PALATALIZATION

PURPOSE: The purpose of this study will be to see if an American English sociolinguistic variable will reveal distinctions on the basis of the variables of age, gender, status, and style and on the basis of the identity of the item to be modified.

FEATURE: We will determine the several types of palatalization which can occur and select ones for this study. They appear to involve any apical (tip of the tongue) consonant which occurs before the sound [y]. For example, if the sound [s] occurs directly before a [y] it may be 'palatalized' to [ß] (pronounced farther back in the mouth). The sentence 'I miss you' may, therefore, sound like 'I mishoo.' The group we will look at includes the following:

[s] ----> [ʃ] as in 'miss you'   [z] ----> [ʒ] as in 'sees you'
[d] ----> [ʤ] as in 'need you' [t] ----> [ʧ] is in 'wants you'

Only stressed items in the first segment and unstressed items in the second will be used. We will not investigate the influence of [y] on [n] and [l]; they are often very difficult to hear. This feature may prove especially interesting since it appears to be a 'casual' or 'allegro speech' phenomenon rather than a 'standard- versus 'nonstandard' one.

METHODOLOGY: You will tape-record an 'interview' with each respondent for a minimum of three minutes. You may discuss any topic you like in these interviews. For example, you might ask your respondent to tell about something they did this last summer, or about their most embarrassing moment — anything, in fact, which will elicit a lengthy sample of speech from them. In addition, we will test this variable along the 'stylistic continuum' in two other places — a reading passage (more formal than the conversation) and a 'phrase list' (more formal than the reading passage).

You will record the following respondents (eighteen in all — 3X3X2):

Age:	15-30 (Y)	30-60 (M)	60+ (O)
Status:	Upper Working (W)	Lower Middle (L)	Upper Middle (M)
Sex:	Male (M)	Female (F)	

ANALYSIS: After the recordings are made, you will identify the palatalized and non-palatalized variants in the interview, the phrase list, and the reading passage from your respondents. We will then combine all the data from the class. We will use "GOLDVARB" for this analysis.

WHY? Why might linguists need to use a statistical program other than one of those widely available? Human language facts do not always satisfy the rigorous demands of some statistical operations. Suppose you want to study English nasals, and you decide that position within the word is important. You can be guaranteed that the slot for word-initial velar nasals will never be filled.

Performance facts also often fail to meet the requirements of statistical programs. Suppose you are examining the shape or presence of relative clause markers. You get a gob of subjects, objects, and indirect objects, but when it comes to clauses whose heads are formed on a genitive (*I know a woman named Alice whose linguistics book was stolen*) or an object of comparison (*I know a lot of people who Sue is shorter than*), you may get very little data.

**VARBRUL SAVES THE DAY!** VARBRUL was designed to take care of both these potential lacks. Performance lacks are particularly likely when data is collected in natural settings, in which control of the output is extremely limited. Sociolinguists are extremely prejudiced in favor of such data, for they feel it is best representative of a speaker's vernacular or most deeply-embedded rule system.

The following excursion into VARBRUL will give you enough ammunition to carry out such a program of research yourself. This demonstration will be limited to the VARBRUL program written for Macintosh, "GoldVarb," available free from

http://individual.utoronto.ca/tagliamonte/goldvarb.htm

where you can also download a Windows version and download manuals, etc….

A RESEARCH SAMPLE
STEPS TO FOLLOW IN CARRYING OUT A VARBRUL ANALYSIS;

#1 TOKEN FILE A data file lies behind what all VARBRUL programs call a TOKEN FILE — a list of the occurrences of the variable of interest arranged for VARBRUL processing. GoldVarb imports data from your data file and converts it into a token file.
Preparation of the data file: The data file (partial) for the research into palatalization looks like the following:

　Here's my palatalization data
　(YWM1SN miss you
　(YWM1DN
　(YWM1ZN
　(YWM2DP
　(YWM2SP etc…

　a) Each line which contains a token must begin with a '(.'
　b) The items are listed after the "(" according to these symbols:
　　Column 1 is Age: 15-30 (Y), 30-60 (M), 60+ (O)
　　Column 2 is Status: Upper Working (W), Lower Middle (L), Upper Middle (M)
　　Column 3 is Sex: Male (M), Female (F)
　　Column 4 is Style: Phrase list (1), Reading passage (2), Interview (3)
　　Column 5 is Item: Z, D, T, S
　　Column 6 is Palatalization (P), not palatalized (N)
　c) Nothing after a space (column 7) is read as part of the token; it can contain notes or context from which the data were taken.

IMPORTANT: Your file MUST be saved in Text-Only (or ASCII) to be imported. You are not in this mode if you have any word-processing junk (e.g., bold, italics, underlining, superscript) in your file. Once your file looks like the one above, you can import it with File -> Import. An imported file is changed into a GoldVarb token file (name.tok) after it is saved. You can create a token file directly in GoldVarb with the command File ->New, but it is not recommended.
Briefly then,　a) If you want to enter data in your own program, do so in a text file and use the File -> Import command to change it to a GoldVarb token file, or
　　b) create a GoldVarb token file with the File -> New command, or
　　c) open an old GoldVarb token file with the File -> Open command.

#2 FACTOR SPECIFICATIONS The 'Factor Specification' window that pops up with a new token file asks you to enter the number of factor groups ('Set No of Groups') and the values in each group ('Legal values'). If you have imported a file, you may do this by using the Tokens -> Generate factor spec's command. (Use "?" for the 'Default value' category.)

#3 CHECK TOKENS Run the Tokens -> Check tokens command to see that you have coded everything correctly. You may ask to see the factor specifications by running the Tokens -> Show factor spec's command. That will create a 'Results' file (name.res).

#4 CONDITION FILE You must now have a 'Condition file' which tells GoldVarb what you want to study. Since we have decided to use Palatalization (P) as the dependent variable, we must make sure it is the first group in the 'Condition file.' In early runs of data, we are normally interested in everything we have coded, so we will want as many of the factors as independent variables as possible. Use the GoldVarb Tokens -> Recode setup… command to produce the following:
(6)
(1)
(2)
(3)
(4)
(5)
)
#5 CELL FILE Run Cells -> Load cells to memory and tell it to make the cell file from the token and condition files you have on the screen. It will name the new cell file for you (name.cel) and you should save it.

#6 APPLICATIONS VALUES: VARBRUL will ask you to choose the applications value. Since there are two conditions in the first (dependent variable) line of the condition file (P and N), we will specify 'PN.' 'P' represents the rule application mode of the dependent variable (palatalization) and 'N' the nonapplications mode (no palatalization).

#7 RESULTS FILE: After that, a 'Results File' (name.res) will be created. You can check overall results and percentages on it and interesting preliminary trends can be identified here. You will also be able to identify any factor groups that cannot be run since they may be 'KNOCKOUTS.'

#8 CROSSTABS: Now you must see if your file contains any interactions. VARBRUL assumes that all your independent variables are noninteractive. The procedure for checking such possible interaction is known as "Crosstabs" and is found on the "Cells" menu. You must run each possible combination of your factor groups for a thorough check. Here, for example, I must check 1 & 2, 1 & 3, etc..., then 2 & 3, 2 & 4, etc.... If a combination of factors interacts, it will show up in the percentages of the Crosstabs calculation. Here's what Group #1 and Group #2 look like when treated this way:

• CROSS TABULATION • 1/23/12 12:27 PM •••
• Cell file: Palatalization.cel • 1/23/12 11:36 AM
• Token file: New Palatalization.tkn • Conditions: Untitled.cnd
Group #1 -- horizontally. Group #2 -- vertically.

```
               Y      %      M      %      O      %       ∑      %
      + - - - - + - - - - + - - - - + - - - - + - - - -
W  P:    52    26:    79    38:    87    47|    218    37
   N:   147    74:   131    62:    97    53|    375    63
   ∑:   199      :   210      :   184       |    593
      + - - - - + - - - - + - - - - + - - - - + - - - -
L  P:    68    29:    55    26:    73    43|    196    32
   N:   165    71:   153    74:    98    57|    416    68
   ∑:   233      :   208      :   171       |    612
      + - - - - + - - - - + - - - - + - - - - + - - - -
M  P:    48    26:    57    32:    58    30|    163    29
   N:   137    74:   121    68:   134    70|    392    71
   ∑:   185      :   178      :   192       |    555
      +---------+---------+---------+---------
∑  P:   168    27:   191    32:   218    40|    577    33
   N:   449    73:   405    68:   329    60|   1183    67
   ∑:   617      :   596      :   547       |   1760
```

There is mild interaction here, but in general, the older, the more palatalization (the rows) and the higher class, the less palatalization (the columns). Just suppose however, that that was not the case; suppose that age radically influenced the direction of palatalization for class. You could do this:

```
(
(6)
(0 (i (AND (COL 1 Y) (COL 2 W) ))
   (h (AND (COL 1 Y) (COL 2 L) ))
   (g (AND (COL 1 Y) (COL 2 M) ))
   (f (AND (COL 1 M) (COL 2 W) ))
   (e (AND (COL 1 M) (COL 2 L) ))
   (d (AND (COL 1 M) (COL 2 M) ))
   (c (AND (COL 1 M) (COL 2 W) ))
   (b (AND (COL 1 M) (COL 2 L) ))
   (a (AND (COL 1 M) (COL 2 M) ))
(? (ELSEWHERE)))
(3)
(4)
(5)
)
```

This should give you some sense of the LISP notation (at least the function of the operator "AND"). In conclusion, our Crosstabs operation has reduced the 6 factor groups to 4 (by combining 1 and 2) and made a more complex factor group of the combination of Age and Class. You can read the above as follows:

    Line 1 = Yong Working Class results
    Line 2 = Young Lower Middle Class Results
    etc…

#9 STEP-UP-AND-DOWN: Next, there is a regression procedure built in to advise us if any of the factor groups we have selected for study is not significant at the .05 level. We will not want to run an analysis that would include entire factor groups that do not significantly contribute to the probability of 'a-prefixing.' In the "Cells" Menu, this procedure is the "Binomial Up & Down." Run it. The procedure reveals, in fact, that all our Factor Groups are significant, but if one had not been, we could simply return to the condition file and delete it.

#10 VARBRUL: Go to the 'Cells' menu and run 'Binomial one-level.' Here's the good stuff.
• BINOMIAL VARBRUL, 1 step • 1/23/12 11:40 AM ••••••••••••••••••••••••••••••••••
Name of cell file: Untitled.cel   Averaging by weighting factors.
One-level binomial analysis...   Run # 1, 205 cells:
Convergence at Iteration 6   Input 0.299

Group	Factor	Weight	App/Total	Input&Weight
1:	Y	0.422	0.27	0.24
	M	0.495	0.32	0.30
	O	0.594	0.40	0.38
2:	W	0.545	0.37	0.34
	L	0.505	0.32	0.30
	M	0.447	0.29	0.26
3:	M	0.547	0.37	0.34
	F	0.451	0.29	0.26
4:	1	0.348	0.21	0.19
	2	0.572	0.38	0.36
	3	0.770	0.55	0.59
5:	T	0.274	0.17	0.14
	S	0.562	0.37	0.35
	D	0.553	0.37	0.35
	Z	0.647	0.42	0.44

Cell	Total	App'ns	Expected	Error
YWM3Z	2	2	1.468	0.724
YWM3T	5	2	1.813	0.030
YWM3S	4	0	2.638	7.746
YWM3D	5	2	3.258	1.394
etc…				

Total Chi-square = 281.0416
 Chi-square/cell = 1.3709
Log likelihood = -985.155

What does it all mean?
a) Convergence at iteration # — how hard the machine had to work to solve your problem. It "fails" after 20 iterations.
b) Input — overall likelihood for the rule to work. In this data run, "palatalization" happens about 30% of the time.
c) Group, Factor, Weight, etc... — the influence of the factor on the likelihood of the rule's application.
d) Cell Total, App'ns, etc...
    a) Cell — identifies the environment
    b) Total — the total number of rule application possibilities in that environment
    c) App'ns — the total number of actual applications in that environment
    d) Expected — the number of applications predicted by the best-fit model VARBRUL could construct for these data
    e) Error — the mismatch between observed data and the best VARBRUL could do. These should be low (under 2.0), or there may be interaction among the factors (more likely between social than linguistic factors). [N.B.: One may ignore artificially high Error scores created by very low numbers of observations.]
e) Total Chi-square — you want insignificance here; that is, you assume the various factors are not interacting. Here there are 10 'degrees of freedom' (Factor Group #1 with 3 factors (3-1=2); #2 with 3 factors (3-1=2), #3 with 2 factors (2-1=1), #4 with 3 factors (2-1=2), and #5 with 4 factors (4-1=3). Therefore, 2+2+1+2+3 = 10.
In a Chi-square table, we find that for 10 df, we must have at least 18.307 to reach significance. Our 281.0416 is well above it (i.e., significant), and we will hope to be able to do something about it later.
f) Chi-square/cell — A low figure here (especially one under 1.00) suggests that you are looking at the right factors to explain the variation you have studied.
g) Log likelihood — we will use this below to examine significance within factor groups.

OTHER CONSIDERATIONS:
How to Reduce Significance: We ran a regression procedure to see if factor groups were significant, but how do we know that the items within a factor group are significantly different from one another? The log-likelihood measure allows us to do a simple test. Suppose that we suspect that the weights of .562 and .553 (for 'S' and 'D' respectively) in Factor Group #5 are questionably close. First, we are statistically justified because the scores are close. Second, we are linguistically justified because there is a similarity between 'S' and 'D' as there is not between either of them and 'T' or 'Z.' Both the 'S' and 'D' forms have opposite values for voice and continuancy ('S' is –voice +continuant, and 'D' is +voice – continuant. 'T' has a – value for both features and 'D' a +. This linguistic justification is even more important than the suspicion that two weights are close. If you simply test all things whose weights are similar, you will be a swell number-cruncher but not much of a linguist.
     Once you are satisfied that your suspicions are linguistically justified, you go to the 'Tokens' menu and check 'Recode setup.' It will ask you to name a new condition file. Do so. Now you will get to check the items you want in the new run. Check 'recode' for Original Group #5 (also Factor Group #5). When the four factors appear on the right hand side, type 'T' for 'T', 'C' for 'S', 'C' for 'D', and 'Z' for 'Z'. This will combine 'S' and 'D' into one group now

known as "C." (It accomplishes this through the implementation of LISP notation; eventually, you will want to learn to do this directly yourself as you do more complex work, since this recode setup in GoldVarb cannot handle everything automatically for you.)

```
(
(6)
(1)
(2)
(3)
(4)
(5  (T (COL 5 T))
   (S (COL 5 C))
   (D (COL 5 C))
   (Z (COL 5 Z)))
)
```

If you do this recoding (which requires a new cell file as well of course) and re-run VARBRUL, you will get the following:

Run # 16, 154 cells:
• BINOMIAL VARBRUL, 1 step • 1/23/12 1:14 PM ••••••••••••••••••••••••••••••••••••
Name of cell file: Untitled.cel   Averaging by weighting factors.
One-level binomial analysis...   Run # 1, 154 cells:
Convergence at Iteration 6   Input 0.299

Group	Factor	Weight	App/Total	Input&Weight
1:	Y	0.422	0.27	0.24
	M	0.495	0.32	0.30
	O	0.594	0.40	0.38
2:	W	0.545	0.37	0.34
	L	0.505	0.32	0.30
	M	0.447	0.29	0.26
3:	M	0.547	0.37	0.34
	F	0.451	0.29	0.26
4:	1	0.348	0.21	0.19
	2	0.572	0.38	0.36
	3	0.770	0.55	0.59
5:	T	0.274	0.17	0.14
	C	0.558	0.37	0.35
	Z	0.647	0.42	0.44

Cell	Total	App'ns	Expected	Error
YWM3Z	2	2	1.468	0.725
YWM3T	5	2	1.813	0.030
YWM3C	9	2	5.898	7.477
ETC...				

Total Chi-square = 199.6476
 Chi-square/cell = 1.2964
Log likelihood = -985.183

Note that a smaller Chi-square has been achieved (and a smaller Chi-square per cell). Is this new configuration acceptable? That is, by conflating these items, have we lost any explanatory value? The answer lies in the Log likelihoods. In this run it is -985.183, but in the first run it was -985.155. Here's what we do.

Subtract:

$$
\begin{array}{r}
-985.183 \\
-\ \underline{-985.155} \\
=\ -000.028
\end{array}
$$

Multiply by 2 = -000.056 and consult a Chi-square table to see if this number is significant or not. How many degrees of freedom do we have here? The answer is the difference between the df's of the first run and this one. Recall that in run #1 there were 10 degrees of freedom. Here there are 9, and 6-5=1. The lowest figure required for significance for one degree of freedom is 3.841. Our result, 000.028, is much smaller. Therefore, we have lost no information by making the combinations we did. But, with 9 degrees of freedom, the Chi-square significance level is 16.919; our score of 199.6476 is still ugly. Why?

When we did the Crosstabs, I was perhaps not completely scrupulous in combining categories which showed even the slightest mismatch in percentages. Perhaps if I had, there would be no residual interaction. On the other hand, perhaps there is simply not enough data here to drive a perfectly good statistical model. The significance of this Chi-square shouldn't annoy us too much.

WHAT ELSE?

Could anything else go wrong? In fact, yes. VARBRUL will not tolerate a factor which predicts either the complete absence (or presence) of the dependent variable. (Note that this is different from the absence of any data, a condition we have already encountered.)

The results file will indicate that a factor which is categorical is a 'Knockout." You simply need to go back into the condition file and remove this factor (by writing the operator "nil" in front of the factor line if you like). Be sure to make a record of this information. You will want to put it in your published tables and discuss it, even though VARBRUL will not include it in the statistical model of your data.

VARBRUL will also not accept a factor group which contains only one factor. It will call it a "Singleton." The cure, of course, is to delete the entire group.

Please note that the GoldVarb Manual and the DOS on-disk documentation contain some detailed instructions, but I have found that you will need a local 'guru' to help you with them unless you are already a semi-hacker.

Any questions? Happy counting!

Please read the following phrases:

HOT CROWD	BAD BLOOD
HIT YOU	HATE CRIME
MSU	LAW BOOK
LEAVE A NOTE	HID YOUR SHOE
FAX THE MESSAGE	SEIZE YOUR PURSE
WET TOY	

Please read the following passage:

Would you believe it? I can't forget the good old days. The first time I saw you was at MSU, and I bet you had no idea we'd become such good buddies. Remember when we tried to tease your roommate, and she got mad and ran out of the house?