

GORDON R. WOOD

REFINEMENTS IN TABULAR MODELS OF VARIATION
IN REGIONAL AMERICAN ENGLISH *

1. INTRODUCTION

Models of the geography of American regional English have been greatly influenced by the work of researchers for the proposed Linguistic Atlas of the United States and Canada. Among the representative publications one should note H. KURATH's *A Word Geography of the Eastern United States* (1949) and E. B. ATWOOD's *The Regional Vocabulary of Texas* (1962), particularly for their use of tables which place selected words in a designated locality and give an indication of relative abundance there. It is these two works which set the pattern for the present computer assisted analysis of the presence of 1,000 regional words in the southern United States. Thus there are resemblances among these models, but there is an essential difference also in that the newer model advances claims about relationships among word choices and thus is implicationally more dynamic than are its predecessors.

2. FIRST STAGES IN MODEL FORMATION

The present study began in the 1940's when the computer was first being praised as the machine for sorting vast bodies of data. At that time the Linguistic Atlas project had not gone forward in many parts of the South, but it had produced a printed questionnaire which asked persons to make choices among synonyms. Here were tools which could be used together as a means of discovering gross features of vocabulary distribution within the relatively unexplored states of Tennessee, Florida, Alabama, Mississippi, Louisiana, Arkansas and Oklahoma. Georgia was included as an eastern anchor, even though the state had been surveyed in part for the Atlas (R. I. McDAVID,

* This study was supported in part by Southern Illinois University, Edwardsville.

1958). Texas was excluded because Atwood's survey was almost ready for publication at that time.

When the questionnaires were returned, many were rejected because the answerer's biography showed that he did not meet certain standards. From the chosen questionnaires the coded data was transferred to punch cards. There were codes for the person's age, sex, education, race, place of birth and longest residence, places of birth and residence of his parents and grandparents; occupation was not coded simply because of the vast range of identification needed. Then the response to each set of synonyms was entered. If, for instance, in item 125, *VEHICLE FOR A SMALL BABY*, the synonym *baby cab* had been selected from the synonym set *baby buggy, baby cab, baby carriage, baby coach*, then it would be punched as the next entry on the card after the biographical data.

The original intention was that the computer would sort the data first under age, then under education, and then under sex. For each sort the computer would print a county by county table showing the actual number of local responses according to a selected heading. Where H. KURATH's tables (1949, p. 12) showed "regular" or "fairly common" or "rare" in Ohio, New York State, and New England respectively, the computer list would show "0" or "2" or some other number for *baby cab* in Hamilton County, Tennessee, and so through each county for each word of the synonym set.

In order to make some comparison with Atwood's tabulation of responses according to age, the first sort of all data was by three age groups - those above seventy, those between thirty and seventy, and those below thirty. A county by county computer listing of 1,000 words as chosen by 1,000 people is enormous. Its very mass made impossible the publication of such a document even for those dialectologists who might want to look at the raw data in its most detailed form. A byproduct, hand drawn maps made by plotting the presence of selected words, permitted a redefining of the technical terms "Midland" and "Southern" dialect as they apply to this region (H. KURATH, 1949; G. R. WOOD, 1963, 1971).

3. REFINEMENTS FOR BOOK PUBLICATION

The computer was next instructed to make total counts of all choices of each synonym within four main divisions of the respective states as

well as for the state itself, and for the entire eight state region. The grand total for each word in the questionnaire was machine sorted according to increasing numerical frequency of response. This computer printout became the source of a table photocopied and placed in a book (G. R. WOOD, 1971, pp. 297 ff.). As a table, it seems at first glance to say what those in Kurath and Atwood also say: here is a regional word count. It shows that roughly 1,000 persons chose or rejected certain words and that their choices were somewhat mixed with regard to the rest of the vocabulary. Actually the claim is different. This is a model of a whole vocabulary as a closed system in which for an entire region the place of each synonym within the system is fixed at a given point on the scale. Obviously a search at a different time or with a different questionnaire or among different people would shift the entries, but it would not alter the systemic design. No such ordered conclusions can be gained from the Kurath and Atwood models since the Linguistic Atlas interview procedure has no way of testing an informant for his range of synonyms.

Given these gross counts, the next stage was to design a table that would display the presence or absence of each word in a smaller geographic setting. It became apparent that statements of percentages would be more informative than would be the presence of raw numbers. Hence a model emerged which listed all of the words in a synonym set and placed them proportionately in each of the eight states. It combined the features of locality found in the Kurath model with the features of relativity in the Atwood one, but it excluded the gradations by age. Furthermore, it refined the statement of relationships by breaking the data into four sub-patterns which answered these questions: What are the relationships in choice of the same word in a section of adjacent states such as those situated east of the Mississippi? Within any state and its major subdivisions what is the relationship of a given response to the largest number of possible responses there? Next, if the statewide response is regularly set at 100 per cent, what are the relationships between the responses in the north to those in the south, or in the east to those in the west? And finally, in each locality what are the relative preferences for each member of the synonym set?

The computer printout of these percentile calculations became the photocopied text of a book (G. R. WOOD, 1971, pp. 64 ff.). An extract from one will suggest the format. The synonyms given in the questionnaire for the concept "a woman who helps at childbirth" were *godmother*, *granny*, *granny woman*, and *midwife*. Consider two choices within

the states of Tennessee and Georgia, and their western-eastern (W E) northern-southern (N S) halves. In the example below, the left column stresses the local relative importance of each word; the state (ST) percentage is not printed but is set at 100 per cent. The right hand column shows the relative preference among synonyms within the full set.

	TENNESSEE			GEORGIA		
GRANNY WOMAN	W	31	4	N	53	6
	E	69	10	S	47	6
	ST		14			12
MIDWIFE	W	45	30	N	63	45
	E	55	37	S	37	27
	ST		68			71

For the word geography of the United States, the refinement called a closed model here has certain advantages. The text of the questionnaire can and should include standard words such as *midwife* alongside those synonyms which have been identified as having regional limits, words like *godmother*, a standard word with a special regional meaning, and *granny woman*, an entirely regional word. The computations, then, show among other things that the regional word and the standard word considered by themselves have variations in regional preference. It is evident that east Tennessee has stronger preference for *granny woman* and *midwife* than does west Tennessee; a similar pattern of major preferences appears in north Georgia (53 and 63 per cent versus 47 and 37 per cent in the tabular extract given above). Obviously the calculations confirm that among synonyms the general word is greatly preferred over a more regional one. In addition they will show what had not been formally identified up to the time of calculation: general words have variations in regional preference as well. The degree of variation leads to the idea that the dialectology of American English must consider this characteristic along with the more easily identified characteristics such as the presence of distinctively local words such as *granny woman*.

REFERENCES

- E. B. ATWOOD, *The Regional Vocabulary of Texas*, Austin, 1962.
- H. KURATH, *A Word Geography of the Eastern United States*, Ann Arbor, 1949.
- R. I. McDAVID Jr., *American English Dialects*, and (with VIRGINIA McDAVID) *Appendix: Maps*, in W. NELSON FRANCIS, *The Structure of American English*, New York, 1958.
- G. R. WOOD, *Dialect Contours in the Southern States*, in « American Speech », xxxviii (1963), pp. 243-56.
- G. R. WOOD, *Vocabulary Change: A Study of Regional Words in Eight of the Southern States*, Carbondale and Edwardsville, 1971.