Inaugural lecture, Harry and Jane Willson Professorship in Humanities (2004)

Bill Kretzschmar, Department of English

--with later editorial comments by Kretzschmar in [ ]

Southern English by the Numbers

I am pleased to be here today among friends at my university. There have been many social occasions off campus when people who don't know me ask me what I do, and when I tell them that I'm an English teacher, they cover their mouths with their hands and blurt out a muffled "I better watch what I say!" Not a good conversation starter. Worse yet, if I tell them instead that I study the way that people talk differently in different places, they sometimes refuse to say anything more to me at all. Dead silence. Nervous looks for other, safer people to talk to. So I often tell people that I work for the university, and if pressed further, I tell them that I am a linguist. They then ask me how many languages I speak, thinking that a linguist must be some kind of a translator. And that's OK with me. At least we can get past my job and go on to pleasantries about the weather. But here, today, you don't have to ask me what I do, and that's good for me, to be together with people who know what I do and who don't feel threatened by it.

Before I launch into the announced topic for today's talk, let me take a few minutes to set it in context. I have to admit to being amazed and astounded by the honor that my colleagues have done me in conferring on me the special professorship which is the occasion for this talk. Let me name it once in full for you, the Center for Humanities and Arts Harry and Jane Willson Professorship in Humanities. My surname has always been the longest one on the list, and I guess it's only fitting that my new title should also

be the longest one that anybody has ever heard of, really pretty astounding in itself. Just to show you that I really am a pedant of an English teacher, to be "astounded" means, in its etymology, to feel as though you have just heard thunder. Here in Georgia we have all experienced a thunderstorm a little more closely than we might like: it makes you stop what you're doing, right now, and look around. And so I want to stop now, to say a public thank you to Jane and Harry Willson for their generosity to the university and to the Humanities (Figure 1).

Figure 1

Harry and Jane Willson Establish CHA Willson Professorship

With a $250,000 gift to the Center for Humanities and Arts, Harry and Jane Willson of Albany, Georgia, have endowed a CHA Harry and Jane Willson Professorship in Humanities. The professorship represents a collaboration between the Franklin College of Arts and Sciences and the Center for Humanities and Arts.

The CHA Willson Professorship, as it will be known, has been awarded to Dr. William A. Kretzschmar, Jr., Professor of English and Linguistics. Kretzschmar will give a CHA Lecture about his research at 4:00 pm on Tuesday, November 9, in 265 Park Hall. The title is "Southern English by the Numbers."

Harry and Jane Willson, although not graduates of the University of Georgia, have supported the University in many ways over the past twenty years. Jane Willson, who earned a BA from Wellesley College in 1945, has served on the UGA Foundation Board of Trustees. Harry Willson earned a BA from Emory in 1941 and a MBA from Harvard in 1943. They are the owners of Sunnyland Farms, the largest mail order pecan products business in the country.

Wyatt Anderson, Jane Willson, Harry Willson, Betty Jean Craige

Kretzschmar served as Director of the Linguistics Program from 1996 through 1999. He edited the *Journal of English Linguistics* for fifteen years and now serves on a number of editorial boards. He is co-editor of the *Oxford Dictionary of Pronunciation of Current English* (2001), and he provides the American pronunciations for the online *Oxford English Dictionary* and for various dictionaries in the Oxford U.S. Dictionaries program. He is author of *Handbook of the Linguistic Atlas of the Middle and South Atlantic States*, published by the University of Chicago Press in 1993, and *Introduction to Quantitative Analysis of Linguistic Survey Data*, published by Sage in 1996.

Kretzschmar has published articles on medieval literature, American English, language variation, and humanities computing. Now he is focusing in his research and in his teaching on American English, language variation, and computer methods for description, analysis, and presentation of language data. With grants from the

(Continued on page 4)

Generosity like the Willsons' is rare, and all the more valuable these days because it is all the more essential as our university seeks to be considered among the best academic institutions in the country. The citizens of the state have put in place the potential for a

great university, and yet it is additional private gifts like the Willsons' that allow Georgia to aspire to compete with Michigan or Chicago—places that have achieved their eminence in large part from the generosity of private donors. I would also like to thank Betty Jean Craige, Wyatt Anderson, and others of my colleagues who participated in the nomination process, for the confidence that they have shown in me. We have a fine Humanities faculty at Georgia, and I know that any of a large number of my peers could be standing up here today. With such a high honor comes an equal responsibility, and I hope to live up to it in my teaching, research, and service for my students and my colleagues. I also hope to live up to my responsibility towards my sponsors, not just Harry and Jane Willson but all of the citizens of Georgia, who have always given their strong support to our university and for public higher education, irregardless of the politics of the day. Finally, I'd like to thank someone here today whose friendship and support over many years have made my work a pleasure: my wife, Claudia, who for 28 years has engaged the Humanities by association if not profession.

As grateful as I am personally for the honor, the Willson Professorship honors the Humanities more than it does any person. When I mentioned the Willson professorship to Richard Bailey, a senior English Language specialist at the University of Michigan, he commented that "This is good for our field," by which he meant English Language studies. I got exactly the same comment from Harold Short at Kings College London, president of the Association for Literary and Linguistic Computing, only for him "our field" is Humanities Computing. However, neither of these areas, neither computing nor the empirical study of the English language, may seem much like the Humanities to some of my literary colleagues. They might ask, what do language and technology have to do

with literary theory or style? And truth to tell, my work does not correspond very well to the Humanities experience of the social acquaintances I talked about in my first words, those people who felt threatened that I would correct their grammar or otherwise cause them acute public embarrassment. English teachers represent the Humanities to most of our citizens, and among all teachers we are thought to possess the special right to make people squirm about grammar, or to feel stupid about what poems and stories mean. Surely *this* is not the Humanities that the Willsons wanted to honor. So, what *is* the Humanities that the Willson Professorship supports? And why should somebody who studies language and develops computer tools be a reasonable choice for that role? Let's spend a few minutes together to explore these questions.

I don't have to remind this audience that the Humanities is not just a part of the university, it is its origin. During the formative period of higher education in the West, beginning in Bologna and Paris and not long after in Oxford, students first and foremost had to master Grammar, Rhetoric, and Logic in Latin, for them a foreign language. This meant learning to read and copy the great Roman writers: orators, polemicists, historians, dramatists, mythologists, fabulists, and others. There was as yet no category of writing that corresponded to our modern sense of literature, and at that time in the Middle Ages all of these writings came under the general heading of "Ethics." Gradually, through the Renaissance, the different genres of writing that we now recognize began to be distinguished, such as the separation of historical writing from fiction or mere polemic. Also at that time, in addition to writings in Latin and Greek, things written in the European vernacular languages became vehicles for study, first in Italian and French, a little later in English. The expansion of university curricula to include serious study of

the natural sciences only began to occur at the end of the nineteenth century. Further great diversification of the university curriculum occurred during the twentieth century.

Now, in the twenty-first century, we have maintained the Humanities as the core of the modern diversified university. Our Franklin College of Arts and Sciences has a Division of Humanities which includes the Departments of Classics, Comparative Literature, English, Germanic and Slavic Languages, Romance Languages, Speech Communication, History, Philosophy, and Religion. These subjects are all clearly related to the historical development that I have sketched. The only change is that study of the content of writings—what we now call literature and literary study—now generally enjoys a higher status than the study of the languages themselves. My literary colleagues thus have good reason to think that what they do characterizes the Humanities, though perhaps some of them may not realize just how recently that status arose, only in the last century, really since 1950, after the first establishment of departments for study of vernacular literatures had occurred as part of the general expansion of the university curriculum in the late nineteenth century. Language study, as I have mentioned, is something that has been essential to the Humanities for a far longer time, though the modern study of philology in the European vernacular languages dates from roughly 1800. The other parts of the curriculum in Franklin College—Biological Sciences, Fine Arts, Social Sciences, and Physical Sciences—are all Johnny-come-latelys of the last century or so compared to the long history of the Humanities in the university curriculum, not to mention all of the subjects considered by the other thirteen schools and colleges of the university which arose even more recently than the Arts and Sciences.

These disciplines cannot exist in an ivory tower (no matter the popular metaphor), and so we must also consider the social development of the Humanities at the university. During the formative period of Western education, the Humanities taught university graduates to become the *literati* who served an illiterate aristocracy in secular affairs and an unlettered public in religion. Gradually over the centuries, the sons of men of means attended the university themselves, and so over time a university education became the hallmark of a gentleman, not just of the gentleman's clerk. And, we must note, in addition to the technical aspects of the discipline, grammar, rhetoric, and logic, the Humanities offered these young gentlemen a system of social values—nothing short of our common Classical Western inheritance, including the ideals of democracy, of justice, of tolerance, and of the necessity in society for citizens with greater means to show liberality to their fellow citizens. The writings that the Middle Ages had all classified as Ethics still taught ethics to these young gentlemen, now from the point of view of history, of literature, and of philosophy. Indeed, as John Cardinal Newman argued in 1852, the ideal of a liberal education in the Humanities had become inseparable from the ideal of a gentleman.

Furthermore, education in the Humanities became the *sine qua non* of high social status. Knowledge of the Humanities was worn as a badge of honor by gentlemen, so much so that their journals and letters were peppered with classical allusions. The failure of the nouveau riche to cite the Greats (as the Classics were called at that greatest bastion of gentility, the University of Oxford) was held against them by real gentlemen as a lack of class, a social shibboleth. Conversely, some ability to cite the Greats could confer respectability upon the user: even the worst scandal sheets among nineteenth-century

newspapers show a high-frequency of classical allusions. The Humanities, therefore, was no longer just a discipline. At the same time that the Humanities embodied the central social values of Western culture, knowledge of the Humanities came to be employed by gentlemen as a social discriminator. The Humanities had thus developed over time within the university from language study of writings that expressed man's best qualities and better nature, to become an important measure of men in society. (I have been saying men here, not men and women, because women were routinely excluded from the university at that time, something we have of course changed for the better.)

So, now we have an answer for the peculiar status of English teachers as representatives to the public of the Humanities. Since the Humanities has become the measure of social status, English teachers have become social gatekeepers. Some of us relish that role: perhaps some of you can remember choice examples of former English teachers. Some of us may even have entered the field to serve as conservators of gentlemanly society, or for language specialists, to be the guardians of linguistic propriety. Whether or not we want that gatekeeper job, we have it anyway because the public gives it to us. For instance, in Georgia we still have to administer the Regents Exam, the writing test required of all university system students for graduation, which is nothing if not a social gatekeeper to make sure that the graduates of the state's universities can write like proper gentlemen and ladies [in subsequent years the Regents Exam has been waived for students at the University of Georgia, but not at all schools in the University System of Georgia--WAK]. Many students fail, not because they are inarticulate or uneducated, but because they have a few too many dispreferred infelicities of grammar. That is the social power vested in us as English teachers. Say it ain't so?

Want to take a chance? Or are those of you without Humanities PHDs already looking for a safe person to talk to after the lecture?

Many of us, however, do not desire to be social gatekeepers and have entered the Humanities for other reasons. These are not hard to imagine. We were attracted to the traditional values of the Humanities, to study of the better nature of humanity and its expression in the languages and writings of the ages. The so-called "New Humanities," including interdisciplinary studies of gender, race, and global and regional culture, is not a radical departure from the traditional core of the discipline but instead a modern rediscovery and contemporary expression of those traditional values. I am proud to say that the Humanities is leading the way for the other disciplines in our inclusive treatment of different genders and ethnicities. My own work with the field of English language and linguistics and humanities computing might also be considered part of the "New Humanities," because it, too, recovers an ancient aspect of the Humanities, and applies new methods including new technology to traditional language study. The Humanities, whether old or New, is not threatening to us its practitioners, and should not be threatening to anybody else.

So, finally, we have an answer for what Jane and Harry Willson honored with their generosity. I don't believe that they supported the Humanities in order to preserve the social hierarchy or social gatekeeping. By their generous action, they themselves enacted one of the chief virtues that the Humanities conserve and transmit, in that they have shown true liberality to their fellow citizens. Those who complain about the "liberals" in the university have the right target and the wrong idea. The contemporary political use of the word *liberal* is what C. S. Lewis called a "verbicide," that is, the

(mis)appropriation of an otherwise valuable word for what he called its "selling power" (1960: 131-32). However, at the university we must remain *liberal* in a more original sense: it is now and has for centuries been our job in the Humanities to be the conservators of our intellectual and cultural inheritance and to transmit those values to the next generation. Harry and Jane Willson have done it themselves, and in so doing they will help us in the Humanities to do our part to keep our university, our state, and our culture vital and strong.

Now, finally, we can turn to the topic for which you have been patiently waiting, Southern English, and we can find out what the New Humanities can tell us about it. Nothing could be more obvious here in Georgia than the fact of Southern accents all around us. Part of that comes from pronunciation, such as the pronunciation of *pin* and *pen* as words that rhyme, or pronunciation of the vowel in *fire* as a near rhyme with *far*. My favorite local example of the latter is the Spell Georgia Cheer at Sanford Stadium, when 92,000 Georgia fans shout G-E-O-R-G-AH-A. Part of sounding Southern comes from words that are traditionally associated with the South, such as *lightwood* or *lighterd* for our Southern kind of kindling for the stove or fireplace, or *grits* for breakfast, or the word *bucket* when others might say *pail.* And part of sounding Southern is the use of verbs and grammatical forms not commonly found in other places, above all the use of *y'all*, but also usages like *might could*, *fixin' to*, and a-prefixing as in *a-hunting* or *a-calling.*

Whenever we hear different Southern voices around us, we hear them as different and interesting if we have not come originally from the South, and we hear them as familiar and comforting if our own roots are Southern. Either way we can tell the

difference, and we charge the difference with significance. It means something to be a Southern speaker (or not to be one), and it means something when we hear somebody speak like a Southerner (or not speak that way). Southern accents mean something in the same way that, when we hear other languages, or the accented English from speakers of other languages, we might think "French is romantic" or "German is disciplined." Southerners hear Southern and think "down home and friendly." Many non-Southerners, unfortunately, hear Southern and think "slow and uneducated" or, perhaps less unfortunately, think "Rhett Butler and Scarlet O'Hara in *Gone with the Wind*." The Humanities clearly has something to contribute when the issue is language and meaning and, more particularly, when the issue is the Southern accent as a kind of sign and what it signifies. Many in the Humanities would be happy to ask whether "Southern" constitutes a mental framework that conditions the thought of those who express themselves in it, and thus it is something that needs to be documented or deconstructed or otherwise dealt with in a principled way. Thus such a discussion falls squarely into the domain of cultural studies, and into areas treated by famous Humanities theorists like Barthes or Benjamin or Derrida or Foucault or Kristeva in their commentary on language.

However, before we jump right in and start that discussion, we need to make sure that we are all talking about the same thing when we talk about Southern English or Southern accents. As it turns out, even though we are all happy to agree about the fact of Southern accents, it is remarkably hard to specify what actual bits of language people have to use in order to be considered Southern or to sound Southern. While we can agree that Southern English does exist, we also have to recognize that there are many flavors of Southern: for example, Zell Miller just doesn't sound like Jimmy Carter. We also don't

know how old it is, whether Southern was there for George Washington and Thomas Jefferson or whether it was even there for the Civil War. And we also can't even agree on where the South is, and thus where "Southern" might be spoken. This latter problem is well known in many disciplines, but asking people where different accents are spoken is a particularly good way to study it--and gives us our first of three cases of looking at Southern English by the numbers. We can begin with perceptions of Southern English, then move on to its history, and finally end up with a technically and statistically complex study, all of which will show us in different ways that you can't really know about Southern English unless by the numbers.

Dennis Preston has asked people to draw maps of where people in America speak differently, and it should come as no surprise that Southern is the area most often drawn. Sometimes it is shown as one area out of many (1989; Figure 2).

Figure 2 (originally Figure 10 in Preston)

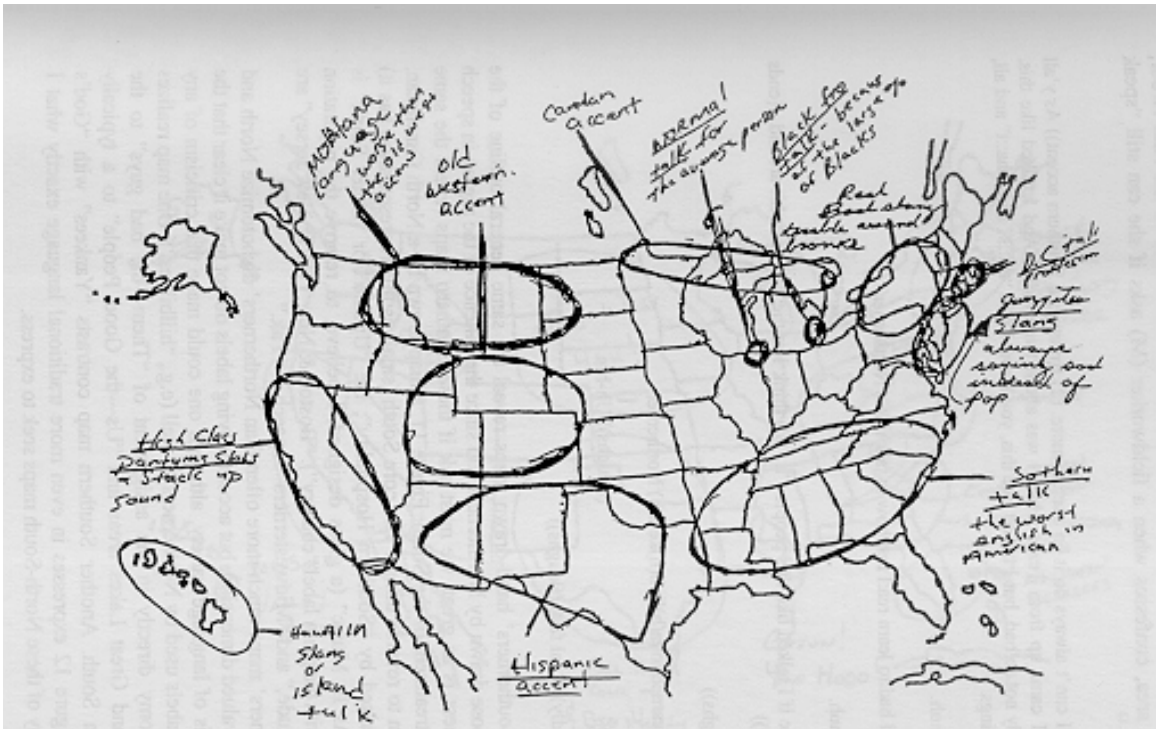


Figure 10: Chicago hand-drawn US dialect map

The labels here tell an interesting story: Chicago, the home of the mapmaker, is called "normal" while Southern is called "the worst English in America." The next figure, by a South Carolina resident, tells a different story (Figure 3).

Figure 3 (originally Figure 12 in Preston)

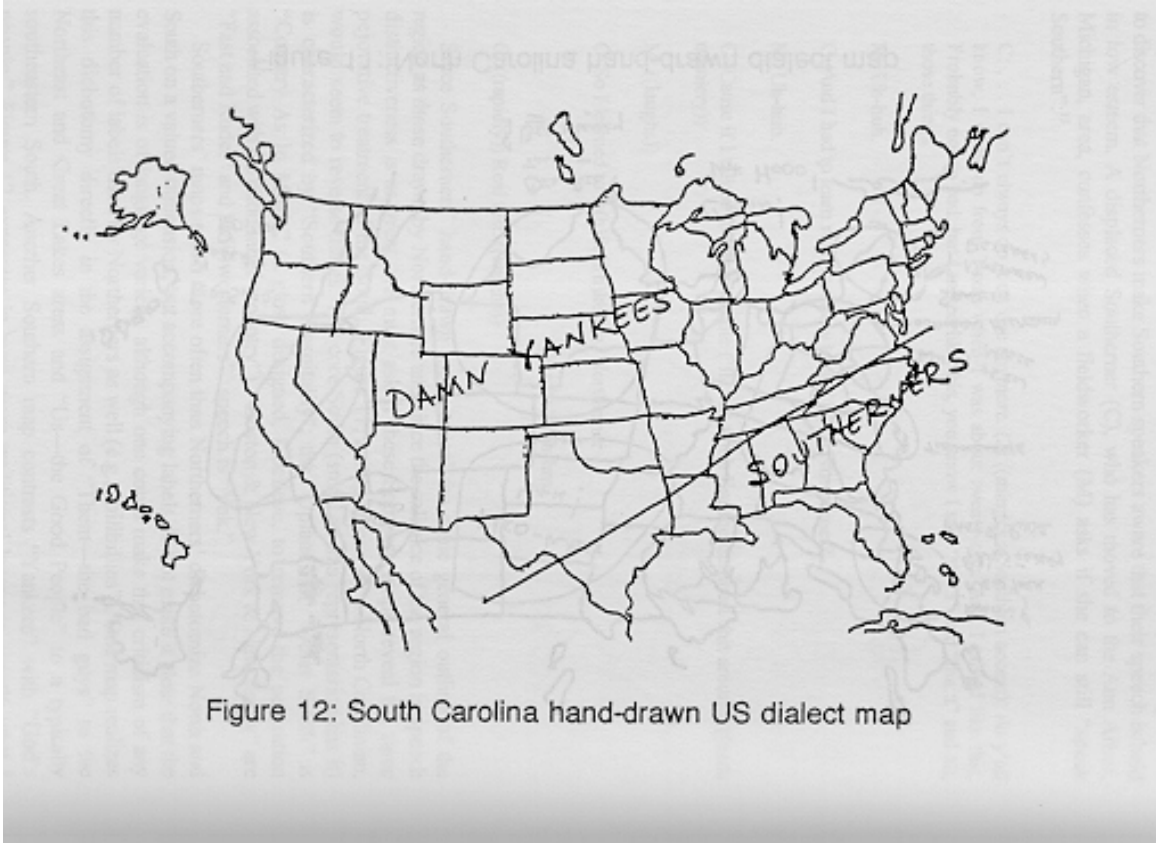


Figure 12: South Carolina hand-drawn US dialect map

Besides having to deal with the evident cultural issues that motivated the labels, Preston had to decide how to reconcile the different geographical areas drawn by all of his mapmakers. For this he needed numbers. This computer-generated map shows the area where his speakers all agreed that Southern English was spoken (Figure 4).

Figure 4 (originally Figure 5 in Preston)

Figure 5: Michigan respondents' core South at the 98% (132 of 136) agreement level

The area is so small that it looks like you have to live in Columbus, GA, or maybe across the river in Phenix City, AL, to speak Southern. This is clearly not right. The next map shows where any of Preston's speakers thought Southern was spoken (Figure 5).

Figure 5 (originally Figure 3 in Preston)

Figure 3: Southeastern Michigan respondents' computer-generalized map, showing where even one respondent outlined an area labeled 'South'.

This map includes even Chicago and the Chicago mapmaker we saw before, and so it can't be right, either. But it's not all wrong: when I was growing up in Milwaukee, WI, I was convinced that those people who lived to the South of me in Illinois had Southern accents! The map of the area where half of Preston's Michigan and Indiana mapmakers located Southern English is better in some ways (Figure 6).

Figure 6 (originally Figure 4 in Preston)

Figure 4: Indiana (outlined — 53 of 105) and Michigan (shaded — 69 of 138) respondents' generalizations at the 50% level of the United States 'South' dialect area

It looks like a compromise that maybe some of us (maybe half of us?) could agree to, though even half of Preston's Michigan group and half of his Indiana group couldn't exactly agree about the borders of the area. Preston's research clearly tells us that, even though just about all of us name the South as a dialect region, we cannot agree where it is. It also tells us that we need numbers and technology to help us understand how we think about the South. Preston's computer-generated maps tell us that the South cannot be adequately described by all or any, but instead we need to think in terms of more and

less. There is no consensus South (unless it is Columbus or Phenix City), and so in any discussion about the South the participants have to guess what the other people are talking about; we all have to estimate at what point between any and all each person we talk to may locate the South. The complexity of the cultural study of the South, whether of Southern accents or other aspects, comes in part from the fact that we usually assume that people know what we are talking about when, in fact, what they think of as the South may be quite different from how we think of it ourselves.

Our own Georgia PhD Susan Tamasi replicated Preston's method with some Georgia respondents (Tamasi 2000)--who also could not agree on boundaries of dialect areas except in probabilistic terms, in Susan's own work with the numbers--and then Susan developed new methods to study the cognitive bases of the perceptual assessments of her research subjects (2001). Susan asked people to sort cards with the names of the states on them into different piles, according to where they thought people talked alike (Figure 7).

Figure 7 (originally Figure E.5 in Tamasi)

Figure E.5: GA5 Dialect Regions

The mottled Southern region in this map indicates that this Georgia respondent does not have a coherent map of Southern speech--the different colors represent the different categories into which the respondent classified the different Southern states. Some of Susan's respondents did classify a number of the Southern states together, but others also created discontinuous classifications of the Southern states. The fact that they could do so shows us that we do not necessarily possess coherent mental maps of dialect areas, and this finding directly addresses the cognitive processes by which we classify the different kinds of speech we hear around us. It has become increasing clear that a significant part of the way that we all manage our relationships with other speakers is through such cognitive classification patterns [see Kretzschmar 2009: Chapter 7 for further discussion of this point]. This is exciting, to begin to uncover the cognitive principles that mediate the decisions we make in conversation, as we interpret what others say to us (and produce our own speech) according to how we perceive the speech

of the people we are talking to. When Susan used statistics to make a generalization from the different responses of her subjects she found, like Preston, that a Southern region did finally emerge, not as some stable notion that we all share, but as a product of the numbers, a statistical abstraction. Preston's and Tamasi's evidence both agrees that we do not share consensus about our mental maps of Southern speech, and that we have to resort to numbers if we are to make any sort of reasonable interpretation of where we might find Southern English. To do otherwise means that we are just talking to ourselves, the intellectual equivalent of singing in the shower, instead of having the kind of serious discussion that we expect in the Humanities.

Let us now consider another question about Southern English where we have to have numbers in order to make any sense. It is unwise to assume that speech habits that we associate with the South have been used there for a long time. Research by Guy Bailey indicates that, among features most commonly associated with Southern English, the pronunciation of the vowel in *fire* as a near rhyme with *far*, the pronunciation of *pin* and *pen* as words that rhyme, and the vocabulary item *fixin' to* are rare before the end of the nineteenth century (1997). This means that all those Civil War movies where the Southerners sound just like they do today are terribly anachronistic. In much the same way, when Shakespeare's plays are delivered in a modern upper-crust British accent, it tends to make the Bard seem foreign to American audiences. In many ways Shakespeare's actual voice would probably have sounded as close to American English as to Modern British English, and my students often hear my reconstruction of Shakespeare's pronunciation as sounding to them like Irish English. Nobody today still speaks Shakespeare's English, and to go looking for Shakespeare's English in remote

Appalachian valleys is the height of folly. It is better to realize that American English began about the same time as Shakespeare lived, and that language change has occurred in both Britain and America, so that there is nothing more authentic about Shakespeare as played on an English stage than on an American one. Moreover, this kind of language change continues today, and other features commonly associated with Southern speech are in rapid decline, such as lack of pronunciation of *-r* after vowels, as in words like *four* pronounced as *foa* or *foe*, or the a-prefix on verbs with *-ing* endings, like *a-running*. Some features of Southern English have been used in the South for a very long time and are still here today, but many other features come and go. So, we cannot just assume that Georgia Washington and Thomas Jefferson spoke Southern English because they came from Virginia, or even that something very close to the modern form of Southern English was the language of the Confederacy--we need to establish with some care just what people said at those times.

To return to Guy Bailey, the way that Bailey can know such things is to compare the talk of older speakers with that of younger speakers. It is simply not the case that old people all talk one way, and younger people all talk another. No government agency put out the word that all children born on or after January 1 of the year 1950 had to talk differently from their parents, any more than that all children with birthdays on or after January 1 of the year 1983 were born knowing how to use personal computers--even though it sometimes seems so. Changes in language are gradual, a matter of trends, of more and less use of particular habits of language over time. Numbers, again. Guy Bailey measured how often older speakers said certain things in comparison with how often younger speakers said them, and from these frequencies he could say which

features of Southern English seemed to be increasing (that is, were more frequent in younger people's speech), and which features of Southern English seemed to be declining (that is, were more frequent in older people's speech).

Another way to find this out is to compare results of older survey research with newer survey research, which is what another of our Georgia PhDs, Ellen Johnson, did in a now famous book on Southern English (1996). When she went back in 1990 to thirty-two of the same places where our Atlas survey had gone in the 1930s, she found that people were using most of the same words as they had sixty years before, but often at different rates. She also found, surprisingly to many, that the number of different vocabulary words in use by the 1990 speakers was larger than the number used in the 1930s. While many people might have thought that regional speech was disappearing because of putative influences of the media and education, Johnson's research tells us that language variation in the South is actually increasing, not going away.

Another of our Georgia PhDs, Allison Burkette, has shown in a detailed article that the relatively infrequent variant words in our Linguistic Atlas materials for the piece of furniture sometimes called a *chest of drawers* actually recapitulate terms found in colonial furniture pattern books or from even earlier in the history of furniture (2001; Table 1).

Table 1

LAMSAS (1930s-40s, 1162 people)		LAMSAS (1930s-40s, 1162 people)	
Responses	Occurrences	Responses	Occurences
bureau	1104	case of drawers	3
dresser	382	dresser drawers	3
chest of drawers	227	stand of drawers	2

chest	44	set of drawers	2
sideboard	34	blanket chest	1
washstand	30	cabinet	1
highboy	27	checkrobes	1
chiffonier	22	chest upon a chest	1
trunk	22	clothes stand	1
drawers	19	clothespress	1
bureau drawers	19	chifforobe drawers	1
commode	17	cupboard	1
dressing table	9	bookcases	1
box	8	cabinet table	1
stand	7	kast	1
lowboy	5	vanity dresser	1
chest on chest	4	wardrobe	1
vanity	4	wash hands stand	1
desk	3		

Johnson data (1990, 32 people)		Johnson data (1990, 32 people)	
Responses	Occurrences	Responses	Occurrences
dresser	18	bachelor's chest	1
chest of drawers	17	chest on chest	1
bureau	5	dressing table	1
chest	5	linen press	1
wash stand	2	press	1
highboy	2	vanity dresser	1
dresser drawers	1		

Trunk, cupboard, and *box* predated the colonial period; the William and Mary period contributed *blanket chest* and *stand*; the Queen Anne period *case of drawers* and *cabinet*; the Chippendale period *chiffonier, lowboy, chest on chest, chest upon chest, press, wardrobe*, and *kast*; the Hepplewhite period *dressing drawers, commode*, and *dressing table*; and finally the Sheraton period, *bookcase,* as a name for a piece that has shelves above the drawer unit. The most common American term for this piece of bedroom furniture is now *dresser*, as Ellen Johnson's data for this question shows, but in the data from the 1930s and 1940s the most common term was *bureau*. Only one term

seems to have been invented recently, *bachelor's chest*. Allison's findings suggest that all the other terms, most of them now mere relics, may have been prominent earlier--but they are still preserved at low levels of use by people today. Her findings are highly significant, because they show us that changes in our language occur by the numbers, as changes in how often people use different terms and other features. Her findings also tell us that our language has a long memory, so to speak, in that the words and other features that historically were popular in use are usually still present and being used by somebody somewhere. In order to do this research, Allison relied on the Museum of Southern Decorative Arts, in Winston-Salem, NC; she had to go to one of our great Southern Humanities institutions to find out about the objects that people were talking about, and to find out about their history. In so doing, she could tell us something important not just about furniture or Southern English, but about how language works over time.

These studies all tell us that, while individual habits of speech--whether words or pronunciations or grammatical usages--are likely to come and go, the tendency to use different habits in different regions will nonetheless continue, and regional variation will persist in much the same geographical patterns even after changes in particular speech habits (cf. Bailey and Tillery 1996). It is thus fair to say that Southern English (along with other varieties of our language) is continuously rebuilding itself, simultaneously dying away with the loss of some speech habits that formerly characterized it, and being reborn with new speech habits. We cannot rightly understand how such a paradox of continuity and change could happen, unless we resort to numbers. We need to see that our language, and by extension the culture that produces it, is something dynamic that accepts the new without entirely throwing away the old. In language, as in many another

aspect of our culture, the framework for our discussion is not so much a matter of right and wrong, or correct and incorrect, as it is a matter of the growth and development of habits and usages and ideas. People continuously invent new usages in language; most inventions never get beyond the inventor, but some of them catch on for a time, and some grow to become what we consider to be normal, whether within regional speech like Southern English or in the wider world of English. At the same time, speakers of the language hardly ever completely lose old usages, so that our language expands, and so that we cannot afford to forget or disregard what came before. In this way, the point of view that the numbers can give us, the model of culture that study of the language sets before us, provides a defense for one of the central activities of the Humanities, the conservation and transmission of our cultural inheritance.

Let us now turn to my last illustration of how the Humanities and the numbers work together to create an understanding of Southern English. This is the most technical and computer intensive application I have used, an implementation of a process known as Self-Organizing Maps (Kohonen 2001). As we have seen so far, neither in our perceptions of Southern English, nor in the production of Southern English features by its speakers, do we find anything like consensus. The situation of what we call Southern English has to be extremely complex given its historical development, its continual change and renewal, which always occurs in somewhat different ways in different places and among various culturally divergent groups of Southerners. The Self-Organizing Map approach to this complexity is to apply computer techniques that are designed to imitate classification processes in the human brain. You may have heard of the process as neural network analysis. I will not describe the statistical processes involved, but you should

know that there is in fact a mathematical model that processes the complex data (as in Table 2).

Table 2 SOM calculations

During the learning process at iteration t, neuron i is updated according to the following rule:

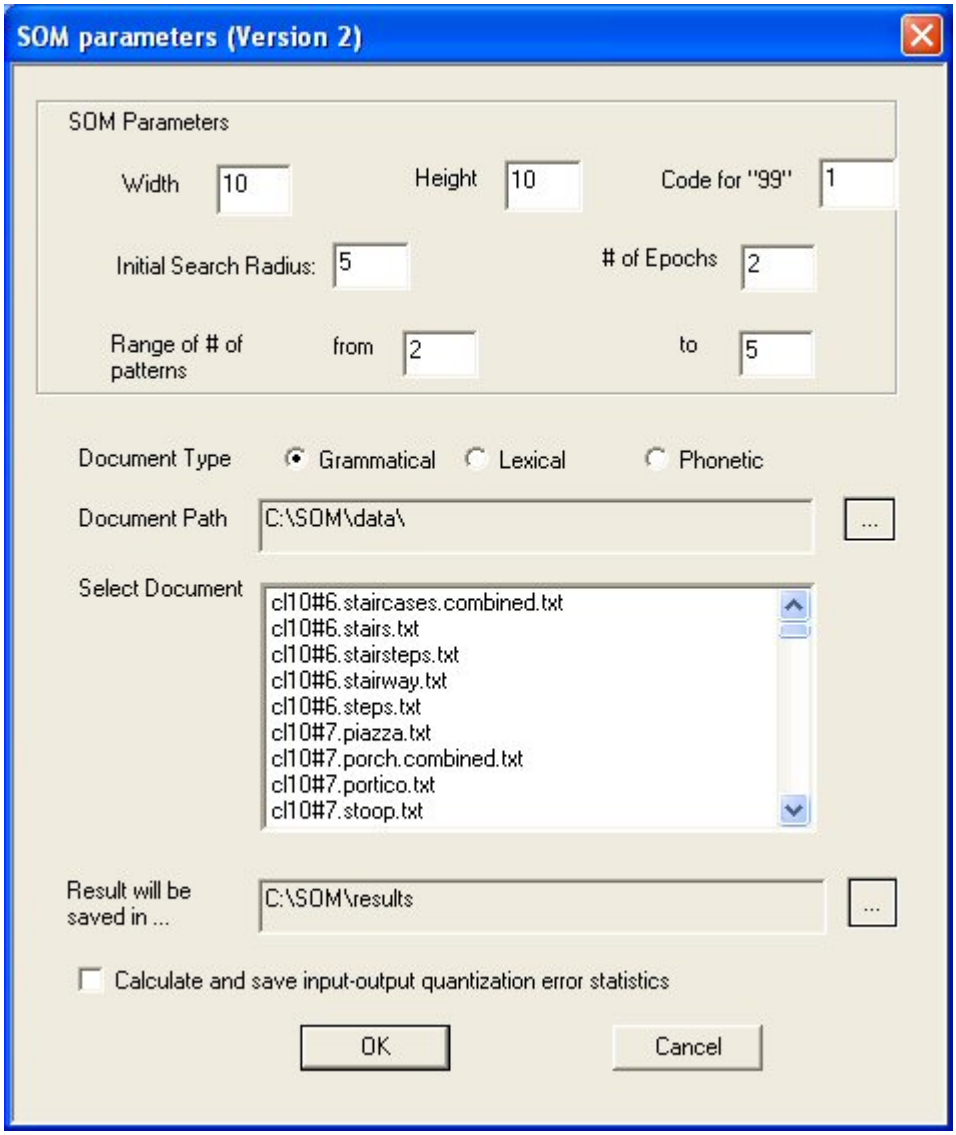
$m_I(t+1) = m_I(t) + h_{ci}(t)\,(x(t) - m_I(t))$,

where c is the index of the winning neuron of input x. In the applied literature, two simple choices of $h_{ci}(t)$ are encountered. The first instance refers to a binary function: $h_{ci}(t) = \forall(t)$ if i 0 $N_c(t)$, o otherwise, where $N_c(t)$ is a neighborhood defined around c and $0 < \forall(t) < 1$ is the learning-rate factor. Both the size of $N_c(t)$ and $\forall(t)$ are decreased monotonically during the learning process to bring it to convergence. . . . Through lateral inhibition, the neighborhood kernel confines the interaction among latent dimensions of a dataset to close neighbors. This fairly unique property is consistent with the premise of SOM that a natural order exists in input data, and that second-order processes dominate this organization. At the end of the learning process, vector m_i corresponding to neuron i contains normalized "weights" that define the mapping of the neuron in the attribute space. The weights quantify the strength of the connections between each input node and neuron i.

The neural network statistics aim at finding abstract self-organizing processes forming “maps” analogous to brain maps. These "maps" in practice consist of matrices of colored circles, each of which relates to some group of data points.

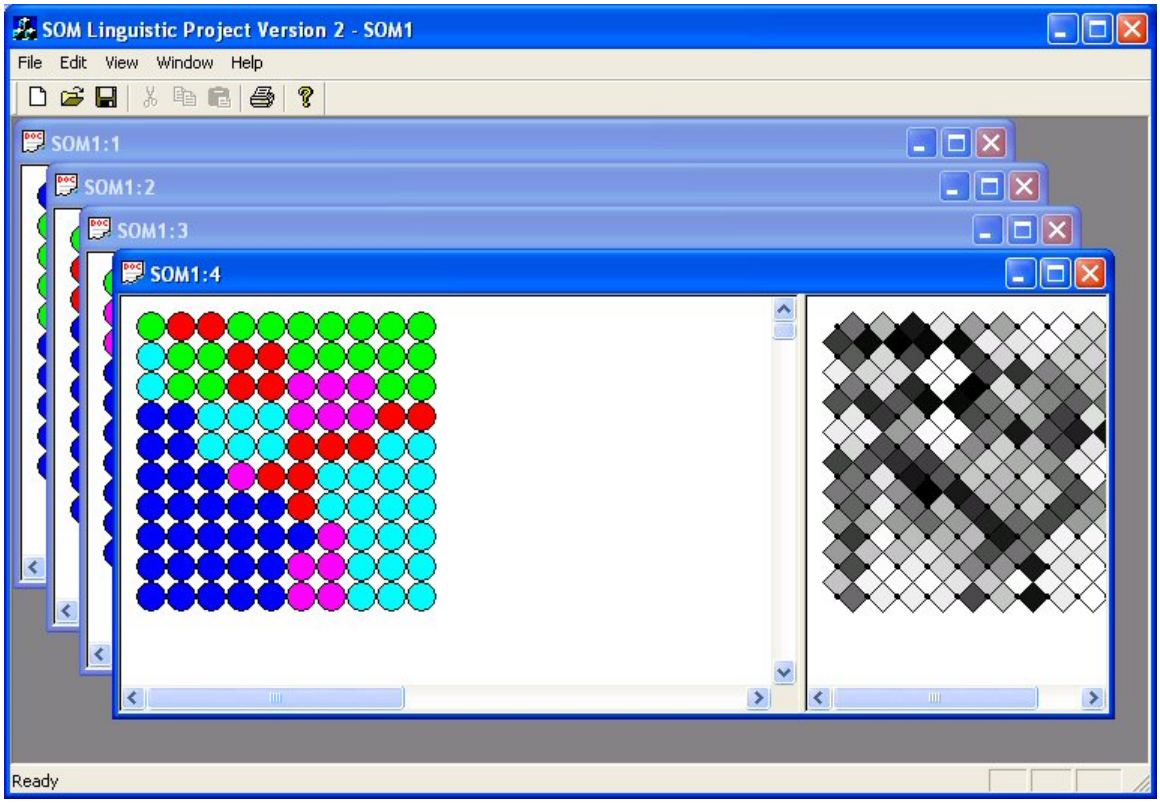
Here's how it works in the computer program written by my colleagues on this project, a team of technical geographers from Buffalo, NY, and here at Georgia (Figure 8).

Figure 8



. First we select for processing a number of questions from our Linguistic Atlas survey data. Here you can see some words for *stairs* and some words for *porches*; we select all of the variants from a number of questions to process, up to around twenty at a time. Then we run the algorithm against the data. This figure shows a program display after selection of six data files for input: 4#5 *quarter of* (for telling time), 5#3 *pretty day* (for good weather), 8#3 *dog irons* (for what you put the wood on in a fireplace), 8#4 *shelf* (for the shelf over a fireplace), 8#6 *kindling* (for the wood used to start a fire), and 9#4 *blinds* (on a roller, for window privacy) (Figure 9).

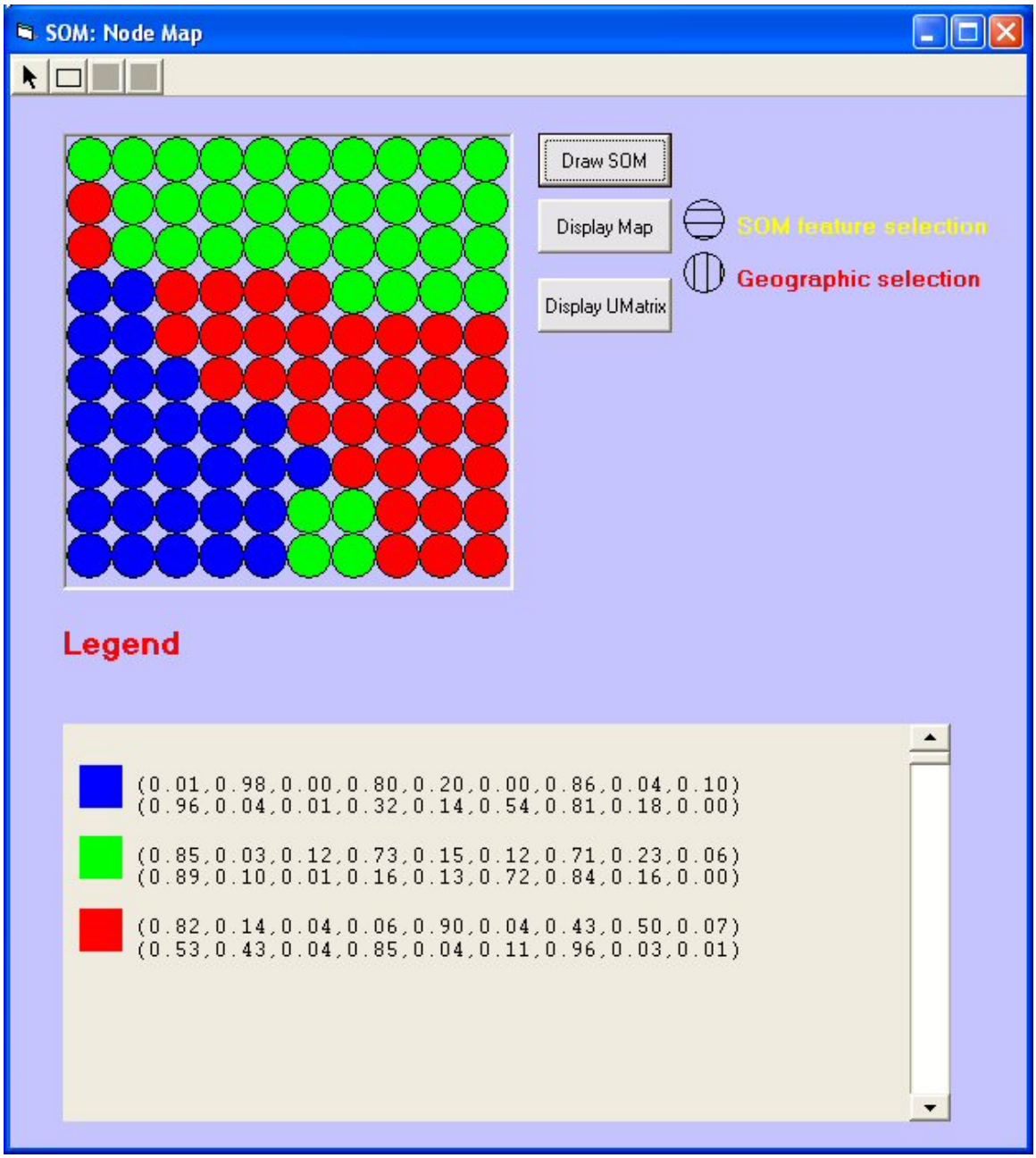
Figure 9



These particular response types (and the others I will discuss below) were selected for this experiment because they all have between 100 and 600 occurrences in the data set, out of 1162 speakers, which we have found to be a rich frequency level for finding regional distributions. The 10x10 matrix of nodes is on the left, and each “page” is a pattern set. The so-called Umatrix on the right is a measure of relatedness between nodes. The different "pages" of the display correspond to the solutions for different numbers of colored patterns among the nodes, from three colors up to a maximum of nine colors.

The next figure shows the node map for the three-pattern solution for this data input (Figure 10).

Figure 10



The average weights for the input words appear at the bottom of the display. The algorithm considers each of the six words that we used as a set of three values, for absence of the feature, for its presence, or for missing data, so the algorithm is measuring similarity for the nodes according to a vector composed of 18 values. The percentages reported in the display represent each value in the vector, and constitute a "weight." It will be appropriate for this experiment to consider quartiles for interpretation of weights and thus of the component words: 75% or better can be taken as selection of the presence or absence state of the word; 50%-75% can be taken as a trend for the presence or absence state of the word; the third number is for missing data, so if it reports a figure above 50% the word should be considered as selected for missing data; and finally, if no number is reported above 50%, the word should be considered as having no trend. So, in this case, BLUE selects (.98) for the presence of *quarter of*, a word with a known Northern distribution; GREEN does not select the presence of any feature; and RED selects for the presence of *pretty day*, a known Southern form. The SOM application also generates a report that lists the weights associated with each separate node. This chart shows the node patterns for several of the nodes in RED, by way of illustration of how the weights work (Table 3).

Table 3

```
Report file of SOM Project version 2

6   /*Number of files in use

cl4#5.quarter_of.combined.txt
cl5#3.pretty-day.combined.txt
cl8#3.dog_irons.combined.txt
cl8#4.shelf.combined.txt
cl8#6.kindling.txt
cl9#4.blinds.combined.txt
```

```
10, 10, 1   /* Height, Width and Code for 99
```

(3 ,1) , (1.000, 0.000, 0.000, 0.000, 1.000, 0.000, 1.000, 0.000, 0.000, 1.000, 0.000, 0.000, 1.000, 0.000, 0.000, 0.842 , 0.085 , 0.073) Midland (pretty day)

(4 ,3) , (1.000, 0.000, 0.000, 0.037, 0.963, 0.000, 0.232, 0.381, 0.387, 1.000, 0.000, 0.000, 1.000, 0.000, 0.000, 0.990 , 0.000 , 0.010) Midland (pretty day)

(4 ,6) , (0.266, 0.210, 0.525, 0.031, 0.453, 0.516, 0.438, 0.492, 0.070, 0.956, 0.039, 0.005, 0.624, 0.012, 0.364, 0.971 , 0.024 , 0.005) Southern (none selected)

(5 ,5) , (0.993, 0.003, 0.005, 0.003, 0.993, 0.004, 0.002, 0.991, 0.007, 1.000, 0.000, 0.000, 0.994, 0.000, 0.006, 1.000 , 0.000 , 0.000) Southern (pretty day + dog irons)

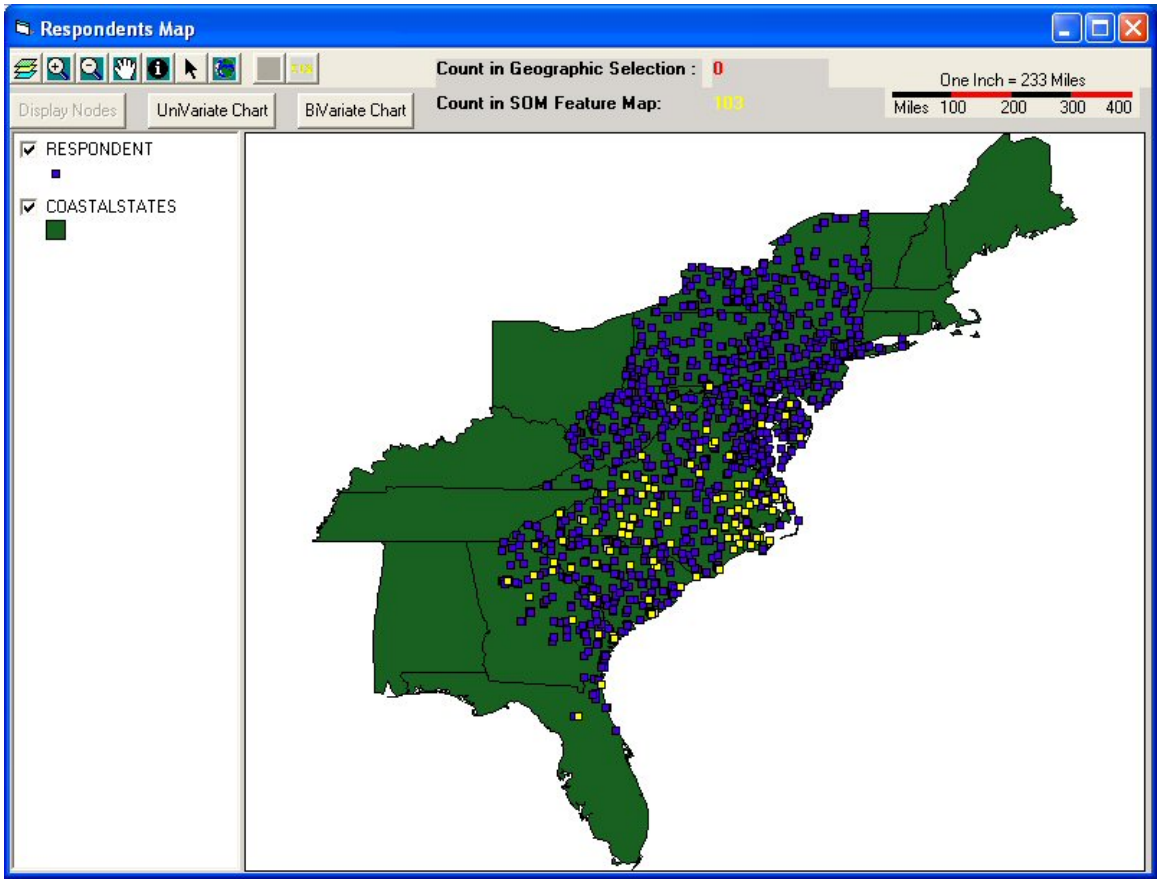
(7 ,10) , (0.856, 0.129, 0.015, 0.054, 0.930, 0.015, 0.594, 0.333, 0.073, 0.008, 0.930, 0.062, 1.000, 0.000, 0.000, 0.870 , 0.102 , 0.028) Midland / Southern (pretty day + shelf)

(10,10) , (1.000, 0.000, 0.000, 0.000, 1.000, 0.000, 1.000, 0.000, 0.000, 0.000, 1.000, 0.000, 1.000, 0.000, 0.000, 0.999 , 0.001 , 0.000) Southern (pretty day + shelf)

I have appended to the weight numbers a description of the general shape of the regional patterns, and I have listed the words selected for their presence. The individual nodes select *pretty day* by itself, or with *dog irons* or *shelf*. It is possible to have multiple nodes with the same word selected, because the application considers the variants offered by speakers for all the other items, too, and so different nodes with the same word selected just have different weights for the other words.

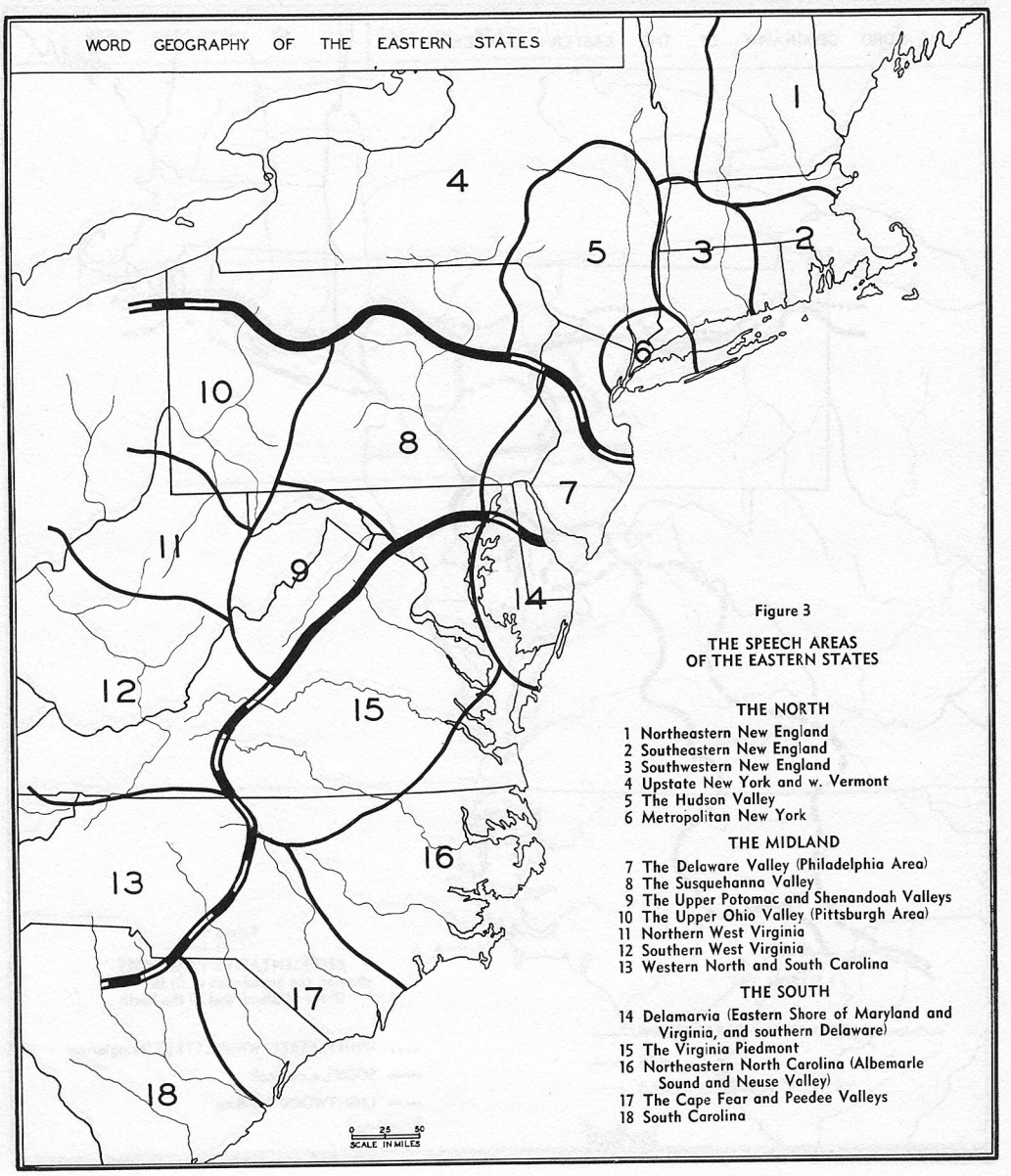
In order to bring all this mind-numbing collection of numbers and abstractions back to reality, here is a map that highlights the speakers selected for a block of nine RED nodes, distributed in a South Midland and Southern pattern (Figure 11).

Figure 11



The terms “South Midland” and “Southern” derive from Hans Kurath’s classic work in dialectology (e.g. Kurath 1949, 1961), as do the terms “North” and “Midland” used above to describe distributions of speakers (Figure 12).

Figure 12

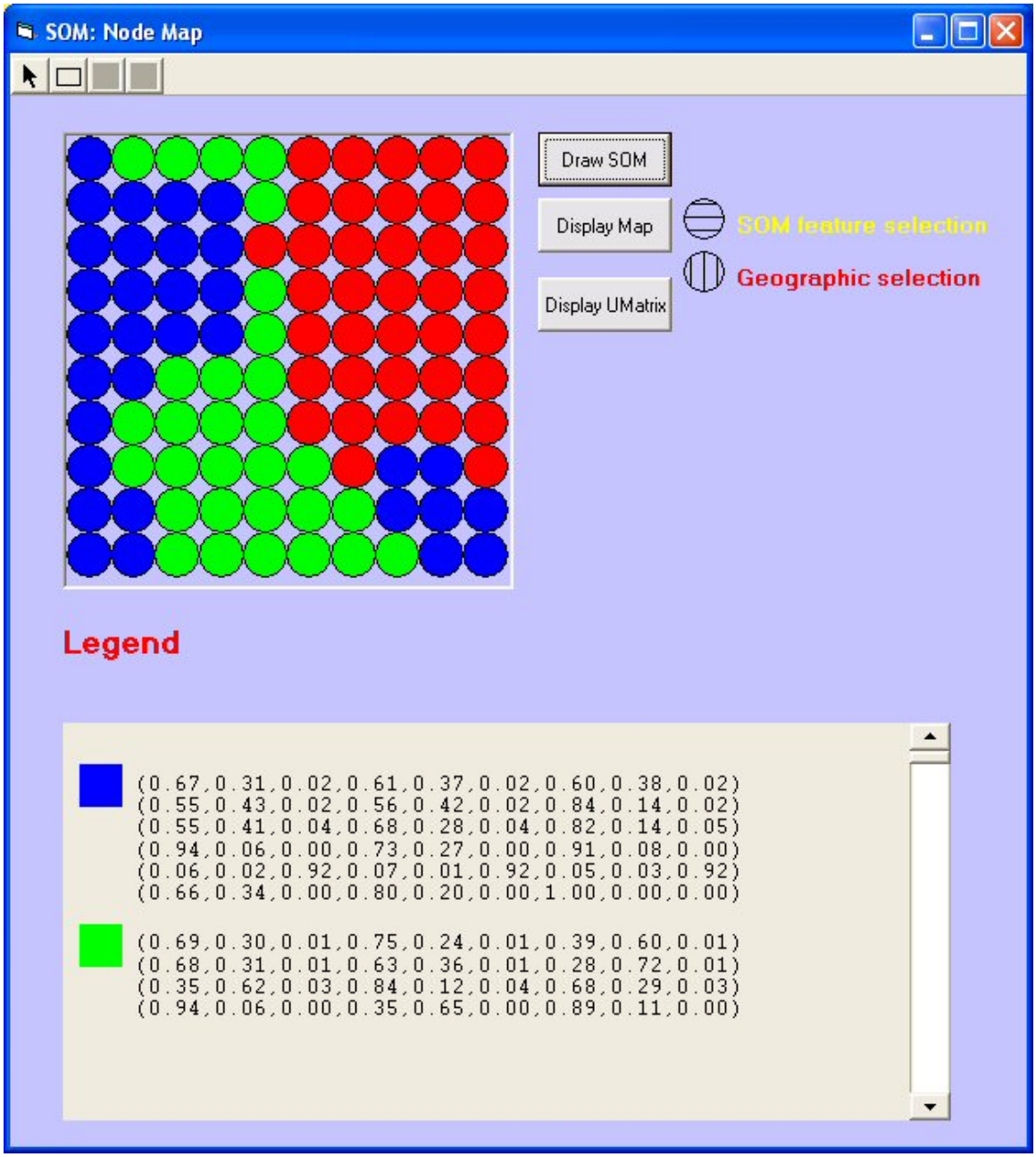


In Kurath's terminology, "Southern" refers to the Coastal and Plantation South, with cultural hearths in Eastern Virginia and Charleston. By "South Midland" he means the Upland South, particularly as it was settled down the Shenandoah River and Great Valley through western Virginia to the Cumberland Gap. Visual validation of patterns using Kurath's terms has the advantage of comparison of the new technique with older findings--more about this later. To return to the RED map, we see that not all of the speakers from the South have been highlighted, but that almost all of the highlighted

speakers do come from the South. Only a few words were included in this experiment in order to be able to keep the illustrations so far simple enough to understand, and so we should not expect too much from it, but even with a few words the algorithm does appear to have selected a group of Southern speakers for one of the major color patterns.

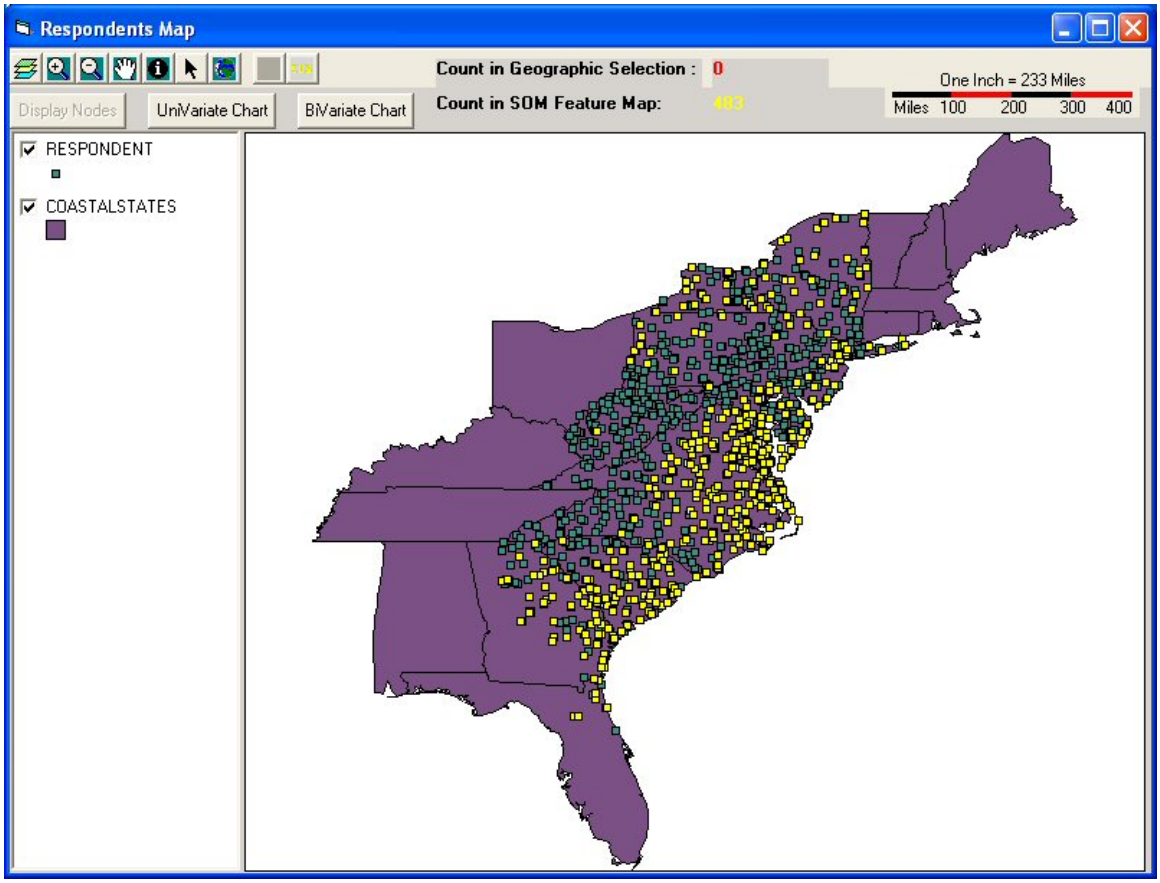
In order to show you a more complete picture, I have made one more SOM experiment with eighteen responses (including those from the foregoing illustrations), and have asked the computer program to produce up to nine patterns instead of the previous five. We will now see the result of the computer algorithm choosing from 3^{18} possible nodes (nearly 400 million nodes), vs only 3^6 possible combinations (729 nodes) before. When the algorithm selects from the millions of possible nodes and arranges them into three patterns (Figure 13).

Figure 13



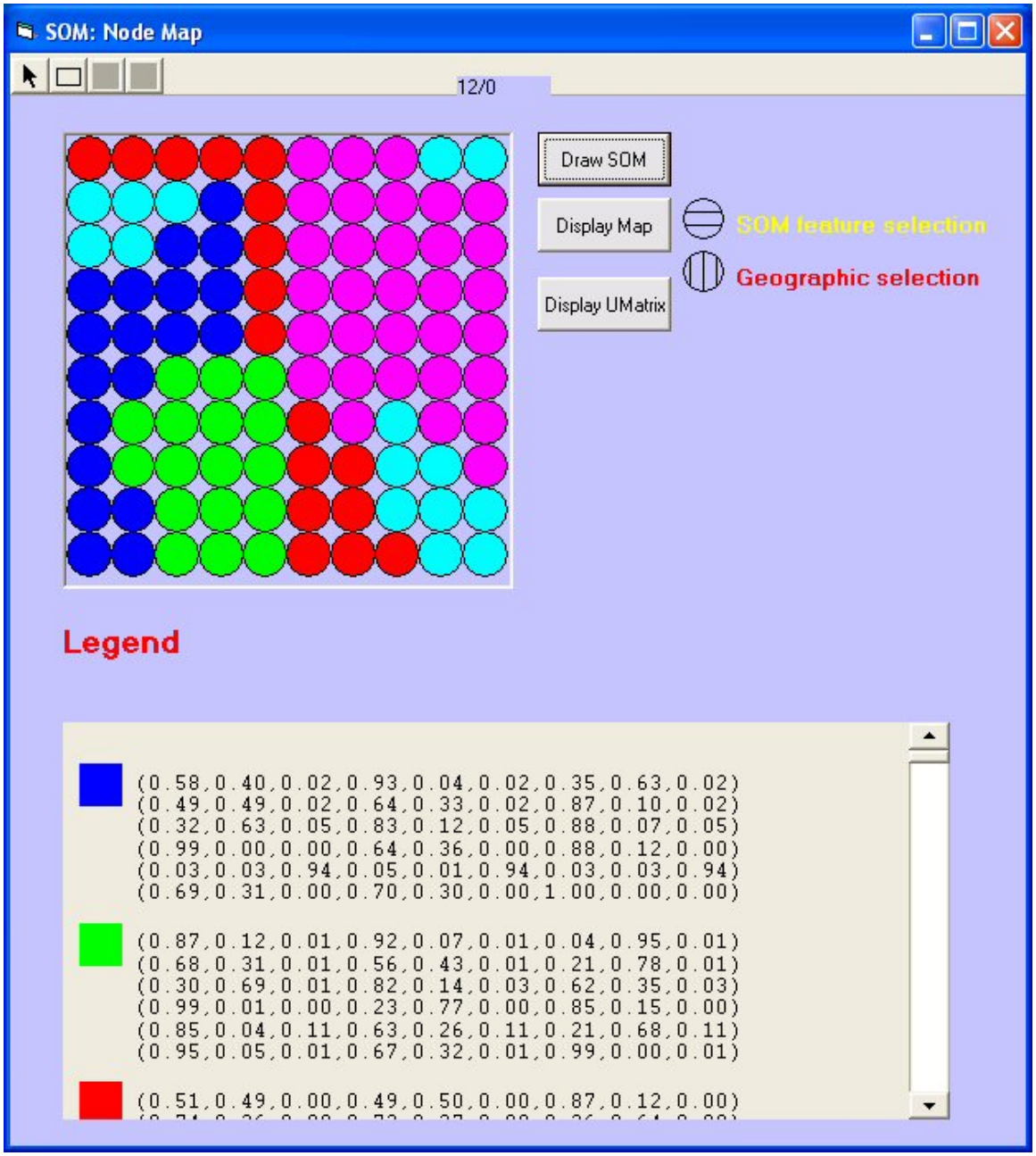
the RED pattern shows a mainly Coastal Southern grouping, though with some leakage in NJ and in NY state (Figure 14).

Figure 14



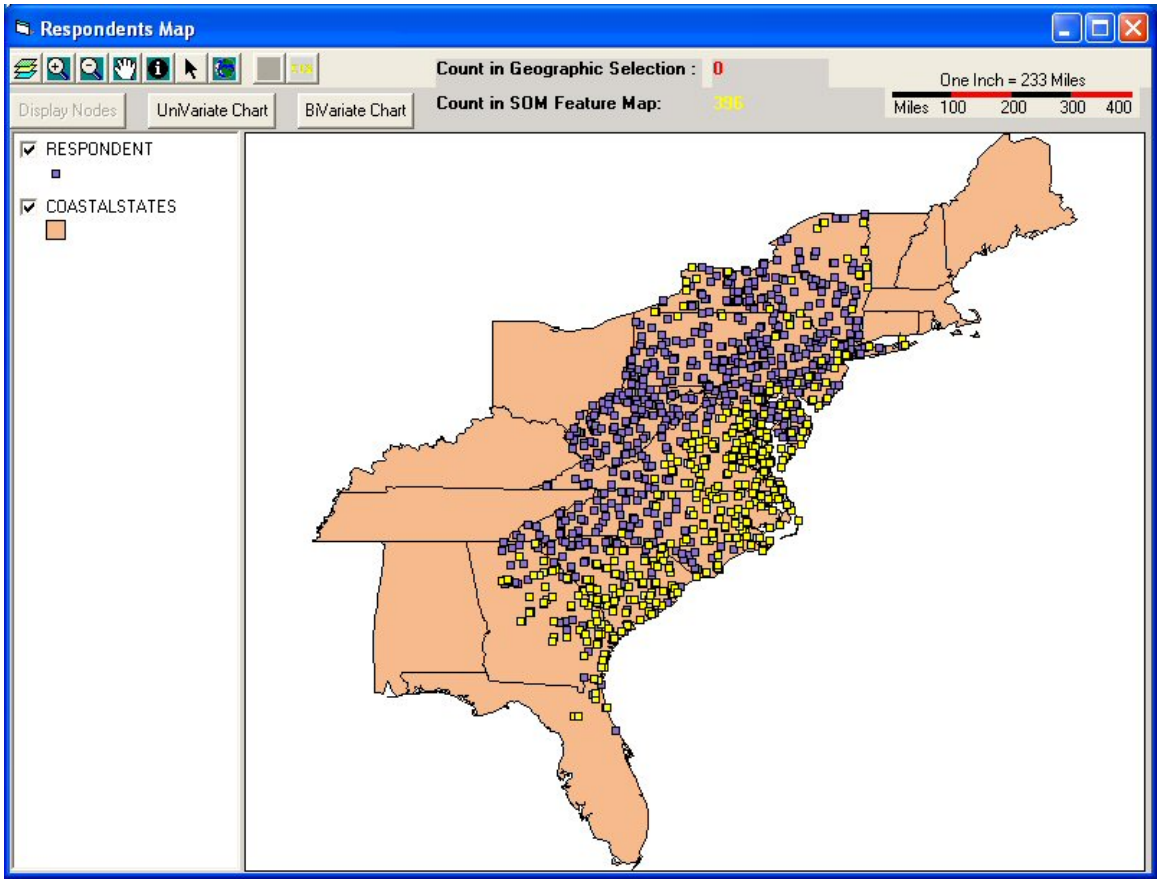
This leakage comes about in part because the field worker who did most of the GA and SC interviews also did upstate NY. Thus we can see that a fact about how the survey was conducted can have a strong effect on the results, something to watch out for. It is more difficult to explain why there are highlighted speakers in NJ. When the millions of possible nodes are selected and arranged into five patterns (Figure 15).

Figure 15



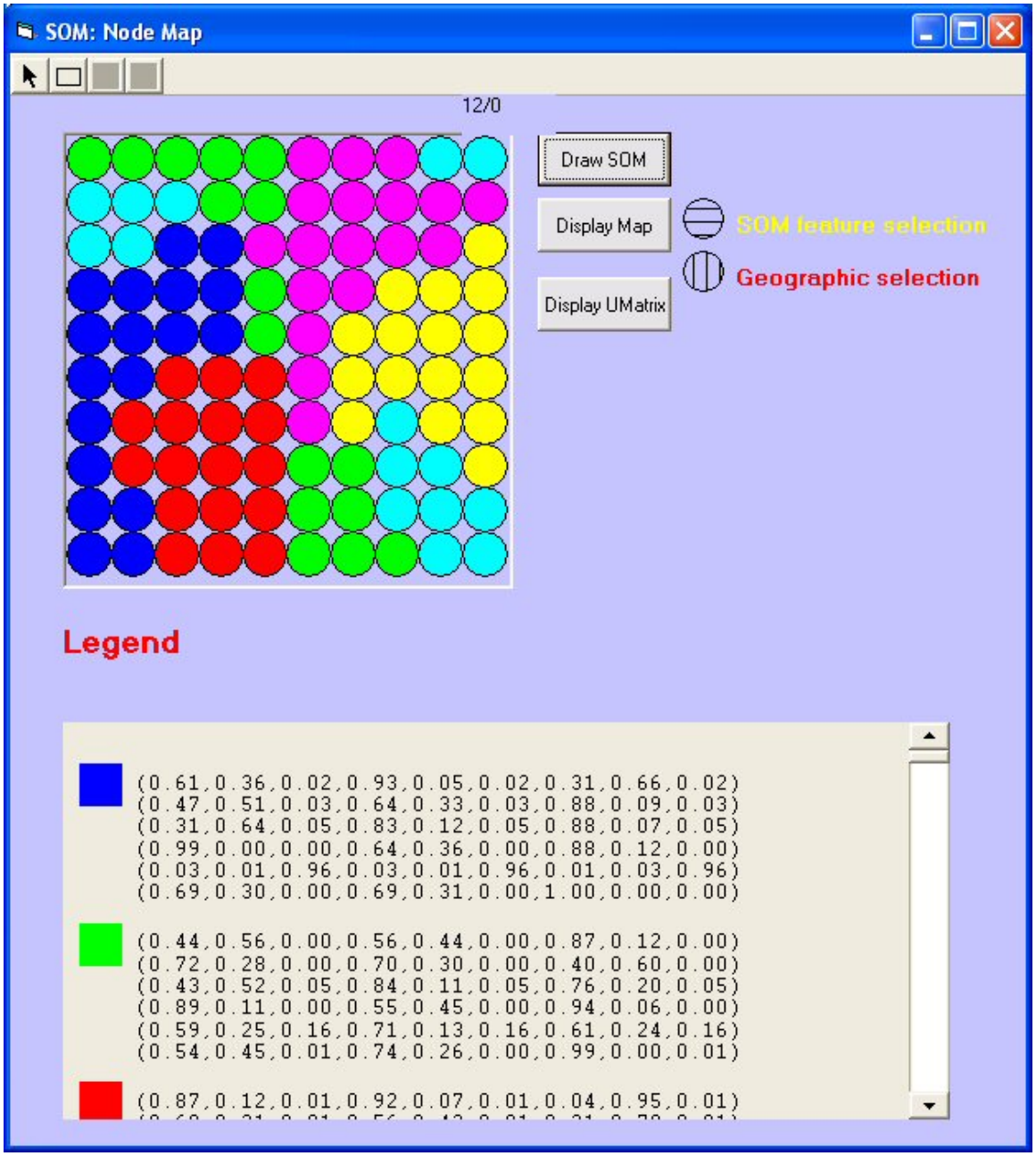
what had been the RED pattern is separated mainly into LILAC and LIGHT BLUE, and the LILAC pattern has a mostly Southern set of speakers (Figure 16).

Figure 16



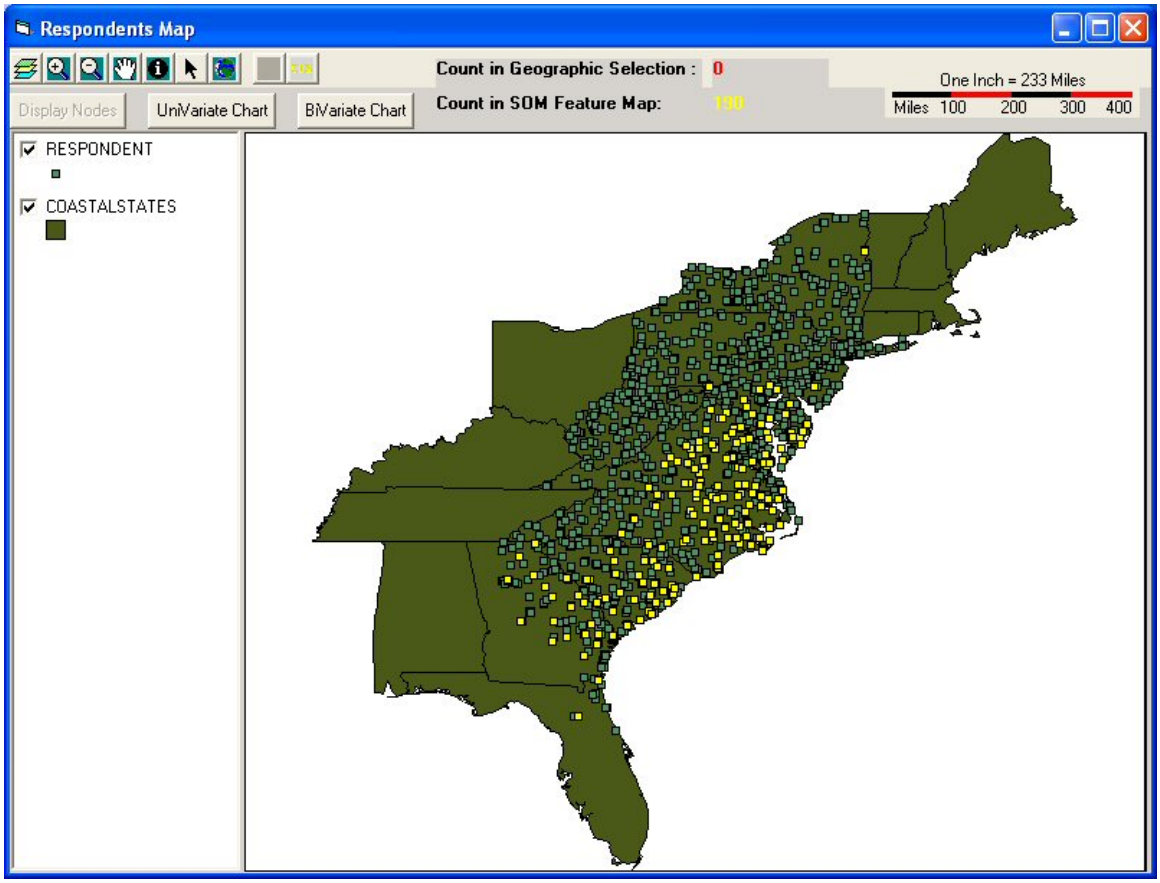
In this arrangement the algorithm has managed to get rid of most of the leakage in NJ, though it still has some in NY. When the nodes are selected and arranged into six patterns (Figure 17),

Figure 17



the best Southern pattern is now the YELLOW one (Figure 18).

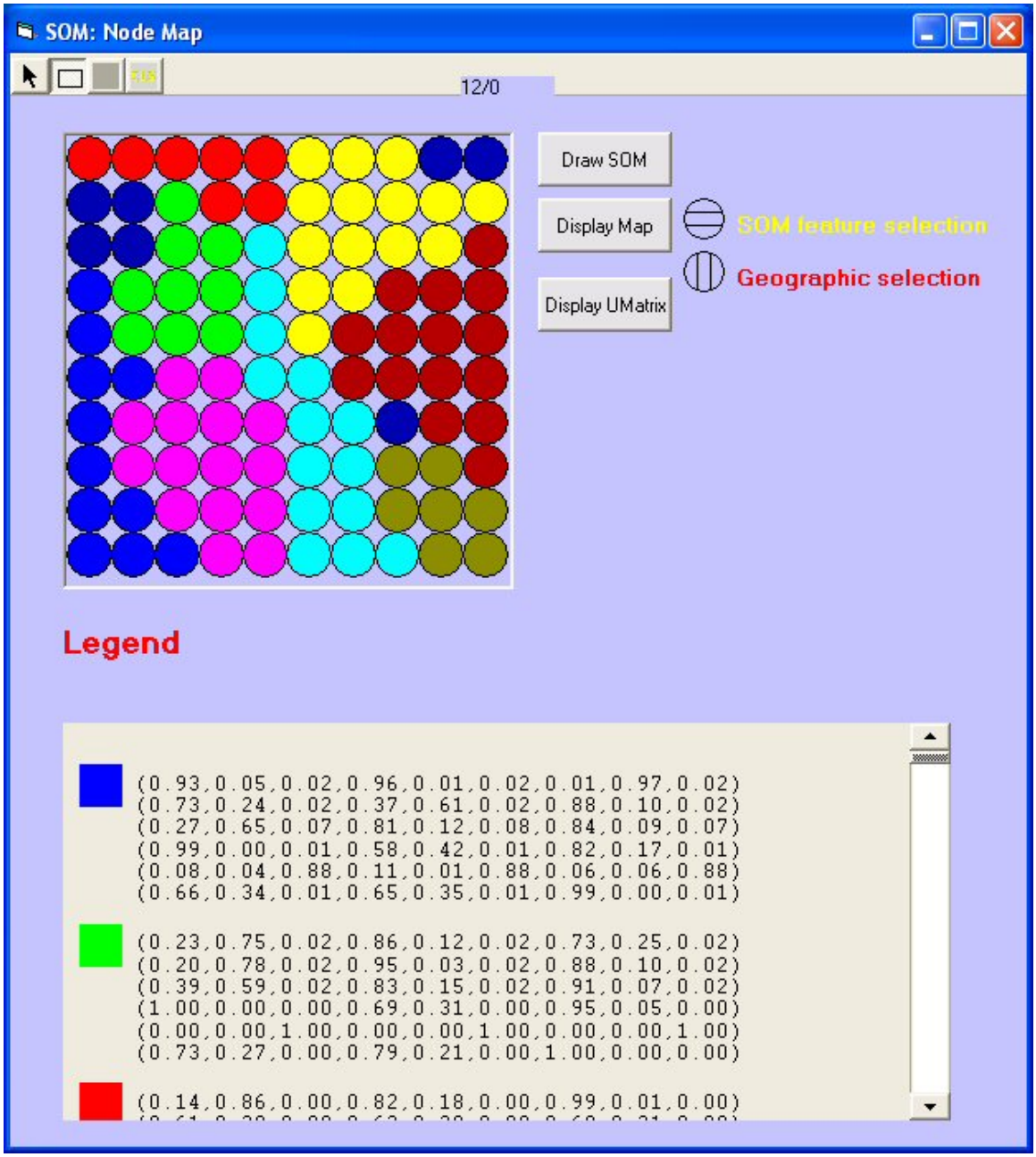
Figure 18



There are practically no more Northern leaks, but the Southern speakers highlighted appear to be somewhat sparser than before.

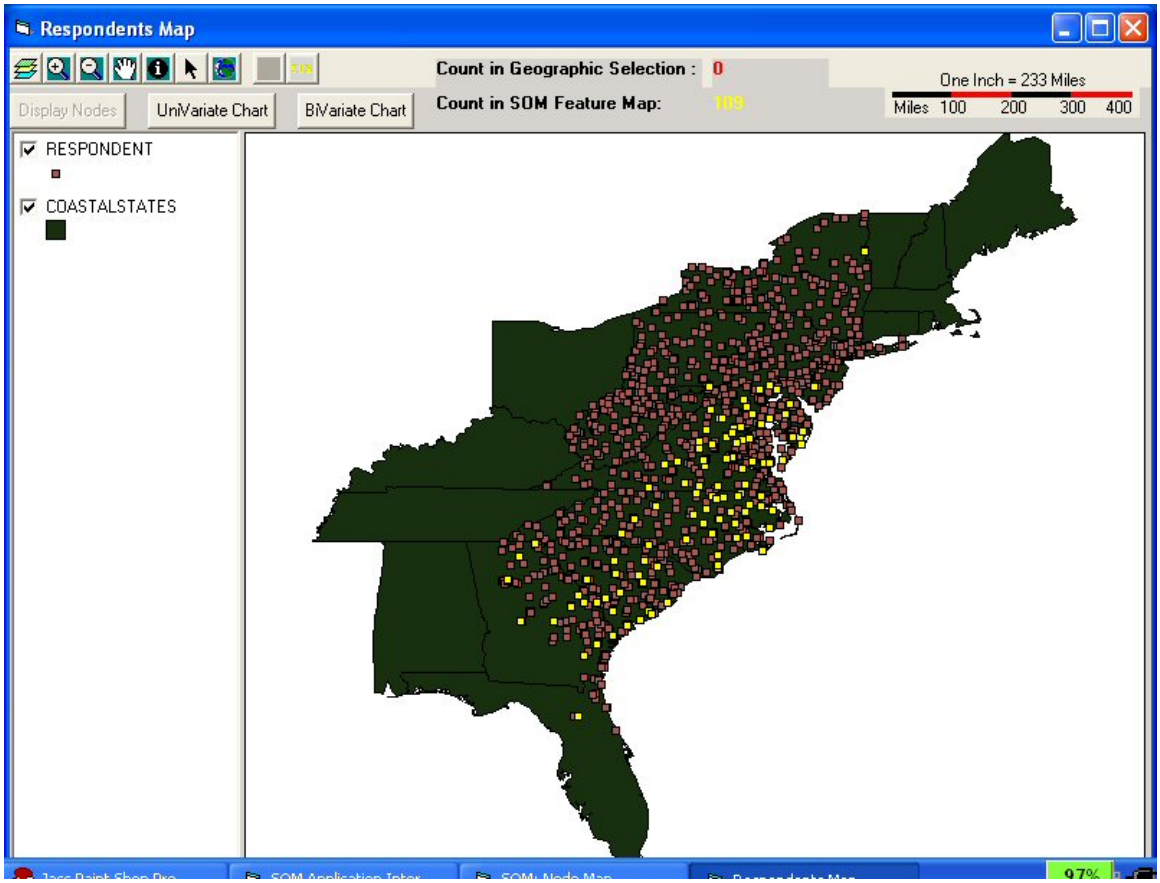
Finally, when the nodes are selected and arranged into nine patterns, the most detailed classification set for this version of the computer program (Figure 19),

Figure 19



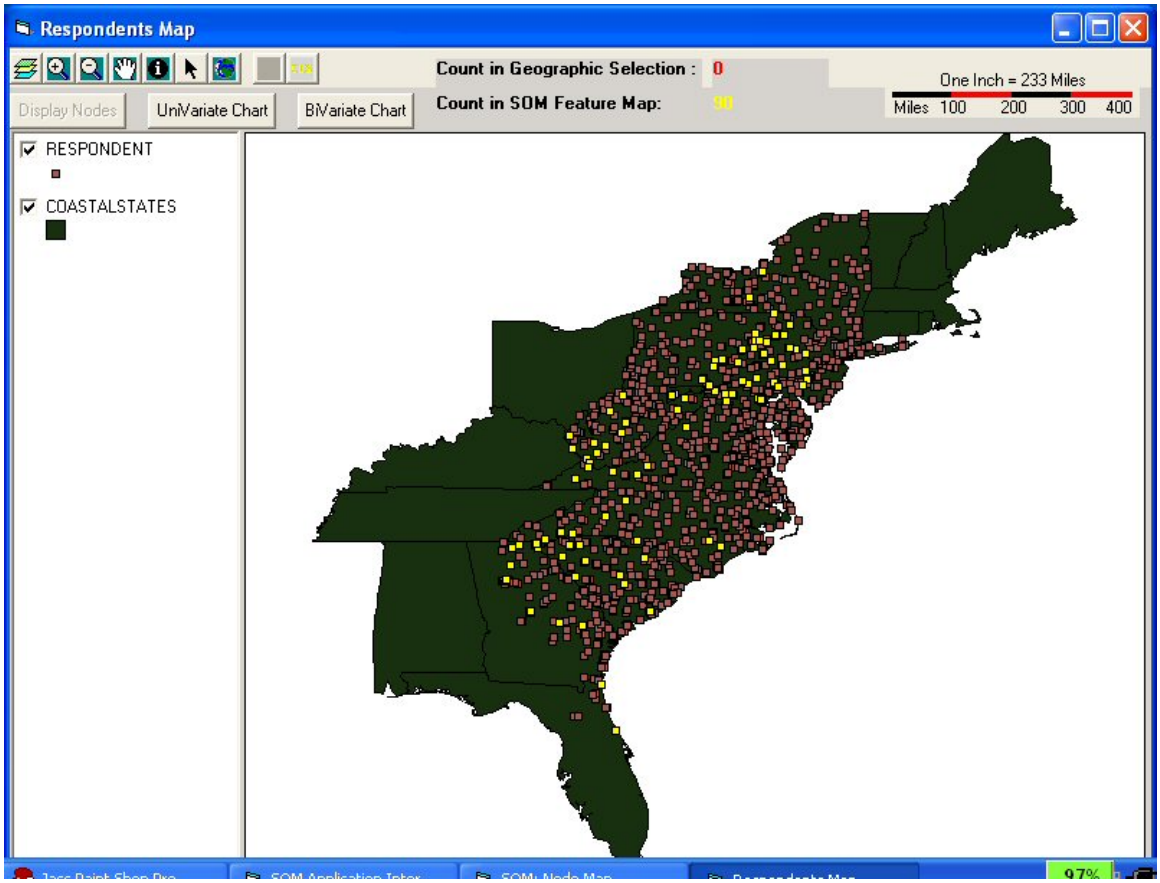
the BROWN pattern shows the best Southern pattern (Figure 20).

Figure 20



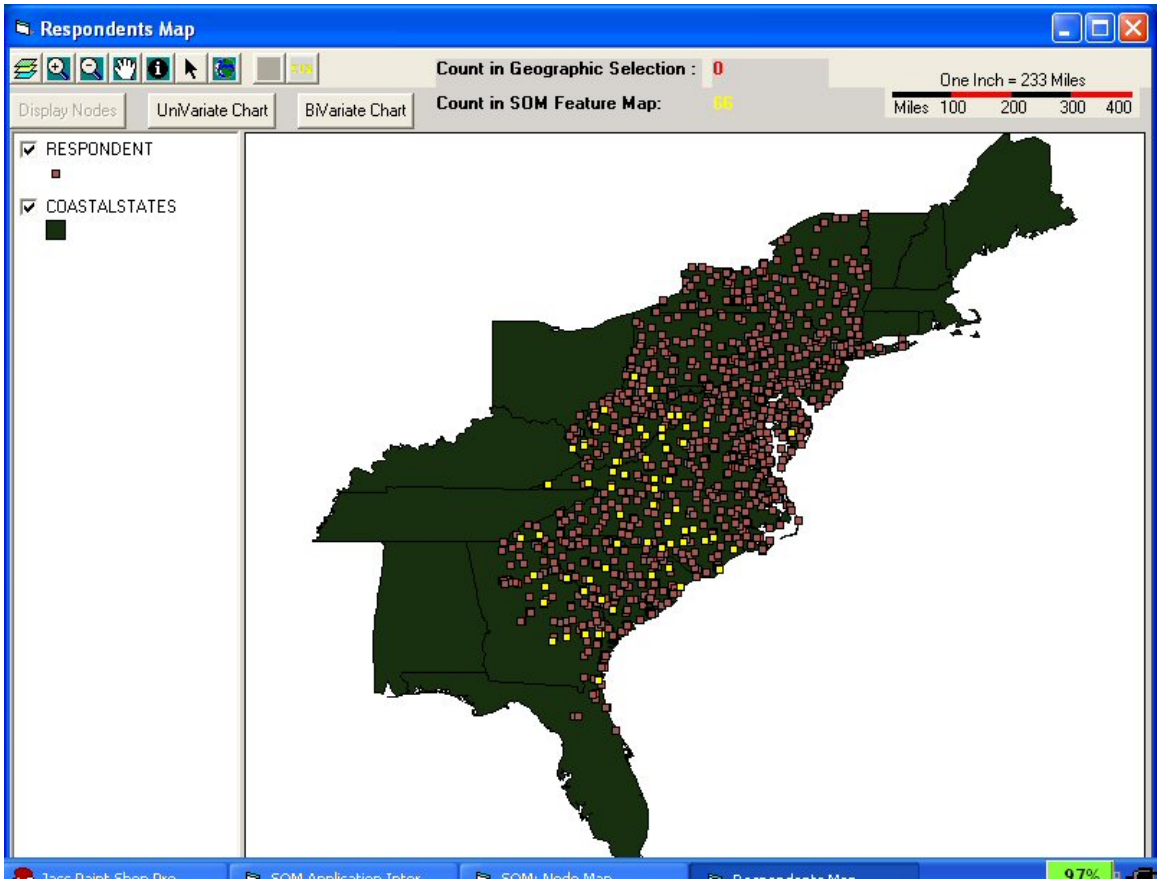
The BROWN pattern is still not far from the Southern grouping we first saw emerge in the six pattern arrangement. However, it is important to note that many Southern speakers are included in other color patterns. Four more colors show patterns with important Southern constituents. BLUE and LIGHT BLUE illustrate what Kurath would have called the South Midland or Southern Uplands. BLUE includes a group of speakers in south-central PA, which supports Kurath's view that the South Midland settlement patterns originated in PA (Figure 21).

Figure 21



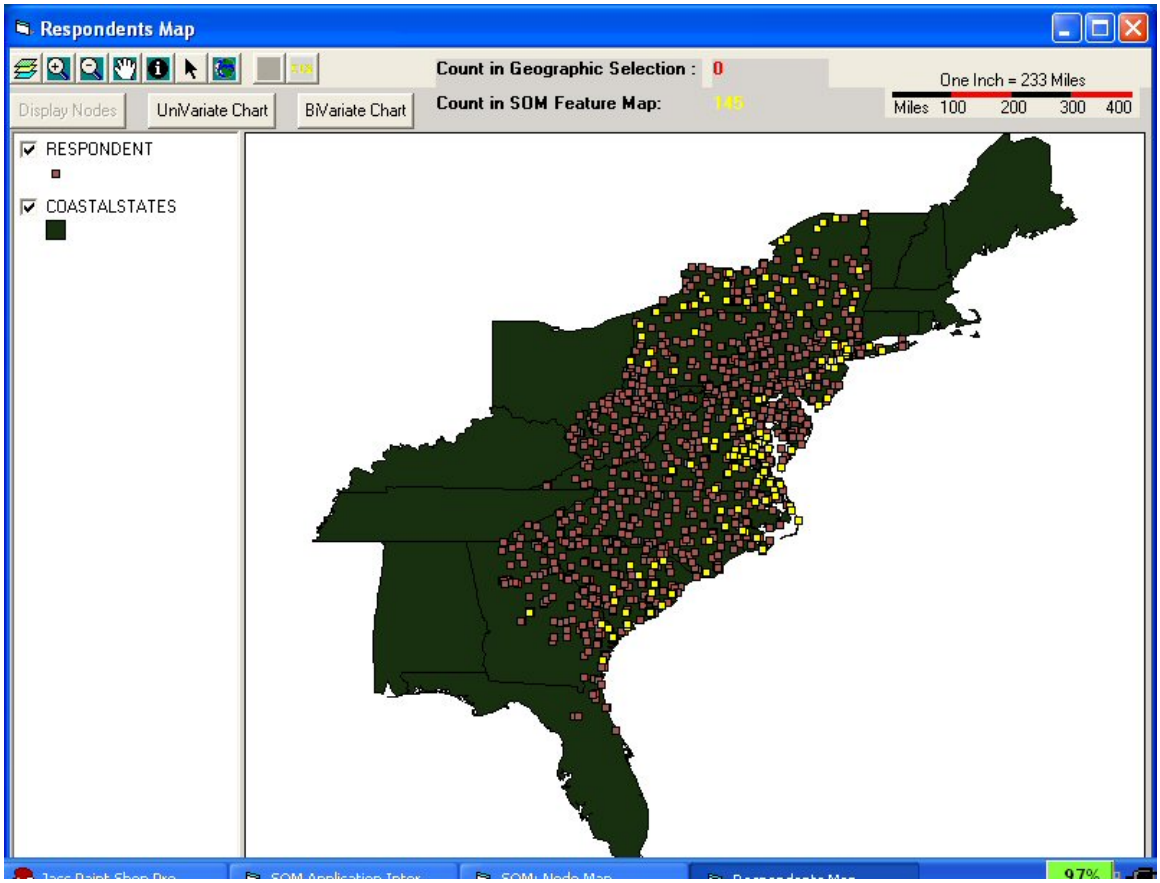
The LIGHT BLUE pattern does not show the PA cluster, but does appear to select more speakers who live closer to the Coastal Southern region (Figure 22).

Figure 22



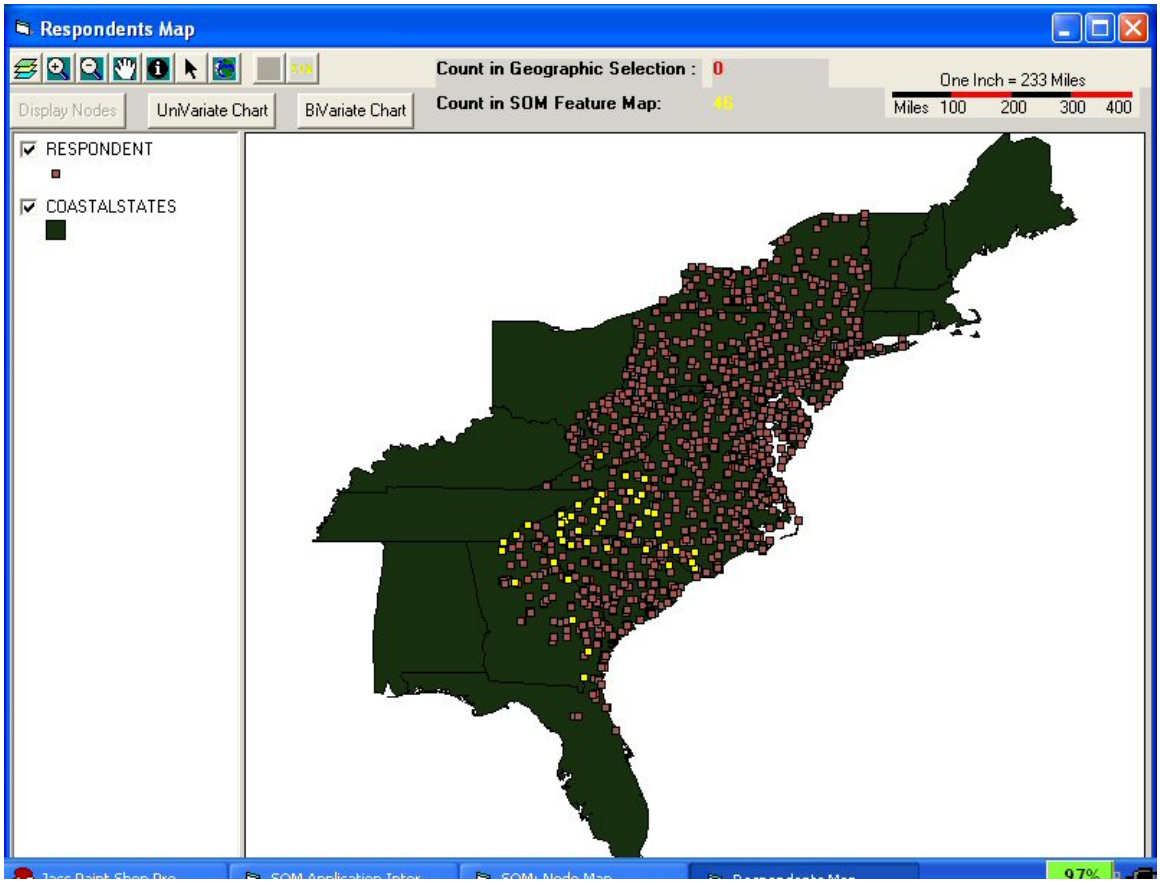
Both BLUE and LTBLUE clusters include many speakers in WV, which separated from VA at the time of the Civil War because of its Northern sympathies, but which shares its historical settlement pattern with the Southern Uplands. The YELLOW nodes show a pattern with clusters at the two Southern cultural hearths, Eastern VA and Lowland SC, with a smaller cluster near New York City and a scattering in Upstate NY (Figure 23).

Figure 23



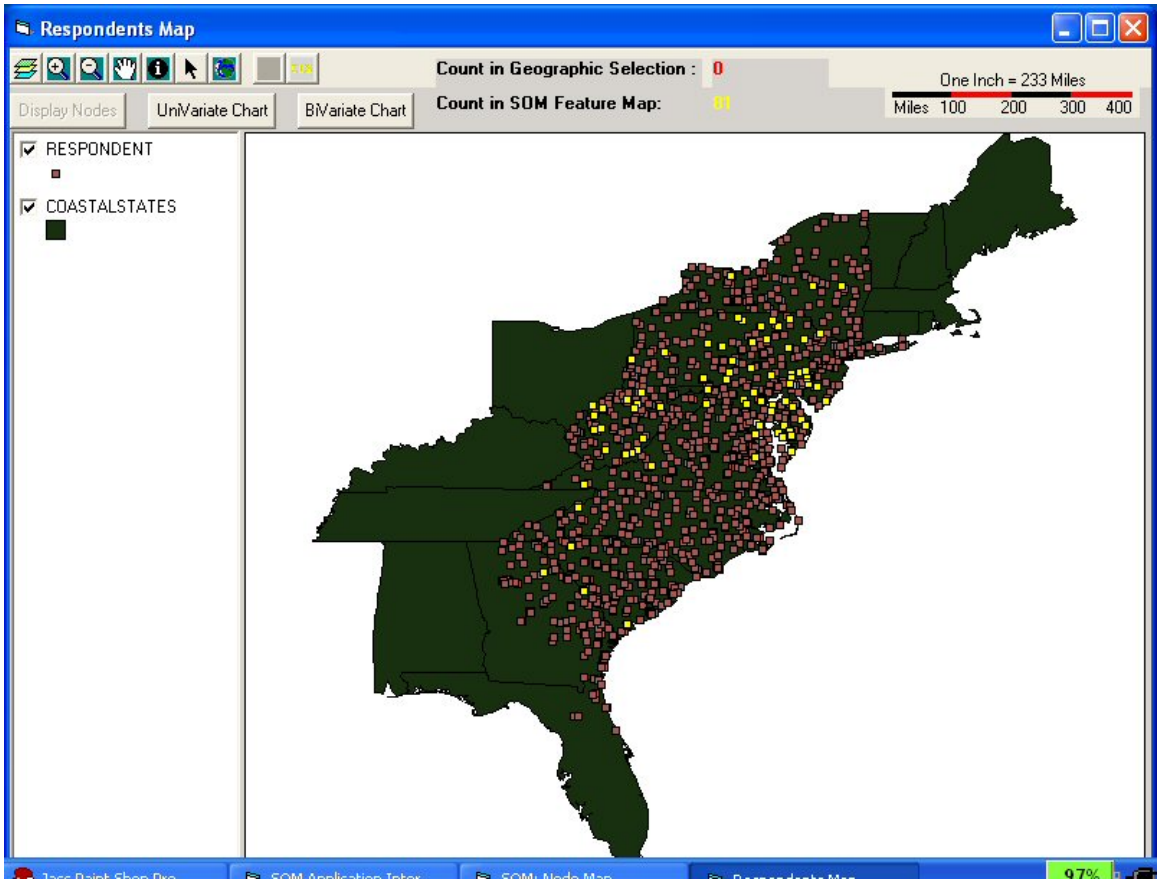
We should remember the field worker who worked in SC and also Upstate NY, but that influence of how the survey was done would not explain why New York City appeared to pattern with the two Southern hearths. Kurath, in a book on pronunciation, claimed that New York City had more in common with the South than one might think, and perhaps we are seeing that here, too. Finally, the LILAC nodes show a pattern of speakers between the two Southern hearths, one that includes the Southern Appalachian region with an extension east into the borderlands between the Carolinas (Figure 24).

Figure 24



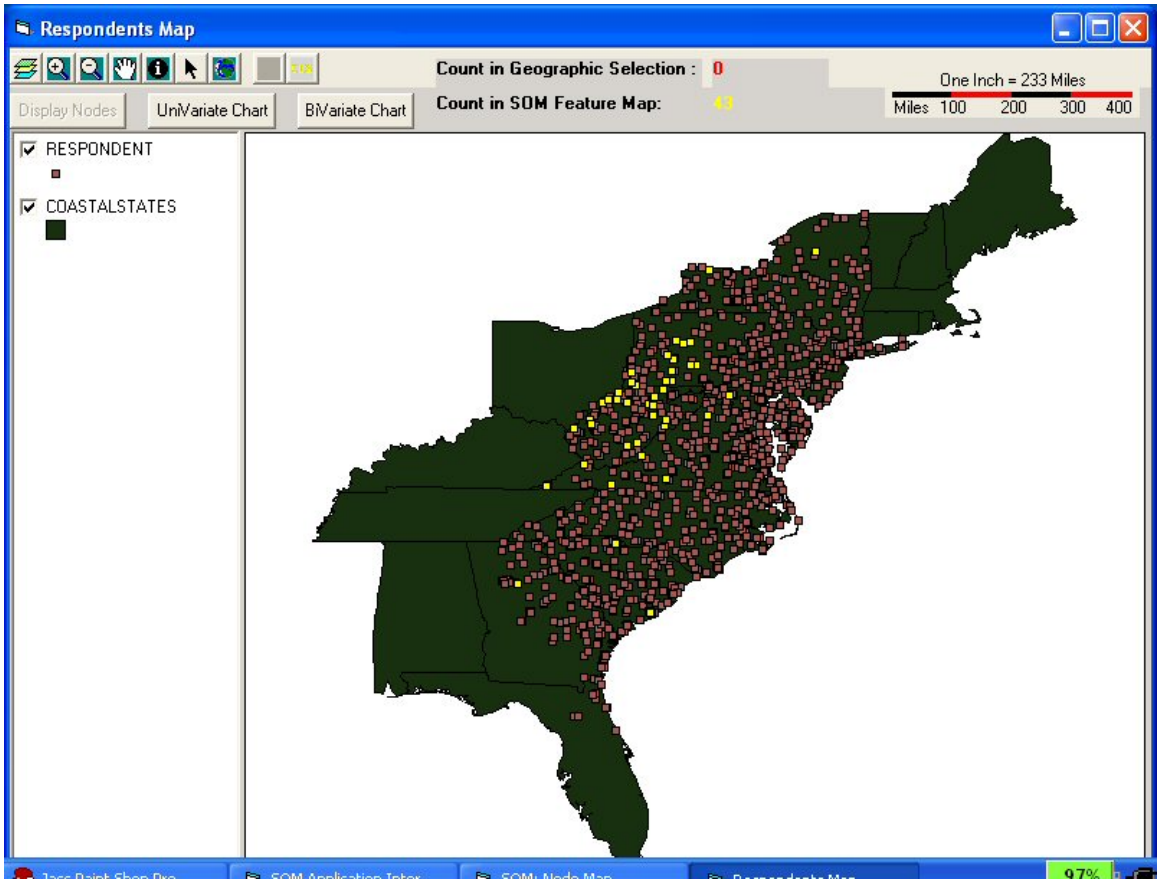
Just for comparison, we should look at the four node patterns from the nine-pattern classification set that are not Southern. The DKBLUE set has mainly Midland speakers, with a number also from WV (Figure 25).

Figure 25



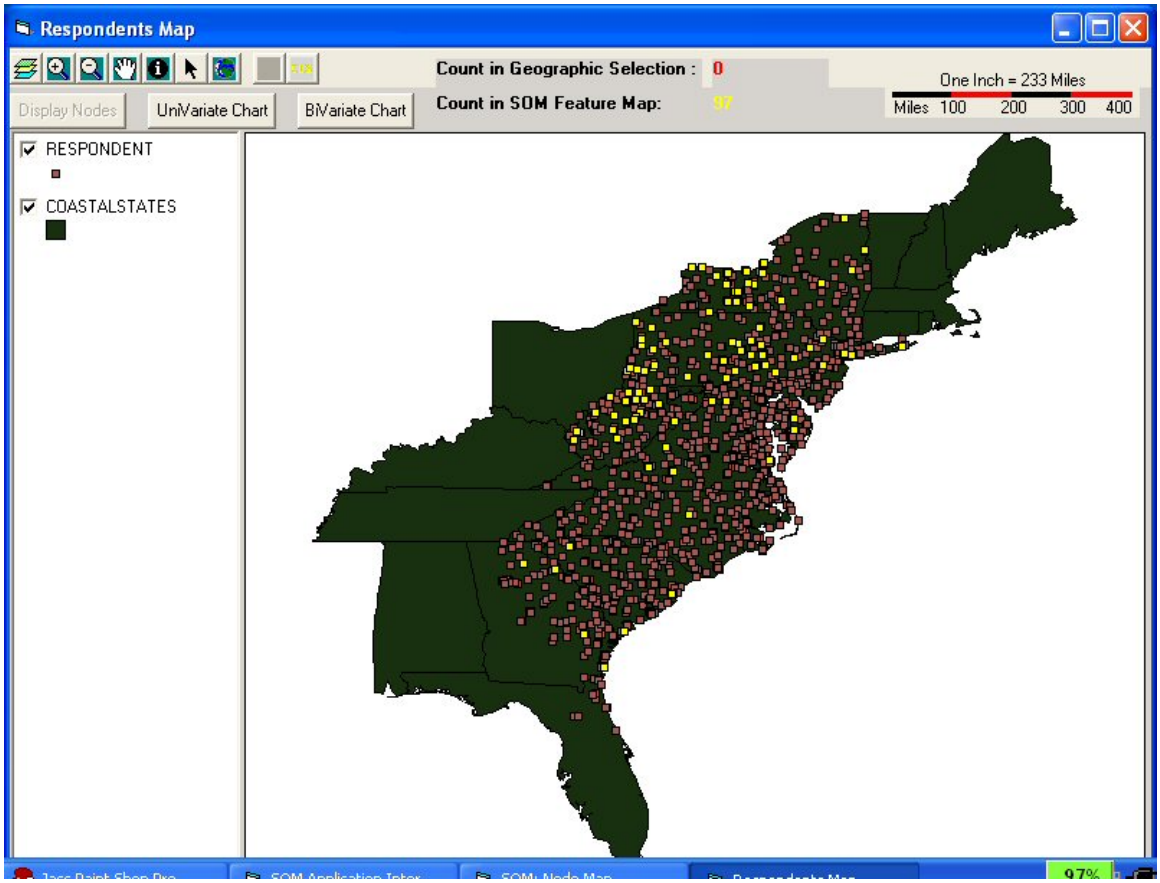
The DKGREEN pattern associates Western PA with WV (Figure 26).

Figure 26



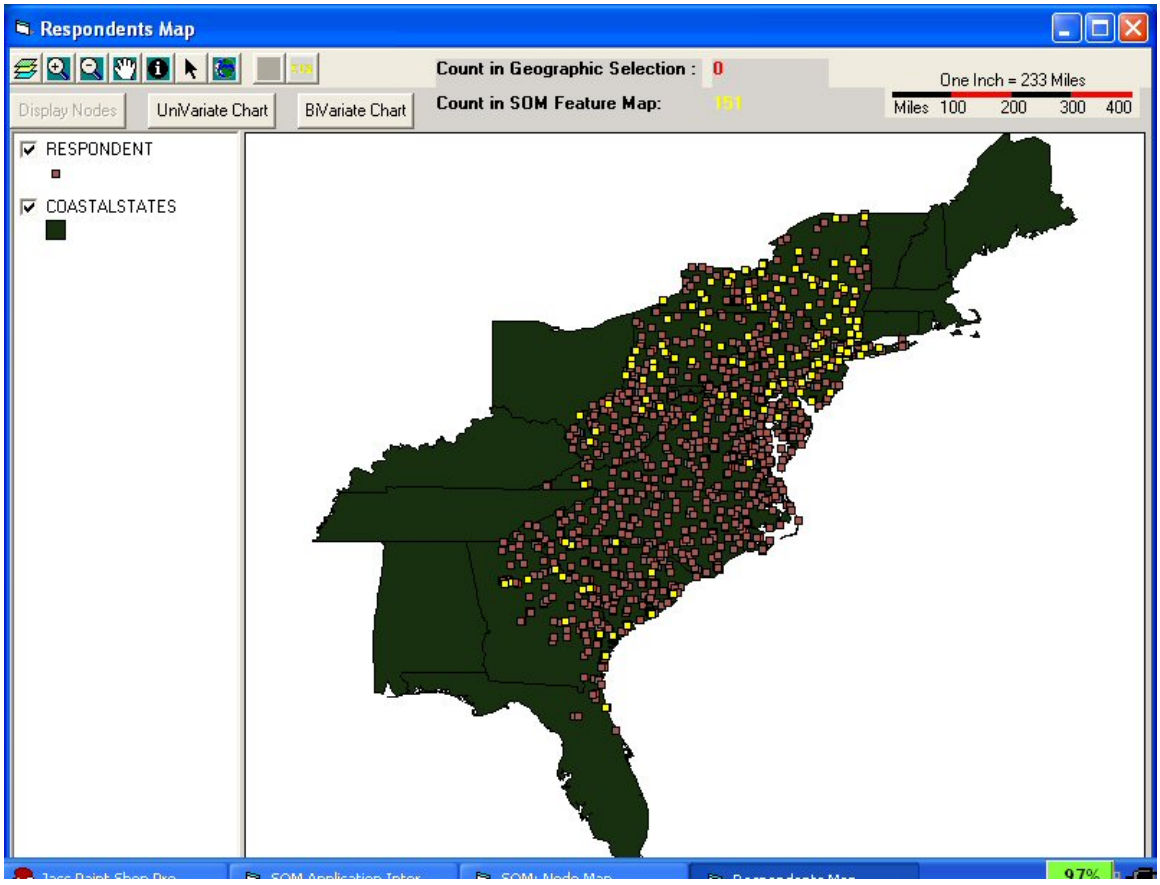
The RED pattern again associated WV with parts of PA, this time also with an area of Upstate NY (Figure 27).

Figure 27



And finally, the GREEN pattern is the only one that shows us a generally Northern distribution (Figure 28).

Figure 28



Each of these patterns has some Southern speakers associated with it, just not so many, in the same way that there were typically some Northern speakers associated with the Coastal and Upland Southern patterns that we saw earlier.

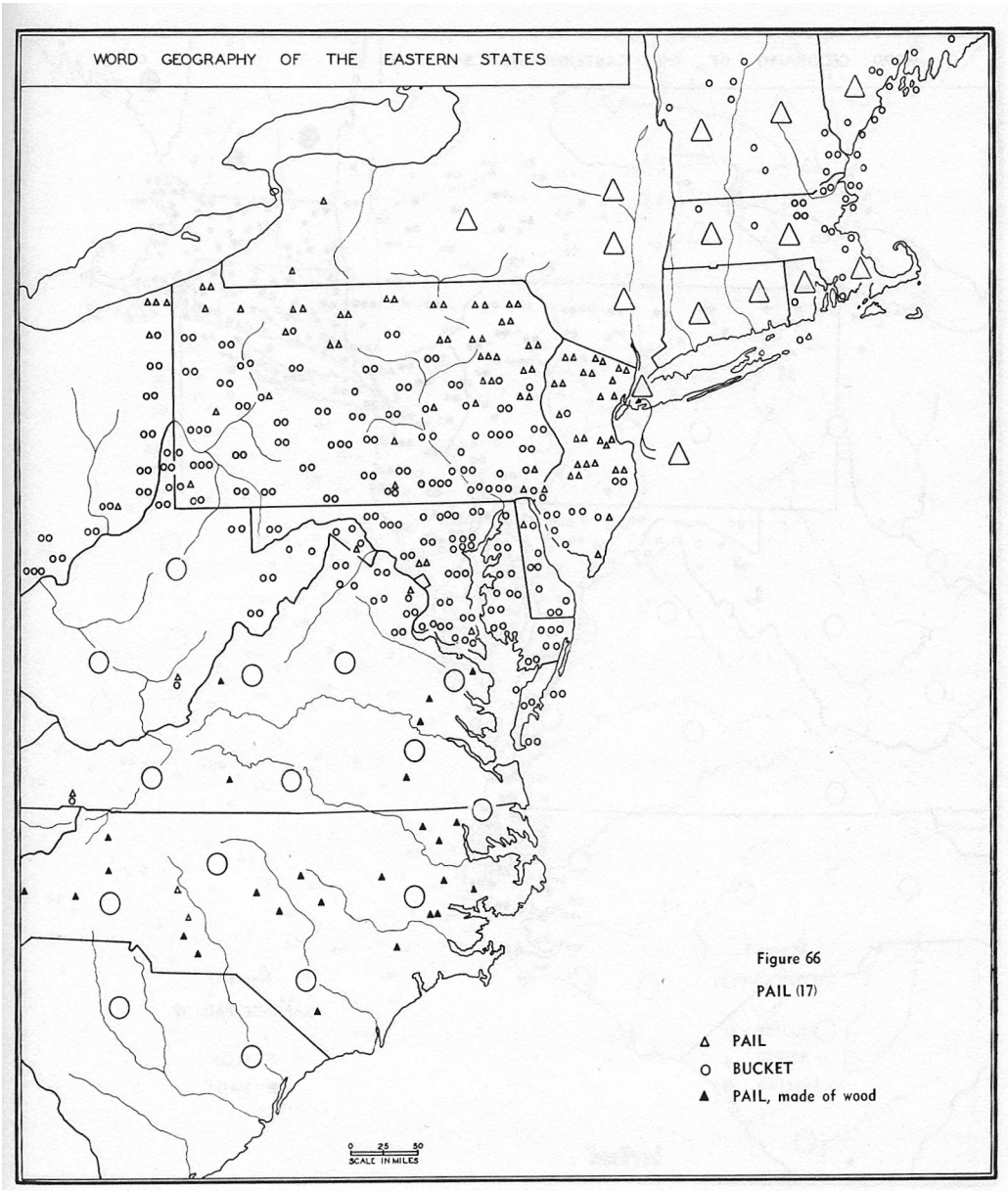
So, what have all these numbers done for us? They have not discovered neat patterns of dialects, whether for Southern English or for another region. Hans Kurath had drawn what appeared to be neat dialect areas in his famous map, and the computer did not exactly confirm those, but this is clearly not a case of Science in triumph over the lesser capabilities of those of us who work in the Humanities. Computers and statistics have not shown us our inadequacies. Our only reasonable way of interpreting the maps generated by the neural network analysis was to refer back to Kurath's research, and it was possible to offer good explanations for most of the maps on that basis. Indeed, one

might say that the most advanced computational and statistical techniques are just telling us what we already knew. But that, too, would not be quite true, because there has been substantial disagreement about how best to think about the obvious fact of dialect differences from place to place in America. The most common belief, at least among non-specialists, is that there really are clear and consistent dialect patterns across America and other places, and people talk about Southern English as though it were something monolithic, as 'pure' in its own way as the sort of English taught in the schools. Our application of computers and statistics does not support that view. [Further work on neural networks and SOM has led to a greater understanding of how the algorithm actually produces results, which offers tantalizing clues to how we each may process language in the brain. See Kretzschmar 2008, 2011.]

What such use of technology and numbers actually does for us is to give us another way of seeing what is before our eyes, or in the case of Southern English, before our ears. We are not all of us wrong to think that there are different voices all around us, and that the differences in those voices are important for our daily lives. But each of us can have only a limited amount of experience with all of those different voices, and we each do the best we can to make good use of that experience. Technology and numbers give us a chance to enlarge our experience, and to think in larger and more systematic ways about all of those different voices. Just how much more and how much better I can illustrate with a historical anecdote from my own experience. When I was a graduate student first learning about language variation in Chicago in the late 1970s, we did not yet use computers in the work. If we wanted to make a map, we took our pencils and

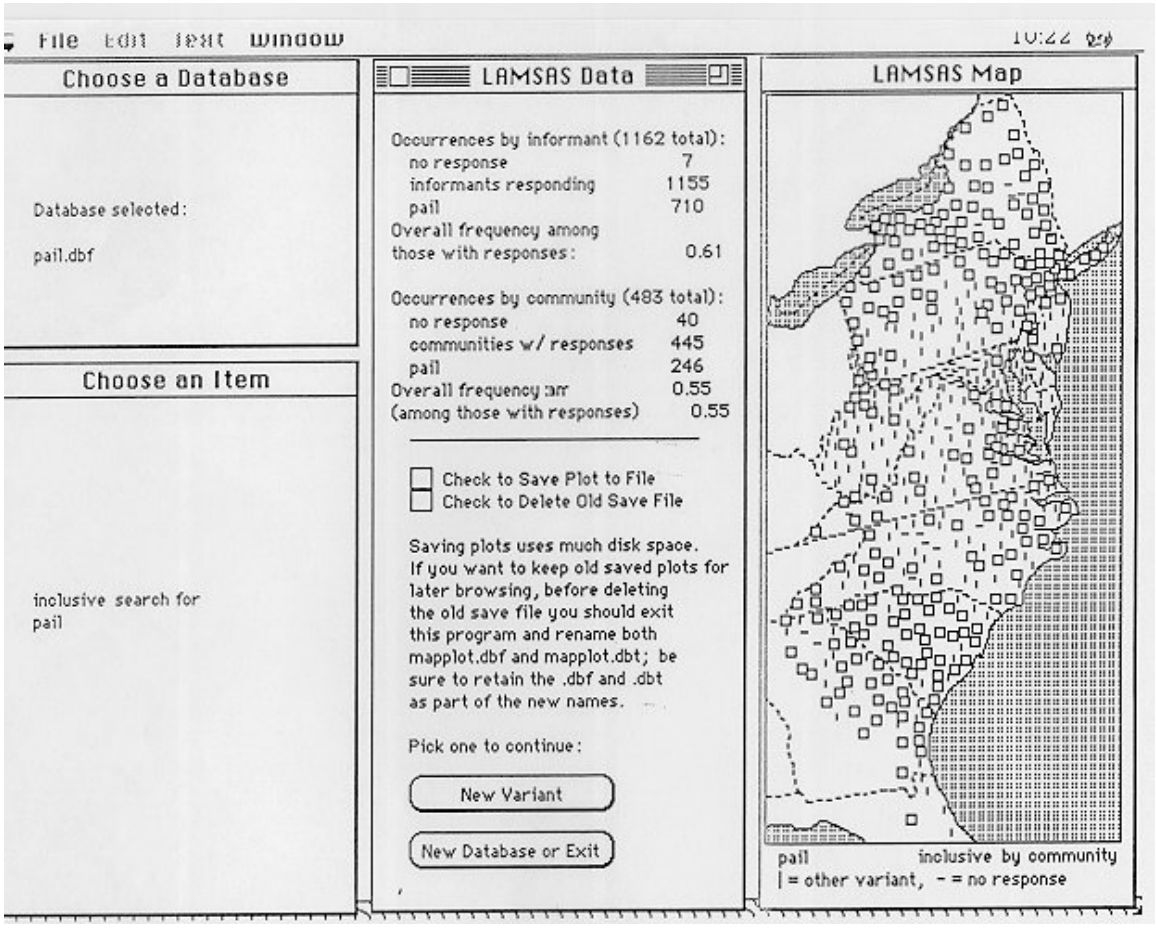
drew little symbols on a printed base map, which would look something like this one (Figure 29).

Figure 29



It took six to eight hours to go through all of the paper field records and make a symbol for what each person had said. Around 1990 I wrote a program to make similar maps by computer, like this one also for the word *pail* (Figure 30).

Figure 30



It took about 90 seconds to make one map, which is about 400 times as fast as the old way. A few years later, we moved the mapmaking process to the Internet [now at www.lap.uga.edu--WAK], so that anybody anywhere could make a similar map in under 10 seconds, another ten-fold improvement in efficiency. Hans Kurath, the famous old dialectologist I have mentioned several times, mapped fewer than 200 words for his

major publications, carefully selected because of the time it took to make each maps. Now my students each term make many more maps than Kurath ever did, online for class exercises. The SOM program, as I have mentioned, compares literally hundreds of millions of possibilities within five minutes on my laptop, and its interface shows us maps at will, practically instantaneously, after the initial processing is done. Our view of language variation and of potential patterns in it is immeasurably better now than it was back in Chicago. Technology and numbers have done that much to enlarge our experience and to improve our vision.

At the end of the day, however, technology and numbers cannot see for us. They can help us to see better, but we still have to see for ourselves. So, as you are aware, the nine patterns from the last SOM run did not explain themselves. They were valuable in large part because they could be connected to historical and cultural information that affects our language, and so receive some kind of interpretation. The SOM displays could be compared to earlier analyses of American language variation, like those by Hans Kurath, and so be interpreted in the light of what we already thought we knew. However, this is no mere "singing in the shower," not a way smugly to tell ourselves how smart we always were, because our new visualizations of the language variation data help us to choose between alternative previous explanations, and to think of new ones because the new visualizations do not conform exactly to what we thought we already knew. While we use Kurath's wisdom to help us interpret what we see, we know that his dialect maps no longer provide an adequate explanation for Southern English or for other regions.

This kind of discussion is just what the Humanities prepares us for, and is just what I take to be our mission in the Humanities, the mission that Harry and Jane Willson

wanted to support. We in the Humanities bring our cultural and intellectual inheritance to bear on problems of interest to us, problems that in my New Humanities work can profit from the application of technology and the use of numbers. Moreover, we in the Humanities don't do this just for ourselves. The training in the liberal arts and Humanities that we provide to students in other divisions of the university will enable them, too, we hope, not to "sing in the shower" but to apply their sciences in the context of history and culture. In so doing, we in the Humanities do our part to keep our university, our state, and our culture vital and strong, and in my view, our use of numbers and technology as appropriate to our own work can help to keep the Humanities vital and strong.

Works Cited

Bailey, Guy. 1997. When did Southern American English Begin? In Edgar Schneider, ed., *Englishes Around the World: Studies in Honor of Manfred Görlach,* Vol 1: General Studies, British Isles, North America (Amsterdam: John Benjamins), 255-76.

Bailey, Guy, and Jan Tillery. 1996. The Persistence of Southern American English. *Journal of English Linguistics* 24: 308-21.

Burkette, Allison. 2001. The Story of Chester Drawers. *American Speech* 76: 139-157.

Johnson, Ellen. 1996. *Lexical Change and Variation in the Southeastern United States 1930-1990.* Tuscaloosa: University of Alabama Press.

Kohonen, T. 2001. *Self-Organizing Maps*. 3rd edition. Berlin: Springer.

Kurath, Hans. 1949. *A Word Geography of the Eastern United States.* Ann Arbor: University of Michigan Press.

Kurath, Hans, and Raven I. McDavid, Jr. 1961. *The Pronunciation of English in the Atlantic States.* Ann Arbor: University of Michigan Press. [rpt. 1982, Tuscaloosa: University of Alabama Press.]

Lewis, C.S. 1960. *Studies in Words.* Cambridge: Cambridge University Press.

Preston, Dennis. 1989. *Perceptual Dialectology*. Dordrecht: Foris.

Tamasi, Susan. 2000. Linguistic Perceptions of Southern Folk. Paper presented at ADS, Chicago.

------. 2001. Cognitive Patterns of Linguistic Perceptions. Paper presented at NWAVE, Raleigh.

Additional references:

Kretzschmar, William A., Jr. 2008. Neural Networks and the Linguistics of Speech. *Interdisciplinary Science Reviews* 33 (2008): 336-356.
Kretzschmar, William A., Jr. 2009. *The Linguistics of Speech*. Cambridge: Cambridge University Press.
Kretzschmar, William A., Jr. 2011. The Beholder's Eye: Using Self-Organizing Maps to Understand American Dialects. In Michael Adams and Anne Curzan, eds., *Contours of English* (Ann Arbor: University of Michigan Press, 2011), 53-70.