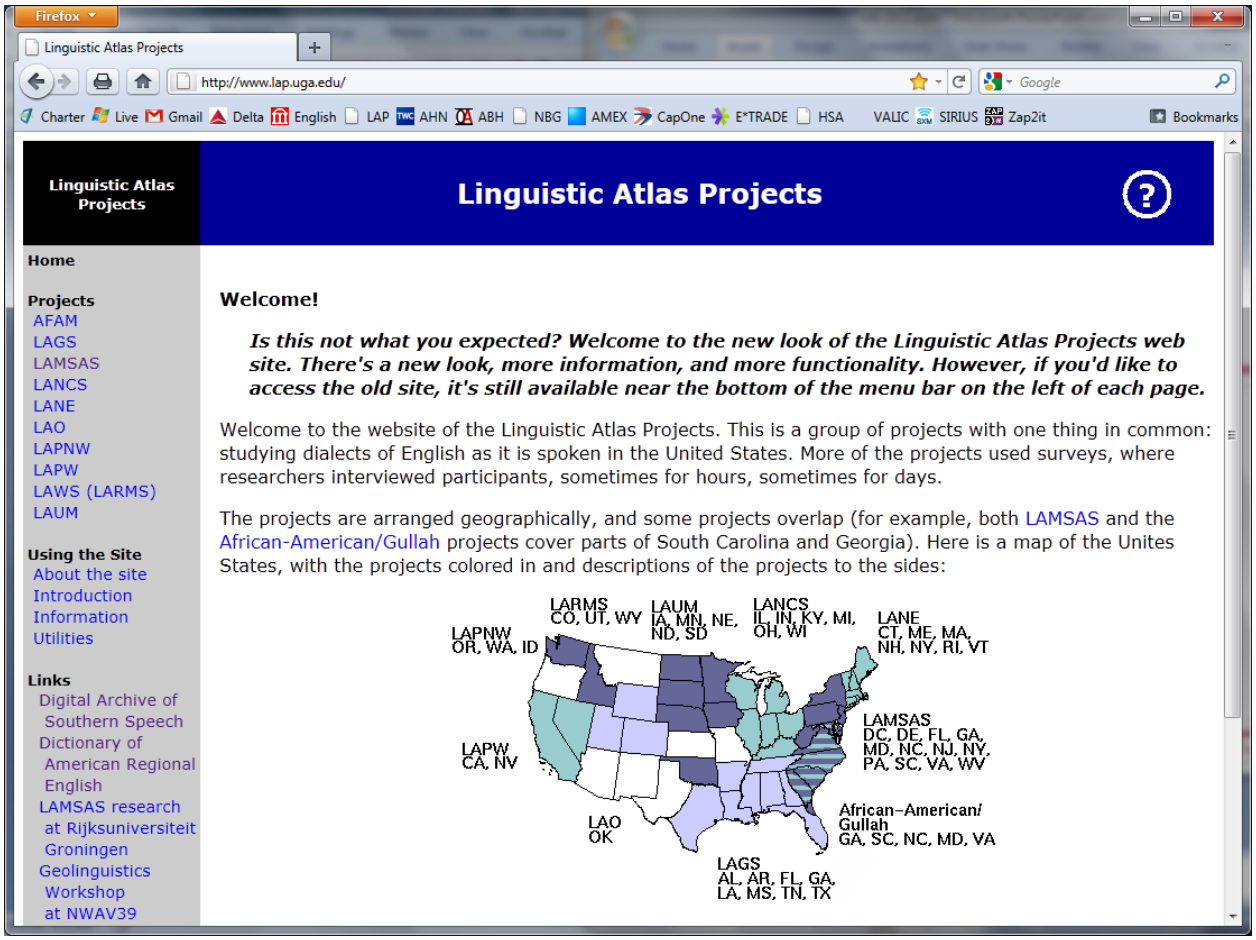
Student Participation in the Linguistic Atlas Project

Bill Kretzschmar, University of Georgia

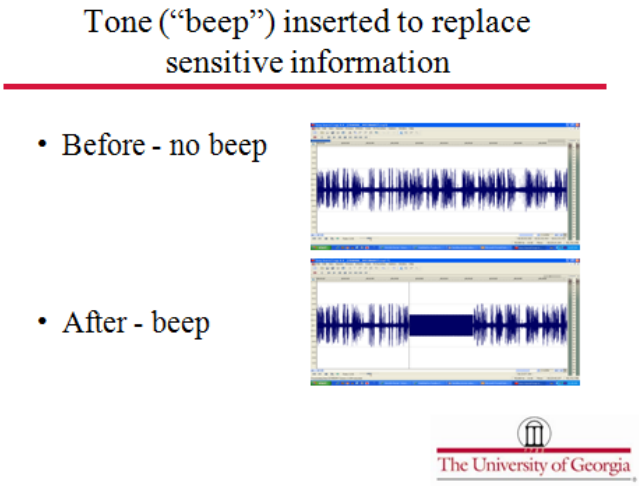
The American Linguistic Atlas Project has always had students as key participants. One need only note that Bernard Bloch, pictured here, famous as the long-time editor of *Language* and a prominent figure in twentieth-century of American linguistics, was one of the original field workers for the Linguistic Atlas of New England in 1931 (O'Cain 1979: 253-254). He was, in fact,

the best field worker on the project overall, having been judged at the top in seven of the nine evaluation categories (Kurath 1939/1973: 52-53). Bloch, however, only finished his dissertation in 1935, a study of rhotacism in New England based on 394 LANE speakers. After Bloch many more of us began doing Atlas work as students, including for example Roger Shuy, Lee Pederson, Joan Hall, Guy Bailey, and yours truly. Some of us were paid participants like Bloch, but many others have contributed to the Atlas in the course of doing their dissertations or have otherwise volunteered their time. Thus students have been intimately involved in both field work and analysis for the Atlas from its very beginnings. Today, however, I would like to talk about the way it works now on the Atlas. We do still employ graduate students, but more and more we make use of undergraduates. This is true both for work on the traditional Atlas, and for contemporary extensions like our Roswell Voices project, a long-term community language study in Roswell, Georgia. As I go along, I will first deal with how students are participating in the traditional Atlas, then move on to Roswell. For both of these, I will talk a little about student activities in the Atlas Office and in the field, then consider training methods and the classroom, and finally give an example of a recent Honors thesis by one of my undergraduate students to show how the work ends up. In the end, I will be able to offer you some concrete conclusions about student participation.



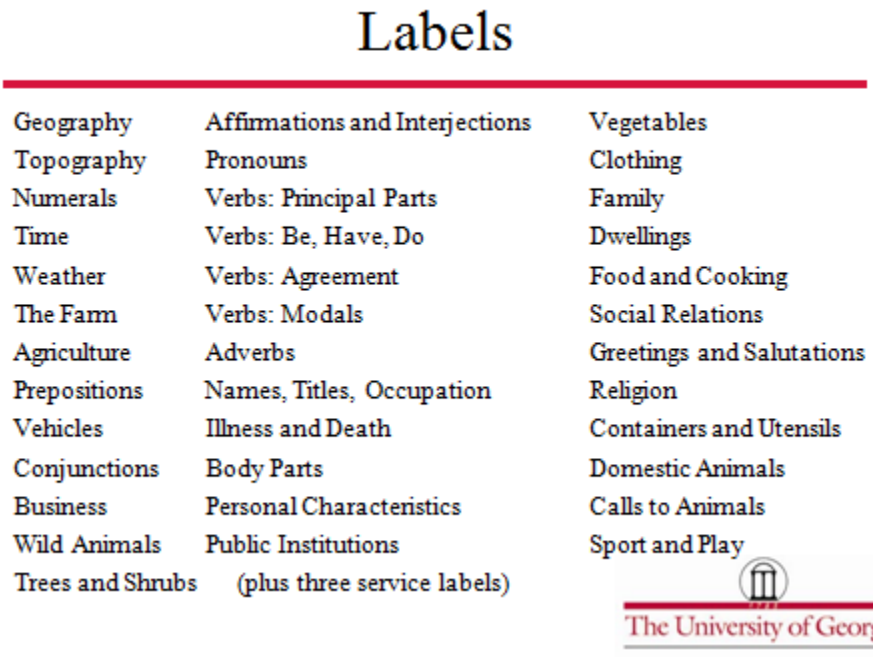
For those of you who don't know about it, the Linguistic Atlas Project began in 1929 with a meeting sponsored by the ACLS, at which a number of linguists prominent at that time gave papers (one of them, amazingly enough, about mechanical sound recording) and also discussed how to construct a survey of American speech that could parallel work in Europe. Field work began in 1931 in New England, under the direction of Hans Kurath, with sponsorship by the ACLS. After New England, the Atlas continued with regional projects under the general direction of Kurath but with immediate management by local researchers. Regions east of the Mississippi were surveyed by about 1980; areas to the west had spottier coverage, and we are still conducting primary interviews in the Western States as field workers, graduate assistants who want to work out West, and resources are available. At the beginning the surveys used

questionnaires of about 100 pages, about 800 elicitation cues, but shorter and shorter questionnaires have been used since then. We have not just employed long-question-short-answer, fixed format questions as in Orton's Survey of English Dialects; field workers were free to get the answers they needed as they thought best, ideally in a conversational manner, and so our tape-recorded interviews contain lots of cultural information the speakers talked about. Today, we have millions of answers to questions, many originally on paper before field recording was practical, many on audio tape, and now a great many in digital format, too, as either text or audio files owing to our efforts over the past two decades to computerize the Atlas.



The digitization process has been accomplished only with great amounts of student assistance. Most recently, we had funding from NEH to digitize nearly 8000 hours of old audio tape, to create an index for what was on the recordings, and to make the sound publicly available. In order to carry this out, we hired a team of three graduate assistants and up to twenty-five part-time undergraduate assistants at any time. Two of the graduate assistants were responsible for supervision of the undergraduates, and one of the graduate assistants served as our technical person to maintain the computer resources for the project (two computers attached to reel-to-reel tape decks to digitize the sound, a dozen desktop workstations for the processing work, and three or four computers attached to a swarm of 1 Terabyte external drives to store and backup our files). Our undergraduates listened to every minute of every interview,

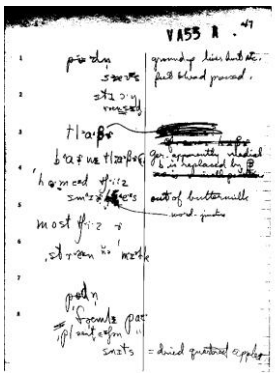
in order to "beep out" sensitive information like names (they inserted a tone for the same length of time as the sensitive speech). Every four minutes or so they inserted a label from our standard list of 40 topics to indicate what the conversation was about at that point, and also indicated whether or not there were any narrative passages in the section of at least one minute of uninterrupted speech.



We then automatically exported MP3 files corresponding to the labels, so that we can offer public versions of entire WAV files from each reel of tape for technical processing, and also post the MP3 files on the Web for general listening. Students only had 80 minutes total to complete the processing of every 60 minutes of raw tape, including copying files and creating the metadata. Our students did an excellent job at these tasks, in part because we had a very specific protocol for every step of the digitization process in the manual created for the project by lead assistant Paulina Bounds, and in part because our graduate assistants were very diligent with their supervision, sampling student output on a daily basis and recommending corrections. I only ever had to "fire" one undergraduate, out of nearly 50 we employed over the course of the grant, because she just could not meet the time and accuracy demands of the work. Our

undergraduate students learned a good deal about digital audio processing, including both how to handle and edit sound files and how to manage metadata. Paulina and I have now published a chapter on legacy data in a new handbook of sociophonetics (Kretzschmar, Bounds, and Palosaari 2010), which shows what sort of "training" the graduate assistants were getting.

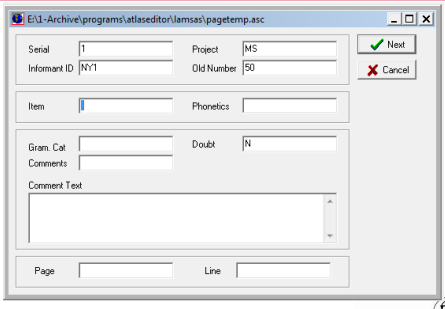
LAMSAS field record



The University of Georgia

When we have not had major funding as we have had recently, we have used both graduate assistants and undergraduates to continue keyboarding paper records from the Atlas as from this field book page, most often two or three students at any one time. This we have been doing for nearly twenty years; there is a great deal of data to enter from all the field records before the days of audio recording. For this task we have used a program for the last decade, written by our former graduate assistant Eric Rochester (now at the humanities computing Scholars Lab at the University of Virginia), that presents a template onscreen for students to type responses into. Students who use the AtlasEditor program see all the variety in responses, whether lexical or phonetic, as they convert the phonetic transcriptions of the field records into standard spelling responses. We have in the past entered the phonetics, too, but at too great a

Atlas Editor program



The University of Georgia

cost for us to keep doing it. Clearly students have to begin with some facility in the IPA, and they develop much more than that over time. An essential part of this process, again, is close supervision, in the form of proofreading with my administrative associate Debbie Vaughn.

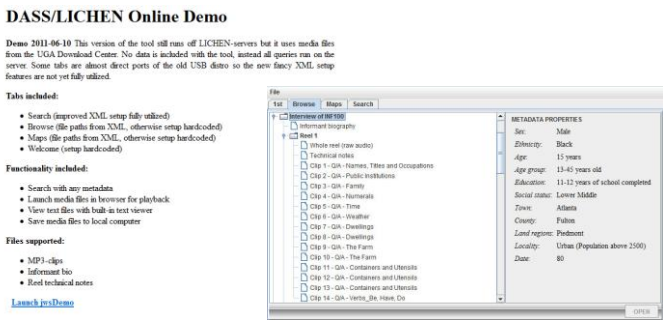
CURO at Georgia

It is no accident that we have been able to find so many good students. About half of the students came from our Honors Program, for which I have been teaching a one-credit "gateway" course called "Introduction to Research in the Humanities and Social Sciences." This course introduced students to research methods, but my examples mostly came from the Atlas: my students visited our offices and read our grant proposals, and many of them decided to try out working for the Atlas. My university has one of the largest CURO programs in the country (i.e. Center for Undergraduate Research Opportunities), which promotes the exposure of undergraduates to research by connecting good students with faculty members doing research. CURO also has a symposium at which undergraduates can present short papers and posters, and students can get academic credit for the research they do on the way to a symposium paper or thesis. So far CURO has been mostly limited to Honors students, but the university is expanding it this year to reach more of the undergraduate population. Besides CURO, I have also served as a faculty mentor for more Honors students, about half a dozen a year, and I advise a large number of English majors. Many of these students have decided to work for the Atlas, too. Most

of the rest of our undergraduates were referred to us as the best students in their Introduction to Linguistics sections, or had taken a course with me and done well in it. During the heaviest periods of undergraduate hiring, we would hold training seminars at the beginning of a new term; at other times, the graduate assistant supervisors would bring students along individually. All of the students, however, already came to us with some familiarity with linguistics, and many had specific knowledge about our project.

Prototype of our new Web site GUI

(in preparation by Ilkka Juuso, University of Oulu, Finland)

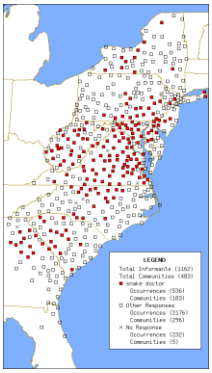


The University of Georgia

The course of mine that attracts the most student interest in doing research on the Atlas is, not surprisingly, American English. In that course I make great use of the interactive GIS section of the Atlas Web site, which is based on the textual data keyboarded from the earlier field records. Two graduate assistants have been responsible for the interactive programming on the site, Rafal Konopka for the initial implementation, and Eric Rochester for the second; we are now developing a major revision of the interactivity of the stie to accommodate our new audio files, again carried out primarily by a graduate assistant, Ilkka Juuso. In the American English course students, both graduate and undergrads, do weekly computer exercises following a 12-part sequence. Students really like the site because it gives them a chance to do work with real research data, and because making maps of who-says-what-where is a bit like playing video games. Making lists is also pretty painless, because our Web site allows subsampling and

automatically produces them. Many students report spending much more time than the 15 minutes I intend for each exercise, and many students end up doing their required writing on the data they discover in the exercises. It is clearly the case that students do not just discover traditional dialect regions in these exercises; indeed, when they look for such regions they are hard to find in the data, as in Exercise 2.

Making maps online



2. Make three maps. Each map should show a word that was found in between 100 and 300 communities [to illustrate Northern, Southern, other distributions]. … Make as many maps as you need to, from as many databases as you need to, in order to create your three maps. This may be harder than you think!

The University of Georgia

Instead, students are learning about the variation that exists in the data, and gaining perspective on a number of different sociolinguistic criteria. We spend most of a class period each week discussing what the students found, because the findings of any one student will not tell the whole story of the sociolinguistic variable that the exercise was about; collectively, the students see enough data that we can make a fair generalization. Because use of the Web site gives students better access to real data than they could get anywhere else, their course papers are often as good or even better than ones I hear at national conferences (though perhaps without the polish that seasoned professionals can apply). Indeed, my student Kristen Frederiksen allowed me to present some graphics from a course paper of hers in my last NWAV presentation, as in this slide. I believe that my students learn more about the realities of American English from using our Web site, and have more fun doing it, than I could teach them from any other source.

Samantha's experiment:
Twelve items Eight subsamples

- **Agriculture/Land Domain**: *hog pen, meadow, swamp*
- **Food Domain**: *cobbler, cornbread, pancakes*
- **Home/Household Items Domain**: *andirons, hearth, pallet*
- **Weather Domain**: *cloudburst, dry spell, steady drizzle*

- By Age: ~16-60, ~60-75, ~75-100.
- By Sex: Male, Female
- By States:
  - Northern Tier:
  - Midland Tier:
  - Southern Tier:

The University of Georgia

Motivated students, and these are often not just the graduate students, download comma-delimited data files from the Web site and then carry out their own analyses, often more complex ones than the Web site alone can support. A case in point is the 2009 undergraduate Honors thesis of Samantha Knoll, which some of you may have heard about at the 2009 SHEL meeting in Banff (Kretzschmar and Knoll 2009). Samantha carried out an experiment in which she listed and plotted the distribution of lexical variants for twelve different questions, both overall and in eight subsamples for each question, in order to test my prediction that such distributions will be non-linear and scaling, a sign of speech as a complex system (Kretzschmar 2009). I am gratified to report that Samantha did indeed find that every distribution she tested, both overall and in every subsample, was non-linear, and that there were three kinds of non-linear "A-curves." Moreover, she was able to advance the work by showing that the speech communities represented by her subsamples were commonly differentiated by the frequency-rank order of the variants. Of course, as her mentor I played some role in Samantha's work, but when I gave her an opportunity she certainly took advantage of it. Her findings are a genuine contribution to the analysis of Atlas data.

Type B order of variants, *cloudburst*

All Type B		Youngest 1/3 Type B	Middle 1/3 Type B	Oldest 1/3	Men	Women	Northern Tier	Middle Tier	Southern Tier	Variants
Cloudburst		Downpour	downpour	cloudburst	cloudburst	downpour	cloudburst	downpour	downpour	2
Downpour		Cloudburst	cloudburst	heavy rain	downpour	cloudburst	downpour	cloudburst	cloudburst	3
heavy rain		heavy rain	heavy rain	downpour	heavy rain	hard rain	heavy rain	hard rain	gully washer	4
hard rain		hard rain	hard rain	hard rain	hard rain	heavy rain	heavy shower	heavy rain	heavy rain	3
gully washer		gully washer	gully washer	gully washer	gully washer	gully washer	hard rain	gully washer	hard rain	2

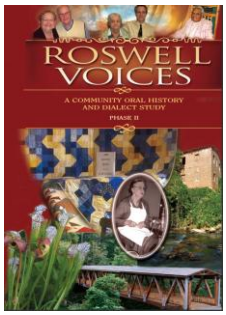
The University of Georgia

Let us now turn to the community language project in the Atlas, Roswell Voices. We were invited into Roswell in 2002 by its Convention and Visitors Bureau, a member of which was familiar with Walt Wolfram's work in North Carolina. The original idea was for us to

discover colorful speechways in Roswell, as Walt had found in Ocracoke and other places, but we have ended up interviewing people, 72 so far, of the oldest, middle, and younger generations, in both the white and historic Black communities there. We have extracted bits of oral history from our guided conversational interviews, and published the selections in two pamphlets sold in the CVB office, which have short audio selections on CD to match the printed transcriptions. We have also made large display posters for the CVB office. The CVB also has some of the audio up on its own Web site, and we plan to mount the Roswell Voices interviews on the pending version of the new Atlas Web site. All of this work has been done on a shoestring, with a little funding from the CVB and a little more from my resources at the Atlas. We have not yet attracted a grant that would permit us to do a thorough job of vetting the interviews, preparing transcriptions, and making good use of the interviews for linguistic purposes.

Roswell Voices

pamphlet poster



Our graduate assistants Becky Childs and Claire Andres have been central figures in both the interviews and work towards the pamphlets and posters. Rachel Votta, who started as one of the Roswell speakers and continued on to do graduate work with me, has collaborated with Claire to publish an article on African American speech in Roswell (Andres and Votta 2009). Undergraduates, too, have played a strong role in Roswell. Two undergrads have received CURO summer fellowships to work on Roswell Voices (a $3000 stipend, so research was their summer job). One of them talked to speakers from the younger generation, the other to business people who had immigrated to Roswell from other countries. The latter student was double-majoring in business and linguistics and so wanted a project that combined the two; that was a good fit for Roswell, a rapidly growing town in the Atlanta metro area, because we need to know more about how people become integrated, or not, with the local community and culture. A cluster of

undergrads from my Sociolinguistics course one year participated as interviewers with several of the younger Roswell speakers. Again, for the undergraduates, supervision is the key to getting the most out of their experience. My student who talked to business people did not do as well as the others, in large part because I was out of the country when she was trying to do her interviews and so I could help her less. She did, however, complete a good presentation for the CURO Symposium after I returned. My other CURO fellow who worked in Roswell, Josh Dunn, completed an Honors thesis on his Roswell subjects (Dunn 2010), and has also co-authored two papers with me, one at NWAV (Kretzschmar and Dunn 2010, which he helped to deliver) and another at the Dialect Society this year (Kretzschmar, Dunn, and Kim 2011)--more about those in a moment. And students continue to volunteer for unpaid work in Roswell, either with back office tasks like transcription or with ideas for field work.

Classroom training for the Sociolinguistics class began with me conducting an actual interview, with Rachel Votta, over the course of two class periods right in front of the class. It was quite a natural interview despite the audience: I arranged the chairs so that Rachel could not see her fellow students, and she relaxed after a few minutes, as is common in interviews, and was able to carry out our guided conversational protocol. Other classroom training for and uses of Roswell Voices focus on the interview recordings. We often play sections of the recordings from the pamphlet CDs, so that students can hear the contemporary variety in voices from the same community. Sometimes the point is practice with type/token analysis on specific features; at other times students are invited to listen for "interesting features" in the recording, which leads to discussion of what they heard in the context of what is known about Southern speech. I have also taught one-credit Honors seminars specifically about Roswell Voices, in which I have provided background about Southern speech, discussed the Roswell Voices project and let the students see grants proposals, discussed what we hear in the interviews, and actually driven them 90 minutes to Roswell to show them the town. These seminars have been quite popular.

We have not used special software for Roswell Voices in the classroom because we have not had the funding to develop any.

We can conclude our survey of student participation on the Atlas with Josh Dunn, whose Honors thesis and two papers with me I promised to talk about. Josh was one of my Honors mentees, having picked me off a list of mentors when he was a freshman. He got his CURO fellowship for the time between his freshman and sophomore years, so he started very early. He interviewed four young Roswell subjects, prepared transcriptions, and did his first CURO symposium presentation on his experience. It was a youthful effort, yes, but it was good enough to be accepted for presentation at the national CURO meeting, too. Josh later interviewed five more young Roswell speakers, and for his thesis focused on type/token counts from a list of twelve Southern features in the literature. His findings show that the younger generation in Roswell has mostly lost the Upland Southern features of the earlier generation, but at the same time most of them still have General Southern features, sometimes at rates higher than their elders. He also applied his count frequencies to my new complex systems model, and we showed in our NWAV paper how it would be possible to be misled about speech in Roswell without sufficient attention to scaling factors. Finally, in our ADS paper Josh's data from Roswell, in combination with some earlier data from UGA students, suggested that the Southern features appeared to have an implicational relationship, so that use of the least common features at some frequency implied some use of the more common features. This suggests how

Dunn's evaluation of features (rounded to nearest 5%, x = rare)

	CF	DF	GF	IF	KF	AM	BM	HM	JM	Older speak
u			10			5		15		30
ai			5		35	45	5	65	x	80
pin/pen		15	15	5	65	45	30	70	x	30
init str		25	35	10	45	40	30	35	20	35
ɔi			x	x				x		
i/ɪ					15			10		10
əou					20			20		10
dawg										10
æu, æi										45
-r unst										10
-r										15
ə ins										10

perception and cognitive habituation are involved in our everyday linguistic choices. Josh was a star, of course, for a student participant on the Atlas, but the extent of his work and the significance of his findings does show what is possible even for undergraduates.

The bigger picture of student participation on the Atlas revolves around whether or not students are paid for their work. When we can pay students, whether undergraduates or graduate students, I feel that we can hold them to a specific work plan and demand a high level of work product. This has been the case in our digitization efforts both with audio and with text. Success rates for such work are directly proportional to the amount of supervision applied. I will admit that with graduate students whose assistantships do not have a particular work plan, I consider my role to be one of providing opportunities and so I give the grad students plenty of room either to make a success or to show that they are not self-directed enough to take advantage. I don't narrowly channel grad students like these into particular Atlas tasks, but instead try to find ways that they can apply their own interests within the Atlas framework. In closely budgeted projects with a specific plan like our digitization work, I have to take a different view. As well as giving attention on a daily basis to particular tasks and problems, we have biweekly staff meetings with my administrative associate and the grad students so that everybody knows what and how we are doing and keeps up. Funded plans also allow for the development and use of new software products. Undergraduate students on a specific work plan have narrowly defined tasks, and so they may not see the bigger picture of what the Atlas does, and instead learn to use particular software as part of the experience of being a member of a research team. However, undergraduates who work with me as volunteers or on CURO projects are treated more individually, like the free-agent grad students except that I try to provide constant supervision for what they are doing. I encourage their combination of interests, and try to help them develop ideas that serve them as students as well as the Atlas project. In the end, then, what I try to do with student participation on the Atlas project is to use the project as

much as possible for the education of the students, and not just use the students as labor for the Atlas.

References

Andres, Claire, and Rachel Votta . 2009. African American Vernacular English: Vowel Phonology in a Georgia Community. In M. Yaeger-Dror and Erik Thomas, eds., *African American English Speakers and Their Participation in Local Sound Changes: A Comparative Study, PADS* 94, 75-98.

Bloch, Bernard. 1935. The Treatment of Middle English Final and Preconsonantal *R* in the Present-Day Speech of New England. Diss, Brown University.

Dunn, Josh. 2010. The Youth of Roswell Voices: A Linguistic and Social Study of Community Definition. Undergraduate Honors Thesis, University of Georgia.

Knoll, Samantha. 2009. Data From the LAMSAS Project: How Language Varies in Systematic Ways. Honors thesis, University of Georgia.

Kretzschmar, William A., Jr., Paulina Bounds, and Naomi Palosaari. 2010. Issues in Using Legacy Data. In M. Di Paolo and M. Yaeger-Dror, eds. *Sociophonetics* (London: Routledge, 2010), 46-57.

Kretzschmar, William A., Jr., and Josh Dunn. 2010. Complex Systems and Sociolinguistics in Roswell, NWAV 39, San Antonio.

Kretzschmar, William A., Jr., Josh Dunn, and Mi Ran Kim. 2011 Implicational Scaling in Southern Speech Features, ADS/LSA, Pittsburgh.

Kretzschmar, William A., Jr., and Samantha Knoll. 2009. How Speech Communities Differ. Paper presented at SHEL, Banff.

Kurath, Hans. 1939. *Handbook of the Linguistic Geography of New England.* Providence: Brown University, for ACLS. 2nd ed., rev., New York: AMS Press, 1973.

O'Cain, Raymond K. 1979. Linguistic Atlas of New England. *American Speech* 54: 243-278.

Abstract:
Over time students have become much more involved in the Linguistic Atlas Project (LAP). Graduate students first became analysts of LAP data in their dissertations, and then became collectors of the data, too, as Lee Pederson and Roger Shuy at Chicago, and later many students at Emory and Georgia, executed interviews. Most recently, the LAP at Georgia has given research opportunities to a great many undergraduate students involved in keyboarding old paper records, and in processing audio-taped interviews to create digital versions for the archive and for public use. These students see firsthand how collaborative research works, and also get to use the research products in undergraduate classes. In our American English course, e.g., use of interactive mapping and data lists from the LAP Web site is a key pedagogical technique, from which students learn about geographical and social distributions in our survey data.