

Social categories and sociolinguistics: applying a realist approach¹

ALISON SEALEY and BOB CARTER

Abstract

In a previous article (Carter and Sealey 2000) we have suggested that sociolinguistics can benefit from recent developments in social theory that are broadly termed "sociological realism." This rests on the claim that social reality is stratified, comprising not only the activities and interactions of human beings (agency) but also the structured social relations arising from these activities and interactions (structure), as well as the products of human consciousness or thought (culture). The present article develops these ideas to argue for a close collaboration between sociolinguistics and social theory, starting with the realist claim that empirical descriptions of the world are always incomplete, since the social world is not fully or directly intelligible to its inhabitants. The article explores some of the implications of this for concept formation and methodology in sociolinguistics, using as examples social categories such as "racial" and "ethnic" groups. It considers how sociolinguistic research can take account of actors' own understandings and mobilizations of such categories while also embedding these within a theoretically adequate description of the social world.

Introduction

Correlating linguistic variation as the dependent variable with independent variables such as linguistic environment, style or social categories, is the primary empirical task of sociolinguistics.

Chambers (1995: 17)

This paper will address several aspects of this claim, which is one not peculiar to Chambers and articulates a well-established position in the discipline. First, we wish to argue that the constitution of independent variables based on social categories is itself a task of theoretical

2. A. Sealey and B. Carter

description. Thus not only is the identification of such categories as "independent" problematic, but it also requires a defence of theoretical description itself. Second, we shall argue that the linear variable correlations often adduced in sociolinguistic research provide an insufficient description of social reality. Where such models are useful, we suggest, is in indicating empirically the "traces" of causal relations, and we shall explore this idea in the context of proposing how a realist epistemology can inform methodology in sociolinguistic research.

Of the "independent variables" listed by Chambers, it is the group identified as "social categories" that will be our main concern. As any introductory text on sociolinguistics will explain, the social categories that are most often the concern of sociolinguistic research include age, sex (or gender), social class, and ethnic or racial group. These categories, not surprisingly, coincide with those that are used in many other areas of research across the social sciences, but in recent years social scientists have begun to question their validity. One line of critique has come from those writing within a postmodernist tradition, where the challenge to a notion of objective social categories is part of a larger attack on any concept of objective truth and scientific rationality. Another source of objection is from those writers, influenced by social constructionism, who are wary of accepting *a priori* categories in the social world, preferring to explore how it — and everything in it — is constructed in texts and talk (e.g. Potter 2000).

The approach we are proposing may perhaps be seen as steering a middle course between the empiricist ontologies that underlie many conventional quantitative studies — if only implicitly — and the strong relativism implied in full-blown constructionism. While we do wish to defend a modest notion of objectivity in social research, we also recognize the limitations of empirical research based on correlations of variables, and we are critical of its tendency to hypostatize social categories and of its linear view of variation. Our perspective is consistent with that of a number of social scientists researching a range of issues in the social world, who have turned to realism and complexity theory to provide answers to the question of how we can know that the variables we name and use represent reality.

Social categories and theoretical description

One of our precepts is that the social world is not fully or directly intelligible to its inhabitants. Social categories such as "class" or "ethnic group" are not, of course, brute facts (Searle 1995). Take, for example,

the categories listed on the "equal opportunities" forms issued by employers to prospective employees. These categories are relatively recent theoretical descriptions that did not exist until they were devised by particular people for specific purposes. This is one reason why debates about ethnic monitoring have proved so durable (and inconclusive): they are about who has the power to define whom and for what purposes, and the resulting categories carry within them a network of theoretical propositions about what is salient for these purposes. For instance, the ethnic monitoring forms currently in widespread use by public service employers in England today distinguish ethnicity in terms of heterogeneous criteria: "White," "Black other" (skin color); "Indian," "Pakistani," "Bangladeshi," "Chinese" (belonging to a nation state — although it is unclear whether this is meant to imply affiliation to, birth in, or parents' birth in, one of these states); "Black African," "Black Caribbean" (skin color plus compound geographical regions). The process of monitoring employment patterns in relation to the responses selected from this list presents a number of problems: is membership of these groups ascribed by others or self-chosen? If people are given the opportunity not to answer, how valid can the exercise be as a measure of patterns of job selection? Since the categories are not of like kind, how can they serve as the basis for a single categorization? We shall return below to the theoretical incoherence of ethnic categories for the purposes of social research, but we introduce some of these problems here because they provide a particularly striking example of the fact that social categories cannot be regarded as given, independent variables to be deployed in the way that correlative models imply.

Of course, it is not only social categories based on ethnicity that are permeated by interconnected theoretical propositions. Another widely used social category in both research and policy is age. Legislation and social practices pertaining to the rights and responsibilities of people at specific ages vary widely, historically and geographically. (We are witnessing one such variation at the time of writing, when there is debate within the UK about whether or not the age of consent for homosexual intercourse should be the same as that for heterosexual intercourse.) Although there is an unavoidable physical, biological dimension to the ageing process, the significance attached to this is not determined by it. Attaining the age of fifty means many different things in different parts of the world and does not automatically entail any of them. In other words, all categories are theoretical descriptions, that is, they are constructed using terms derived from a coherent theory. In this important sense, theories are underdetermined by the empirical world: reality does not instruct us about what categories to employ in describing it.

Two kinds of social category

Few social researchers, including sociolinguists, would dispute the claim that the social categories deployed in research cannot be taken as presented to us by the social world. Our realist argument will start by drawing a distinction between two kinds of social category: those constituted by involuntary characteristics, and those characterized by some degree of choice on the part of the people who belong to them.

Following Greenwood's useful distinction, we shall refer to the first kind of category as "social aggregates" (Greenwood 1994). These are groups — such as the poor, the unemployed, women over 50 — whose only common feature is the property identified as salient by whoever is employing the category — such as those responsible for benefits policy or for selling insurance to over-50s. Aggregates do not imply shared conventions and norms to which people can be party. Of course, it is undeniably the case that being poor or unemployed may induce some common habits, forms of recreation, states of health, and so on, but this cannot be taken as a commitment to a shared culture of being poor or unemployed. And again, although the feature in question has an objective quality — lack of a job or lack of income or having lived for a certain number of years — the grouping of people having this feature in common is a consequence of a particular theoretical description, one that realizes some purpose in aggregating people in this particular way.

The extent to which the poor or the unemployed or those over 50 wish to identify with such a definition of themselves, the extent to which belonging to such an aggregate is seen as expressive of what they are, cannot be ascertained from membership of the category itself. There are no shared rules and conventions about how to belong to the culture of the poor or unemployed. Nevertheless, the objective features of people's situation that are the basis of their ascription to the category (low income or no job) will seriously constrain potential efforts to be described in terms of other categories (as "upper class," say, or as "wealthy").

The second approach to categorizing groups of people identified by Greenwood is with reference to their membership of "social collectives." These are descriptions of those groups in which members must be party to sets of conventions and norms. There are two implications that follow from the use of the term "party to." First it signals that membership of such groups is indicated by an awareness of, and therefore some kind of commitment to, the conventions that constitute the group in the first place. To use Greenwood's example, to belong to the group "married people" you must have understood certain rules about the institution of marriage, have performed a recognized ceremony, have a spouse, and

so on. Marriage is, properly speaking, a conventional arrangement; one needs some cognisance and understanding of what it is in order to undertake it (or to reject it). Being unemployed or over 50 are not conventional arrangements, although there may well be all manner of conventional ways of responding to the fact of being unemployed or over 50.

Second, the term "social collective" simultaneously avoids what Archer (1989) has termed "the myth of cultural integration." That is, we may employ the social category "married people" without reading back from it, as it were, that all those falling into this category share an *identical* commitment to the meaning of "being married." This conflationary impulse is particularly marked in the discussion of groups based on notions of race or ethnicity (Caner 2000) as we shall see.

Aggregates, collectives, and sociolinguistic research

Let us now consider the relevance of these two kinds of social category for research in sociolinguistics, taking age-based categories as an example. One well-researched topic is the occurrence of distinctive features in the speech of adolescents (e.g. Cheshire 1982; Eckert 1988; Kerswill 1996a; Romaine 1984). Studies on this topic usually start with the researcher identifying as salient the property of falling into a given age range, so from this perspective "adolescents" constitute an aggregate category. Sociolinguistic variables at the levels of phonology, lexis, or syntax are then analyzed, and correlations are established between membership of the category and the tendency to use one of the variants. These findings often point to significant patterns, but they remain *descriptions*, as opposed to *explanations*, of the correlations found. A finding that often emerges is of "adulthood [as] the focal point for linguistic innovation and change" (Chambers 1995: 176), but while the majority of speakers in the group studied display this tendency, there are invariably counter-examples. This points to the fact that there is no entailment (logical necessity) involved in belonging to the category "adolescent" and, for instance, using /t/ rather than /ʔ/ (a tendency that is found even when style and context are accounted for, Kerswill 1996b). It is at this point in the process of looking for explanations of sociolinguistic phenomena that the aggregate category "adolescent" ceases to be useful, precisely because it is simply an aggregate of contingent features.²

If we want to explain *why* many adolescents are in the vanguard of linguistic change, and why some are not, we need to introduce a strong notion of social agency — an acknowledgement that people have some degree of choice over what they do, including how they speak. This in turn

requires the use of social categories that recognize the relevance of actors' own understandings. Those findings that appear at first to be anomalies in the statistical correlations then become more explicable. The researcher must reconstitute the theoretical descriptions of social categories as collectives rather than aggregates. Adolescents who subscribe to the value systems of either "jocks" or "burnouts," for example (Eckert 1988), do so on the basis of their own decisions (even if there is not an infinite range of options from which they can choose, and even if they see these as the only options available).

As research produces descriptions that are based on collective (as opposed to aggregate) social categories, it reveals properties of the interaction between structure, culture, and agency. Thus, an investigation of why some young people use marked variants while others do not entails a recognition of the role of the speakers themselves in this variation. However, moving from an aggregative view of social categories presents the researcher with a different order of epistemological problem (which may be one reason why people frequently opt for an aggregative approach). The core of this problem is that the identification of collective categories involves the researcher in making judgements about who is "party to" the norms and rules that constitute the category in question.

Such judgments will be less difficult in some cases than in others. To return to Greenwood's example, cited earlier, it is reasonable to assume that those people who are married have some understanding of what this description means (although, as we also pointed out earlier, this by no means entails that they will all have the same understanding). Constituting "married people" as a collective category in this way does not preclude its employment as a social aggregate category. For instance, if we want to research the relationship between being married and having a mortgage, we may decide that a quantitative study is the most reliable way of discovering whether there is any sort of correlation at all between being married and having a mortgage. We are advocating methodological relativism here: which sort of category we employ will depend on what it is we are trying to find out.

Thus it should be clear that in defining aggregate categories, less attention is paid to actors' own understandings. The feature defining membership of an aggregate category must be one that is contingent, external, and measurable. You may see yourself as the youthful Adonis, but you've still been around for 50 years, and the researcher (or the insurance salesperson) will go by the birth certificate rather than rely on the personal PR. Analytically, then, with aggregate categories, priority is given to the analyst's own depiction of that category.

Concept formation: the case of "ethnic" groups

While some social categories can therefore be constituted by researchers as both social aggregates and social collectives, others are more troublesome. This is particularly the case when we are dealing with aspects of social reality that do not lend themselves readily to empirical measurement and where, moreover, actors' interpretations are critical. Adolescents may be defined aggregatively by the researcher as all individuals between the ages of 13 and 18. Researchers may take the decision to focus on this age group because of a belief (derived from previous research) that adolescents tend to use language in particular ways. However, using language in these ways is not simply an irresistible stage in "sociological maturation" (Kerswill 1996b); it involves a degree of choice. This is why actors' understandings are a central element in the theoretical description of social collectives. "Jocks" and "burnouts" are descriptions that rely on social actors' adoption, and maintenance, of distinct linguistic features. Their adoption of this form of social action suggests that the individual is not only "party to" the norms and rules relevant to being a member of the group but has also made some assessment of the potential benefits and losses attaching to group membership.

It is considerably more difficult to disentangle aggregate and collective categories when we use the theoretical description "ethnic groups." Ethnicity is not an empirically observable feature of human beings. How, then, are people to be placed in "ethnic" categories? If this is done by skin color (which is problematic as it is not a categorical variable), or place of parents' birth, "ethnic group" is being employed as an aggregate category. It therefore implies nothing about commitment to sets of norms or conventions (of language, dress, custom, or habit, although all or some of these may be observable among many members of a particular category). If people are classified into "ethnic groups" with reference to some notion of identity or culture, then the term is being employed as a collective category. The researcher then faces the familiar problem of making judgments about who is party to which sets of norms and conventions, and who is not.

We would claim that this problem arises because it is not possible to distinguish between ethnic groups as aggregate categories and ethnic groups as collective categories, and that this inevitably leads to confusion about what it is that is actually being measured.³ Researchers are then unlikely to be able to operationalize concepts with the precision they need. They may seek to sidestep the issue by simply adopting commonsense terms and definitions, or by deploying the self-definitions used by

their informants, but, as we argue below, such a move raises a number of problems for the analysis of any findings.

To take one recent example, Fought (1999), in her study of "Chicano English," is concerned with a sound change that is taking place in both the "minority" and "majority" communities of California. Her article classifies speakers with reference to ethnic categories — as it must do, given her commitment to "the need to analyze variation within the context of those social categories that are of particular significance to the specific community being studied" (1999: 5). These categories, like those from the UK discussed above, are of several kinds and are resistant to empirical verification. They include "speakers of Anglo ethnicity," "Anglos," "Anglo speakers," "the majority Anglo community," "the California Anglo community," "Puerto Ricans," "Puerto Rican speakers," "a Latino community," "Latino young adults," "Los Angeles Chicanos," "the minority Mexican-American community," "speakers of other ethnicities," "minority communities," "minority groups," "minority speakers," "African-Americans," "African-American local communities," "the black and white communities."

Our disagreement with Fought, let us be clear, is not about her study as a whole, but rather with the use of ethnic categories in this kind of research, of which her study is but one example. In order to establish correlations between use of a specific linguistic feature (in this case, /u/-fronting) and membership of a social group, the researcher identifies the group — and those groups with which it contrasts — *a priori*. Since the membership qualifications of "ethnic" categories are notoriously difficult to pin down, as we have argued, the boundaries of the categories to be compared are inherently unstable.

For speakers themselves, this analytical problem may be of very little significance. In order to go about one's social business on a daily basis, it is quite satisfactory to define oneself and others — when it is relevant for particular types of interaction — in terms such as "Anglo," "black," "Puerto Rican," or "Latino." And indeed such self-definitions are important for the analyst in interpreting such phenomena as strong identification with one social group — and the language associated with membership of it — rather than another. However, from our realist perspective, it is quite clear that there can be no logical necessity between belonging to such an imprecise category as "Anglo" or "Chicano" and using language in particular ways. This is not only because the category is inherently imprecise (Whose definition of it is to be authoritative? Which attribute[s] of the category are to count as definitive?), but also because speakers are not bound by their membership of any social category to use one set of language features rather than another. For us, the use of

social categories based on "ethnicity" is conflationary, in that it fails to distinguish between culture, structure, and agency.

Structure, culture, and agency

People are routinely classified by institutions, such as government agencies, as belonging to particular groups. There are clearly identifiable patterns in the way membership of these groups correlates with opportunities for employment, housing, health care, and so on. These patterns of structured social relations tend to be stable and persistent, providing the contextual social conditions into which individuals are born. Such conditions are, of course, logically antecedent and temporally prior to the individual, generating anterior distributions of resources within which individuals are involuntarily placed by the accident of birth. Being born poor in a capitalist society, or female in a sexist one, will profoundly affect your chances of one day becoming a millionaire or the Vice Chancellor of a major university.

These probabilities, however, are always conditioning rather than determining, and each individual has some degree of choice about how to respond to the contexts in which she finds herself (Layder 1997; Carter and Sealey 2000). One of the areas in which a degree of choice is exercised is in relation to the cultural resources that are a feature of the context of social existence — including language. Language is preeminent among these cultural resources, being not only a resource in its own right, but also the means of accessing these resources. Now there are clearly constraints on people's access to these linguistic resources. There are obvious difficulties for an individual who seeks to become a fluent speaker of a language not widely used in the locality; adopting the speech style of those who are widely seen as members of a different social category is likely to be viewed as a form of "crossing" (Rampton 1995), which may attract social censure.

However, such deviations from the supposed norms are possible, and we would argue that this is because language, as a cultural resource, exists in what Popper has identified as the "world 3" of human cultural resources. (For Popper, "world 1" comprises the physical world, "world 2" comprises the subjective conscious experiences emerging from the engagement of human consciousness with world 1, and world 3 comprises the objective products of human consciousness, embodied in libraries, books and other media — Popper 1972.) This tripartite model allows Popper to develop a crucial distinction between subjective knowledge — knowledge possessed by a knowing subject — and objective

knowledge — the logical contents of our theories, conjectures and guesses. Crucially for Popper, almost all our subjective knowledge (“world 2” knowledge) depends upon “world 3,” that is to say on (at least virtually) linguistically formulated theories (Popper 1972: 74).

World 3 is an emergent product of the engagement of human agency with the physical world, and it is through interaction between ourselves and world 3 that objective knowledge grows. It is this emergent quality that imparts to world 3 its autonomy. World 3 resources, such as theories, ideas, and propositions, are objective knowledge in the sense that they are “... totally independent of anybody’s claim to know” and are also “... independent of anybody’s belief, or disposition to assent, or to assert or to act. Knowledge in the objective sense is *knowledge without a knower*; it is *knowledge without a knowing subject*” (Popper 1972: 109). World 3 knowledge therefore does not have a color or a gender; it does not belong to one social class or another. For the very same reasons, it has an independence from individual users and its elements can be used by any individual or group possessing the appropriate symbolic code.

Furthermore, as is demonstrated particularly by corpus analysis, and patterns of colligation and collocation, language as a system constrains its speakers in various ways, pointing to its partial autonomy from its speakers. De Beaugrande (1999: 131) describes the interplay between language and agency thus:

The *standing constraints* persisting on the plane of the system (e.g. the English article going before the noun) interact with *emergent constraints* being only decided on the plane of the discourse (e.g. the lexical choices appropriate for a job interview).

Reducing language to simply what speakers say (as in various forms of constructionism) eliminates its world 3 structural features. Reducing language to its systemic, logical features (as in various forms of structuralism) eliminates its world 3 emergent features. These include the fact that language change is brought about by the engagement of human purpose with an independent and antecedent body of linguistic resources. In both cases the conflationary impulse results in the neglect of the interplay between linguistic and other cultural resources, and the processes by means of which human beings, in seeking to modify or maintain aspects of their social worlds, modify or maintain those resources. When this interplay is overlooked the outcome is often too close an association, in the researcher’s account, between language and its most typical speaker. Cultural coherence is then assumed by the researcher on the basis of typical speech use, and typical speech use is adduced as the basis for a claim for cultural coherence (Archer 1989).

A particularly striking example of this kind of conflation is provided in the various labels that have been assigned to a variety of American English, namely “Black English,” “African-American Vernacular English,” and, latterly, “Ebonics.” Wardhaugh (1992), for example, claims, “... we can be sure that there is such a variety of English as Black English in the United States” (1992: 326). In the preceding passage, he has described the linguistic features that contrast with standard English, providing empirical evidence for his claim that the variety exists. He continues, “Those who speak it recognize that what they speak is something different from the varieties employed by most non-blacks” (1992: 336–337). This sentence provides examples of two of the problems we have been discussing. First, the perceptions of the speakers themselves are presented as equivalent to the theoretical description of the variety by the linguist. As we have said, we believe speakers’ perceptions to be important sources of sociolinguistic evidence (cf. Preston 1998; Cameron 1995), but we would wish to distinguish speakers’ own accounts from those of the analyst, since the two kinds of description serve different purposes. The second problem is signalled by the need for Wardhaugh to qualify “non-blacks” by the word “most.” The passage continues,

Most Americans are prepared to categorize someone who contacts them by telephone as either black or non-black using speech alone as the criterion, and most such categorizations are correct. In cases of mis-categorization, it is usually because of special circumstances: a black person has been brought up very closely with non-blacks, or a non-black has been brought up very closely with blacks (1992: 337).

Now at this point, from our perspective, the alarm bells should begin to ring for the researcher, prompting a reevaluation of the categories “blacks” and “non-blacks.” If the variety really is “Black English,” then *all* those speakers and *only* those speakers who belong to the category “black” will use that variety. If, as both Wardhaugh and any number of studies and personal experiences will testify, there are exceptions in both directions (black speakers who don’t use the variety and non-black speakers who do), then it behoves the researcher to analytically detach the variety from the speakers, and to seek a more satisfactory theoretical description of the relationship between agency (speakers) and culture (language varieties).

Let us be clear about this. Our objection to Wardhaugh’s use of the term “Black English” is not that it imperfectly captures the empirical usage of the language form denoted by the term. Few descriptive terms will have a nil leakage of this sort (Bell, personal communication). But our reservations about “Black English” as a social category for research

purposes — social or linguistic — stem from the fact that it conflates aggregative features (the number of people who use this particular language variation) and collective features (the extent to which individuals using the variation are “party to” norms and conventions that define membership of that group). This conflation obscures the social contexts in which the mechanism of being dark-skinned in a color-racist society is triggered to produce the particular outcome in which certain distinctive language forms become expressive, and constitutive, of particular forms of social identification.

One significant study in which issues like these have been explicitly addressed is the research into Creoles in the Caribbean by Le Page and Tabouret-Keller (1985). These authors do distinguish between language and speaker, claiming, “We should constantly remind ourselves that languages do not do things; people do things; languages are abstractions from what people do” (1985: 188). Le Page and Tabouret-Keller found that their informants were likely to hold stereotypical views about the various “communities” found in their locality, and that their beliefs about both ethnic categories and language varieties were influential on their behavior. However, from the researchers’ point of view, these beliefs were highly problematic, for many of the reasons we have outlined above (such as the variable salience given by different informants to skin color, family descent, religious affiliation, and so on). Le Page and Tabouret-Keller’s account, in contrast with the traditional correlational studies that they criticize, emphasizes the role of agency:

the individual creates for himself the patterns of his linguistic behaviour so as to resemble those of the group or groups with which from time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished (1985: 181).

These “acts of identity” assume a high priority, and individual agency is paramount: “A community, its rules, and its language only exist in so far as its members perceive them to exist” (1985: 205). Now although we are sympathetic to several aspects of Le Page and Tabouret-Keller’s account of “acts of identity,” we would wish to take issue with this particular claim on several counts. First, it denies the world 3 status of language itself, and the properties pertaining to language as a world 3 object, namely its logical, grammatical, and lexical patterns and relations. Second, it subjects language to the tyranny of the present tense: structures of language become the effect of contemporary action and perception. It then becomes difficult, on both counts, to explore the interplay of language with agency, with the efforts of people to modify or maintain some aspect of social

reality. To reiterate, we regard it as essential both to recognize and to maintain a distinction between (a) social actors’ own purposes and choices from among linguistic varieties, and how they perceive these from their own experience;⁴ (b) the structural enablements and constraints that condition the choices open to them; and (c) the properties of language itself, as an element in the cultural system. It is to an elaboration of methodologies consistent with this that we turn in the concluding section.

Methodological implications

Our argument falls into three parts. First, the social world is not directly accessible to individuals: social categories, therefore, are unavoidably theoretical descriptions. That is, they are fashioned out of theoretical networks, and they are devised for particular purposes. The positing of social categories is not the exclusive preserve of researchers: the social categories that we all use in our everyday interactions are also theoretical descriptions (cf. Carter 2000). This makes problematic any efforts to define, for research purposes, social categories like “adolescents” — and especially those like “race” or “the Latino community” — in such a way that they can function as “independent variables.”

The recognition that all social categories are theoretical descriptions, or, as Pawson (1989: 287) puts it, that “all measurement is an act of translation,” also raises the question of how to discriminate between the often competing accounts provided by these descriptions. Our second claim is that social researchers must provide a defence — involving an explicit and conscious scrutiny — of the theoretical descriptions they employ. (We assume at the very least that researchers believe that what they produce offers a different understanding from the understandings of participants, as nonresearchers, of whatever aspect of the social world is being investigated.) In other words, they must at least claim some form of epistemic authority (irrespective of whether they wish to define this as “social science”; cf. Brumfit 1997). Theoretical descriptions necessarily draw on other theoretical descriptions, and these other descriptions, as world 3 entities, are necessarily independent of the people who deploy them, in the important sense that they did not produce them. As propositions about reality, these descriptions can be assessed in terms of their truth or falsity, their consistency, their evidence, and so forth. As descriptions produced by other minds, and upon which subsequent descriptions to some extent rest, they act as guarantors against the solipsistic fallacy that the reality we seek to describe is the product of

the theories we hold about that reality. This is the basis for our defence of a modest notion of objectivity within social science.

The realist epistemology we are advocating offers one way of defending a notion of social science without excluding the perspectives of the researched. This is because, and this is our third point, it is not methodologically prescriptive and always seeks to account for the interplay between structure, culture, and agency.

Of course we recognize that these ideas are by no means original to us. Sociolinguistic questions are necessarily (as our opening quotation signifies) questions about individual speakers, linguistic resources, and the wider society, including the groups to which people belong. Sociolinguistic researchers have provided a range of descriptions of the ways in which these different elements may be related, and our argument here is intended simply as a contribution to the explicit conceptualization of the ontological status of culture, structure, and agency and of the mechanisms involved in the interactions between them. We have suggested that some theoretical descriptions of these phenomena locate individual speakers at the center of a nest of concentric circles. Speakers then voluntarily proclaim their identification with certain social groups rather than others, subject to the pressures exerted on them by immediate peer groups — the gangs to which they belong, the tight social networks of which they are members (Milroy and Milroy 1992) — and the more remote contextual influences such as social class. Chambers (1995) summarizes aspects of this idea as follows: "... the difference between social networks and social classes as norm-enforcement mechanisms has to do with their proximity to the individual, or the immediacy of their influence" (1995: 68), and "... the difference between class and network seems to be one of degree rather than kind" (1995: 71).

The problem with such accounts, from our realist perspective, is the linear model on which they are based. In this, the linguistic practices adopted by actual speakers are regarded as deviations or deflections from the probabilistic patterns identified by the correlations of independent variables. Actual speech that does not conform to the pattern predicted by the correlated variables becomes a remaindered leftover, to be explained by factors exogenous to the variables originally considered, such as local networks, political attachments, and so on.

The methodological program we propose takes a different stance toward the complexity of the natural and social world. There is a growing interest in the utility for social research of the concepts of "emergence" and "complexity," based on the recognition that linear models of the sort we have discussed do not represent reality adequately. This is principally because of the emergent nature of the social world: structured social

relations are emergent entities (that is, they are the products of earlier engagements between people and the world they encounter). They are irreducible to people, while people are not the puppets of structures because they have their own emergent properties (Archer 1995). Byrne has summarized this realist insight thus:

... the effect of two or more variable causes acting together is not simply the sum of their effects taken separately. Instead we find that there are complex emergent properties. The relationship among the variables alters their causal propensities (1998: 173).

One important methodological implication for sociolinguistic research seems to us to be as follows. Rather than defining the problem as the correlation of linguistic variation with variations in speakers' membership of (theoretically defined) social categories, researchers might start from the case. That is, the initial focus is on the language produced by speakers, and "investigation becomes case not variable driven" (Williams 2000: 11). This kind of approach was adopted by Le Page and Tabouret-Keller when they collected data from a random sample of children who had immigrated to Britain from the Caribbean. They identified a series of linguistic variables and did not correlate these with preselected social categories but subjected their findings to cluster analysis. Byrne (1998: 170) defines clusters as "types, qualitative sets, which 'emerge' from the application of computation to large multi-variate data sets." From their results, Le Page and Tabouret-Keller inferred varying self-identifications by the children with others in their "multi-dimensional social space" (1985: 127).

Now as we have already explained, this account runs the risk of giving too great a weight to self-definition, although for some research purposes self-defined social categories may be the most appropriate. Alternatively, the researcher can follow Williams' (2000: 11) suggestions for those engaged in survey research and defer the identification of categories until a later stage of the analysis:

The conjectural character of the data collection leads to a flexibility about both the definition and the measurement or "variables," indeed the variables themselves are simply outcomes, or "traces" (Byrne 2000) of yet unidentified (though possibly hypothesised) mechanisms. The only thing that we know is "real" is the case itself and the operationalisation of the variables is deferred to the identification of antecedent case characteristics.

In other words, analysts allow for the possibilities not only that neither commonsense social categories nor social scientific ones will correlate categorically with the linguistic variables they have identified, but also

that the mechanisms bringing about linguistic variation and change may be complex and multiplicative, rather than linear and additive. This kind of model takes account of the insight from complexity theory that

... small changes in causal elements over time do not necessarily produce small changes in other particular aspects of the system, or in the characteristics of the system as a whole. Either or both may change very much indeed, and, moreover, they may change in ways which do not involve just one possible outcome (Byrne 1998: 14).

One branch of the linguistic research enterprise where evidence is found of counterintuitive correlations, and of challenges to the boundaries of existing linguistic categories (Sinclair 1991; Stubbs 1996), is corpus linguistics. As banks of authentic language data, the components of a corpus represent "cases" of speakers actually using language. As de Beaugrande (1999: 132) puts it, "... sociolinguists who work with abundant real data are referring the language back to the real community who produced or received those data." Now we are not necessarily advocating the use of corpus approaches for sociolinguistic research, although we are committed to methods that collect data from authentic interactions as opposed to experimental and artificial contexts. ("By its own criteria," writes Byrne [1998: 65], "experimentalism only works when the world is linear and when causes are simple and single.") For research in sociolinguistics, where spoken data is crucial, an awareness of measurement as an act of translation (Pawson 1989) is essential. Transcription is always theoretically inflected (Ochs 1979) and decisions about which features the transcriber is to encode in variation studies are likely to prove theoretically challenging, as well as practically time-consuming. However, these problems notwithstanding, the real data held in large corpora are susceptible to objective analysis, in the sense of objectivity we defined above. Corpus analysis rests not only on theories about language use but also on theories of computational science (rather as Galileo's observations of planetary movement relied on theoretical developments in the understanding of optics as well as astronomy [Stubbs forthcoming]). So, whether sociolinguistics' real data is tied to corpus work or not,

Real data also indicate that much of the socially relevant variation within a language does not concern the phonological and syntactic variations that sociolinguistics has investigated in terms of "features" and "rules," but rather discursive variations in selection and combination that serve the varying and sometimes conflicting interests of diverse social agents (de Beaugrande 1999: 133).

To conclude: we suggest that developments in realist social theory have a direct relevance for sociolinguistics. They offer the opportunity for a defense of sociolinguistic knowledge as objective knowledge and so may be used to reinforce the claims of sociolinguists to epistemic authority when investigating and explaining language practices. Further, we have sketched the outlines of an alternative methodological approach appropriate to a realist sociolinguistics, one that challenges the linear correlation model in favor of one based on the concepts of emergence and complexity. Finally, we hope to have indicated some of the potential generated by this particular crossing of the borders between sociolinguistics and social theory. While it would be unrealistic to expect the traffic to increase sufficiently to provide a unified theory of how language and society are related (Wardhaugh 1992; Hudson 1996; Coupland 1998), we at least hope to have indicated that much is to be gained from making more than the occasional visit.

*University of Reading
University of Warwick*

Notes

1. This article was presented as a paper at Sociolinguistics Symposium 2000, University of the West of England, April 2000. We are grateful to participants for their comments, and to Allan Bell and Martyn Hammersley, who offered constructive comments on an earlier version of the written article.
2. One explanation sometimes given is that these changes are brought about by this generation of speakers in the service of the evolution of language itself, with its adaptive function, but this account is teleological, since language cannot have purpose or intention.
3. "Ethnic" has largely replaced "racial" as the term for these kinds of category as used in sociolinguistic studies, some of which appear to use the terms interchangeably (e.g. Rampton 1995), but very much the same kinds of argument apply to both — cf. Carter (2000).
4. Potter (2000: 23) has argued that "the danger from the proposed enterprise of realist sociolinguistics is that it forces the analyst away from taking participants' activities and orientations seriously." We hope it is clear from the foregoing that realist sociolinguistics does not entail any such neglect.

References

- Archer, M.S. (1989). *Culture and Agency: The Place of Culture in Social Theory*. Cambridge: Cambridge University Press.

- (1995). *Realist Social Theory: A Morphogenetic Approach*. Cambridge: Cambridge University Press.
- Brumfit, C. (1997). How applied linguistics is the same as any other science. *International Journal of Applied Linguistics* 7(1), 86–94.
- Byrne, D. (1998). *Complexity Theory and the Social Sciences*. London: Routledge.
- Cameron, D. (1995). *Verbal Hygiene*. London: Routledge.
- Carter, B. (2000). *Realism and Racism: Concepts of Race in Sociological Research*. London: Routledge.
- ; and Sealey, A. (2000). Language, structure and agency: what can realist social theory offer to sociolinguistics? *Journal of Sociolinguistics* 4(1), 3–20.
- Chambers, J.K. (1995). *Sociolinguistics Theory: Linguistic Variation and its Social Significance*. Oxford: Blackwell.
- Cheshire, J. (1982). *Variation in an English Dialect*. Cambridge: Cambridge University Press.
- Coupland, N. (1998). What is sociolinguistic theory? *Journal of Sociolinguistics* 2(1), 110–117.
- de Beaugrande, R. (1999). Linguistics, sociolinguistics, and corpus linguistics: ideal language versus real language. *Journal of Sociolinguistics* 3(1), 128–139.
- Eckert, P. (1998). *Jocks and Burnouts: Social Categories and Identity in the High School*. New York and London: Teachers College Press.
- Fought, C. (1999). A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics* 3(1), 5–23.
- Greenwood, J.D. (1994). *Realism, Identity and Emotion: Reclaiming Social Psychology*. London: Sage.
- Hudson, R. (1996). *Sociolinguistics*. Cambridge: Cambridge University Press.
- Kerswill, P. (1996a). Children, adolescents and language change. *Language Variation and Change* 8(2), 177–202.
- (1996b). Milton Keynes and dialect leveling in south-eastern British English. In *English: History, Diversity and Change*, D. Graddol, D. Leith and J. Swann (eds.). London: Routledge.
- Layder, D. (1997). *Modern Social Theory*. London: UCL Press.
- Le Page, R.B.; and Tabouret-Keller, A. (1985). *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
- Milroy, L.; and Milroy, J. (1992). Social network and social class: toward an integrated sociolinguistic model. *Language in Society* 21, 1–26.
- Ochs, E. (1979). Transcription as theory. In *Developmental Pragmatics*, E. Ochs and B.B. Schieffelin (eds.). London: Academic Press.
- Pawson, R. (1989). *A Measure for Measures: A Manifesto for Empirical Sociology*. London: Routledge.
- Popper, K.R. (1972). *Objective Knowledge: An Evolutionary Approach*. London: Oxford University Press.
- Potter, J. (2000). Realism and sociolinguistics: response to Carter and Sealey. *Journal of Sociolinguistics* 4(1), 21–23.
- Preston, D.R. (1998). Myth 17: They speak really bad English down south and in New York City. In *Language Myths*, L. Bauer and P. Trudgill (eds.). London: Penguin.
- Rampton, B. (1995). *Crossing: Language and Ethnicity Among Adolescents*. Harlow: Longman.
- Romaine, S. (1984). *The Language of Children and Adolescents: The Acquisition of Communicative Competence*. Oxford: Blackwell.
- Searle, J.R. (1995). *The Construction of Social Reality*. London: Allen Lane.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell.
- (forthcoming). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Wardhaugh, R. (1992). *An Introduction to Sociolinguistics*. Oxford: Blackwell.
- Williams, M. (2000). Can we be realists? Realism and the problem of operationalisation in survey research. Available through <http://sociology.plymouth.ac.uk>.