

Delimiting the Sydney Speech Community

Author(s): Barbara Horvath and David Sankoff

Source: *Language in Society*, Vol. 16, No. 2 (Jun., 1987), pp. 179-204

Published by: [Cambridge University Press](#)

Stable URL: <http://www.jstor.org/stable/4167835>

Accessed: 16/03/2011 15:46

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=cup>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Cambridge University Press is collaborating with JSTOR to digitize, preserve and extend access to *Language in Society*.

Delimiting the Sydney speech community

BARBARA HORVATH

*Department of Linguistics
University of Sydney*

DAVID SANKOFF

*Centre de recherches mathématiques
Université de Montréal*

ABSTRACT

Quantitative analyses of large data sets make use of both linguistic and sociological categories in sociolinguistic studies. While the linguistic categories are generally well-defined and there are sufficient tokens for further definition based on mathematical manipulation, the social characteristics such as socioeconomic class or ethnicity are neither. The familiar problem of grouping speakers by such sociological characteristics prior to quantitative analysis is addressed and an alternative solution – principal components analysis – is suggested. Principal components analysis is used here as a heuristic for grouping speakers solely on the basis of linguistic behaviour; the groups thus defined can then be described according to sociological characteristics. In addition, by naming the principal components, the major linguistic and social dimensions of the variation in the data can be identified. Principal components analysis was applied to vowel variation data collected as part of a sociolinguistic survey of English in Sydney, New South Wales, Australia. (Sociolinguistics, variation studies, quantitative methods in linguistics, dialectology, Australian English, role of migrants in language change)

In *The social stratification of English in New York City*, Labov contrasted two approaches to the study of social variation in language (1966:209–11). The first approach, which we will call *social grouping*, assigns speakers to analytically defined social groups according to demographic or sociological criteria and uses group membership (or any of its defining criteria) as an explanatory variable in the statistical analysis of linguistic behaviour; for example, by first averaging values of linguistic variables across the speakers in each group and then comparing averages across groups. In the second approach, which we will call *linguistic grouping*, the overall distribution of the linguistic variables is first determined

and then the speakers exhibiting each pattern of linguistic behaviour are identified as to their social characteristics. The terms social and linguistic grouping do not mean that sociological considerations predominate in one approach and linguistic concerns in the other, but only refer to the temporal order in which they enter into the statistical analysis.

There is every reason to believe that both approaches will converge to essentially the same account of the variation given an adequate social description and an adequate linguistic description. Nevertheless, although he does not reject either approach and in fact uses both approaches in the New York City study, Labov does present some arguments for preferring social grouping to linguistic grouping, and it is this approach that has become more widespread in sociolinguistic studies.

In this paper we would like to reassess these analytical options, to investigate some of the advantages of the second approach, and to illustrate linguistic grouping by using data taken from a survey of vowel variation in Sydney. Where the social grouping approach, with its a priori identification of the possible explanatory variables, is methodologically most compatible with multiple regression or analysis of variance types of statistical analysis (as in "variable rule" methods), the linguistic grouping approach, based in the first instance only on a comparison of linguistic behaviour, is better adapted to methods for the statistical analysis of pairwise similarities and differences among individuals, namely those drawn from classification theory and cluster analysis. In this paper we make use of a basic method of this type, principal components analysis. Principal components analysis not only gives the distribution of the linguistic variables and identifies speakers associated with that distribution, but also handles a large number of variables more naturally and easily than the regressions used by quantitative sociolinguists. Since this analytical technique may be new to linguists, a fairly detailed though nontechnical description will be given.

SOCIAL GROUPING

We begin by reviewing some of the advantages of social grouping. Labov says this approach "seems more fundamental and more closely tied to the genesis of linguistic differentiation" (1966:210). Few would dispute that linguistic variation is but a part of broader processes of social differentiation and generally reflects social and demographic divisions rather than causing them. Thus regression models, with the independent variables being social ones and the dependent variables linguistic, seem eminently natural. Labov also points out that this approach would "avoid any error which would arise from assuming that a group of people who speak alike is a fundamental unit of social behavior" (1966:211). Finally, the actual methodology associated with social grouping, group averages

DELIMITING THE SYDNEY SPEECH COMMUNITY

in the first instance, refined to regression or correlation in more detailed analysis, is straightforward and relatively easy to carry out and to interpret.

The advantages are not universally pertinent, however. Were social and linguistic categories always well-defined in advance of the analysis, were there enough individuals and not too much interaction among the social categories, and were the type of mathematical form of the relationships (e.g., linear or logistic linear) of the connection between social and linguistic variables known in advance, then the regression approach based on social grouping would indeed be the only logical choice. The first two conditions tend to break down all too often, as we will discuss in the ensuing sections. And based on present knowledge, the third condition cannot really be satisfied either.

The best articulated model for the genesis of social conditioning in linguistic variation has been worked out over the years by Labov in New York City and Philadelphia. The key role here is played by the adolescent peer group as the social locus of dialect differentiation. It is at this age and in this social context that the subtle sociolinguistic patterning is developed. (See, for instance, Labov's discussions of differences between native and nonnative New Yorkers [1966:174-78].)

Unfortunately, aside from the ethnolinguistic tour de force represented by the Harlem study carried out by Labov's own team, little quantitative empirical characterization of the role of the adolescent peer group is available (see Cheshire 1982; Poplack 1979). And though regression is a useful tool in detecting and measuring the linguistic correlates of sociodemographic factors, it does not at this time represent a formal mathematical model of sociolinguistic differentiation. (Indeed, it is far more adequate as a model of grammatical contextual conditioning of linguistic variables than it is of social conditioning.)

LINGUISTIC GROUPING

The priority of the linguistic grid

In analyzing quantitative sociolinguistic data, an important goal, and one difficult to achieve, is the elaboration of the most meaningful and revealing linguistic categories possible for the classification of the data tokens. Linguists' understandable preoccupation with this goal means that the *linguistic* categories are generally established prior to final statistical analysis. Of course, there is often feedback among preliminary statistical assessments of the data, the actual coding of a number of data tokens, and the choice of relevant linguistic distinctions to be used in the final analysis. The conceptual ordering of linguistic categorization before statistical analysis, however, in no way implies that all results of strictly linguistic interest must be discovered or predetermined before the statistical analysis. On the contrary, strong quantitative co-occurrence relationships among

cross-cutting linguistic categories, the degree or lack of conditioning of one form or structure by another, and the hierarchical order of members of a paradigm according to the strength of their interactions with the linguistic context, are all types of "purely linguistic" facts which are revealed only through statistical analysis. The point is simply that some sufficiently fine analytical grid, based on phonological or syntactic concepts that are theoretically well-defined, is usually a prerequisite to numerical manipulation.

Problems with extralinguistic categories

The extralinguistic categories most frequently used by linguists are socioeconomic class, age, gender, education, and race or ethnicity. Early work by Labov (1966, 1969), Cedergren (1973), and by many others since has clearly demonstrated the insights into language change processes to be obtained in trying to correlate a variety of speaker characteristics with the frequency of usage of linguistic variants. Some recourse to these sociological or demographic categories is obviously necessary from the outset in order to ensure that a sample of speakers is chosen which taps the whole range of possible variation in the speech community or subgroup under study. In contrast to the linguistic categories, however, there is less certainty about the types of sociological categories that are pertinent in classifying speakers' linguistic behaviour and concern about whether or not the sociological categories are well-defined. In addition, there is also the possibility that there will be no sociological studies available for the speech community to be studied, as was the case when the survey of Sydney was started.

We have only to do a cursory examination of socioeconomic class to know that often the sociological categories as well as linguists' use of such categories are open to question. The concept of socioeconomic class or status is certainly debated within sociology, political economy, and related disciplines. Although much use has been made of stratificational or pluralist sociological methods in assigning speakers to "classes" or to positions on a socioeconomic scale according to some combination of status, income, occupation, and education, these may be only indirectly relevant to a speaker's position with respect to the relations of production, which Marxists, at least, would claim to be more fundamental determinants of attitude and behaviour – if one less readily operationalized. Moreover, sociological scales and indices can only be applied in a rather arbitrary way to large segments of the population. Women, for instance, present a difficult problem. Generally, they are assigned the class of their husbands, but sometimes their class is determined by their own occupations. Children at some indeterminate point are no longer assigned to their parents' socioeconomic class. Assigning class to upwardly or downwardly mobile speakers who have, in effect, changed their class positions is yet another problem. Another is the nontransferability of "objectivized" scales and categories among different social formations. Not only do class-language relationships differ among the large urban areas studied to date in England, the United States, and Quebec, but these studies are

now being reproduced in rural communities and in preindustrialized and industrializing countries, where the relations of production are quite different, at least on the local level, and so may be their sociolinguistic reflections. It was for such reasons that efforts were made to adapt Bourdieu and Boltanski's (1975) notion of the linguistic marketplace (Sankoff & Laberge 1978), but even this does not guarantee that speakers who are grouped together will exhibit similar linguistic behaviour nor, as sociolinguistic research has made increasingly clear, will any method based on purely extralinguistic criteria rather than performance itself.

These problems exist not only with socioeconomic groupings but also with groupings based on education, ethnicity, or even demographic factors such as age, gender, and geographical location. The years of schooling a speaker has had may have different significance for his or her linguistic behaviour depending on whether the schooling took place in the inner city or in suburbia (in the United States), in State, Catholic, or private schools (in Australia), and pre- or post-1960 (in Quebec – see Kemp 1981). In migrant situations, ethnicity may continue to be pertinent to speech patterns for a number of generations (Labov 1966; Laferriere 1979) or may become irrelevant with the first generation of native-born speakers (Labov 1972:281). Where speaker's sex or age are factors, these may interact strongly with each other or with other factors in their effects on linguistic performance. Establishing distinctions among age groups can become a problem when speakers are grouped into categories such as "children," "adolescents," "young adults," and so forth. These are quite likely to be culturally/historically specific categories, not universal, ahistoric ones.

There are basically two types of problem in all of this. One has to do with the very different nature of extralinguistic factors when compared with linguistic factors. It is generally less clear at the outset of an analysis which extralinguistic factors may be pertinent in differentiating among patterns of linguistic performance than is the case for intralinguistic factors. There is generally no natural and easily identifiable set of discrete distinctions or levels within an extralinguistic factor as there is for phonological or syntactic factors. There are not enough speakers sampled (a few dozen or a hundred) to allow a systematic process of feedback between statistical analysis and sociological categorization, in comparison with the hundreds of thousands of linguistic contexts generally available for statistical comparison. Finally, where intralinguistic factors characteristically show little statistical interaction in their effects on linguistic behaviour, extralinguistic factors often exhibit a great deal, and while it may be detected and measured, small sample size precludes its sociological interpretation in any conclusive way. All this means that the choice and definition of a set of socioeconomic and demographic factors and distinctions involves a considerable degree of arbitrariness, which might obscure the real basis of linguistic differentiation within the speech community, and which, because of small speaker sample size, may not be easily traced or remedied in the course of statistical processing.

The second type of problem stems from the inherently less deterministic type of conditioning or predictability imposed by extralinguistic factors. It is not uncommon for one linguistic element to categorically require or categorically preclude the presence of another; it is rare that a sociological subcategory, within a speech community, will be categorically associated with the presence or absence of a linguistic entity.¹ Furthermore, whereas for any linguistic context, if enough data are collected (from a single speaker), the results will eventually fall in line, quantitatively speaking, with the pattern of related contexts. No amount of extralinguistic conditioning can completely determine the behaviour of a speaker in comparison with the rest of the community. There are always a considerable degree of individual variation and usually a number of speakers whose behaviour diverges completely from what their extralinguistic parameters would predict. These facts result from the common practice of grouping speakers according to extralinguistic criteria and aggregating data within each group for purposes of statistical analysis (social grouping). Since the speech of individuals evinces many discrete and categorical patterns, which reflect the basic structures of their individual grammars, and since different speakers within the same community may differ in the details of these patterns, inappropriate aggregation of tokens from speakers according to their sociodemographic characteristics might involve the grouping of data generated by qualitatively distinct grammars. The statistical analysis would then result in a blurring of the discrete differences among a number of individual systems, producing an overall "average" grammar representative of no one speaker and devoid of any of the strong co-occurrence restrictions proper to their individual competences.

Avoid data aggregation

When Labov chose to aggregate individuals' linguistic behaviour, his purpose was, as he said, to avoid errors that might arise from the assumption that people who spoke alike were a fundamental unit of social behaviour. In the discussion thus far we have tried to detail some of the difficulties associated with identifying that fundamental unit of social behavior. We would like to consider now analytical techniques which do not, at least initially, require aggregating individual data but which first rely on individual data to group people whose linguistic behaviour is similar.

Though in recent years both types of problem just discussed in connection with extralinguistic factors have been shown empirically to be more potential than real, there has been a widespread tendency to carry out as much of the statistical analysis of sociolinguistic data as possible without aggregating the tokens from individual speakers. The "sociological" part of the analysis becomes a second step, after the linguistic details have been worked out.

For example, in the multiple regression analysis of linguistic variation data, tokens of each variant of a variable are initially classified according to features present in the phonological, syntactic, and/or lexical context. The data from each

DELIMITING THE SYDNEY SPEECH COMMUNITY

speaker may then be analyzed separately to estimate the effect of each feature in favoring one or the other variant (e.g., Guy 1980). Alternatively, under the assumption that speakers share the same linguistic constraints (feature effects) but differ by their degree of overall preference for one variant over another, the data from all speakers may be used together in the feature effect estimation, with the addition of one extra parameter per speaker to represent their individual overall preferences (e.g., Sankoff & Thibault 1977; Kemp 1979; Laberge 1980).

In both cases, the sociological component of speaker comparison is carried out using the previously calculated linguistic constraints, by comparing speaker parameters with sociological measures either informally or through a second regression analysis (e.g., Sankoff & Labov 1979: Fig. 2). Though it is thus postponed to a second step, recourse is still made to predefined categories for grouping speakers.

Still another approach, intermediate between an entirely separate set of feature effects per individual speaker and only one distinct parameter per individual speaker, allows for the existence of only a limited number of distinct sets of feature effects within the community, without specifying a priori which speakers share a similar pattern (Rousseau & Sankoff 1978). In this approach, the statistical analysis itself produces well-defined groupings of speakers according to similarity in their context-determined choice of variants.

This idea of using individual's data to group people whose linguistic behaviour is similar, and only after these groups are constituted to determine whether they consist of individuals with any social characteristics in common, is also a motivation for using principal components analysis on linguistic variation data (e.g., Poplack 1979: Ch. 6).

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis is a computer-assisted analytic technique that can be used to address the problems associated with linguistic grouping versus social grouping. It is a tool widely used in the social sciences for structuring large data sets into interpretable patterns and is available in all the major statistics packages (e.g., the Statistical Package for the Social Sciences [SPSS]). When used in linguistic studies, only individual linguistic data are used as input. It is very difficult to explain how principal components analysis works without resorting to a technical description that would be inappropriate here; furthermore, such descriptions are readily available (e.g., Taylor 1977). Our approach will be first of all to give a short conceptual description and then to demonstrate the use of the technique in an analysis of the sociolinguistic variation of five vowel variables in Sydney English.

Principal components analysis is the simplest of the data reduction methods (some others are factor analysis, multidimensional scaling, and analysis of correspondences) which enables us to visualize the major relationships in a multivari-

ate data set. It is normally used when the data set consists of ten or more variables and can be used with more than 100 variables. To begin to conceive of what principal components analysis does, we will start with the familiar notion of a scattergram. This is usually two dimensional, the two axes often being, in sociolinguistics, social class and an index for some linguistic feature. Each speaker is assigned a point on the scattergram, whose coordinates are determined by his/her values on the two axes. Patterns often emerge which can be described and interpreted in terms of some hypothesis about the relationship between language and social structure.

At this point we should ask why any two axes are chosen for such scattergrams in the first place. The analyst may have graphed out a whole range of different variables and rejected many for which no interpretable patterns emerged, but it is more likely that the theory within which the linguist is working suggests the axes that would be most likely to reveal important patterns. However, one cannot be sure that another pair of axes would not have proved more insightful.

Principal components analysis is a technique which allows one to arrive at the most meaningful axes in a mathematically well-defined sense. It will determine not just two but a number of axes and also arrange these axes in a hierarchy which indicates which ones are of most importance in explaining the patterns. The determination of the axes is a purely mathematical one, based on the correlations among the linguistic variables only. The success of such an approach to data description lies in the analyst's being able to name the axes, once they have been found by the mathematical procedure. In other words, the same theory that "suggested" the axes to use in the scattergram example is employed in interpreting the axes arrived at mathematically. What the principal components program does is to systematically determine all the relevant axes and point out the axes which are most important in accounting for the variance in the data set. Since these axes are all arrived at using only linguistic data, it will allow us to examine the question whether people who speak alike have any social characteristics in common rather than making the assumption that because they have some social characteristics in common, such as social class, gender, or ethnicity, they speak alike.

The precise way the technique does this is too complex to go into here. Briefly, it analyzes the multidimensional space whose coordinate axes are the variables of the study (a twenty-dimensional space in the present case). For the same space, any number of sets of twenty axes may be defined. The program finds one such set which has several important properties. The axis which is called the first principal component is the line through the space which comes as close to all the data points as possible, thus it accounts for the largest proportion possible of the variance in the data. The second principal component is that line through the space perpendicular to the first, accounting for the largest amount of the remaining variance, and so on. Each variable is given a loading on each principal component; one principal component can be plotted against another,

DELIMITING THE SYDNEY SPEECH COMMUNITY

revealing the relationship among the variables. In addition, each speaker is given a component score on each of the principal components and then two-dimensional graphs can be plotted. Such a process leads to the identification of the linguistic groups we are after. In the case to be described here, four principal components were identified, yielding six graphs. The procedure followed for interpreting the graphs was first to outline clusters of speakers (linguistic groups) and then to examine the original data and describe both the linguistic behaviour and social characteristics of the speakers in the linguistic groups in order to name the four principal components.

We will leave further explanation of the technique to the demonstration of how it is used.

THE SYDNEY STUDY

We applied principal components analysis to data taken from a sociolinguistic survey of English conducted in Sydney, Australia (Horvath 1985). The collection and analysis of the data were accomplished in a manner that has become quite standard in such surveys, following closely the work of Labov, Shuy, Wolfram, and Fasold. Before the second world war, Sydney was a fairly homogeneous society of Anglo-Celtic origins, in which most people spoke English as a native language. Since the war, however, there has been migration from many parts of the world with the earliest and now largest groups coming from Greece and Italy. One could characterize the migrant population linguistically as follows: bilingual adults approximately forty to fifty years of age who probably learned English around age twenty; a proportion of their children who are teenagers and who also may be bilingual but who probably learned English at a very early age, either from siblings or perhaps upon entering school.

The importance of ethnicity as a social correlate of linguistic differentiation and change was first brought to our attention by Labov in his New York City study. He notes: "Since the relations of Jews, Italians and Irish (and now Negroes and Puerto Ricans) have formed one of the principal themes of social dynamics in New York City, we can believe that such linguistic correlates can lead to generalized changes in the speech of the city as a whole" (Labov 1966:392).

Including ethnic diversity as part of a sociolinguistic sample of a speech community, however, is not without its problems, especially when the ethnic diversity is of such recent origin, as it is in Sydney. Previous studies of Australian English have eliminated speakers from migrant communities from the sample, not only prejudging the relationship of these people to the rest of the speech community but also being unable to assess their role in linguistic change. If ethnic adults are to be included, there is likely to be a wide range of abilities in English that will be netted in: Some speakers may have quite heavy accents, while those who arrived as children or who were perhaps born in Australia may have no accent at all. The decision was made in the Sydney study to cast a wide

net and then to devise a way of defining a speech community which would be able to indicate the relationship of the members of the ethnic groups, especially the adults, to that community.

The sample included 117 speakers from Anglo-Celtic and Italian background; other social characteristics represented were socioeconomic class, gender, and two age groups – teenagers and adults. Although there was some guidance in assigning socioeconomic class from Congalton's (1969) scale for Australia, the scale itself was certainly inadequate in its categorization of occupations to handle the diversity in the sample. In general, class was determined as follows: The division between upper and lower working class reflected that between skilled and unskilled labor. The middle class included professional occupations and managers. These categories, of course, were used for sampling and interpretation only, not for statistical analysis.

Several phonological variables were studied, but only four vowel variables (iy, ey, ow, ay) were included in the principal components analysis to be described here. Previous studies of Australian English (Mitchell 1946; Mitchell & Delbridge 1965; Bernard 1970) had identified variation in vowel articulation to be of major importance in structuring linguistic differences among Australian English speakers, although Mitchell and Delbridge found no social correlates of this linguistic variation. In addition, these earlier studies claim that particular phonetic variants of the vowels are systematically related to one another; these have been traditionally called Broad, General, and Cultivated varieties of Australian English. We have expanded this inventory to include a fourth phonetic category required to describe the data adequately; as a consequence we will refer now to Cultivated, General, Broad I, and Broad II. The phonetic values of these labels are given in Table 1. Whereas it is typical in Australian English studies for the speakers to be judged as Cultivated, General, or Broad, in this study each vowel was judged separately. Twenty instances of each vowel for each individual were coded into these categories.

The inclusion of speakers who learned English as young adults presents a problem for coding the vowel data. We are interested in the effects of the large influx of migrants into the Sydney speech community on Australian English and if we exclude these people, we may miss just those speakers who could provide the most insight into how a migrant community can affect the linguistic patterns of their new speech community. However, we are not interested in describing interlanguage varieties of English in this study. Hence, there is a fifth category, essentially unanalyzed, for variants of each of the vowels which might be described as accented English. In other words, they are variants of the vowels likely to be affected by interference from the first language and unlikely to have any effect on the development of Australian English. We have labelled this category Accented.

The input to the principal components analysis consisted of, for each speaker, the number of occurrences of each variant of the four vowels. Four principal

DELIMITING THE SYDNEY SPEECH COMMUNITY

TABLE 1. *Phonetic variants of the vowel variables*

Cultivated	General	Broad I	Broad II
(iy) [iɪ]	[əɪ]	[ə̃ɪ]	—
(ey) [ɛɪ]	[ʌɪ]	[ʌ̃ɪ]	[ãɪ]
(ow) [ʌ̃ʊ]	[ãʊ]	[aʊ]	[ãʊ]
(ay) [aɪ]	[ãɪ]	[ɔ̃ɪ]	[ɔɪ]

Note: Except for those variants labelled Broad II, the phonetic descriptions are based on Mitchell & Delbridge (1965) and Bernard (1970).

components were identified and speakers' component scores were plotted on a series of graphs. In what follows, the groupings will first be outlined, the linguistic behaviour of the speakers in those groups will be discussed, and the social characteristics associated with those groups will be identified. Since the graphs generally show the speakers to be arrayed along a continuum, the initial identification of the linguistic group will have to be somewhat arbitrary, as it must always be no matter what approach is taken. How useful the divisions are can be determined later when we look at the linguistic and social features associated with the groups outlined. For our purposes a group will be defined as a cluster of points (individual speakers).

In the first graph, Figure 1, where the first component is plotted on the horizontal and the second on the vertical, two clusters emerge, a crescent-shaped one and a moon-shaped one. The shapes themselves are unimportant; labels are used only as a matter of convenience. It is clear (and not unexpected since we know our sampling procedure has netted in some speakers who are not native speakers) that there are some individuals whose linguistic behaviour sets them apart from the rest. This graph leads us to expect to find at least two varieties of Sydney English which are fairly distinct. On the other hand, the two clusters may each have subgroups that could be identified, but if they do, we would expect the subvarieties to be much more closely related – differences of degree rather than kind.

A plot of the second and third components (Figure 2) spreads out the speakers in the crescent shape more and allows us to identify three subgroups.² These subgroups form more of a continuum, as one might expect to find within a speech community, indicating the likelihood of more overlap in the linguistic behaviour of the speakers. Only reference to the original data can show that for sure, however. Seven speakers are outside the crescent space.

Figure 3 is a graph with the first component on the horizontal and the third component on the vertical. Since the first component separated the two quite different speech groups, we will again concentrate on only the crescent-shaped group. A division into upper and lower clusters is indicated by heavy lines and a possible subdivision of the upper cluster is shown by the dotted line. Whether this latter division reveals any interesting patterns will be seen later.

THE SYDNEY SPEECH COMMUNITY

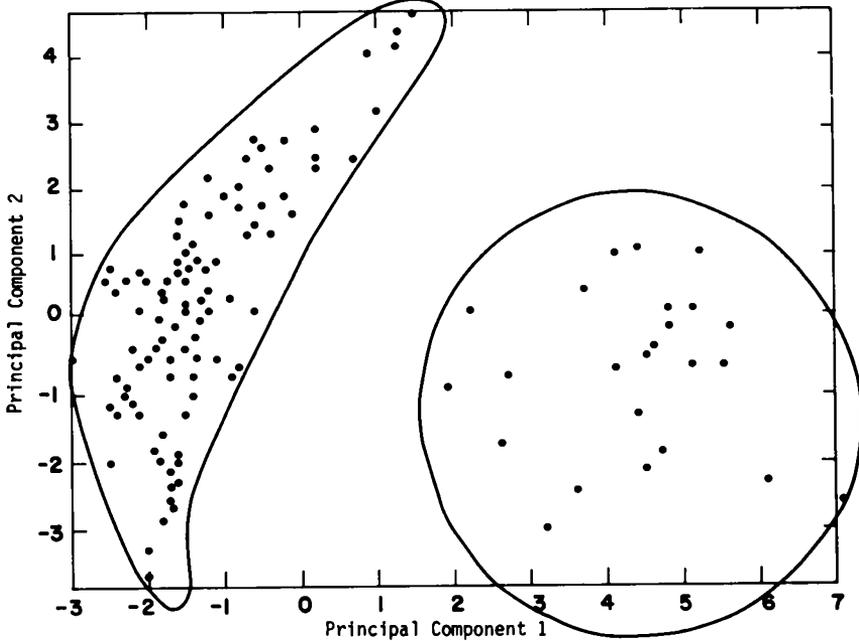


FIGURE 1: The Sydney speech community.

In Figure 4 the first component is again plotted horizontally but this time against the fourth component. This component appears to separate the moon shape into two clusters, one fairly tight cluster and the other rather scattered.

Given the graphs associated with these four principal components, we have identified several possible linguistic groupings:

1. a division of the speakers into two major groups with one, the crescent shape, having far more speakers than the moon-shaped cluster;
2. a division of the crescent shape into at least three subgroups;
3. another division of the crescent shape into at least three subgroups; and
4. a division of the moon shape into two groups.

A total of at least five and possibly as many as eleven distinct speech patterns have been identified, depending on the amount of cross-cutting of the divisions in Figures 2 and 3.

Table 2 contains the loadings of each of the twenty variables on the four principal components. This will be of help in interpreting Figures 1-4. Concentrating first of all on the moon-shaped group identified in Figures 1 and 4, we find that all of the twenty-four speakers in the cluster are Italian adults. This is substantiated in Table 2, where we see that all the accented variants are highly

DELIMITING THE SYDNEY SPEECH COMMUNITY

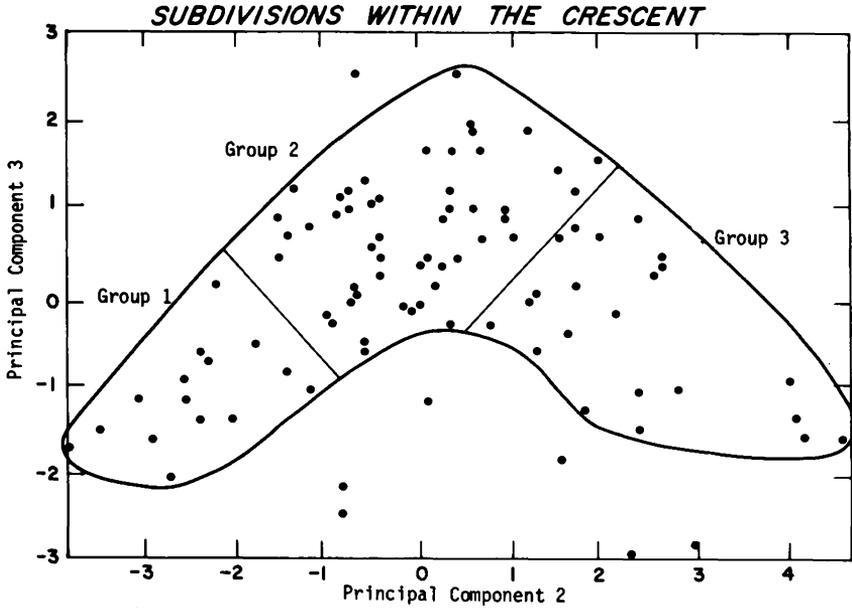


FIGURE 2: Subdivisions within the crescent shape.

associated with the first principle component, and to a lesser extent so are the Broad II variants. The variants which contrast most strongly with these are the General variants, the dominant form among the Anglo speakers.

The distribution of the vowel variants among the two Italian subgroups is given in Figure 5. We can now see that the two groups are primarily differentiated on the basis of the use of Broad II variants and in the use of Accented variants in all but (iy), where the groups are undifferentiated. This is confirmed in Table 2, where the Broad II forms are uniformly strongly associated with the fourth principal component, while the accented forms, except for the case of (iy), are strongly but negatively associated with it. Note that there is little relationship between most of the other forms with this component, and no systematic relationship contrasting Cultivated, General, and Broad I forms with respect to this component. We may then name the two subgroups of the moon shape the Broad II and Accented Varieties, bearing in mind that these names do not preclude the presence of large numbers of other variants of the vowels.

There appears to be no differential association between membership in one of these groups and either socioeconomic class or sex. Clearly, what is interesting about this group is the lack of internal patterning socially. There is, however, a dramatic age effect: All of the members of these groups are adults.

Turning to the crescent-shaped group, which we now know consists of all the

FURTHER SUBDIVISION IN THE CRESCENT

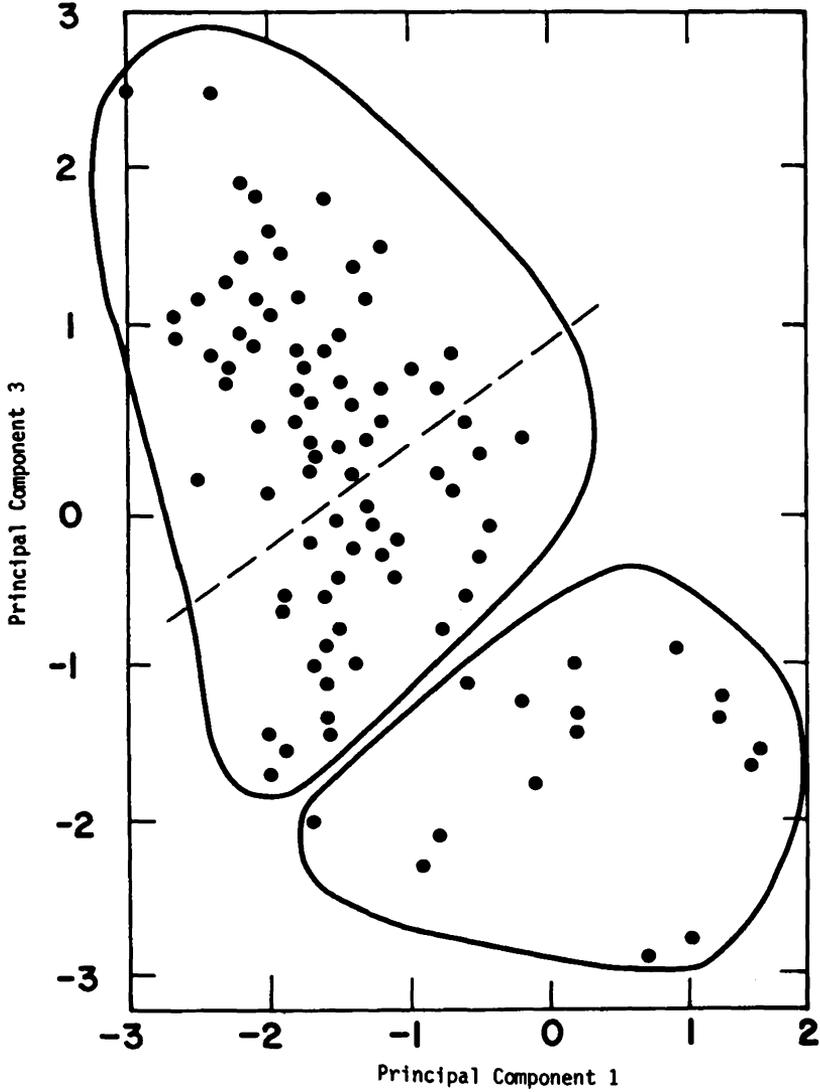


FIGURE 3: Further subdivisions within the crescent shape.

Anglo adults, three of the Italian adults, and all of the teenagers, we will try to describe the linguistic features associated with the three subgroups. One possibility is, of course, that these will simply reflect the tripartite division that was coded in the data: Cultivated, General, and Broad I. As Figure 6 shows, howev-

DELIMITING THE SYDNEY SPEECH COMMUNITY

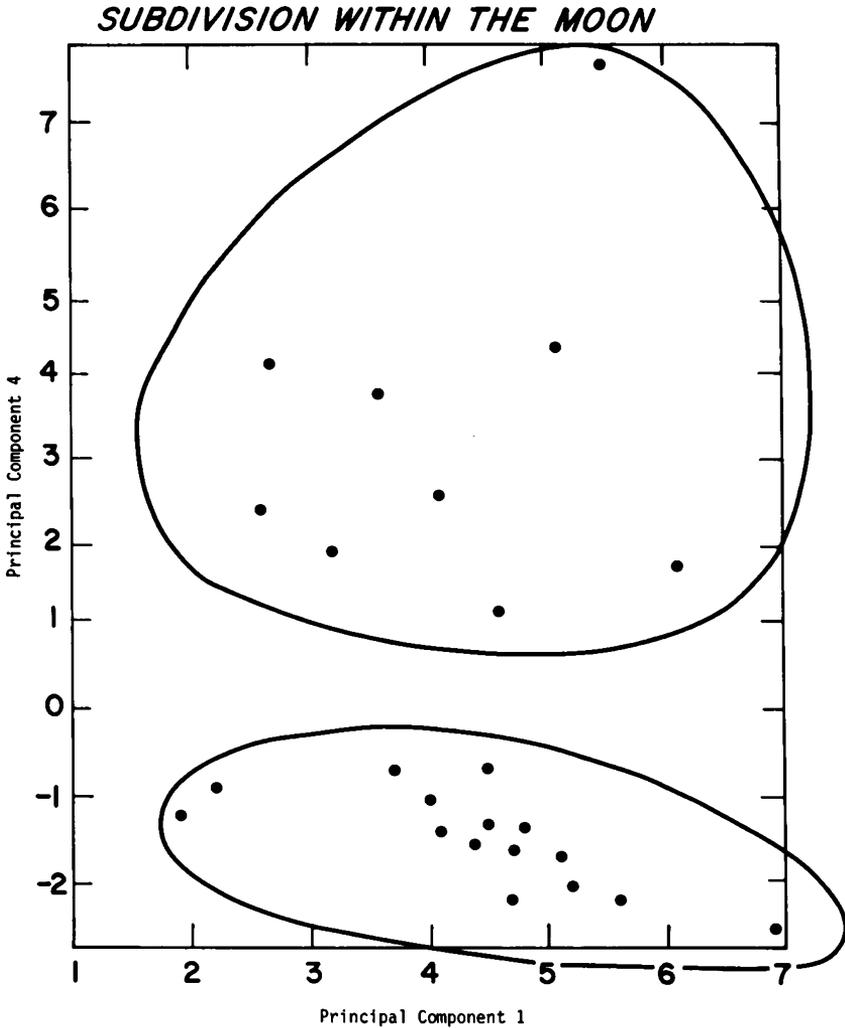


FIGURE 4: Subdivisions within the moon shape.

er, speakers in all three groups use all variants of the vowels; they are only differentiated by the relative proportions of Broad I, General, and Cultivated variants of the vowels in their speech. In Group 1, fewer than half (43.1%) of the vowels were classed as Broad I; for Group 2, 63.4 percent were General; and for Group 3, 50 percent of the vowels were classed as Cultivated. The three groups, then, can be labelled Broad I (Group 1), General (Group 2), and Cultivated (Group 3) but only on the understanding that these labels merely differentiate

TABLE 2. *Loadings of twenty vowel variants on the first four principal components*

Vowel	Variant	First component	Second component	Third component	Fourth component
(iy)	Accented	.35	-.16	.14	.06
	Cultivated	.09	.37	-.35	-.10
	General	-.30	.01	.28	.09
	Broad I	-.16	-.30	-.13	-.09
	Broad II	.08	.00	.04	.24
(ey)	Accented	.29	-.09	.27	-.26
	Cultivated	.09	.40	-.18	.04
	General	-.33	-.01	.21	-.02
	Broad I	.00	-.32	-.31	-.07
	Broad II	.15	-.12	-.01	.53
(ow)	Accented	.33	-.11	.21	-.15
	Cultivated	-.06	.33	.07	.02
	General	-.31	.07	.14	.02
	Broad I	-.17	-.20	-.38	-.20
	Broad II	.13	-.07	-.24	.48
(ay)	Accented	.30	-.14	.20	-.23
	Cultivated	.16	.38	-.21	-.12
	General	-.29	.08	.26	.19
	Broad I	-.16	-.33	-.29	-.04
	Broad II	.18	-.09	.01	.39

speakers quantitatively, not qualitatively. That is to say, speakers in the Cultivated group use more Cultivated variants of the vowels than speakers from the other two groups, General speakers use more General variants, and so on.

Figures 7a-h depict the distribution of the speakers in the crescent shape (along with seven marginal speakers who were not clustered, but who were closer to the crescent shape than to the moon shape) according to various social characteristics. The overall pattern indicates that over half of all speakers in the crescent shape are classed as General and the rest are distributed evenly between Broad I and Cultivated. Figure 7a identifies all the speakers by ethnicity; the most interesting difference between the Anglos and the Italians is the fact that few of the Italians are found in the Broad I group.

In Figure 7b the Anglo adults and Anglo teenagers are identified. Age appears to play some role in the distribution: Adults outnumber teenagers in the Broad I end of the crescent shape and at the most Cultivated end, whereas teenagers predominate in the General area. Notice in this figure, where only the Anglos are displayed, that the decision made earlier to divide the second from the third group is given some further justification. The space here is quite apparent, whereas when the Italian teenagers are included this is less evident.

By examining the adults and teenagers separately, we can get a better sense of

VOWEL VARIATION WITHIN MOON SUBGROUPS

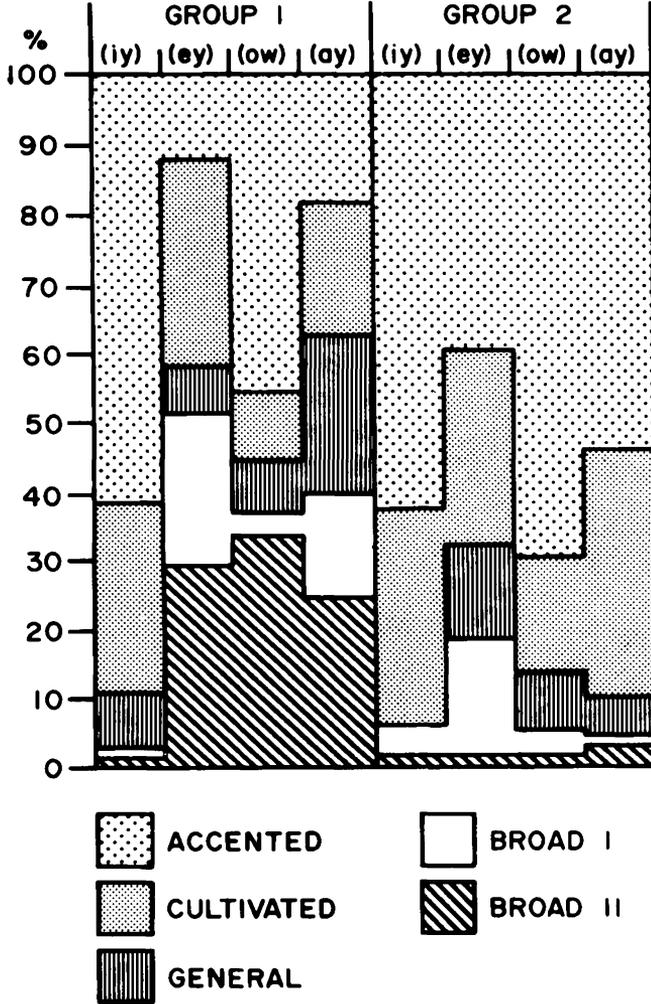


FIGURE 5: Vowel variation within moon subgroups.

the relationship of age, ethnicity, sex, and socioeconomic class to the linguistic groups. Figures 7c and 7d show the distribution of Anglo adults according to socioeconomic class and sex. Notice once again that the three-way division of the crescent shape is supported. There is a strong association between lower working class and Broad I (70%), upper working class and General (60%), and

VOWEL VARIATION WITHIN CRESCENT GROUPS

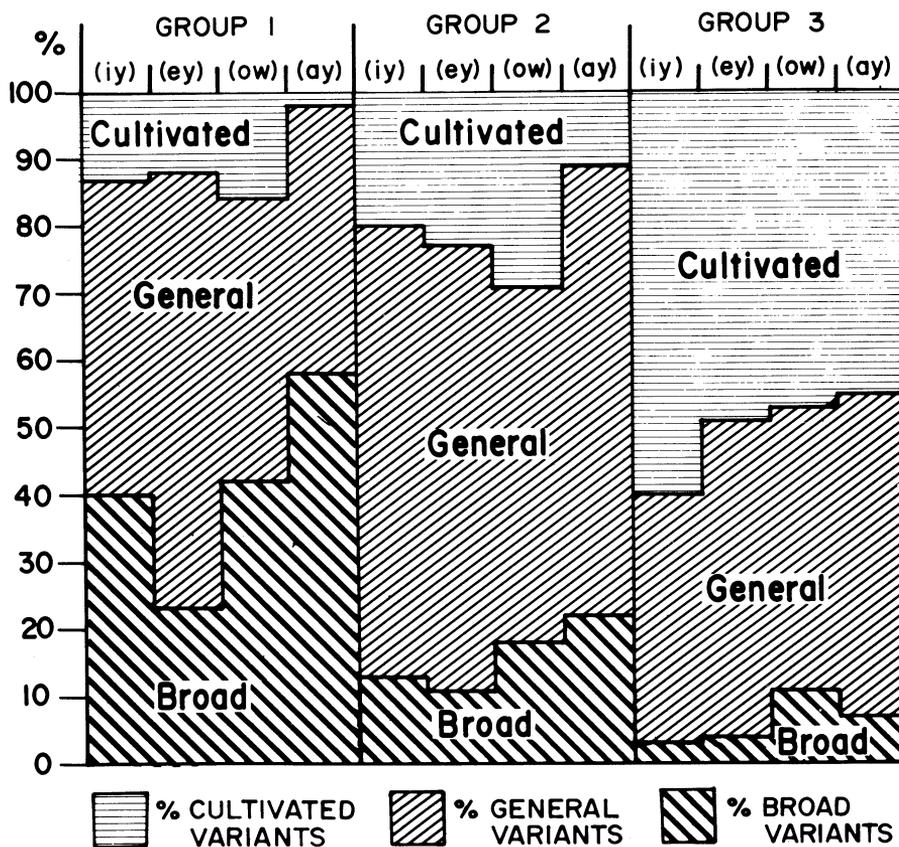


FIGURE 6: Vowel variation within crescent groups.

middle class and Cultivated (55%). The distribution of the speakers by sex is more striking; over half of the Anglo women adults are grouped as Cultivated, whereas only one male speaker falls within that group.

Anglo teenagers, displayed in Figures 7e and 7f, do not reproduce the adult patterns. Over half (62%) of all the Anglo teenage speakers from all three socioeconomic classes are classified as General, compared with just 38 percent of the Anglo adults, who were classed as General. If we look at the middle class teenagers, 80 percent of them fall in the General category. Sex also reveals some quite striking patterns: All but two males (85.7%) fall within the General and there are no male teenagers identified as Cultivated. Although the female teenagers are more evenly distributed across the groups, they also do not reproduce

DELIMITING THE SYDNEY SPEECH COMMUNITY

the adult female pattern except for the relatively low number in the Broad I. They do parallel the boys in that there is an increase in the number of speakers classified as General among the teenagers. For the males, there is a decrease in the number of Broad I speakers.

Figures 7g and 7h show how the Italian teenagers are distributed within the crescent shape. The most noticeable feature of this distribution is the fact that the speakers are even more highly concentrated toward the center, that is, they do not range along the whole continuum. In fact, looking first at the socioeconomic class distribution, we see that of the twenty-five speakers, only one (4%) is classed as Broad I and five (or only 20%) are classed as Cultivated. A total of 76 percent of the speakers are classed as General. The distribution by socioeconomic class is interesting because it does not follow the Anglo pattern, that is to say, the number of lower working class teenagers who speak the Cultivated variety equals that of the middle class speakers and the one speaker of Broad I is from the middle class. If we look at the Italian teenagers distributed by sex, we see that the one Broad I speaker is male and that the Italian males, unlike their Anglo peers, are represented in the Cultivated group.

The only pattern that has yet to be described is the one revealed in Figure 3, where the third component is plotted against the first. It is this third component that seems to divide the speech community into two or possibly three groups. In Figure 8 the age distribution has been indicated, revealing that the third component divides the upper group into generational varieties. If we divide this group in two at the dotted line, we find that the age distribution shows that the majority of the teenagers (66.7%) cluster above the dotted line and the adults are less clustered but the majority fall below the dotted line (72.4%). In examining the loadings on Table 2, we see that the third component strongly distinguishes the General from both the Broad and Cultivated forms. We have already seen the concentration of teenage Anglos and teenage Italians in the middle group of the crescent shape in Figures 7e-h, a group which has on the average higher scores on the third component than the other two groups in the crescent shape and, therefore, the contrast between the General variants and the Broad I and Cultivated variants on the third principal component is to be expected.

Thus we have used the principal components analysis to initially identify groups of speakers who behave linguistically alike, and we have then interpreted the clusters both linguistically and socially. We can now name the four principal components in linguistic and social terms as well as indicate the amount of variance accounted for by each of the components.

PRINCIPAL COMPONENT 1 (accounts for 31.9% of the variance)

Linguistic Interpretation: Australian English vs. Migrant Australian English. The Sydney speech community is like that described for most urban communities; its speakers can be arranged along a continuum. However, this speech community has a satellite speech community which, while sharing some charac-

BARBARA HORVATH AND DAVID SANKOFF

<p>ALL SPEAKERS IN THE CRESCENT (ETHNICITY)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>ANGLO (N=58)</td> <td>25.8</td> <td>50.0</td> <td>24.1</td> </tr> <tr> <td>ITALIAN (N=25)</td> <td>4.0</td> <td>76.0</td> <td>20.0</td> </tr> </tbody> </table> <p style="text-align: right;">7a</p>		BROAD	GENERAL	CULTIVATED	ANGLO (N=58)	25.8	50.0	24.1	ITALIAN (N=25)	4.0	76.0	20.0	<p>ANGLO ADULTS VS ANGLO TEENAGERS (N=58)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>ADULTS</td> <td>15.5</td> <td>19.0</td> <td>15.5</td> </tr> <tr> <td>TEENAGERS</td> <td>10.3</td> <td>31.0</td> <td>9.0</td> </tr> </tbody> </table> <p style="text-align: right;">7b</p>		BROAD	GENERAL	CULTIVATED	ADULTS	15.5	19.0	15.5	TEENAGERS	10.3	31.0	9.0				
	BROAD	GENERAL	CULTIVATED																										
ANGLO (N=58)	25.8	50.0	24.1																										
ITALIAN (N=25)	4.0	76.0	20.0																										
	BROAD	GENERAL	CULTIVATED																										
ADULTS	15.5	19.0	15.5																										
TEENAGERS	10.3	31.0	9.0																										
<p>ANGLO ADULTS (SOCIOECONOMIC CLASS)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>LOWER WORKING CLASS (N=10)</td> <td>70.0</td> <td>20.0</td> <td>10.0</td> </tr> <tr> <td>UPPER WORKING CLASS (N=10)</td> <td>10.0</td> <td>60.0</td> <td>30.0</td> </tr> <tr> <td>MIDDLE CLASS (N=9)</td> <td>11.1</td> <td>33.3</td> <td>55.5</td> </tr> </tbody> </table> <p>(% of those speakers identified in the sample as LW, UW, or M) 7c</p>		BROAD	GENERAL	CULTIVATED	LOWER WORKING CLASS (N=10)	70.0	20.0	10.0	UPPER WORKING CLASS (N=10)	10.0	60.0	30.0	MIDDLE CLASS (N=9)	11.1	33.3	55.5	<p>ANGLO ADULTS (SEX)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>MALES (N=14)</td> <td>35.7</td> <td>57.1</td> <td>7.1</td> </tr> <tr> <td>FEMALES (N=15)</td> <td>26.7</td> <td>20.0</td> <td>53.3</td> </tr> </tbody> </table> <p style="text-align: right;">7d</p>		BROAD	GENERAL	CULTIVATED	MALES (N=14)	35.7	57.1	7.1	FEMALES (N=15)	26.7	20.0	53.3
	BROAD	GENERAL	CULTIVATED																										
LOWER WORKING CLASS (N=10)	70.0	20.0	10.0																										
UPPER WORKING CLASS (N=10)	10.0	60.0	30.0																										
MIDDLE CLASS (N=9)	11.1	33.3	55.5																										
	BROAD	GENERAL	CULTIVATED																										
MALES (N=14)	35.7	57.1	7.1																										
FEMALES (N=15)	26.7	20.0	53.3																										
<p>ANGLO TEENAGERS (SOCIOECONOMIC CLASS)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>LOWER WORKING CLASS (N=9)</td> <td>33.3</td> <td>55.5</td> <td>11.1</td> </tr> <tr> <td>UPPER WORKING CLASS (N=10)</td> <td>30.0</td> <td>50.0</td> <td>20.0</td> </tr> <tr> <td>MIDDLE CLASS (N=10)</td> <td>0</td> <td>80.0</td> <td>20.0</td> </tr> </tbody> </table> <p style="text-align: right;">7e</p>		BROAD	GENERAL	CULTIVATED	LOWER WORKING CLASS (N=9)	33.3	55.5	11.1	UPPER WORKING CLASS (N=10)	30.0	50.0	20.0	MIDDLE CLASS (N=10)	0	80.0	20.0	<p>ANGLO TEENAGERS (SEX)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>MALES (N=14)</td> <td>14.3</td> <td>85.7</td> <td>0</td> </tr> <tr> <td>FEMALES (N=15)</td> <td>26.7</td> <td>40.0</td> <td>33.3</td> </tr> </tbody> </table> <p style="text-align: right;">7f</p>		BROAD	GENERAL	CULTIVATED	MALES (N=14)	14.3	85.7	0	FEMALES (N=15)	26.7	40.0	33.3
	BROAD	GENERAL	CULTIVATED																										
LOWER WORKING CLASS (N=9)	33.3	55.5	11.1																										
UPPER WORKING CLASS (N=10)	30.0	50.0	20.0																										
MIDDLE CLASS (N=10)	0	80.0	20.0																										
	BROAD	GENERAL	CULTIVATED																										
MALES (N=14)	14.3	85.7	0																										
FEMALES (N=15)	26.7	40.0	33.3																										
<p>ITALIAN TEENAGERS (SOCIOECONOMIC CLASS)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>LOWER WORKING CLASS (N=8)</td> <td>0</td> <td>75.0</td> <td>25.0</td> </tr> <tr> <td>UPPER WORKING CLASS (N=7)</td> <td>0</td> <td>85.7</td> <td>14.3</td> </tr> <tr> <td>MIDDLE CLASS (N=10)</td> <td>10.0</td> <td>70.0</td> <td>20.0</td> </tr> </tbody> </table> <p style="text-align: right;">7g</p>		BROAD	GENERAL	CULTIVATED	LOWER WORKING CLASS (N=8)	0	75.0	25.0	UPPER WORKING CLASS (N=7)	0	85.7	14.3	MIDDLE CLASS (N=10)	10.0	70.0	20.0	<p>ITALIAN TEENAGERS (SEX)</p> <table border="1"> <thead> <tr> <th></th> <th>BROAD</th> <th>GENERAL</th> <th>CULTIVATED</th> </tr> </thead> <tbody> <tr> <td>MALES (N=13)</td> <td>7.7</td> <td>76.9</td> <td>15.4</td> </tr> <tr> <td>FEMALES (N=12)</td> <td>0</td> <td>75.0</td> <td>25.0</td> </tr> </tbody> </table> <p style="text-align: right;">7h</p>		BROAD	GENERAL	CULTIVATED	MALES (N=13)	7.7	76.9	15.4	FEMALES (N=12)	0	75.0	25.0
	BROAD	GENERAL	CULTIVATED																										
LOWER WORKING CLASS (N=8)	0	75.0	25.0																										
UPPER WORKING CLASS (N=7)	0	85.7	14.3																										
MIDDLE CLASS (N=10)	10.0	70.0	20.0																										
	BROAD	GENERAL	CULTIVATED																										
MALES (N=13)	7.7	76.9	15.4																										
FEMALES (N=12)	0	75.0	25.0																										

FIGURE 7: Distribution of speakers within the crescent by social characteristics.

teristics of the main speech community, lacks the patterning, both linguistic and social, of the main speech community as well as having features in their speech which the main speech community does not have. This may well be a typical pattern for cities with large migrant populations.

Social Interpretation: Ethnicity. Although it is obvious that ethnicity is the most important factor in separating the main speech community and the satellite, ethnicity must be seen as a quite complex category. For instance, only bilingual adults who arrived in Sydney at approximately twenty years of age, not the second generation, are found in the satellite community. We might want to characterize these speakers as having a higher degree of ethnicity.

PRINCIPAL COMPONENT 2 (15.0% of the variance)

Linguistic Interpretation: Three Varieties of Australian English: Broad I, General, Cultivated. Although this component allows us to separate the speakers

DELIMITING THE SYDNEY SPEECH COMMUNITY

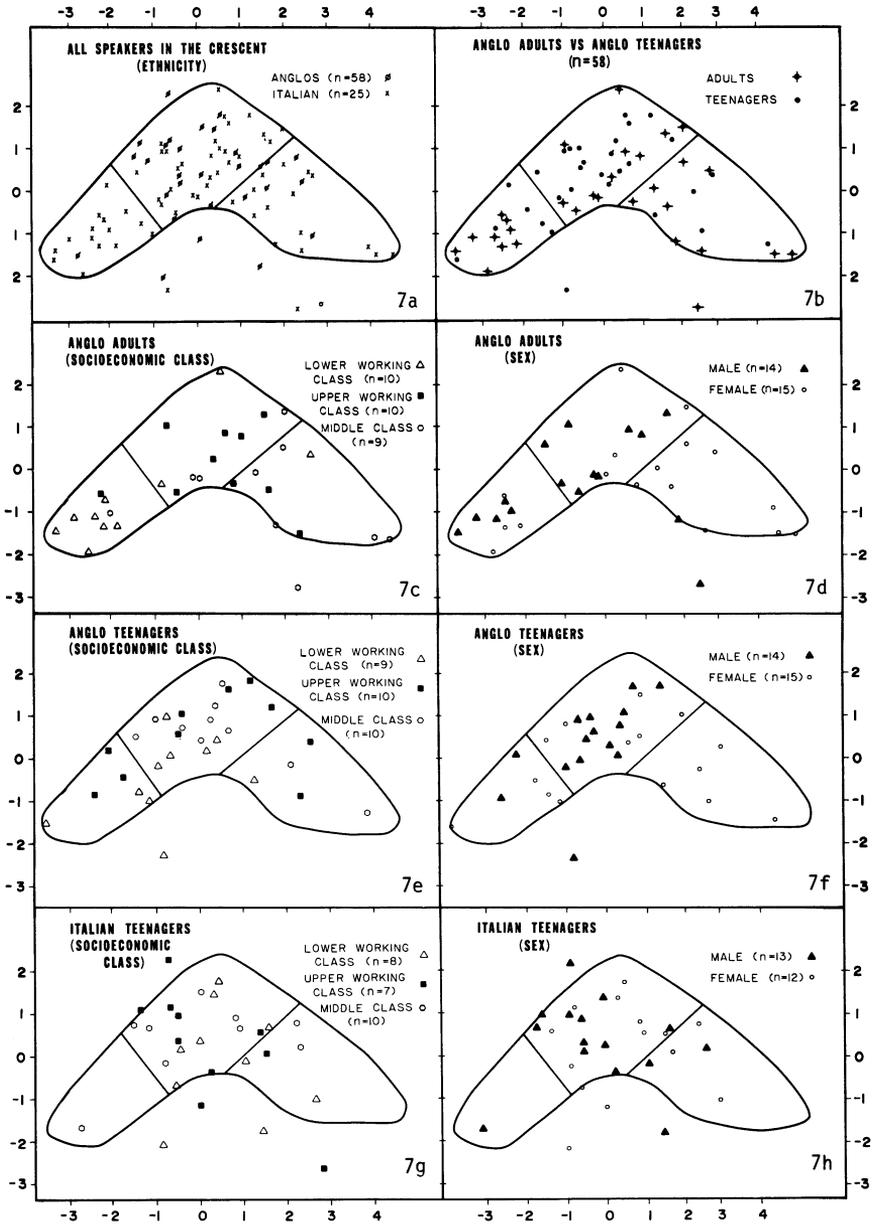


FIGURE 7: Distribution of speakers within the crescent by social characteristics.

AGE DISTRIBUTION IN THE CRESCENT

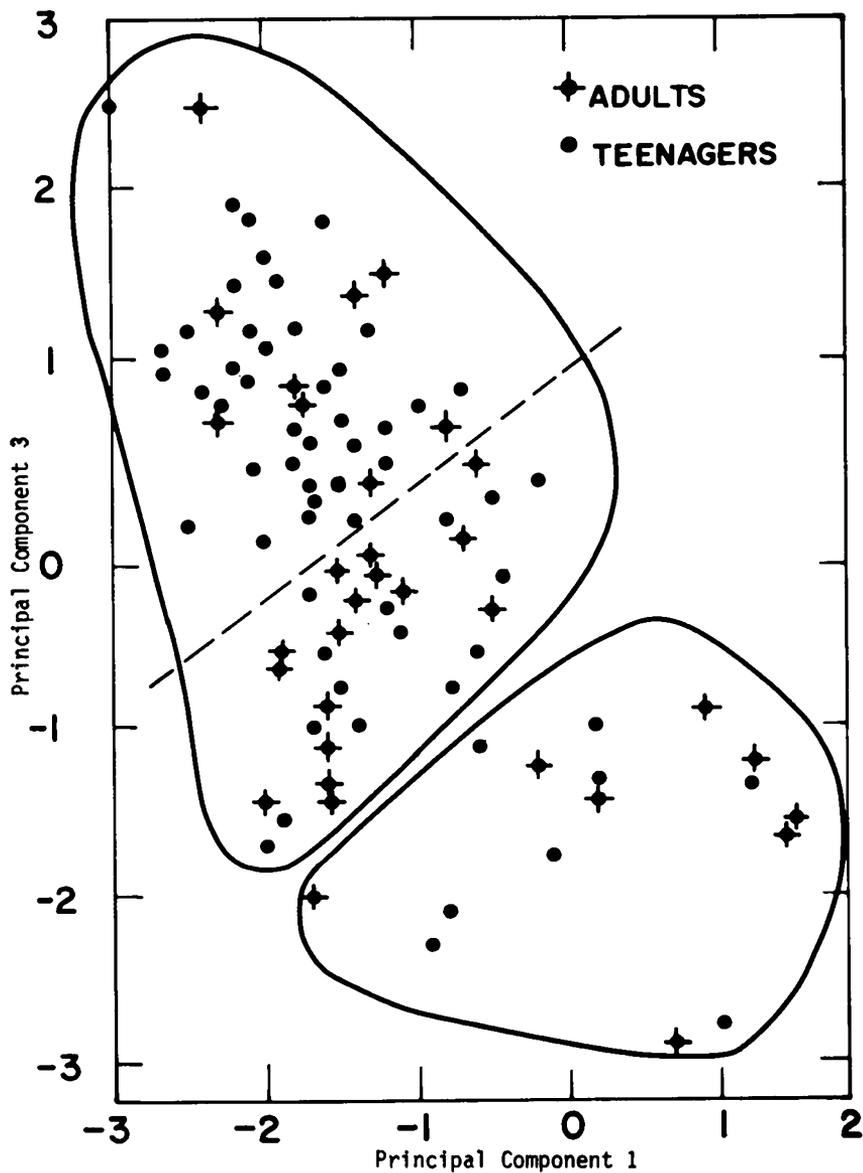


FIGURE 8: Age distribution in the crescent.

DELIMITING THE SYDNEY SPEECH COMMUNITY

into three groups, it is most important to recognize the high proportion of General vowel variants used by all three groups.

Social Interpretation: Sex and Socioeconomic Class. When viewed as the whole speech continuum, male/female differences by and large account for the range of the continuum; however, socioeconomic class also varies along this component.

PRINCIPAL COMPONENT 3 (8.6% of the variance)

Linguistic Interpretation: Generational Varieties. There are some important differences in the speech community when the time dimension is considered. This may represent a maturation sequence, in which teenagers have not yet adopted the socially distinctive traits of their elders. However, it may well be evidence of language change in progress, indicating that the General variants of the vowels are coming to be even more dominant than they have been in the past.

Social Interpretation: Age and Ethnicity. Although age is the critical characteristic here, ethnicity cannot be ignored. The fact that the Italian teenagers use more features of General than the Anglo teenagers, who in turn use more than Anglo adults, may indicate that the Italians are more advanced in the change than the Anglos. This is a tentative conclusion, however; more age groups need to be added to the sample before anything definite can be said about change in progress.

PRINCIPAL COMPONENT 4 (8.1% of the variance)

Linguistic Interpretation: Presence or Absence of Broad II Variants in the Satellite Community. The satellite community itself exhibits differences in the amount and type of Australian English features in their speech. The most striking difference has to do with the presence or almost complete absence of the Broad II variants. Those speakers without these vowel variants formed a much tighter cluster than did speakers having these variants in their speech.

Social Interpretation: Nature of Interaction with the Core Speech Community. Although this component is somewhat more difficult to identify, it is clear that neither socioeconomic class nor sex has anything to do with the differences in linguistic behaviour in the satellite community. Some other factors, perhaps degree of interaction with speakers in the core speech community or with speakers of particular varieties in the core speech community, are probably at work here. However, no pertinent information was collected in the interviews.

DISCUSSION

Although there are many further observations that could be made about Australian English, let us return instead to the central issue in this paper: the basis for grouping speakers in sociolinguistic studies of language variation. In all of the sociolinguistic surveys that have been done so far, the investigators have begun by identifying a language and a geographical region (English in the Lower East

Side of New York City, or Detroit, or Norwich; French in Montreal; Spanish in Panama) and then proceeded to design a sample that would in one way or another ensure that all varieties of the language would have a chance of being represented in the analysis. In many instances, decisions are made to exclude some speakers of the language, for example, in studies of Australian English, it has been common practice to exclude migrants and often the children of migrants from the sample based on the belief that these speakers are not part of the speech community. Exclusions of this type can affect the adequacy of the descriptions of the speech community; important varieties may be left undescribed and possible sources of language change in progress may go undetected.

Some critical delimitation of the community under study is usually necessary for practical reasons. However, whenever possible, it is important to start with the broadest possible definition of what constitutes membership in the speech community (i.e., a speaker speaks the language under investigation and lives in the region of concern). The linguistic analysis should then indicate whether or not particular speakers or particular groups of speakers are core or peripheral members of the speech community. As our analysis of Australian English indicates, it is closer to the truth to say that some speakers are more or less in or out of the speech community. Even the adults, some of whom spoke a fairly heavily accented variety of English, still shared features of Australian English. If, for instance, one were to do an analysis of Australian, American, and British varieties of English, it would be safe to say that the Italian adults would fall within or be more closely associated with the Australian varieties than to either the American or British. If one of the characteristics of a speech community is that speakers share the same social evaluation of speech, then the Sydney survey yielded some anecdotal evidence that the Italian adults knew what "proper" English (General-Cultivated) was and could be quite critical of "ocker" English (Broad I). Many claimed to watch only the ABC (the noncommercial television channel) because they wanted to learn "proper" English.

If, however, such problematic speakers are included in sociolinguistic surveys, there must be some means of determining whether any given speaker is central or peripheral to the main speech community. This is precisely what principal components analysis makes possible. Quite clearly, most but not all of the Italian adults are peripheral to the Sydney speech community.

But the importance of the principal components technique lies primarily in its usefulness in the identification of speakers who *do* fall clearly within the speech community and for allowing groups to be determined on linguistic behaviour, not social characteristics. The question that is central to sociolinguistics is not what kinds of social characteristics combine to differentiate groups in society, but what social characteristics are linguistically relevant. In the beginning of this paper, for instance, it was suggested that ethnicity might or might not be relevant; as it turns out, it is of utmost relevance in that it separates almost all the Italian adults from the rest of the speech community. However, the Italian

DELIMITING THE SYDNEY SPEECH COMMUNITY

teenagers are clearly part of the core speech community. But closer inspection reveals that they behave linguistically unlike Anglo teenagers, so ethnicity is still relevant for this generation.

Although the main advantage of the principal components analysis is that speakers are grouped according to linguistic criteria, a further advantage is that speakers are never "lost," since individuals are plotted on the graphic displays. It allows us to examine more closely those individuals whose linguistic behaviour is deviant. In the Sydney study, for instance, there are cases of speakers who fall outside of the speech community as a whole, as well as cases of speakers who are in the speech community but who are deviant with respect to other members of their social group. With the social grouping approach, these speakers are not so easily identified.

In summary, then, we have clearly distinguished five linguistic varieties, each of which can be characterized socially as well: the Broad I variety used mostly by adult, lower working class Anglo males; the Cultivated variety used largely by female Anglos of middle or upper working class; the General variety spoken by upper working class Anglo adults, the large majority of Anglo teenagers, and almost all the Italian teenagers; and two varieties used exclusively by Italian adults, these varieties being distinguished by whether or not Broad II variants of the vowels are used.

NOTES

1. Of course, geographical boundaries and ethnic boundaries, such as black and white divisions in many northern cities in the United States, can be associated with categorical differences among the speakers. In this case, sociolinguists usually carry out separate statistical analyses in the different subcommunities. Our discussion remains pertinent *within* each of the communities.
2. More than three groups would be delineated – there are perhaps as many as five. Once again, whether three or five is more useful can be determined by a closer look at the linguistic data themselves as well as a look at the social characteristics of the speakers.

REFERENCES

- Bernard, J. (1970). Toward the acoustic specification of Australian English. *Zeitschrift für Phonetik* 23(2/3):113–28.
- Bourdieu, P., & Boltanski, L. (1975). Le fetichisme de la langue. *Actes de la recherche en sciences sociales* 4:2–32.
- Cedergren, H. J. (1973). The interplay of social and linguistic factors in Panama. Ph.D. dissertation, Cornell University.
- Cheshire, J. (1982). *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.
- Congalton, A. A. (1969). *Status and prestige in Australia*. Melbourne: F. W. Cheshire.
- Fasold, R. W. (1972). *Tense marking in Black English: A linguistic and social analysis*. Washington, D.C.: Center for Applied Linguistics.
- Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (ed.), *Locating language in time and space*. New York: Academic. 1–36.
- Horvath, B. M. (1985). *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.

- Kemp, W. (1979). L'histoire récente de *ce que, qu'est-ce que, et qu'osque* à Montréal: trois variantes en interaction. In P. Thibault (ed.), *Le français parlé: études sociolinguistiques*. Edmonton: Linguistic Research. 53-74.
- (1981). Major sociolinguistic patterns in Montreal French. In D. Sankoff & H. Cedergren (eds.), *Variation omnibus*. Edmonton: Linguistic Research. 3-16.
- Laberge, S. (1980). The changing distribution of indeterminate pronouns in discourse. In R. W. Shuy & A. Shnukal (eds.), *Language use and the uses of language*. Washington, D.C.: Georgetown University Press. 76-87.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45:715-62.
- (1972). *Language in the inner city: Studies in the black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Laferriere, M. (1979). Ethnicity in phonological variation and change. *Language* 55:603-61.
- Mitchell, A. G. (1946). *The pronunciation of English in Australia*. Sydney: Angus and Robertson.
- Mitchell, A. G., & Delbridge, A. (1965). *The speech of Australian adolescents*. Sydney: Angus and Robertson.
- Poplack, S. (1979). Function and process in a variable phonology. Ph.D. dissertation, University of Pennsylvania.
- Rousseau, P., & Sankoff, D. (1978). A solution to the problem of grouping speakers. In D. Sankoff (ed.), *Linguistic variation: Models and methods*. New York: Academic. 97-117.
- Sankoff, D., & Laberge, S. (1978). The linguistic marketplace and the statistical explanation of variability. In D. Sankoff (ed.), *Linguistic variation: Models and methods*. New York: Academic. 239-66.
- Sankoff, D., & Labov, W. (1979). On the uses of the variable rule. *Language in Society* 8:189-222.
- Sankoff, D., & Sankoff, G. (1973). Sample survey methods and computer-assisted analysis in the study of grammatical variation. In R. Darnell (ed.), *Canadian languages in their social context*. Edmonton: Linguistic Research. 7-64.
- Sankoff, G., & Thibault, P. (1977). L'alternance entre les auxiliaires "avoir" et "être" en français parlé à Montréal. *Langue française* 34:81-108.
- Shuy, R. W., Wolfram, W. A., & Riley, W. I. (1968). *Field techniques in an urban language study*. Washington, D.C.: Center for Applied Linguistics.
- Taylor, P. J. (1977). *Quantitative methods in geography*. Boston: Houghton Mifflin.
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech*. Washington, D.C.: Center for Applied Linguistics.