An Information Theoretic Interpretation of the Variable Rule

Matthew R. Sorenson
WICAT Education Institute

'Information' has been used informally and as a technical term in several disciplines, but has not been well-developed in the social sciences. This paper will look at information theory and its relevance to linguistics, trying in particular to show how it illuminates the nature of the variable rule.

Claude Shannon, John Von Neuman, and Norbert Weiner were the major figures in the postwar interdisciplinary science of systems. Von Neuman contributed game theory, dealing with strategic decisions. Weiner founded cybernetic control theory and focussed on how systems regulate their own behavior to meet goals. And Shannon established a mathematical definition of information that encompassed both communication and thermodynamics.

Although Shannon is the recognized father of information theory, it is the main contention of this paper that the ideas contributed by Weiner and Von Neuman are essential in a correct application of it. It is because decision and control have been neglected by linguists that information theory has not been found fruitful.

By 1953, when Charles Hockett published a review of Shannon and Weaver's Mathematical Theory of Communication, there had already been some interest by linguists in information theory. Hockett lists three references and adds that

'It is not certain that all these references are based on adequate understanding of the theory.' (Hockett 1953: footnote 2)

Hockett's (1953) review of Shannon and Weaver was tentative and ambiguous. He approached the topic with the attitude that it must be valuable to linguists, but he failed to show how.

The review's first section describes the essential concepts of the theory. The following are paraphrases of Hockett's main points.

Signals are chosen out of a set of possible alternatives. Information is transmitted when the variety in the alternatives is narrowed down by selecting one of the possibilities. If there is no choice, there is no information transmission.

If the variety in the set of alternative signals increases, then the system has a greater informational capacity. A system with 'yes', 'no', and 'maybe' can transmit more information than a system with only 'yes' and 'no'.

The basic terms apply to the capacity of an information source, but not to single instances. Information theory is concerned with averages.
The source, channel, and receiver of information are distinguished, but the signal at any point in this sequence may be transformed (or transduced) into various material forms. Information is independent of mechanism.

The code, or the syntax and semantics of the signal set, may be manipulated for the sake of efficiency or redundancy, depending on the demands on the system. The two primary means of manipulation are adjustments on the frequency and on the interdependence of signals.

Shannon's first theorem says that codes can be made maximally efficient, if desired, so that channels can always be used at their maximum capacity. Shannon's second theorem is a generalization to cases involving a noisy channel, and says that any degree of noise can be offset by the use of redundancy.

Having established these basic notions, Hockett begins a discussion of how they may apply to linguistics. His first point is that the continuous nature of speech and the discrete structures of language are not inconsistent and may be mathematically transformed into each other within information theory. Speakers (and field linguists) learn through experience to act as transducers to interpret continuous signals into discrete language units.

Then Hockett addresses the extreme redundancy in language. Stress, for example, is redundant because it can be transmitted simultaneously with other signals. Writing is estimated to be about 50% redundant since we can read text in which half the letters have been removed.

Problems of indeterminacy and transformation in morphology, and encoding through writing are muddled through in information theoretical terms, but as Hockett admits,

'Since there is currently no way in which all this can be disproved, it does not qualify as a scientific hypothesis; it is merely a terminology.' (p. 42)

Hockett's final section outlines 'general implications' and touches on issues that would arouse discussion for years to come: the confusion of information and meaning, the tempting parallel between information and energy, and the idea that structuring of energy processes is accomplished by informational 'triggers'.

The understanding of information theory in this review is unfortunately structural instead of regulative. The new ideas are seen as ways of describing the way language is, instead of the way it works or how it is used. For example, Hockett raises the idea that social information can be transmitted by variation in language, but immediately dismisses the possibility as uninteresting (Hockett 1953: 39).
After the Hockett review, information theory was not absorbed by linguistics. Bar Hillel and Carnap (1952) made a valiant but uninterpretable effort at extending information theory into semantics. Greenberg (1956) made a well-conceived but superficial application in historical linguistics. Hockett continued to be interested and his 1977 volume contains early articles on several related topics. Zelig Harris tried early to examine syntactic structure in terms of statistical distribution of elements, and apparently is still at it (Harris 1987). Jakobson edited a volume on mathematical linguistics containing a piece by him on information and communication (Jakobson 1961). The same volume also included a chapter by Rulon Wells referring to Church and Wittgenstein as the conceptual forebearers of information theory.

Although cluttered and unsystematic, On Human Communication by Colin Cherry (1957) was for years the best effort at applying information theory to language. It suffered, however, from too much engineering jargon and not enough appreciation of the human dynamics of communication.

In 1966 Alfred Smith edited a volume on Communication and Culture. Chapter two brought together articles which outlined the basic concepts. The section on social interaction included Goffman. Chapter six discussed networks, and seven dealt with noise. The sections on feedback and control and on redundancy and equilibrium included some rich ideas by great thinkers like William Powers, Ross Ashby and Karl Deutsch. Sections on semantics and pragmatics pushed into frontier areas.

However, the bulk of the work in Communication and Culture was either suggestive or peripheral. The book was designed with a clear idea of the interdisciplinary theory that could be applied, but the content never did rise to the concept. Particularly disappointing was the section on linguistic theory (chapter four). Greenberg, Gleason and Pike, pillars of the old guard, presented a very static and structural view. They did not understand the new ideas, and linguistics was not ready to deal with problems of decision and control.

Also included was the classic Chomsky piece called Three Models for the Description of Language. We must remember that as revolutions impose order, they destroy variety. The unflowered bud of new scientific ideas are uprooted along with lots of dead wood when a new order is ushered in. Perhaps the most compelling reason for the failure of information theory to take hold in linguistics was that it was associated historically with the structural and quantitative world view that was swept away in the rationalist fervor of the sixties.

Reminiscent of the early interdisciplinary days, a recent symposium discussed the development of information theory throughout the thirty years since Shannon and Weaver (Machlup and Mansfield 1983). The physicists at the symposium were well ahead of the pack. The linguists still failed to understand the nature of information theory and how it could be applied to human communication.
Recently some efforts have been made to take the rigorous notions of information and entropy from physics and apply them to language (Campbell 1982 and Eigen and Winkler 1981). Unfortunately, these writers assumed that all there was to linguistics was Chomsky, and their presentation suffered.

Presently there are several areas on the 'fringes' of linguistics where information theory is being either implicitly or explicitly applied, including semiotics, formal semantics, discourse theory, and variable rule analysis. We will focus on the variable rule.

The variable rule

As variation involves the probabilities of selecting signals out of sets of alternatives, it can be described in the language of information theory. Linguistic variables that can assume a minimum range of values (say two) have the capacity of transmitting the minimum quantity of information (one bit). Variables that accomplish a greater degree of variety reduction have a greater informational capacity.

A serious question that must be addressed at this point is whether information theory can apply at all to language or social interaction. It has been argued (Waddington 1977:140) that if we don't know (i.e. can't list) the set of signal possibilities, then we can't quantify how signals are selected, and therefore information theory is inapplicable in any interesting human domain.

A reasonable response is given by Dretske, who suggests that it is valuable to use information theory comparatively rather than absolutely.

'Such comparisons can be made without ever determining absolute values for either magnitude. That is, one can use these formulas in the way one would use a piece of string that is not marked off in inches and feet. One can use the string to determine whether A is longer than B without determining the length of either A or B.' (Dretske 1981:54)

Another qualitative approach to using the originally quantitative notions of information theory is proposed by Ashby (1972). He suggests using set theory to describe the relationships between ensembles of alternatives, replacing probability with membership as the central construct.

Anyway, to the extent that we can specify the range of values that can be taken on by a linguistic variable, we can use absolute informational measurements. In other cases, there are other means.

The variable rule, as developed by Labov and others is a generalized statement of the distribution of the probabilities of alternative patterns, as taken from a measured corpus. It is therefore a statement of informational capacity of that variable and its tokens as sociolinguistic signals.
It is clear that the frequency of occurrence of a variable (or the frequency of application of a variable rule) is both monitored and meaningful among speakers of a community. The signals are transmitted, received and have interactional consequences.

Since the variable can occur within many different kinds of words, or sentences, or situations, there is not a single probability for its occurrence. The constraints on the probability of application of a rule (called 'factors' in the variable rule literature) may be due to phonological, grammatical, or interactional environment. This requires us to use conditional probabilities in calculating the informational capacity (Cherry 1957:182ff. and Lerner 1972:63ff.). This does not affect the applicability of the informational approach, but is just a means of relativizing for context.

Factored into the machinery of the variable rule is an 'input probability', which is a sort of baseline probability, independent of all environmental constraints. The other factors are adjustments on this baseline.

By using measurements of either conditional probability or input probability, the informational capacity of all the signals in the set of alternative tokens can be established. The tokens with little probability of occurring will have the greatest capacity. Highly probable tokens will have less capacity.

Linguistic variables contribute to style (Labov 1972: chapter 3). As style changes, by cooperative shifting to a new set of 'contextual cues' (Gumperz 1982: chapter 6), the frequencies and expectations for signals are 'recalibrated'. The new becomes the norm. The variables that define a style become expected and can carry lower amounts of information. There is a sort of self-organized 'bootstrapping' going on between speakers when signals structure style, and style in turn structures the informational capacity of those very same signals.

Guy (1980) discusses whether the variable rule should apply to individuals or groups. Guy concedes to the Bickerton-Bailey axis that the variable rule applies where variation is ubiquitous and homogeneous. In communities where individual patterns are stable, but there is stratified variation between speakers, the averaging effect of the variable rule analysis obscures the facts. The important point here is that variation is meaningful only where there is a conventional code for its use. Variation, because of its informational nature, tells us about interaction, not about the gross structure of the speech community: about style, not stratification.

Guy (1980) also discusses the problem of how big the corpus should be for a good variable rule analysis. The method of the Labovians has been to think up strategies for enriching the quality of the interview. If there were a greater appreciation for the role of variation in regulating interaction, we would focus on more natural conversation. If the data is artificial and static, lots of it is going to be needed.
So just as the proper appreciation of information theory requires the consideration of decision and control, so the proper appreciation of the variable rule requires that we apply it to natural conversation where speakers are deciding, and controlling the speech situation.

Frake showed how asking for a drink in Subanun required competence far beyond the grammatical (Frake 1964).

'If messages were perfectly predictable from a knowledge of the culture, there would be little point in saying anything. But when a person selects a message, he does so from a set of appropriate alternatives. The task of an ethnographer of speaking is to specify what the appropriate alternatives are in a given situation and what the consequences are of selecting one alternative over another.'

The selections made in generating messages reduce possibilities, and thus transmit information. The content of the information is the concern of semantics and ethnography. The form and quantity of the information is characterized in the variable rule.

The selections may be grammatical/semantic (which modifier? which quantifier?) or may be less referential (which form of address? which dialect?). Some sets of alternatives are more easily specified than others. A choice between tu and vous (Brown and Gilman 1960) is more discrete than a choice about how to formulate place (Schegloff 1972). But in any case, there are choices made by speakers with consequences for the subsequent discourse.

The variable rule has too often been used in a way that describes only the static structure in a speech community. Language variables have been portrayed as passively responding to external aspects of the situation (see the various applications in Labov 1980). Bell reports dissatisfaction with this view. He discusses both 'responsive' and 'initiative' style shifts and concludes that

'Language is to be seen not merely as a dependent variable, manipulated by non-linguistic factors. It may be an independent variable which itself influences the situation.' (Bell 1984:183)

Brown and Levinson also ascribe to this view, in which variables are seen as manipulated by speakers for social-interactional purposes.

'... language usages are tied to strategies rather than directly to relationships, although relationships will be characterized by the continued use of certain strategies' (Brown and Levinson 1978:286)

Accommodation and control

One of the reasons that speech communities are coherant systems is because the members actually do 'cohere'. Language behavior can contradict the natural tendency toward entropy as speakers put more and more constraints on their selections.
'Variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the social dimension. This cause and effect relationship holds on three levels. First it operates synchronically for an individual speaker who, in specific situations, shifts style to sound like another speaker. Second, it operates diachronically for individual speakers who, over time, shift their general speech patterns to sound like other speakers. Third, it operates diachronically for an entire group of speakers which, over time, shifts its speech to sound like another group.' (Bell 1984:151)

Here Bell has suggested accommodation as the basic mechanism of language dynamics for both individuals and groups. Accommodating speakers who move into a new mode of interaction have changed selection patterns, but not grammars. This is indicated by the possibility of reverting back to older patterns, or employing 'inappropriate' forms for the sake of changing frame, testing reactions, or making jokes.

The 'work' that is constantly being done in speech communities is this negotiated convergence to specific sets of conventions. When interaction is weak, the conventions decay and expectations are fuzzy. But the highly predictable patterns of usage are reestablished as speakers accommodate. This is not to say that there are no conventions for interacting with strangers. But the more intimate or specialized the relationship is between speakers, the more discriminations must be made, which involve greater variety reduction, and greater information flow.

Information serves control, the manipulation of response toward some established goal. Inasmuch as speakers have goals and process information, they are able to control the speech situation. According to Ashby (1964) the basic principle of control in information systems is the need for requisite variety. In order to control the environment a system must have as many effective options to respond as there are effective constraints imposed on it.

If speakers share goals, their accommodation reduces the variety in the set of alternatives. They require less variety to control their shared environment, and so the code is restructured accordingly.

If speakers do not share goals, then they have different ideal states toward which they are trying to steer the situation. In this case, the requisite variety for maximizing control must be generated by increasing the range of alternative responses. Variation will increase and speakers will tend to diverge from shared conventions. Thus, Roloff suggests this general principle.

'As a relationship escalates from non-interpersonal to interpersonal, the variety of communication strategies will decrease; as a relationship de-escalates from interpersonal to non-interpersonal, the variety of communication strategies will increase.' Roloff (1976:189)
Speakers must have a soft touch on the brakes, however, and they generally negotiate their position between the extremes, like Schopenhaur's porcipes (Eigen and Winkler p. 153), which must stay close enough for warmth and far away enough for comfort. The signals are neither categorical nor discrete. Instead they are like the impulses that build up in a neuron until a firing threshold is reached.

In living systems, small changes can have large effects, and individuals can structure groups. In a speech community, accumulated fluctuations can result in shifts from one macroscopic state to another. In his review of Shannon, Hockett describes the effect of such a cumulated informational input as 'trigger action' (Hockett p. 47).

Variable rule analysis can tell us about the informational inputs to such triggers. But the variable rule must be a part of a larger view of interaction and social structure, including considerations of decision and control. Helpful for this larger view is an understanding of entropy and nonequilibrium.

Entropy

'... entropy is a macroscopic property (i.e. one which does not apply to individual atoms or molecules but only to appreciable aggregates of them) and is thus comparable to other macroscopic properties such as temperature and pressure.'

(Denbigh 1975:67)

Contrary to the accounts of some writers (Waddington 1977:143), information is not the negative of entropy (negentropy), and it should not be said that information decreases as entropy increases. It may be true that when entropy increases, an observer's ability to predict its behavior is decreased. But that is not to be confused with the capacity of signals (events) in the system to carry information. Entropy is a structural property of a system whereas information is a relation between elements or processes within a system.

This confusion is handled by Gatlin (1972:48ff) in her distinction between 'potential information' (which increases with the entropy or potential variety of a system) and 'stored information' (which increases with the degree of variety reduction of a signal). The latter is what Shannon calls 'redundancy'.

Gatlin (1972:chapter 2) also makes a distinction between two different types of stored information, or redundancy. The first is what she calls 'divergence from equiprobability'. She defines and contrasts a higher level of redundancy as 'divergence from independence'. Both can be measured in DNA and can vary independently of each other. Her measurements suggests that lower organisms attain informational complexity by increasing divergence from equiprobability. Vertebrates, however, have evolved to the highest levels of complexity by increasing divergence from independence, i.e. by imposing structure not only on the probability of DNA substructures, but of ordered sequences of these substructures (Gatlin 1972:80).
Non-equilibrium

'At equilibrium molecules behave as essentially independent entities; they ignore one another. We would like to call them 'hypnons,' 'sleepwalkers.' Though each of them may be as complex as we like, they ignore one another. However, nonequilibrium wakes them up and introduces a coherence quite foreign to equilibrium.' (Prigogine and Stenger 1984:181)

Equilibrium is the extreme independence and equiprobability of microstates. Strictly speaking, it is impossible to find in social systems. But we can discuss different social systems, or speech communities, in terms of how far from equilibrium they appear to be functioning. This degree is a matter of interaction between parts of the system itself, and interaction between the system and its environment.

'So far we have discussed isolated systems... However it is possible to extend Boltzmann's explanation to open systems that interact with their environment ... Equilibrium is the result of competition between energy and entropy. Temperature is what determines the relative weight of the two factors. At low temperatures, energy prevails ... At high temperatures, however, entropy is dominant and so is molecular disorder.' (Prigogine and Stengers 1984:125-6)

At high temperatures, interaction between elements is disrupted. The regularity or predictability of the ongoing relationship between any two given parts of the system is lost. You can't tell who you are going to bounce up against next, or what the nature of the interaction will be. Urban anonymity may be a social reflection of this high temperature disruption.

In these circumstances, it takes more communicating to maintain the same degree of structure. Thus it is the informationally more complex systems that can regulate their behavior and protect their integrity at the higher levels of energy flow.

The Second Law applies to closed systems. Speakers and speech communities are open systems, however, and receive informational (and energetic) input from a rich, living environment. The inputs feed the constant processes of structuration that offset the tendency toward a spread of the system into more and more equally probable, independent microstates. So certain events (and certain correspondences between events) become more predictable and law-like, to the effect of maintaining systemic identity. This structuration of probabilities is the 'life work' of living systems.

'We can isolate a crystal, but cities and cells die when cut off from their environment. They form an integral part of the world from which they draw sustenance, and they cannot be separated from the fluxes that they incessantly transform ... It is difficult to see how Boltzmann's order principle can be applied to such situations ... It is the opposite of disorder, a privileged state to which only a comparatively small number of complexions may correspond. In Boltzmann's terms, it is an 'improbable state'.' (Prigogine and Stengers 1984:127)
As speakers interact, make selections, accommodate, and transmit information by strategic use of significant social variables they are organizing their community away from equilibrium. In such systems variety is high, variety reduction is high, information flow is high, signals are not equiprobable, and they are highly interdependent. We find redundancy chasing variety in a kind of inflational spiral of ever-increasing informational capacity.

**Individuals**

This improbable, high-flow, non-equilibrium sort of order is the result of the aggregation and amplification of purposeful individual behavior. In speech communities it is constructed from the situated interaction of speakers, negotiating their identities and relative advantages.

'A system far from equilibrium may be described as organized not because it realizes a plan alien to elementary activities, or transcending them, but, on the contrary, because the amplification of a microscopic fluctuation occurring at the 'right moment' resulted in favoring one reaction path over a number of other equally possible paths. Under certain circumstances, therefore, the role played by individual behavior can be decisive.' (Prigogine and Stengers 1984:176)

Above, we referred to the bootstrapping by which style is created by manipulating variables, which in turn are constrained by style. This constructive connection between levels of functioning according to Prigogine, is characteristic of all living systems.

'One of the most important problems in evolutionary theory is the eventual feedback between macroscopic structures and microscopic event: macroscopic structures emerging from microscopic events would in turn lead to a modification of the microscopic mechanisms. Curiously, at present, the better understood cases concern social situations.' (Prigogine and Stengers 1984:191)

The integration of the macro and micro levels of analysis is at center stage in sociolinguistics and discourse analysis (Knorr-Cetina and Cicourel 1981). In this connection, we see the need for an understanding of information theory, and of the variable rule, that emphasizes decision and control. The traditional structural interpretation won't do.

**Actuation**

Labov talked about how our active manipulation of the frequency of the redundant sociolinguistic variables can define styles. Minor variations don't count, but patterns do.
'If contrast exists between casual and careful styles, and the variables which we are using play a significant role in that contrast, they do not seem to operate as all-or-none signals. The use of a single variant ... does not usually produce a strong social reaction; it may only set up an expectation that such forms might recur, so that the listener does begin to perceive a socially significant pattern. Every speaker occasionally begins a (dh) word with a sharp onset, which can be interpreted as an affricate, [dθ]. However, in the prestige form of speech, these forms recur so seldom that they are negligible. Any pattern of expectation set up by them dies out before the next is heard.' (Labov 1972:108)

David Lewis long ago described how small fluctuations can grow into stable structures, noting that the source of the fluctuation is unimportant.

'A convention is produced when a big enough fluctuation meets strong enough amplifying forces.' (Lewis 1969:86)

What sorts of social interaction can amplify variation? How does the structure of the community affect the interactional dynamics of groups of speakers in it? How big of a fluctuation is big enough? These questions have long intrigued Labov and he has written of them as the 'actuation problem' (Weinreich, Labov and Herzog 1968, Labov 1980).

In discussing the social origins of sound change Labov relies on four explanations for why change does or does not result from fluctuations: population influx, imitation, standardization, and interaction networks (Labov 1980:260-64).

The first consideration is the influx of new speaker patterns. Usually such an influx serves as a source of fluctuation and the efforts of the incoming groups to retain their identity may amplify these distinct patterns. In this case, the openness of the system allows for greater variety and differentiation.

'...it is the entrance of new ethnic and racial groups into the community that provides the motivating forces behind this renewed diversification [in Philadelphia].' (Labov 1980:263)

Speakers who share identity and solidarity will tend to imitate the core members of their group. The 'emblematic function' of a pattern of variation (Labov 1980:262) encourages speakers to value certain peculiarities. The identity of a group is expressed in these patterns, as is the membership of individuals in that group. Such autocatalytic (self-perpetuating) accommodation will amplify fluctuations to the point where even the wider community grammar can be restructured.

'Once we are willing to refine our notion of prestige to give full weight to the local prestige associated with the Philadelphia dialect, Tarde's laws of imitation gain in respectability. But we must be ready to recognize that such a local prestige ... is powerful enough to reverse the normal flow of influence, and allow the local patterns to move upward to the upper middle class and even to the upper class.'
Variations, however, also tend to be dampened by the system. This is what Labov referred to in the above passage as the 'normal flow of influence'. The power of macro-level prestige standards to inhibit participation in divergent patterns is well known. Labov provides an example.

'Young black speakers do not participate at all in the evolution of the vowel system that is described here; instead, they clearly show their allegiance to a nationally based black English vernacular that is extraordinarily uniform in all the cities of the North.' (Labov 1980:263)

Fourth is the reference to the nature of the communication networks of innovating speakers.

'We have approached the problem of why sound changes take place at a particular time by searching for the social location of the innovators: asking which speakers are in fact responsible for the continued innovation of sound changes, and how their influence spreads to affect the entire speech community.' (Labov 1980:252,261)

He goes on to confuse innovation with hypercorrection, but the valid point is that the channels between people and groups in the speech community will certainly have effect on whether changes and variables are going to be amplified or dampened. The general principle is this: to the extent that boundaries are effective, internal fluctuations will be protected. On the other hand, where boundaries are ineffective, the effects of the larger system, or the environment, will be overpowering.

Thus in Philadelphia, as in Labov's earlier study of Martha's Vineyard, there were clearly bounded and rather autonomous subsystems that tended to pursue changes and amplify the variants that marked their identities.

But in situations where subgroups are highly integrated into the larger system, fluctuations will be dampened. This has been the case with immigrant speech communities that disappeared into the American 'melting pot' (Ferguson and Heath 1981:Part III). The parallel in chemical systems is described by Prigogine.

'... theoretical studies and numerical simulations show that the critical nucleus size increases with the efficacy of the diffusion mechanisms that link all the regions of systems. In other words, the faster communication takes place within a system, the greater the percentage of unsuccessful fluctuations and thus the more stable the system. This aspect of the critical-size problem means that in such situations the 'outside world', the environment of the fluctuating region, always tends to damp fluctuations. These will be destroyed or amplified according to the effectiveness of the communication between the fluctuating region and the outside world.' (Prigogine and Stengers 1984:187-88)
There are various forces at work in the speech community, dampening and amplifying the fluctuations. The balance between them determines the state of the system. Influx from outside can be a source of new fluctuations. The imitation that arises out of in-group prestige will amplify them. Standardization imposed by the larger system will dampen them, by means of rich channels of communication with the subgroups.

All these mechanisms depend on information flow, generated by selecting signals out of a range of variation. The attitudes and interactions that prevail between groups are cumulative results of these same informational processes. The degree of variation at different points in the actuation process can be measured using the concepts of information theory.

Summary

We have looked at the history of information theory in linguistics and tried to draw a picture of a new view of language and communication. The frame for this picture has been borrowed from the 'hard' sciences, but it has been painted using the palate of linguists themselves. Our approach has been an exercise in the creative mixing of metaphors, revealing the common concepts between how Prigogine describes self-organizing chemical systems and how Labov has described the dynamics of the speech community.

The informational interpretation of the variable rule has been used to unify discussions of style shifting, accommodation, variation, language change and the relation between individual and community grammar. This approach enriches the issues and provides sociolinguistics with tools, both conceptual and instrumental, for productive research.

To provide examples of the application of these ideas to actual cases of variation would demand another paper. But first the ideas need to be laid down. If the concepts are sound, the numbers solve themselves. The constructs of information theory are not just terminology, but are empirically testable in the field.

Sociolinguistics amounts to more than just using a tape recorder instead of intuition. By looking at how people use language, we arrive at a different idea of its essential nature. As in all sciences, the future lies in an appreciation of how order arises out of the interaction of indeterminate elements. Compare these two statements about science; one a prediction, the other a partial fulfillment.
'Some things never happen in the physical world because they are impossible, others because they are too improbable. The laws which forbid the first are the primary laws: the laws which forbid the second are the secondary laws. It has been the conviction of nearly all physicists that at the root of everything there is a complete scheme of primary law governing the career of every particle or constituent of the world with an iron determinism ... One would not be surprised if in the reconstruction of the scheme of physics, which the quantum theory is now pressing on us, secondary law becomes the basis and primary law is discarded. In the reconstructed world, nothing is impossible though many things are improbable.' (Sir Arthur Eddington 1958:75,98)

'... our vision of nature is undergoing a radical change toward the multiple, the temporal, and the complex. We were seeking general, all-embracing schemes that could be expressed in terms of eternal laws, but we have found time, events, evolving particles ... A new unity is emerging: irreversibility is a source of order at all levels. Irreversibility is the mechanism that brings order out of chaos.' (Ilya Prigogine 1984:292)

I would like to offer this stochastic, indeterminate, and informational view of language as an alternative to the static view of the formalists.
References.


Miller, George A. 1956. The magical number seven plus or minus two: Some limits on our capacity for processing information, Psychological Review, 1956, 63, 81-97.


