The East-West New England Dialect Boundary: Another Look at the Evidence

Lawrence M. Davis, Charles L. Houck and Veronika K. Horváth

Abstract

This paper analyses twenty-one list manuscripts from the Connecticut portion of the Linguistic Atlas of New England (1939-43) (LANE) to examine further the concept of dialect boundaries. We chose Connecticut because of Kurath's (1939) claim that the Connecticut River forms a clear east/west boundary for New England. A second reason for choosing Connecticut is that, further north, almost all of New England falls in LANE's eastern dialect area.

So far as we know, no one before us has questioned Kurath's (1939) claim in the Handbook of the Linguistic Geography of New England that 'New England has two major dialect areas, an Eastern and a Western', and that 'the "seam" between these two settlement areas runs straight north from the mouth of the Connecticut River (between [subjects] 30 and 31) through Connecticut . . .' (p. 8). Given our findings, we think that the time has come to take another look at that analysis and the assumptions behind it.

The analysis involves a multidimensional scaling (MDS) procedure to analyze the lexical and phonological responses in the LANE records of the sixty-seven Connecticut subjects. We believe that the results from this analysis will contribute further to the continuing discussions regarding what exactly we mean when we use the term dialect boundary.

This article analyses twenty-one list manuscripts from the Connecticut portion of the *Linguistic Atlas of New England* (1939-43) (LANE) to examine further the concept of dialect boundaries. We chose Connecticut because of Kurath's (1939) claim that the Connecticut River forms a clear east/west boundary for New England. A

second reason for choosing Connecticut is that, further north, almost all of New England falls in LANE's eastern dialect area. Figure 1 shows a map of the sites sampled by LANE in Connecticut:

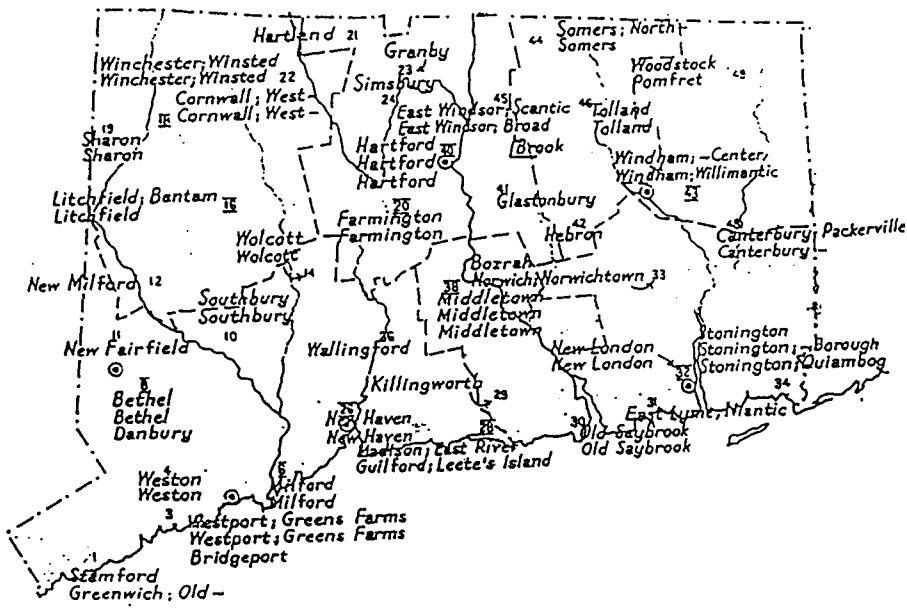

Figure 1

So far as we know, no one before us has questioned Kurath's (1939) claim in the *Handbook of the Linguistic Geography of New England* that 'New England has two major dialect areas, an Eastern and a Western', and that 'the "seam" between these two settlement areas runs straight north from the mouth of the Connecticut River (between [subjects] 30 and 31) through Connecticut . . .' (p. 8). Given our findings, we think that the time has come to take another look at that analysis and the assumptions behind it.

This article will present the results of a multidimensional scaling (MDS)[1] procedure to analyze the lexical and phonological responses in the LANE records of the sixty-seven Connecticut subjects. We believe that the results from this analysis will contribute further to the continuing discussions regarding what exactly we mean when we use the term *dialect boundary*.

Method

First we had to choose a method of analysis that would provide a descriptive statistical tool that helps to discover underlying structures, relationships, or affinities in data which contain considerable variation and error. MDS is such a tool. We view the major advantage of MDS over other statistical techniques used on dialect data to be that one can use MDS to analyze simultaneously the similarities and differences between the individual subjects' lexical and phonological inventories on the one hand, and similarities and differences between the behavior of the lexical and phonological items themselves on the other.

MDS analyses give two sets of scores: the object scores and the category quantifications. The object scores are assigned to the subjects so that those who show similar patterns of presence and absence of the selected lexical and phonological items will receive similar scores. For example, two subjects who show exactly the same pattern of presence or absence of certain items will receive exactly the same object score. A subject whose responses differ only by the presence of one lexical or phonological item will receive a score closer to the first two subjects than one who differs by the presence of two different lexical or phonological items, etc.

The category quantifications provide information on the actual dialect terms in the subjects' speech. These scores characterize the presence or absence of lexical items. For example, if two lexical items, like *bucket* and *gunny sack* or *gutters* and *tassel* are present in the inventory of the same subjects, the category 'presence' for the two items will receive similar quantifications, i.e. similar values. Both the object scores and the category scores can be plotted, and can be further analyzed using correlation coefficients. MDS also calculates a Variable Dimension Score for each variable. This score indicates the relative importance of every variable in the overall solution, and thus provides a way to quantify the results of the item-based results we get from traditional dialectology.

All this means that multidimensional scaling, like other forms of multivariate statistics, allows dialectologists to group similar subjects together rather than responses. This capability is not insignificant since, at least in the real world, we tend to think of dialects as groups of people speaking similarly rather than groups of unrelated responses. For example, the methods of traditional linguistic geography can tell us whether a Survey of English Dialects subject said [θundə] or [θʌndə], but, if the former pronunciation was recorded, it cannot tell us whether he or she also said [butə], or whether he or she uses other northern forms such as *beasts* vs. *cattle*. While we are not the first dialectologists to use multivariate statistics to group subjects

(Linn 1981, Linn and Regal, 1985,1988, Wachal 1986, Cichocki, Péronnet and Babich 1988), we would argue that, given modern statistical models, dialectologists should undertake more efforts in that direction.

Of course, this is not the method employed in traditional American dialectology. That method involved making list manuscripts of each item on the questionnaire (i.e., the different terms for the dragon fly, the pronunciation of *four*, and so on). The dialectologist would then map these terms, and, where patterns seemed to exist, he or she would draw an isogloss between them, much as Orton and his colleagues did in *A Word Geography of England* (1974) and *The Linguistic Atlas of England* (1978). At that point, the American and English methods diverged: the English, of course, mapped the isoglosses only, while the Americans went further and also tried to find major 'bundles' of isoglosses in order to establish major and minor dialect boundaries. Figure 2 shows one such bundle of isoglosses for Connecticut (Kurath et al. 1939: 30):

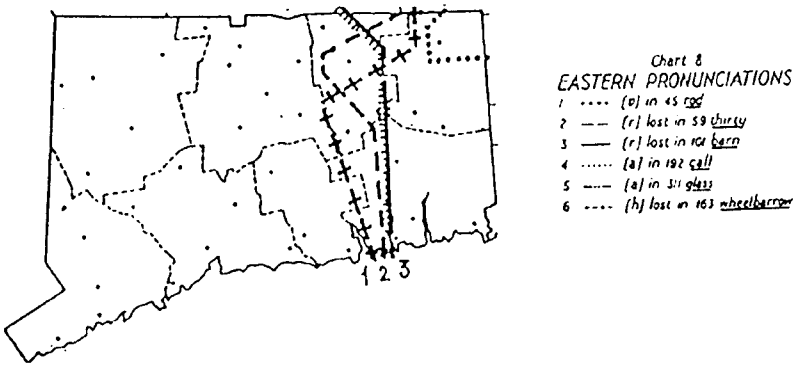

Figure 2

For our study, we first had to select and code the data. Table 1 lists the ten phonological items used by Kurath to show the presence or absence of constriction of postvocalic /-r/:

barn
father
afternoon
girl
water
beard
morning
corncrib
this year
thirty

Table 1

Table 2 lists the remaining selected four items with their respective contrasts which, according to Kurath, also evidence distinct isoglosses:

rod [ɑ] versus [ʊ]

calf [æ:] versus [ə]

glass [æ:] versus [ə]

tassel [æ:] versus [ə]

Table 2

Table 3 lists the lexical items which Kurath used:

1. sour milk/lapp(bb)ered milk/curdled milk/curdled klabber/bonny kapper (milk)/ klabber/klabber(ed) milk/thick milk
2. wheat bread/bread/loaf bread/raised bread/light bread/raised wheat bread
3. quilt/bed quilt/comfortable/tak comfortable/comforter, comfort quilt/ comfort/batwork comfort/batswork quilt/feather/down comfortable/patchwork quilt/puff
4. stone boat/stone board/stone drag/stone drug, drog/drag/drag board/drog, drug/sled
5. tassel, tossel/corn tassel, tossel/topgallant/top out/top/tip/pole /corn top/spindle/blow/tassel/tossel out/hound's ears
6. stovepipe(bibe)/smokepipe/stove funnel/funnel/pipe/
7. griddle cake/pancake/flap over/flapjack/slapjack/flannel cake/griddles
8. gutter(s)/eave(s) trough(s)/gutter pipe/trough(s)/conductor(s)/eave(s) pipe(s)/spout

Table 3

The presence or absence of each variant given in response to a lexical or phonological question was coded as a separate variable. Since MDS works best when all the categories contain at least ten percent of the possible answers, we excluded items such as *bread, loaf bread, raised bread, light bread,* and *raised wheat bread,* but were able to include the more ubiquitous *wheat bread.* More specifically, since the maximum possible frequency for any one form was sixty-seven – the number of LANE subjects in Connecticut – we included in the analysis only the items which had at least six occurrences. As a result, we were able to use twenty-three lexical items and fourteen phonological ones. In fact, the frequency range for the lexical responses was from six to sixty-one, meaning that as few as six subjects used the least frequent form and as many as sixty-one used the most frequent one.

Results

Our first task was to examine the object scores to ascertain whether Kurath was justified in proclaiming such a definitive east-west boundary along the Connecticut River. Looking at the scores for the constriction of postvocalic /r/, we did find it to be every bit the east-west New England marker that Kurath claimed it to be. Indeed, with one lone exception east of the river, Connecticut subjects were r-less in at least some of their responses, and, generally, the further east that they reside, the more r-less they became. The situation west of the river is more complex than indicated by Kurath's claim. First of all, Hartford, the state capital, is r-less, and lies just west of the river. Middletown and Old Saybrook, also just west of the river, are mixed, but the former is only weakly so. New Haven, the site of Yale University, is also mixed, and lies some thirty-five miles west of the river. Wallingford, some twenty miles west of the river, and Milford, over forty-five miles west, both evidence clear r-less responses. In addition, when one looks at the Variable Dimension Scores, there is no doubt that this phonological feature has discriminating power. The scores range from a high of .875 to .545, with *beard* the highest, and *corncrib* the lowest. One can say as Kurath did, however, that in general speakers west of the river are r-full.

In order to run a regression analysis on the data, we arbitrarily chose eleven sites in central Connecticut[2] and used their object scores for the analysis.[3] Where there were two object scores, we took the average, but these scores were close enough so as not to skew the results. The Spearman Rank Correlation analysis on the relationship between the subjects' geographical location and their pronunciation of constricted postvocalic /r/ revealed a very high correlation between the two (0.932) where the

probability of getting such results by chance are less than one in 100 ($p < .01$). Figure 3 presents a scattergram of these findings:

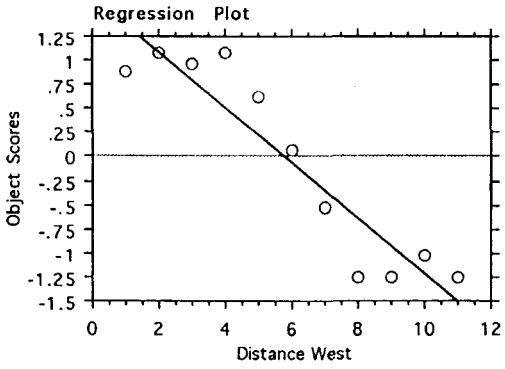

Figure 3

This inverse correlation between the absence of postvocalic /r/ constriction and distance west, coupled with the high Variable Dimension Scores (.875 to .545) leaves little doubt as to the importance of postvocalic /r/ in the LANE. It does suggest, however, that there is no clear, sharp boundary between eastern and western Connecticut; rather, there is basically a gradual lessening of r-lessness as one moves westward.

The other phonological items that Kurath used to establish his east-west boundary are not very discriminating at all, in spite of maps such as Figure 2 noted earlier. The [æ:/a] contrast in *glass*, *calf*, and *tassel* does not show any central distributional tendency. Both [æ:] and [a] occur quite frequently on both sides of the Connecticut River. Ninety-one percent of all the subjects have [glæs], seventy-seven percent have [kæf], and fifty-two percent have [tæsəl], and in no case is there an east-west distribution of these terms. The latter item, *t[æ]ssel*, is probably nondiscriminating since that distinction was carried well west of the Mississippi, let alone west of the Alleghenies. The [rɑd / rɒd] distinction is almost as mixed and nondiscriminating as *tassel*. Sixty-seven percent have [rɒd] and twenty-three percent have [rɑd]. Later dialectologists have in general considered this more of a northern/southern distinction than east/west. These high percentages of incidence for *glass*, *calf*, and *tassel* among Connecticut speakers indicate that [æ] is just likely to occur east of the river as is [a] and cannot be used in any definitive way in establishing an east-west dialect boundary, despite the fact that eleven occurrences of [a] are in fact located east of the river. Four

occur west of the river as well.

The presence or absence of postvocalic /-r/ appears to be the only east-west marker, and the Connecticut River seems to be the boundary, only because it is in the middle of the state. Our analysis indicates a clear linear relationship for postvocalic /r/ rather than a clear dialect boundary. Too many r-less speakers exist in the LANE records west of the Connecticut River. R-lessness, at least in 1931-33, was probably enough to geographically mark Connecticut speakers, even though what seemed to be other related markers were probably more perceptual than actual. They certainly were not as discrete as the Kurath (1939) statements make them out to be.

The lexical object scores, which measure how similar were the subjects' responses, reflect the east-west boundary to some extent, but overall the scores are much lower. For example seventeen of the lexical scores are below .50, while only four subjects evidenced r-object scores that low. This suggests, of course, that the lexical data lack much similarity or agreement.

The category quantifications of the lexical items themselves are even less revealing. The values are very low, with only two showing any discriminating power at all. These two lexical items are *lapp(bb)ered* milk with a value of .474 and *stone boat* with a value of .331. The first can be compared with two other members of this lexical set: *curdled milk* with a very low value of .002 and *bonny klapper* with a value of .235. The value of *stone boat*, .331, can be compared to that of *stone drag* and *drag* with their values of .219 and .048. The only other somewhat high value was the ubiquitous *pancake* with .224, hardly a discriminating term. And it is only relatively high when matched against *griddle cake* (.020), *flapjack* (.032), and *slapjack* (.004).

The category quantifications are all relative. The **highest** category quantification score for the lexical items was .474, while the **lowest** score for postvocalic /-r/ constriction was .575. Since the highest score for postvocalic /-r/ constriction is .875, it seems clear enough that /-r/ is a much better indication of regional difference than any of the lexical items. Furthermore, a look at the actual LANE maps confirms these results. As Harold Orton liked to say, there it is – right on the ground. No lexical item we examined came even close to being representative of regional distribution as was *corn crib*, which had the lowest category quantification score of all the examples of postvocalic /-r/ in our study.

Conclusion

Our conclusion for the LANE Connecticut results is that there is no set of

lexical and/or phonological features which, seen together, can justify Kurath's positing the Connecticut River as the east-west boundary. What does appear, however, is the undeniably strong east-west correlation for postvocalic /r/ constriction. The other phonological items that Kurath used to establish the east-west boundary proved to be less than helpful, showing little or no distributional patterning. The distribution of the lexical object scores also fails to support Kurath's east-west division. Furthermore, the lexical results were not isomorphic with the scores for postvocalic /r/; that is, a subject with a high r-less object score may only have a lexical object score half that of the r-less object score. In addition, the weak category quantification scores for the lexical items do not allow us to place great stock in their discriminating power. We suspect that where Kurath found dialect boundaries, there are only occasional isoglosses. In no instance did we find the necessary bundling of isoglosses that would indicate a major or even a minor dialect boundary. Kurath's Connecticut isoglosses in Figure 2 here and in the *Handbook of the Linguistic Geography of New England* do not provide a categorical division between lexical items or pronunciations any more than his *wheelbarrow* isogloss did so further north (see Davis and Houck 1995: 380-81). Many r-less speakers, for example, are found west of the Connecticut River as well as east of it.

These findings should be considered in the context of earlier studies of ours. Our studies of the upper midwest (Davis and Houck 1995) and Iowa (Horvath and Houck 1996) as well as the paper we presented at Methods IX on the dialect situation here in England (Davis, Houck and Upton 1997) give us real reason to question the traditional American notion of dialect boundaries. In all three of these works, we tried to show that certain so-called major dialect boundaries are a function of which forms the dialectologist chooses to combine and to map. In the Methods IX paper we noted earlier (Davis, Houck and Upton 1997), we quoted from Gaston Paris' (1888) 'Les parlers de France,' who in turn quoted from Paul Meyer. The following is probably truer today for the United States and England than it is for much of western Europe, including the area that Paris was discussing:

> . . . dans une masse linguistique de même origine que la nôtre, il n'y a réellement pas de dialectes; il n'y a que des traits linguistiques qui entrent respectivement dans des combinaisons diverses. . . (163)

Given the findings of our previous work and this one as well, we must conclude that dialect boundaries in both England and the United States have both a psychological and a perceptual reality that typically are very difficult to verify

objectively. It is difficult to say at what point someone becomes dialectally different from another – what composite of phonological and lexical features permits us to recognize another as being dialectally different from ourselves. We all make these distinctions, but how we do so is not obvious from our data or from any other English or American data we have seen.[4]

The Connecticut isogloss for postvocalic /-r/ is not at all easy to draw, and it may not be possible to draw it at all. It is conceivable that one might have more success if one were to use a statistical model such as that suggested by Kretzschmar (1992), a method which uses statistical techniques very different from multi-dimensional scaling to establish isoglosses. It would be most interesting to use Kretzschmar's (1992) method on the Connecticut data to see if his results would replicate ours. We suspect that they would, given the actual numbers of occurrences for both postvocalic /-r/ constriction and the lexical items examined here. To use Harold Orton's expression negatively, the Connecticut boundary is not at all apparent on the ground.

In *American Tongues*, a film produced some years ago and still used at a number of American universities, Roger Shuy tells the story that, as one passes from western to eastern New England, one goes from /park jə kar/ to /pak jə ka/. Our evidence suggests that things are far from that simple. To complicate matters further, we found that r-lessness occurred more frequently, but not statistically so, before nasals and in open position, so *your* and *car* would evidence more r-lessness than with *park*.

Michael Linn has observed (personal communication) that sometimes just one perceptual isogloss, if it is important enough, might be considered a dialect boundary. Charles Houck tells the story of riding southward in the Midlands with Stanley Ellis when Ellis asked, 'Did you feel that bump?' After Houck replied that he had felt nothing, Ellis responded with, 'Well, we just crossed the [bugə/[bʌgə] line!' Unfortunately, this particular distinction was not published in Orton's *Linguistic Atlas of England*, but we can be fairly certain that it would not match completely with the lines marking variant pronunciations of *thunder* and *butter*.

Bloomfield (1933: 328-29) recognized this problem when he noted that the [hus] / [hys] boundary in Dutch was different from the [mus] / [mys] boundary. He, like Gilliéron before him, concluded that 'every word has its own history'. The multidimensional scaling analysis of the Connecticut data certainly lends additional support to this hypothesis. It is altogether fitting to end this paper by pointing out that we Americans probably should have long ago taken our clue from English dialectologists, and should have just observed, recorded, and mapped the regional differences. Our research indicates that, whatever discoveries future dialectologists

might make about English in England and America, they will not involve finding bundles of isoglosses justifying major and minor dialect boundaries. Harold Orton and Eugen Dieth, on this question, surely knew what they were about.

NOTES

We would like to thank William A. Kretzschmar, Jr. and Michael D. Linn, who read earlier versions of this article and made some very candid and very helpful suggestions for its improvement.

[1] The MDS algorithm used in this paper is the Homogeneity Analysis via Alternating Least Squares or HOMALS, an SPSS program developed by the Department of Data Theory, University of Leiden.

[2] The towns sampled were, from east to west, Canterbury, Windham, Norwichtown, Hebron, Glastonbury, Middletown, Farmington, Wolcott, Southbury, New Milford, and New Fairfield. We ran similar correlations through both the northern and southern parts of the state. Since our findings for all three of these correlations were essentially the same, we decided to present the results for the central part of the state here.

[3] We decided to use Spearman's Rank Correlation Coefficient, a non-parametric correlation coefficient, because the data are ordinal in nature.

[4] While isoglosses may not bundle into clearly-defined dialect boundaries in England and the United States, Guillaume Schiltz has noted (personal communication) that a very different situation obtains in much of western Europe. He notes, for example, that the *Schwarzwaldschranke* 'runs over the northern and middle Black Forest and divides the Low-Alemanic dialects in the West from the Swabian in the East.' In addition, Klausmann 1997 demonstrates that the Allemanic-Franconian dialect boundary, a function of the Medieval boundary between Franconia and Swabia, is still alive and well, evidenced by a number of isogloss bundles.

REFERENCES

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart, and Winston.

Cichocki, Wladyslaw. 1988. The use of dual scaling in social dialectology: multidimensional analysis of vowel variation. In Alan Thomas (ed) *Methods in Dialectology*, pp. 187-99. Clevedon and Philadelphia: Multilingual Matters.

Cichocki, Wladyslaw, Louise Péronnet and Rose Mary Babich. 1988. Atlas linguistique du vocabulaire maritime acadien: a final progress report. In Alan Thomas (ed) *Methods in Dialectology*, pp 109-19. Clevedon and Philadelphia: Multilingual Matters.

Davis, Lawrence M., and Charles L. Houck. 1995. What determines a dialect area?: Evidence from the Linguistic Atlas of the Upper Midwest. *American Speech* 70: 371-86.

Davis, Lawrence M., Charles L. Houck, and Clive Upton. 1997. The question of dialect boundaries: the SED and the American atlases. In Alan Thomas (ed) *Issues and Methods in Dialectology*, pp. 271-83. Bangor: University of Wales.

Horváth, Veronika K. and Charles L. Houck. 1996. The Northern-Midland Boundary in Iowa Again?: A Global Approach to Dialectology. Midwest American Dialect Society. Chicago, November.

Houck, Charles L. 1967. A computerized statistical methodology for linguistic geography: a pilot study, *Folia Linguistica* 1: 80-95.

Houck, Charles L. 1969. A Statistical and Computerized Methodology for Analyzing Dialect Materials. Diss. University of Iowa.

Houck, Charles L. 1985. The Iowa Northern-Midland Boundary Revisited: A Computerized and Statistical Analysis. Midwest American Dialect Society. St. Louis, November.

Houck, Charles L. 1986. Multidimensional Scaling as a Statistical Analytical Procedure in a Traditional Dialect Study. Midwest American Dialect Society. Chicago, November.

Klausmann, Hubert. 1997. Der Ellwanger Sprachraum – ein Ostschwäbisches Randgebiet. In A. Ruoff and Peter Löffelad (eds) *Syntax und Stilistik der Alltagssprache: Beiträge der 12. Arbeitstagung zur alemannischen Dialektologie*, pp. 65-83. Tübingen: Max Niemeyer Verlag.

Kretzschmar, William A., Jr. 1992. Isoglosses and predictive modeling, *American Speech* 67: 227-49.

Kurath, Hans. 1939. *Handbook of the Linguistic Atlas of New England.* Providence, RI: Brown University for the American Council of Learned Societies.

Kurath, Hans. 1939-43. *Linguistic Atlas of New England.* 3 vols. in 6 pts. Providence, RI: Brown University for the American Council of Learned Societies.

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

Linn, Michael D. 1981. A statistical model for classifying dialect speakers. In H. Warkentyne (ed) *Papers from the Fourth International Conference on Methods in Dialectology*, pp. 244-53. Victoria, B.C.: University of Victoria Press.

Linn, Michael D. and Ronald Regal. 1985. Numerical taxonomy as a tool in dialect research. In H. Warkentyne (ed) *Papers from the Fourth International Conference on Methods in Dialectology*, pp. 245-62. Victoria, B.C.: University of Victoria Press.

Linn, Michael D. 1988. Verb analysis of the Linguistic Atlas of the North Central States: A case study in preliminary analysis of large data sets. In Alan Thomas (ed) *Methods in Dialectology*, pp. 138-55. Clevedon and Philadelphia: Multilingual Matters.

Paris, Gaston. 1888. Les parlers de France, *Revue des patois gallo-romans*. 2: 161-75.

Wachal, Robert S. 1986. An Investigation into the Use of Inverse Factor Analysis for Forming Clusters of Informants Based on Linguistic Variables. Midwest American Dialect Society. Chicago, November.

Addresses

Lawrence M. Davis
Department of English
Wichita State University
Wichita
KS 67207
USA
<davis@wsuhub.uc.twsu,edu>

Veronika K. Horváth
Department of Computer Science
Southern Illinois University
Carbondale
IL 62901
USA

Charles L. Houck
Department of English
Ball State University
Muncie
IN 47306
USA