported for the levels of a predictor to the overall significance of the predictor.

Significance. Tests of statistical significance are estimates of the likelihood of getting the observed data from some distribution by random stochastic variation. Typically this is phrased comparatively, in terms of a 'null hypothesis': what is the likelihood of the data that is characterized by predictor X being randomly drawn from the same distribution as the data characterized by predictor Y? This is opposed to the 'experimental hypothesis', which is that X and Y have different effects, and the data they characterize are not drawn from equivalent distributions with only stochastic sampling differences. A variety of statistical methods are available for estimating significance, but they are generally mapped onto a **p** value.

Reports of statistical significance should be cognizant of these issues. The paper must indicate what specific test was used, and should make clear what the test is comparing; for example, as noted above in the discussion of reference levels, the significance values compare a given level of a predictor with whatever level was selected as the reference. In Goldvarb, the significance values reported in the step-up procedure compare the quality or goodness-of-fit of the model (measured by log-likelihood) obtained when a given factor group is included with the quality obtained when it is excluded. Hence this provides a measure of the significance of the predictor (the factor group) as a whole.

Reporting p values should respect the guidelines given above regarding precision. It is usually adequate to give asterisks indicating various criterial levels, e.g., *p<.05, **p<.01. Alternatively, p values may be reported with 2-3 significant digits. For small p values (p<.001) there is no need to give the number of zeros preceding a non-zero values. Reporting values like p=3.47 E-12 is pointless overkill, implying that you can be accurate to within one in a trillion, despite the fact that you are likely working with only a few hundreds or thousands of data points. Such a value is not in any meaningful way 'more' significant than p=.0001. If you have selected .05 or .01 as your criterial value, the important thing to report is whether or not the test clears that hurdle.

5. Convergence.

Regression analyses proceed iteratively, refining estimates at each successive iteration, testing whether that iteration has significantly improved a goodness of fit measure, continuing until a best fit is achieved. This is called convergence. But with a complex model, there are very many possible adjustments to parameter estimates that can refine the model. Sometimes it is impossible, within practical limits of computational power, to find a unique best fit model. This is nonconvergence.

Nonconvergence is usually due to an imbalanced distribution of data across predictors, such as a collinearity or non-othogonality of predictors. For example, if most tokens of predictor x are also tokens of predictor y, it may be impossible to uniquely attribute their joint effects to one or the other. Since, as we have noted, natural linguistic data is almost always imbalanced in some dimensions, this is not an uncommon problem, especially when many predictors with multiple levels are being tested. If a model fails to converge, it is essential to explore why this has happened and attempt to resolve it. This may be achieved by restructuring the model so as to eliminate the problematic imbalance; for example, one might need to eliminate or combine some predictors, collapse levels, etc.). In any case the paper must report any issues of nonconvergence.

6. Mixed effects models.

The 'variable rule' model (Cedergren & Sankoff 197x, currently implemented in Goldvarb Tagliamonte 2xxx), which was long the principal tool in variationist research, is a 'fixed effects' model, meaning that every factor is hypothesized to be correlated with the dependent variable. In such a model each of the predictors must be orthogonal to the others: e.g., each following context of a variable can co-occur with each preceding context, and both can be produced by speakers of any age or social class. The 'mixed effects' models that have become prominent in the past decade are multileveled. They take into account the fact that some of the observations in the data set are clustered into larger groups – a hierarchical rather than an orthogonal data structure. In variation and change studies, the most common example of this is the individual speakers that compose social groups. An analysis of a gender effect might contrast female speakers with male speakers, with, say, Jane and Sally comprising the female group, and Pete and Harry comprising the male group. These are not orthogonal characterizations of the speakers, however: Pete can never be observed as a female, and Jane is never in the male group. Consequently a fixed effects model like Goldvarb cannot use both such analyses at the same time: one can run an analysis using gender as a predictor but not the individuals, or an analysis that uses the individual speakers but not the gender categories. This means that possible generalizations about the groups or about the behavior of specific individuals may be overlooked.

Another frequent use of mixed models addresses lexical idiosyncrasy. An atomistic phonological or syntactic analysis may have predictors like monosyllabic vs. polysyllabic words, or nouns vs. verbs, but this does not address the possibility that individual words may differ from other with the same number of syllables or the same word class. Again, the specific word and its phonological or syntactic status are not orthogonal.

A mixed effects model seeks to address this issue by treating the higher level groups as fixed effects, systematically correlated with the dependent variable, while simultaneously treating the items that compose those groups (in our examples, the individual speakers or words) as 'random effects', meaning that they are assumed by the model to vary randomly around the central value for their group, in a way that is not systematically correlated with the dependent variable. This provides richer information in the model: every token in the data set is identified not only as having been uttered by a male or female speaker, but also by precisely which speaker uttered it, and every word is treated not only as an intersection of certain phonological and syntactic properties, but also as a specific word.

Random effects. The point of using random effects is to test a hypothesis that individual speakers or specific words vary idiosyncratically, in ways that are not adequately explained by their social characteristics (for speakers) or morphosyntactic and phonological characteristics (for words). For speakers, this is always a reasonable hypothesis; it is self-evident that individuals have idiosyncrasies, even if linguistically they share remarkable levels of grammatical detail with other speakers of the same language and dialect. But for words, this hypothesis runs up against the long history of theoretical debate in linguistics about lexical idiosyncrasy, beginning with the 19th century argument about whether each word has its own history or whether sound change is regular across all words. Therefore, investigations of random word effects should be approached with caution.

A principle pitfall of using random effects is neglecting to consider whether data quantity is adequate. As with all regression analysis, each predictor needs to be supported by a reasonable number of observations. This means that each speaker given a random factor should produce a reasonable number of tokens and each word that is considered a random factor should occur reasonably often in the corpus. What is 'reasonable'? A specific answer is difficult to give, but rules of thumb can be given. Fifty tokens per random factor should give adequately reliable results in most cases (although some authors suggest that 100 tokens per speaker is better; cf. Guy 1980). Random factors supported by less than about 30 tokens should be considered suspect, and items with Ns less than 20 should not be analyzed individually; rather, they should be combined in a residual category.

Achieving reasonable Ns for individual speakers is usually feasible, but lexical items present a serious problem. Lexical frequency notoriously follows at Zipfian distribution, with a handful of high frequency and a very large number of words with very low frequencies. Hence a typical study that seeks to use random factors for lexical items will ordinarily have adequate Ns for just a few words. Words that occur only one or a few times cannot be meaningfully tested with a random factor; rather, they must be pooled, which obviates the point of investigating individual words. Furthermore, numerous studies have shown that lexical frequency itself has a systematic

effect on certain phonological properties. Consequently, it is likely to be very difficult to distinguish a random effect associated idiosyncratically with a lexical item from the lexical frequency of that item. Hence it is unwise to routinely treat word as a random effect. If the analyst has a specific hypothesis about lexical idiosyncrasy, it should be a focus of the research, treating the word as a primary (i.e. fixed) effect, not a collateral element in a model organized around other categories that group words together.

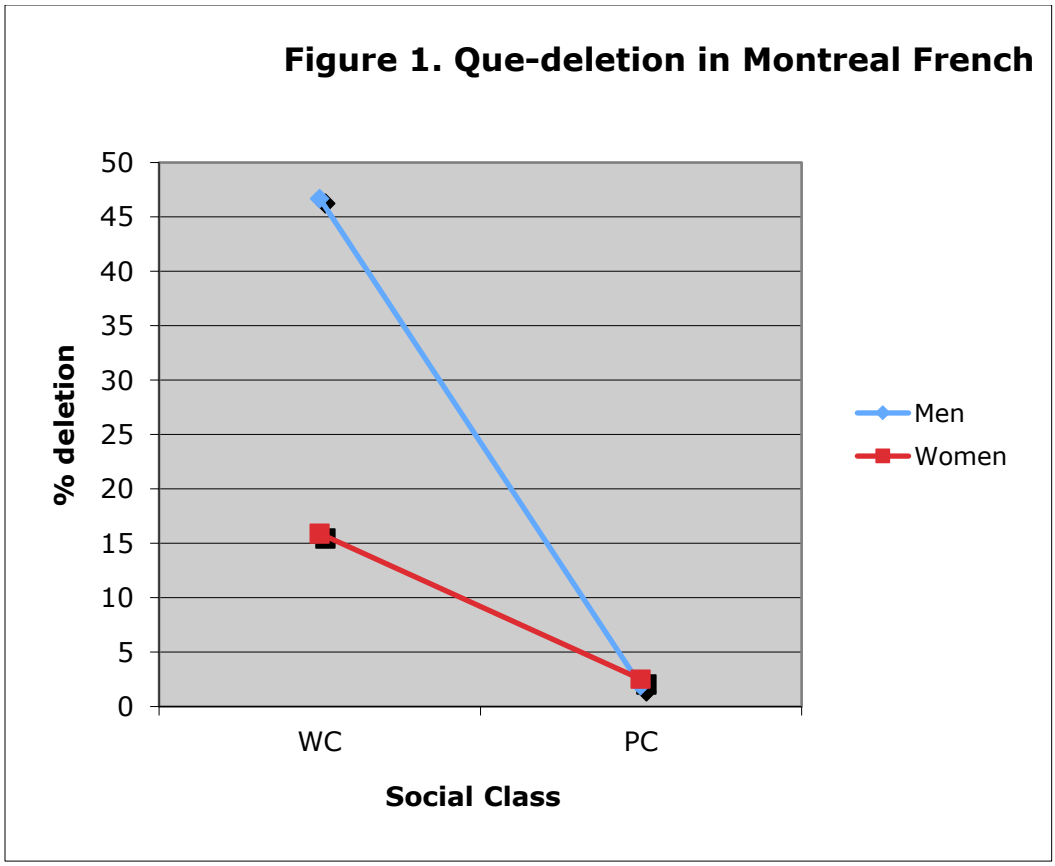
A related problem with random factors is that most models assume that they have a normal distribution. This assumption is more likely to be satisfied in a large population, but with a small number of speakers, it may well be that there are a few idiosyncratic individuals to one side of the distribution (a skewed distribution), or that there is a bimodal distribution, etc. With words, the Zipfian frequency distribution makes it very difficult to assemble a set of words for each level of the fixed effects that can reasonably approximate a normal distribution. These deviations from normality will impair the validity of the analysis. Ideally, to satisfy the assumption of normality and achieve robust reliability, the number of speakers belonging to each social group, and the number of words (with adequate token numbers) in each lexical category should be in the range of 20-30. This ideal is rarely approached in practice, and caution should be exercised in interpreting the results of an analysis which falls short of this goal.

In any case, when mixed effects models are used, all the random effect estimates should be reported, along with the Ns and summary statistics for each item (speaker or word). Paolillo (2017) further argues that "standard deviations for all random effects should be reported". The paper should discuss the results for the random factors, addressing all the same questions that are raised for main effects: Are they significant? Are some words or speakers in fact substantially different from others? Are they fairly normally distributed? The report should also indicate how much of the variance the random factors explain; if it's a lot, the fixed effects are not doing much – i.e. the variability is not very orderly. If the random factors are doing little to improve the model, there may be no point in including them. Recall the scientific injunction known as Occam's Razor: use the fewest explanatory principles necessary. Including a random factor for every speaker and every word drastically proliferates the number of 'explanatory principles' in the model, and defies the classical objective of linguistic analysis, the search for 'linguistically significant generalizations' (Chomsky 1965).

7. Interactions.

Interactions between predictors are cases where their effects are not completely independent and orthogonal; rather, the effect of predictor A is different in the presence of another predictor B, compared to its effect when B is not present. These cases are important to identify, but often hard to interpret and understand. Wherever the analyst has reason to expect interactions, they should certainly be tested for, but this should be done purposefully and selectively. A general search for many or all possible interactions should always be avoided; it vastly complicates the analysis and greatly increases the likelihood of getting false positives for significance (with a .05 criterion, randomly varying data should show a false in one of every 20 tests). The report should give an account of the reason – the hypothesis being tested – for each interaction test, as well as indicating the result of the test. Three-way interactions are particularly difficult to interpret; the motivation for such a test should be clearly explained, along with the direction and meaning of the interaction.

It is often the case that a graphic display conveys an interaction more effectively than an interaction term in a statistical analysis. Consider for example the case of que-deletion in Montreal French shown in figure 1 (data from Sankoff, Cedergren and Sarrasin 1971).



These data clearly indicate a major effect for class (professional class speakers delete much less than working class speakers), and show a big gender effect in the working class, which disappears in the professional class. This is an interaction; if gender and class did not interact, the red and blue lines would be parallel, so that a gender difference was observed in both classes. A multivariate analysis with an interaction term would return a significant interaction effect, but the graphic display already gives a clear demonstration of the nature and magnitude of the interaction.

8. Examples.

To illustrate good reporting practices, consider the following two examples, Table 1 from Bouchard 2017 (a study of null subjects in São Tomé Portuguese) and Table 3 from Tamminga 2016 (investigating English t,d deletion). Bouchard presents a Goldvarb analysis and a LME analysis with R side-by side. The table title indicates that the figures refer to the occurrence of a null subject, not a filled pronominal subject. Ns and percentage of null subjects are given for the total data, and for each level of each predictor. Factor weights from the Goldvarb analysis are presented for every factor, and which means that the reference levels in the R analysis are clearly indicated and documented for their summary statistics. The reference levels are mostly

associated with large Ns. With Ns mostly in the hundreds or thousands, she conservatively gives at most three significant digits for the percentages and effect estimates, and two digits for the factor weights and p values. Criterial levels of the p-value are highlighted with asterisks. Effect estimates are reported for all predictor levels, regardless of the significance value.

Table 1. Significant linguistic factors for the use of null subject

(Intercept=1.48; N=4512; [Ø]=68.5%)

	Estimate	p-value	Factor weight	%[Ø]	N-total
Type of clause (vs. coordinate clause)			0.52	68.7	470
main clause	0.08	0.50	0.54	71.2	3170
subordinate clause	-0.35	<0.01**	0.44	58.4	872
			range: 0.10		
Priming effects (vs. no priming)			0.47	64.8	1009
full NP	0.17	0.17	0.52	63.1	604
null subject	0.43	<0.001***	0.58	76.9	1975
overt subject	-0.18	0.08	0.43	58.0	924
			range: 0.15		
Morphological regularity (vs. irregular)			0.54	71.1	1822
regular	-0.29	<0.001***	0.46	66.7	2690
			range: 0.08		
Semantic content (vs. external activity verb)			0.55	72.1	2773
mental activity verb	-0.08	0.46	0.53	64.3	603
stative activity verb	-0.49	<0.001***	0.43	61.8	1136
			range: 0.12		
Person and number (vs. 1st person singular)			0.46	64.5	1595
2nd person singular	-0.06	0.86	0.45	58.7	46
3rd person singular	0.31	<0.001***	0.54	73.5	2037
1st person plural	-0.02	0.86	0.46	54.7	406
3rd person plural	0.50	<0.001***	0.59	72.4	428
			range: 0.14		
Animacy (vs animate)			0.30	67.4	4309
inanimate	1.73	<0.001***	0.70	91.1	203
			range: 0.40		
Coreferentiality (vs coreference with subject)			0.72	77.8	2491
Switch, coreference with IO	-2.04	<0.001***	0.25	33.3	21
Switch, coreference with DO	-0.82	0.01*	0.54	61.7	47
Switch, coreference with OO	-0.90	0.05*	0.52	60.9	23
Complete switch	-1.03	<0.001***	0.48	57.1	1930
			range: 0.47		

Bouchard 2017

Tamminga's Table 3 presents a GLMM model, and her title informs the reader that the numbers refer to the occurrence of retention (rather than deletion) of T/D. The basic statistics, Ns and percentages, are given for every level of every predictor, including the reference levels. Estimates and standard errors (SE) are given for each predictor, as well as a significance figure.

With token counts in the thousands, Tamminga conservatively reports just two or three significant digits for all the values given.

The predictors in this table are usefully divided into 'Predictors of Interest' – those that address the research questions that the paper poses – and 'Control Predictors, which are contexts that prior research has amply documented to affect this variable: preceding and following segment and the morphological makeup of the word. This analysis is also noteworthy for incorporating two continuous predictor variables, the lag between the priming context and the target token, and the frequency of the lexical item.

	Estimate	SE	$Pr(>\|z\|)$	N	% retained
Predictors of interest					
Prime variant (vs. deletion)				(del.) 2839	46
Retention prime	-0.083	0.14	.56	2942	55
Morphological match (vs. mismatch)				(mismatch) 2075	58
Match	-0.43	0.13	.0011	3706	47
Log lag	-0.05	0.069	.50	N/A	N/A
Interaction terms					
Prime x matched	0.64	0.18	$< .001$		
Prime x lag	0.17	0.09	.065		
Matched x lag	0.20	0.08	.020		
Prime x matched x lag	-0.39	0.12	.0013		
Control predictors					
Target morphological category (vs. monomorph)				(mono) 4075	40
Polymorpheme	-0.43	0.13	.0011	1706	75
Preceding segment (vs. liquid)				(liquid) 540	59
Nasal	-0.61	0.12	$< .001$	2032	46
Obstruent	0.20	0.13	.13	1181	76
Sibilant	-0.58	0.12	$< .001$	2028	38
Following segment (vs. approximant)				(appx.) 1070	40
Obstruent	-1.36	0.11	$< .001$	1286	16
Pause	0.82	0.090	$< .001$	1568	62
Vowel	1.09	0.088	$< .001$	1857	70
Log frequency	-0.13	0.015	$< .001$	N/A	N/A
Speaker gender (vs. female)				(F) 3202	53
Male	-0.29	0.087	$< .001$	2579	47
Intercept	1.26	0.23	$< .001$		

Table 3: GLMM predicting retention by morphological match in TD data, N=5781. AIC = 6251.5.

References

Bouchard, Marie-Eve. 2017. *Linguistic variation and change in the Portuguese of São Tomé.* Doctoral dissertation, Department of Linguistics, NYU.

Cedergren, H. & D. Sankoff 1974. Variable rules: performance as a statistical reflection of competence. *Language* 50: 333-355.

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge MA: MIT Press.

Guy, Gregory R. 1980. Variation in the group and the individual: the case of final stop deletion. In W. Labov, ed., *Locating Language in Time and Space*, 1-36, New York: Academic Press.

Paolillo, John. 2017. Proposed Guidelines for Reporting Mixed-Effects Models.

R-project for statistical computing. 2018. *R*. https://www.r-project.org/

Sankoff, D., S. Tagliamonte & E. Smith. 2015. Goldvarb Yosemite: a multivariate analysis application for Macintosh. University of Toronto.
http://individual.utoronto.ca/tagliamonte/goldvarb.html

Sankoff, G., Cedergren, H. J. & Sarrasin, R. (1971). Quelques considerations sur la distribution sociolinguistique de la variable QUE dans le francais de Montreal. Paper read at the 39th Annual Meeting of the Association canadienne-francaise pour l'Avancement des Sciences, Sherbrooke. [Need to verify!]

Tamminga, Meredith. 2016. Persistence in phonological and morphological variation. *Language Variation and Change* 38:335-356.

Extras, questions:

1) Blurb for NWAV workshop session (can serve as a basis for an abstract or preface?):

Variationist research has experienced a substantial proliferation in models and methods for quantitative statistical analysis. Results obtained from general-purpose statistical software largely designed for other fields must be adapted to the study of language variation and change, and accommodate the data types encountered in language. With a focus on regression analysis for natural production data, this brown bag will suggest author guidelines for providing sufficient information in tables and figures for reviewers and readers to evaluate the quantitative argumentation. Presentation by Greg Guy will be followed by Q & A.

2) Do we need to incorporate Paolillo's remarks about significance, complete models, use of disparate analysis types, etc?