Culture

Social Interest, Linguistic Indifférence

David Sankoff

Citer cet article

Sankoff, D. (1994). Social Interest, Linguistic Indifférence. *Culture*, *14*(2), 27–37. https://doi.org/10.7202/1083528ar

Résumé de l'article

En tenant compte des facteurs humains qui sous-tendent l'analyse et la description sociolinguistique, l'auteur utilise les données et la méthodologie pour établir un rapport entre les impératifs épistémologiques de l'étude de la variation syntactique et les préoccupations théoriques pour la polyvalence forme-fonction. Il montre que le noeud de toutes les positions syntactiques ou sociologiques de nature théorique ou méthodologique dépendent des intentions de communication des locuteurs à un moment de l'énoncé où il existe un choix entre des structures référentielles ou fonctionnelles équivalentes. L'auteur situe ce moment par une critique socio-scientifique des idéologies linguistiques.

Social Interest, Linguistic Indifference

David Sankoff *

Based on the human interests underlying sociolinguistic description and interpretation, the author traces the epistemological imperatives of the study of syntactic variation through data and method to the theoretical preoccupation with form-function polyvalence. The ultimate locus of all syntactic or sociological claims of a theoretical or methodological nature derives from the communicative intentions of a speaker at the moment in dicourse where more than one referentially or functionally equivalent structure is accessible. We situate this moment in a social scientific critique of linguistic ideologies.

En tenant compte des facteurs humains qui sous-tendent l'analyse et la description sociolinguistique, l'auteur utilise les données et la méthodologie pour établir un rapport entre les impératifs épistémologiques de l'étude de la variation syntactique et les préoccupations théoriques pour la polyvalence forme-fonction. Il montre que le noeud de toutes les positions syntactiques ou sociologiques de nature théorique ou méthodologique dépendent des intentions de communication des locuteurs à un moment de l'énoncé où il existe un choix entre des structures référentielles ou fonctionnelles équivalentes. L'auteur situe ce moment par une critique socio-scientifique des idéologies linguistiques.

In reflecting on the motivation and impact of the Montreal French Corpus of 1984 (Thibault and Vincent, 1990) and its predecessor of 1971 (Sankoff and G. Sankoff, 1973), it is perhaps instructive to review the synthesis of some issues in the theory of linguistic variation formulated almost ten years ago and published in Newmeyer's Cambridge Survey of Linguistics (1988). The original research motivations leading to the establishment of a computerized corpus of the French spoken in Montreal were explicitly based on the sociolinguistics of variation as an object of quantitative scientific study, with a strong, if implicit, notion of the appropriateness of doing this work in the Montreal speech community. Within a short time, corpus research emerged as a focus of debates both on paradigms of linguistic analysis and on the quality of Quebec French. Over a longer period of time, many of the analyses carried out on the corpus, notably on *avoir/être* (G.Sankoff and Thibault, 1977), on *on/tu,vous* (Laberge, 1978) and on *ponctuants* (Vincent, 1983) became archetypes for the variationist position on form-function polyvalence, especially in syntax and discourse. These issues: the role of the researcher in the speech community, the ideology of "objective" scientific research in social science, linguistic normativism and prescriptivism, structural versus generative paradigms, and questions having to do with functional equivalence of distinct forms, might seem disparate and unrelated, but in fact each one can only be fully understood in the context of the others.

* Centre de recherches mathématiques, Université de Montréal, P.O. Box 6128, station Centre-ville, Montréal, Québec, H3C 3J7

My basic thesis concerns the extension of the notion of inherent variability from phonology to syntax, the lexicon and discourse. Labov's *oeuvre* has made credible to linguists the idea that two or more different articulations of a given phonological form may occur in the same word or affix, in the same contexts. Which form will occur at a given point in time can only be predicted in terms of a probabilistic model, whereby the effects of the linguistic and extra-linguistic context can be ascertained with accuracy, but where the output of the analysis remains just a probability. The choice of form always contains a component of pure chance, though this is precisely delimited. The locus of the random component lies in the cognitive processes responsible for behavioural variability in general (see literature on probability learning).

These phonological alternations are not generally concomitant with changes in the denotational value (referential meaning) of a lexical item, nor the syntactic function of an affix or particle. The different variants may, however, have different social and stylistic connotations, being explicitly or implicitly associated with the social or demographic group which uses them most frequently, and with the social context or type of interaction in which they typically occur.

Alternation among forms also occurs at the syntactic, lexical and pragmatic levels. It would be advantageous to be able to analyze variation at all these levels within a common framework. There are, however, fundamental differences between variation at the phonological level and the others. The immutability of referential or syntactic function in the presence of phonological variation is uncontroversial, but syntactic and pragmatic equivalence and lexical synonymy are the subject of much debate: two different lexical items or structures can almost always have some usages or contexts in which they have different meanings, or functions, and it is even claimed by some that this difference, subtle though it may be, is *always* pertinent whenever one of the forms is used. This position on the uniqueness of the link between form and function is common both to strict distributionalists who rely only on the distribution of surface forms for grammatical argumentation and proof, and to functionalists from various traditions who also admit elicited, interpreted or introspected data about function. I use the terms "functionalist" and "distributionalist" in an idealized way to summarize two epistemological attitudes to the data for linguistic analysis, as summarized in Table 1.

Variationism rejects the uniqueness of the form-function relationship, however. While it is indisputable that some difference in connotation may, *upon reflection*, be postulated among so-called synonyms whether in isolation or in context, and that for a number of competing syntactic constructions, they may be acceptable in somewhat different contexts, there is no reason to expect these differences to be pertinent every time one of the variant forms is used. Indeed, underlying the study of syntactic variation within a framework similar to that of phonological variation is the hypothesis that for certain identifiable sets of alternations, these distinctions come into play neither in the intentions of the speaker nor in the interpretation of the interlocutor. We may say that *distinctions in referential value or grammatical function among different surface forms can be neutralized in discourse.* This is the source of the phenomenon of "equivalence" and the justification of the syntactic variable that have so preoccupied sociolinguists and their critics. It is the fundamental mechanism of non-phonological variation and change. By the phrase "in discourse", I refer not to discourse analysis but rather to speakers' sustained and repeated exercise of their linguistic facilities in producing connected text or discourse. In what follows, what we exemplify with syntactic variation pertains as well to lexical variation and variability in discourse structure.

There is no universally acceptable test for discriminating between the distributionalist and variationist position, since they do not accept the same types of data or proof criteria. For variationists, the systematic study of competing forms requires not only the identification of these forms, but also of the individual contexts in which differences between them are neutralized. This constitutes the *interpretive* component of variationist methodology. Analysts must in effect be able to infer the meaning or function of each token (occurence). In the most favorable situation, they do this in their capacity as ingroup members familiar with the particular individuals and interaction being studied, drawing on the intersubjective understanding of co-members of the same speech community. In the more usual situation, the linguist has to "know" enough about the speech variety and to "understand" enough about what is transpiring in the particular discourse, to be able to infer speakers' intentions. Thus a basically hermeneutic task is combined with more mechanical distributionalist procedures prior to any statistical analysis. Interpretation of this sort is not new in linguistics and literary studies, but in variation

studies it takes on a particular character, because of its application to large samples of tokens. Another, more important aspect of sociolinguistic variation practice is the sociological implication of judging the function of tokens, some of them socially stigmatized in discourse. We will return to this aspect.

The notion of neutralization-in-discourse is implicit in most work on syntactic variation, and thus from this viewpoint a purely distributionalist methodology cannot suffice. Distributionalists would not agree, however, since the inevitable existence of *some* complementarity of distribution of forms (however marginal) would, for them, carry some nuance into every context of variation. Nor would most functionalists agree with the neutralization hypothesis, since intense enough reflection about two forms will eventually identify a distinction between them, not only as a result of theory-driven introspection, interpretation or elicitation, but also in the course of naïve, norm-influenced reflection.

I claim, however, that we have no more direct access to speakers' intentions than through their utterances themselves, nor to how hearers decode these than through their responses, particularly in natural situations. Analysts may be motivated by theoretical, normative or critical considerations to discern intentions, or to deny them, whether or not these interpretations are accurate. Even the speakers themselves may incorrectly believe, upon reflection, that their linguistic choices were prompted by certain intentions, when these intentions are nothing but *a posteriori* artifacts of linguistic introspection or afterthoughts inspired by linguistic norms. There is no "objective" way of telling whether one form was used instead of its alternate because of the desire to convey some subtle distinction or whether a free choice was made among two or more equally serviceable alternatives, and any use of interpretations is inevitably bound up with ideology, at both the societal and theoretical (linguistic) levels.

Table 1 summarizes the distinctions between the approaches. The link between the linguistic debate and social issues is found in the corpus-based nature of variationism, and the critical aspect of token interpretation, to be elaborated on in the following paragraphs.

Table 1

Distinctions between approaches in linguistic analysis

Approach	Method for identifying meaning or function	Assumptions
Distributionalist (Objective)	From similarities, differences in distribution of forms. Usually example-based.	Unique form-function hypothesis. Existence of context-independent component of meaning. All functional distinctions recoverable from surface.
Functionalist (Intuitive)	Elicitation, interpretation, introspection. Usually example-based.	Unique form-function hypothesis. Traditional or theory-derived categories socially neutral.
Variationist (Critical)	Interpretation of forms in context. Corpus-based.	Critical interpretation dominates prescriptive and theoretical biases.

Having identified the crucial distinction between the variationist position on form-function relationships and that of other approaches, what are the consequences at the level of final results of studies carried from these different perspectives? In linguistic interaction involving variable behaviour, variationists can hope to establish the functional equivalence of socially stratified forms, where distributionalists or functionalists will necessarily assume that there is a concomitant differential in function. Working class variants often tend to be syntactically and morphologically reduced, trading off the redundancy and clarity appropriate to the written language and to interaction in formal and technical domains, in favour of efficiency and intersubjectivity among (usually intimate) participants, often with compensatory elaboration at the level of the lexicon and discourse mechanisms (see Slobin's 1979 charges to language "Be clear" versus "Be quick and easy"). This tendency holds not only for the working class, but for the spoken language in general in familiar and intimate circumstances. Sociolinguistic methodology, however, characteristically succeeds in tapping the vernacular more easily with working class subjects than with middle class or bourgeois speakers. Distributionalism or functionalism then inevitably infer that working class language is functionally reduced. This type of result does not arise from variation studies. This is due not only to the recognition of the possibility that distinctions are neutralized in discourse, but to the origins and interests of variationist sociolinguistics as a way of studying language.

In what follows, I adopt a some of the ideas and language of Habermas' critical theory to understanding (i) the social interests underlying linguistic research paradigms, (ii) the origins of variation theory in colonialized and minority language communities, (iii) the type of data which must be accounted for in these communities, (iv) the particular kind of analytical problems and theoretical questions pertinent to these data, and (v) the form-function problem we have discussed for syntactic variation.

* * *

The variationist viewpoint on language is determined first by a scientific interest in accounting for grammatical structure *in discourse* (i.e. in sustained linguistic production) —be it spontaneous natural conversation, formal narrative or argumentation, or various written genres— and second, by a preoccupation with the polyvalence and apparent instability *in discourse* of linguistic form-function relationships.

When scientifically accounting for an entire speech sample or corpus, striking and widespread regularities may emerge which pertain solely to the relative frequency of occurrence or co-occurrence of various structures, rather than to their existence or grammaticality. The origin of this interest, and its connection with form-function polyvalence, are crucial to understanding quantitative syntax, variation theory and sociolinguistics.

The internal linguistic conditioning that preoccupies variationists, and the methods they have developed to study it, can be amply exemplified in the speech of a single individual (see Guy, 1979), and does not depend on some highly formalized notion of "community grammar". The same holds for form-function polyvalence. Nevertheless, it is of critical importance that the theory and methodology we shall be discussing evolved during research on multi-speaker samples from sociologically or ethnographically well-defined speech communities such as in the Montreal French project; it is in this type of research site that we find the origins of the concern for sustained discourse.

It is essential to a critical understanding of why certain topics are studied, which data are used, and which methodologies are applied in scientific investigations, that we attempt to characterize the interest underlying this enquiry. By "interest" we do not mean the intellectual predilections of the individual researcher nor even the explicit or hidden objectives of those who fund research programs, but rather the social and technological projects propelling the historical evolution of societies. Habermas (1972) contrasts two fundamental human projects: the emancipation from material constraints, i.e. achieving mastery of nature, and the emancipation from social constraints, i.e. the identification and dismantling of repressive mechanisms in the social order. The former project is carried out through labour, guided by a positivist science involving controlled experimentation, physical measurement, abstract formulation, and objective criteria for consensus. The latter project is mediated by interaction, and its science is reflexive, interpretive, and antithetical to the constraints involved in controlled experiments and formalized language. It bases its consensus on intersubjective understanding not limited by external, "objective" criteria.

Furthermore, Habermas (1971, Ch. 6) observes that the dialectic between the two projects currently sees the extension of positivistic criteria into social science, the methodology of prediction and control

of nature being applied to individual and social behaviour. This displaces the role of social science in unmasking repression in the social order by an opposite orientation reinforcing the ideological justification of existing social, political and economic configurations.

We can discern three divergent research paradigms in linguistic science: the introspective-generative, experimental-evaluative, and descriptive-interpretive. (Note that these paradigms do not coincide with the approaches to data discussed earlier: functionalist and distributionalist components can co-exist in generative research, for example.) Insofar as linguistics is an academic discipline, research in all three paradigms does not evolve solely according to its own inner logic but is subject to the processes determining the role of knowledge production and distribution in society. In addition to these common influences, the experimental-evaluative and descriptive-interpretive approaches are specifically inspired by further influences.

In the case of the experimental-evaluative approach, the key external impetus for linguistic research is first-language teaching to speakers of non-standard dialects and second-language teaching to immigrant speakers of minority languages. There is an explicit, widely-accepted goal motivating the kinds of teaching falling under these rubrics: the transmission of linguistic capacities to those who do not have them, but should, in the view of the dominant society (or, more accurately, the substitution or displacement of one set of linguistic behaviours by another). Such goal-oriented activity is mediated by purposive-rational logic, in Habermas' terms, and inevitably gives rise to research on efficient teaching methods, optimal conditions for learning, and explanations and remedies for learning problems. The methodology for this type of enquiry, in the fields of educational linguistics, psycholinguistics, language evaluation, etc., necessarily involves controlled experimentation, laboratory conditions, questionnaire survey methods, proficiency testing, and a conceptual apparatus borrowed from the physical and biological sciences and developed for the prediction and control of natural processes. This apparatus is also used in neurolinguistics, experimental phonetics and even foreign language teaching, though these do not share the socially evaluative component characteristic of much educational and psycholinguistic research carried out in working class, immigrant and other minority contexts. Here the standard or the majority language is desirable; everything else is incorrect or interference.

The descriptive-interpretive current which includes variationism also has roots in non-standard dialects and the language of minorities, but the interests underlying this research are quite different. Its origins are to be found rather in the liberal and other progressive counter-attacks of the 1960's on racial and cultural stereotypes and on the paternalistic and repressive social policies which these stereotypes serve to justify. Individual variationists may be deeply concerned with improving education, but this does not account for the roots of variationist linguistics in class and minority struggles, its divergence from dialectology and other branches of linguistics, and the internal logic of its development.

The premiss that no natural language is logically or aesthetically superior to another has been little questioned in linguistics. The attack on linguistic stereotypes was not part of any debate within linguistics, rather it formed part of the generalized, if uncoordinated, assault on conventional values and norms, and the social hierarchies they ratify. There was no need within the discipline of linguistics to prove that non-standard dialects were full-fledged languages, but there was a social need to demolish linguistic stereotypes through their study in a rigorous, scientific way.

This, in part, is why the florescence of variationism and its emergence as a paradigm distinct from dialectology, ethnolinguistics, traditional pidgin and creole studies, etc., dates from 1969, with the publication of Labov's major study of copula contraction and deletion in Black English, rather than from his earlier work in Martha's Vineyard (1963) or in the Lower East Side (1966), which were not particularly pertinent to any current social movement. It also explains the early epistemological rapport between the Montreal French researchers and Labov and his colleagues and students.

The descriptive-interpretive approach typically sees the researcher deeply immersed in the speech community and intent on reducing the effects of his or her own role as an expert on and/or native speaker of (a more standard version of) the language under study, and as a (usually petit bourgeois intellectual) member of the wider society, with concomitant preconceived notions about communicative behaviour. Awareness of how the researcher's own theoretical biases and social ideology impinge on the details of scientific functioning, and the struggle to transcend them, are the essence of a critical methodology.

This type of research characteristically generates coherent, explicit and compelling critiques of class-based, race-based, or other dominant ideologies of language, with their normativisms, prescriptivisms and stereotypes about logic, aesthetics and intelligence. Furthermore, this work inevitably has social repercussions for the wider community, provoking media attention, intellectual debate and hostile criticism from the educational and literary establishment and other professionals of language, thus unmasking an interest in maintaining a repressive status quo. It is in engaging in this conflict of ideologies that linguistics may have a socially emancipatory role. Notions that the technically scientific aspects of linguistics may themselves be instrumentally relevant specifically to the working class or minorities are fundamentally mistaken. Technological progress, including a deeper understanding of the structural, psychological or physiological properties of language, forms part of the project of control over the external world—language being treated here as a formal object. As such, it will be appropriated by the classes who generally benefit most from science and technology (Emonds, 1976: xii). It is only the critical social scientific reflection upon language use in its communicative function, with the obligatorily interpretive dimension to its "method" we have discussed, which can have a role in an emancipatory project.

* * *

What is the appropriate data for linguistic research with the kind of social orientation we have been discussing? First, introspection by speakers of non-standard dialects is notoriously unreliable. This is partly because the censure or stigma attached to non-standard forms suppresses them, whether the speaker is conscious of this or not. It is also partly due to *categorical perception* which works in the opposite direction, the existence of a non-standard form entailing the perceived exclusion of the standard form from the dialect, though in fact it may be relatively common.

It is next to impossible for a speaker in a linguistically stratified community where there are pervasive linguistic norms (whether these are well-formulated or not, realistic or not, accepted or not) to systematically make accurate judgments about which forms belong to which variety, though this same speaker might be an ideal informant when it comes to grammaticality of forms *invariant* within the community.

The controlled elicitation and testing methods of psycholinguistics and educational linguistics are even less informative about non-standard usage, given the close association between the test situation and the stigmatization of non-standard forms versus the approbation attached to the "correct" normative answer. In addition, controlled experimentation and questionnaires characteristically require a pre-established inventory of responses, inevitably strongly coloured by the contrast between prescribed usages versus deviant or erroneous behaviour. This bias is hard to avoid in any work on non-standard dialects, but the experimental-evaluative approach runs counter to any type of heuristic search for patterns and structures having no direct counterparts in the standard variety. Finally, it is well-known that vernacular usage and bilingual or bidialectal behaviour are extremely sensitive to the communicative situation. They tend to be absent from formal interviews and in some cases can be observed only in highly unobtrusive ethnographic work. They are unlikely to be manifested during an examination or a laboratory experiment.

Consideration of the types of data available through the introspective and experimental-evaluation approaches, then, leads to the realization that we can rely neither on how speakers think they behave nor on how they think they ought to behave. For non-standard, minority or colonialized speech varieties, direct observation of language *use* is essential, in as natural a communicative interaction as possible. At the very least, we require recordings of relatively lengthy conversations, even if they be between the linguist and the speakers, and they should not take the form of elicitation sessions. It is preferable, of course, for conversation to be between two speakers of the same vernacular, or for the recording to be made of natural interaction, rather than an interview, but the minimum requirement is to obtain some sample of the speaker's actual speech.

The imperative to deal with language use, rather than reflection about language use, as basic data, can now be seen to be derived from the interest in research on non-standard speech varieties free from the misleading effects of stereotypes, from contamination by the norm and from categorical perception, and designed to be able to detect and handle principles or organization different from those of the standard language. This interest, which must be seen as emancipatory in the social context in which it emerges, contrasts sharply with the interest in control and prediction underlying the experimental-

evaluative approach. Sociolinguistics and psycholinguistics thus constitute profoundly different approaches, though superficially they share a number of features: concern with non-standard dialects, use of non-introspective data and statistical methodology.

*　*　*

In the study of extensive speech samples or of other types of sustained discourse, what are the substantive grammatical questions that emerge? As with the introspective-generative approach, the presence or absence of certain forms or co-occurrences invite description, generalization, comparisons and explanations. But what is equally important, and often more so are quantitative patterns of occurrence relatively inaccessible to introspection or even testing methodology. Regular, complex relationships may exist at the quantitative level among a number of structures, but upon introspection, all we can say is that they are all simply "grammatical". Quantitative regularities may be vaguely guessed at through introspection, but may not be characterized with anything like the precision with which intuition-based methods can establish categorical relationships.

The quantitative facts are not minor details of linguistic behaviour. Universal hierarchies and co-occurrence constraints not manifested in terms of grammaticality versus ungrammaticality for a given language, are nonetheless often present in clear and well-developed form in usage frequencies. The classical examples are constraint hierarchies for the expression of certain allophones (or the application of optional phonological and morphophonological rules), but it is also true of syntax, in the study of variable rule order, optional movement or deletion rules, and in preferences among semantically or functionally equivalent phrase structures.

Moreover, it is these variable aspects of grammatical structure which are always the locus of linguistic change. Change virtually always requires a transitional period, often very lengthy, of variability, competition among structures and divergence within the speech community. The detailed nature of linguistic change and of its synchronic reflex— dialect differentiation — cannot be understood without coming to grips with quantitative relationships.

The tools for studying these relationships are necessarily very different from those used in the introspective-generative paradigm. Frequency counts of forms in contexts are not just quantitative

refinements of judgements of grammaticality and have even less to do with acceptability. Counts of 0% are analogous to judgements of ungrammaticality, but not identical to them. Non-occurrence does not necessarily indicate a prohibited form. It may simply be the result of a complex combination of features which could be perfectly grammatical but unlikely to appear in any reasonably sized corpus. Conversely, intuitively ungrammatical forms may appear systematically and at a non-negligible rate in spontaneous speech through the interaction of the grammatical facility with processing constraints. Though Labov has introduced methods of "natural experimentation" to heighten the rate of occurrence of certain complex forms, these techniques are not nearly as easy to use as the generativist's use of intuitions about sets of sentences involving any number of combinations of syntactic features.

The ease with which grammaticality judgements are made may be seen as one of the motivations, or encouragements, for investigating the finer distinctions between syntactic theories. In contrast, the virtual absence from actual linguistic usage of any of the key contexts for resolving them, diminishes these issues in importance for the variationist, who has many highly frequent phenomena to account for. These latter phenomena are in turn of no interest to the generativist, who does not encounter frequencies in the course of his or her analyses.

That generativists and variationists focus on different questions about language thus does not derive immediately from differences in the explicitly stated overall goals of the two paradigms, but rather from the different data each must account for, and the tools which each regards as valid.

It has often been argued that the key role of variation studies is to shed a new kind of light on specific issues which arise in generative theory. In fact, the contribution of variationism has this aspect only occasionally. It is rather in the investigation and solution of its own internally-generated problems that it has most contributed to the understanding of language. The major variationist insights into the structure of language, how it is used and how it evolves, were not in general motivated by issues in generative theory (nor did they have any impact on that theory), often despite authors' explicit claims to that effect, but ensued rather from the internal logic of the descriptivist-interpretive paradigm in which they were made.

*　*　*

The insistence of descriptivist-interpretive research on physical recordings and transcribed speech corpora, on counting occurrences of forms, and on statistical methodology have led some critics to label it, inaccurately, as positivist and/or scientistic. An exclusive dependence on observed facts, on objective evidence, is indeed diagnostic of a positivistic orientation, but the reliance on observation in our case pertains only to linguistic form; as we have seen there is a strong interpretive component to the analysis of linguistic function.

A critique of positivism in the social science and humanities can not in any case be meaningfully justified simply because of the use of some type of data or analytic technique. Positivist science is better characterized as *excluding* certain types of data or interpretation, such as the subjectivity of participants. In linguistics, this attitude is to be found in approaches which are strictly distributionalist, where all analytical groupings and distinctions must be made on the basis of the sharedness or the complementarity of the observed distribution of surface forms: sounds, particles, words or syntactic constructions. Thus, insofar as hypotheses about linguistic structure (however arrived at) must be verified against the grammaticality or ungrammaticality of surface strings of words, and insofar as syntactic and semantic theory are oriented to account for structures thus determined, generative syntax is effectively positivist. That generative linguists use intuitions about grammaticality does not detract from this fact. These "yes - no" intuitions in no sense constitute interpretive or reflexive science, but simply substitute for external observations of linguists' own behaviour as expert "native speakers". It is in this sense that we consider generative method to be a type of positivist distributionalism.

Scientism is a somewhat vaguer label; it refers to the use of the experimental, mathematical and quantitative apparatus of physical science to study concepts apparently pertinent to the social sciences or to the humanities, but which are in fact oversimplified, poorly operationalized or of little relevance to the real issues in these domains. In psycholinguistics, this is exemplified by the Bernsteinian paradigm (Bernstein, 1964), which sets up categories of behaviour qualified as "restricted" or "elaborated". Once this is done, "scientific" studies of the working class versus middle class vernaculars may be undertaken. However the fundamental question of justifying the application of the "restricted" versus "elab-

orated" labels to particular forms is not accessible to quantitative or other formal methodologies so the apparently scientific nature of this approach is illusory. Indeed we would argue that, as is typical of research in the experimental-evaluative paradigm, the use of these categories stems from an uncritiqued class-based normativist ideology of language (despite the often explicitly emancipatory aspirations of the practitioners).

The descriptivist-interpretivist approach cannot be accurately portrayed as positivistic or scientistic. As we have seen, the distribution of forms is only one of the two major types of data in any variationist study, the other being the identification of the linguistic function of each form. Aside from phonological studies, this identification of function has an unmistakably hermeneutic, or interpretive, component which is antithetical to positivist criteria. And as we have also seen, it is the fundamental issues of linguistic change and variation in the speech community which require that we come to grips with the form-function problem and force us into the sociologically critical and essentially non-positivist analysis of function. Were we to content ourselves with the statistical analysis of surface forms, this might justify the term scientistic, but the very fact of counting or using statistics does not, since they are used within the framework of a broader attempt to account for both components of the form-function relationship.

Most variationist work also involves data of an extralinguistic nature and the statistical correlation of these with linguistic data. This too has also provoked the label of scientism— incorrectly again, since such correlations are properly used not as explanation or as indication of causality, but in conjunction with other types of analysis: sociological, ethnographic, historical and critical—in order to understand the processes of linguistic differentiation at the community level.

Macroscopic sociodemographic categories do not directly affect the performance of individual speakers; implicit in any correlational study is the existence of mediating processes or intervening mechanisms which lead from extralinguistic factors, through conscious intent and/or unconscious tendencies, to actual behaviour.

* * *

Quantitative, statistical and probabilistic notions have been introduced into linguistics many times, and they have long been standard in related fields such as lexicology and acoustic phonetics. It is only Labov's work, however, that they have become credible in phonology, and to some extent, in other areas of grammar.

The role of statistics in the study of variation has been the subject of much criticism. In response we must emphasize our interest in accounting for large corpora which contain many tokens of a limited number of forms in a variety of comparable contexts. The universal experience in corpus-based research is that the structure of communication in the speech community, the structure of variation and change, is realized through recurrent choices being made at various interactional and grammatical levels by speakers. This is where the form-function problem is originally confronted. Many "functions" can be carried out by several different "forms" and the questions of who, when and why become immediately pertinent in accounting for those actually used.

Now, whenever a choice can be perceived as having been made in the course of linguistic performance, and where this choice may have been influenced by factors such as the nature of the grammatical context, discursive function of the utterance, topic, style, interactional context or personal or sociodemographic characteristics of the speaker or other participants, then it is difficult to avoid invoking notions and methods of statistical inference, if only as a heuristic tool to attempt to grasp the interaction of the various components in a complex situation.

It is not a requirement that the choice mechanism itself have any particular linguistic or sociological interpretation. Statistical methods are indifferent to the origin of the variability of the data, whether it be in the grammatical generation of sentences, in processes of production and performance, in the physiology of articulation, in the conscious stylistic decisions of speakers, or even simply as an analytical construct on the part of the linguist. The linguistic significance does of course depend on the nature of the choice process, but this question must be addressed prior to the statistical analysis (in the collection and coding of the data) and/or afterwards, in the interpretation of the results.

Much of the debate over the use of statistics has had to do with the notational representation of sta-

tistical regularities within a grammar. Objection to formalisms containing numbers or numerical parameters are often phrased as distaste for the notion of a numerical component in the mental grammatical facility. That mental processes may involve systematic tendencies which are non-categorical even in the most highly specified circumstances is commonplace, however, and linguistic behaviour follows suit, independent of the fact that linguistic competence may also include types of structures which have no counterparts in other domains of mental activity. Furthermore, it is a fallacy to think that numerical parameters at the notational level must correspond to some specific stored numerical value at the cognitive level, any more than the hierarchical structures of phrase structures must have a neurological representation involving direct counterparts to the lines and nodes of a tree diagram.

Indeed, where statistical regularities are found in linguistic performance, they are important as properties of language independent of whether they are consequences of:

- the physiology of articulation, in phonology;

- processing considerations in syntax;

- social or biological universals, as in the competition of tense and aspect inflections with periphrastic constructions based on verbs for standing, sitting, going, etc., or in the competition of modals with verbs for volition, ability, desire, etc.;

- panlinguistic typological tendencies ("parameters") which may or may not be coded in some innate form on the individual level; or

- some punctual actualization of the individual's grammatical facility.

There are many types of causes of statistical regularity, and which one or ones are pertinent to a given linguistic pattern remains an empirical question.

* * *

The way in which syntactic variation emerges through neutralization of distinctions in discourse is not a process which fits in well with current formalisms for synchronic descriptions of syntax. A basic assumption in linguistic theorizing is that the syntactic component of language is in large measure autonomous. It may have certain well-defined input and output interfaces with the phonological, lexical, se-

mantic or pragmatic components, but otherwise the processes and constraints which constitute syntax interact essentially among themselves without reference to non-syntactic factors in determining the grammatical sentences of the language. Though some students of communicative behaviour may criticize this postulate, it would be intellectually counterproductive to pretend that the study of autonomous syntax has not been highly successful in discovering, explaining and accounting for this complex and subtle aspect of linguistic structure. In accomplishing this, however, modern syntax has excluded from its purview concepts and phenomena which might be (and some have been) considered syntactic in nature, and has consigned them to the lexical, semantic or pragmatic components of language. These include most of the equivalent-in-discourse relationships of the type we have been discussing. The forms which enter into contextual, stylistic, or social complementarity of distribution do not generally originate as closely related syntactic structures. Rather, they have something in common on the referential or pragmatic level only and participate in entirely different syntactic structures. This has led to characterizations of syntactic variationism as being technically naive for identifying variants of a variable through their "having the same meaning" or "carrying out the same function". On the contrary, it is only by refusing to limit the range of possible variants to the categories of a particular formalization of autonomous syntax, that we can have access to the origins of syntactic variation. If the domain of variability in discourse expands, however, and/or one form tends to displace another in a wide range of contexts, the equivalent-in-discourse relationship must eventually have repercussions at the purely syntactic level — the variation or change in question must be grammaticalized. To understand the origin of this type of syntactic change, then, we must look beyond syntax. During the process of grammaticalization, of course, properly syntactic considerations may predominate more and more, but it is precisely at the blurred margin between the syntactic and the extrasyntactic that the study of syntactic variation is particularly revealing and has the most to contribute.

It is natural, within a theory of autonomous syntax to confine syntactic change and variation to small changes in a single feature, condition or parameter. This models the perceived gradualness of change analogous to feature-by-feature phonological change, or to morphological change which affects

least "salient" (Naro, 1981) forms or members of a paradigm first. This does occur on the syntactic level too, of course, but in change through neutralization in discourse, gradualness is achieved by the incremental spread of the contexts where the neutralization occurs, while the difference in form generally remains unattenuated.

***

Based on the interests underlying the descriptive-interpretive approach to language, we have traced the epistemological imperatives of the study of syntactic variation through data and method to the theoretical preoccupation with form-function polyvalence. We have identified the ultimate locus of all syntactic or sociological claims of a theoretical or methodological nature as being the communicative intentions of a speaker at the moment in discourse where more than one referentially or functionally equivalent structure is accessible. And we have attempted to situate this "moment" in a social scientific critique of linguistic ideologies.

In a theory where "interaction" is as basic a category as "work" (Habermas, 1972, Ch. 3), the dominant ideology is not merely an epiphenomenon of some mechanistic economic process. Rather it plays a crucial role in the justification of the existing social order and its support by imposing guidelines for individual behaviour in interaction. This ideology is itself generated and reinforced in social praxis, i.e. in interaction, where the existing configuration of power, prestige and wealth appears normal and inevitable according to all the criteria of this same ideology. Positivist science with its pre-defined categories and rejection of subjectivity is limited to quantifying and formalizing existing relationships. Critical social science on the other hand, through its focus on the intersubjectivity of participants in interaction and its historical scope, can penetrate the appearances of inevitability and seek the social interests which actually determine both action and ideology.

Not only is positivist methodology ideologically imposed by the same interests which propagate normativism and prescriptivism, thus "confirming" stereotypes of working class and minority language, but this ideological basis is hidden behind a "scientific" rationale which claims universality for positivist criteria. Thus Lavandera (1978), in the name of scientific rigour, criticizes Laberge's (1978) recognition of functional equivalence of second person pronominal forms *tu/vous* with *on* in contexts of indefi-

niteness in French. Laberge had warned against analyses which equated the loss of *on* with a loss of the corresponding referential distinction. Lavandera dismisses the "social conviction" behind this warning and calls for more "empirical" methods for proving that the distinction is not lost, or that if it is lost, then it does not imply reduction at the cognitive level. In our view, this faith in "empiricism" and refusal of the hermeneutic aspect of the analysis is itself eminently ideological. It is precisely the hermeneutic recognition of equivalence that allows Laberge to avoid the conclusions predetermined by a normative ideology in the guise of a supposedly universally valid positivism.

References

BERNSTEIN, D.
1964   Elaborated and Restricted Codes: Their Social Origins and Some Consequences, *The Ethnography of Communication*, John Gumperz and Dell Hymes (eds.), Special Issue of American Anthropologist, 66(6.2).

EMONDS, J.
1976   *A Transformational Approach to English Syntax*, New York, Academic Press.

GUY, G.
1979   Variation in the group and the individual, *Locating Language in Time and Space*, W. Labov (ed.) New York, Academic Press, pp. 1-36.

HABERMAS, J.
1971   *Toward a Rational Society*, J.J. Shapiro (trans.), Boston, Beacon Press.

1972   *Knowledge and Human Interests*, J.J. Shapiro (trans.), Boston, Beacon Press.

LABERGE, S.
1978   The changing distribution of indeterminate pronouns in discourse, *Language Use and the Uses of Language*, R.W. Shuy & A. Shnukal (eds.), Washington, D.C., Georgetown University Press.

LABOV, W.
1963   The social motivation of sound change, *Word*, 19: 273-307.

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: 715-762.

LAVANDERA, B.
1978   Where does the sociolinguistic variable stop? *Language in Society*, 7: 171-182.

NARO, A.
1981   The social and structural dimensions of a syntactic change, *Language*, 57: 63-98.

NEWMAYER, F. (ED.)
1988   *Linguistics: the Cambridge Survey*, Cambridge, Cambridge University Press

SANKOFF, D.
1988   Sociolinguistics and syntactic variation, *Linguistics: the Cambridge Survey, Vol. IV, Language: the sociocultural context*, F. Newmeyer (ed.), Cambridge, Cambridge University Press, pp. 140-161.

SANKOFF, D. and G. SANKOFF
1973   Sample survey methods and computer-assisted analysis in the study of grammatical variation, *Canadian Languages in their Social Context*, R. Darnell (ed.), Edmonton, Linguistic Research Incorporated, pp. 7-64.

SANKOFF, G. and P. THIBAULT
1977   L'alternance avoir/être dans le français parlé à Montréal, *Langue française*, 34: 81-108

SLOBIN, D.I.
1979   *Psycholinguistics* (2nd ed.), Glenview, Il., Scott, Foresman and Company.

THIBAULT, P. and D. VINCENT
1990   *Un corpus de français parlé*, Quebec, Université Laval, Collection Recherches sociolinguistiques.

VINCENT, D.
1983   *Les ponctuants de la langue*, Ph.D. dissertation, Université de Montréal.