DE GRUYTER MOUTON
DOI 10.1515/sem-2013-0004 — Semiotica 2013; 193: 67–82

Vern S. Poythress

Information-theoretic confirmation of semiotic structures

Abstract: Information theory indirectly confirms some fundamental structures in semiotics. By offering quantitative criteria for efficient transmission of data, it suggests by analogy ways of thinking about efficient communication in language and other media. The criterion in information theory for maximal capacity for information at the source leads to preference for independent data, which can be generalized to the semiotic principle of approximate independence among many kinds of emic units. This independence is closely related to what Kenneth L. Pike's tagmemic theory has called *distribution*. The criteria in information theory for faithful transmission of data lead by generalization to the semiotic principles of contrast and variation. Together, the aspects of contrast, variation, and distribution constitute fundamental structures characterizing the whole field of semiotics. They also lead to the development of three interlocking views of communication, the particle, wave and field view, which enable us to explain a number of more complicated phenomena in communication. These tools for semiotics receive confirmation from the quantitatively more specialized concerns of information theory.

Keywords: semiotics; information theory; language; tagmemics; efficiency; noise

Vern S. Poythress: Westminster Theological Seminary. E-mail: vpoythress@wts.edu

We propose to show that information theory (Pierce 1980; Yeung 2002; Cover and Thomas 1991) indirectly confirms some of the fundamental structures in semiotics.

1 Information theory in relation to semiotics

What is information theory? Information theory is a mathematical theory that studies quantitatively the communication of data. We should appreciate both the strengths and weaknesses of the theory. It has the strength of offering a rigorous mathematical analysis that focuses narrowly on the problems of transmitting data. But its narrow focus leads to built-in limitations (Nauta 1972; Pierce 1980:

4, 8–9; Scarrott 1993; Cropley 1998a, 1998b). Its concern is with the signifier, whether that takes the form of oral speech, graphical letters and words, computational data or data about the colors in a picture. The signified, the dimensions of meaning found in the signs transmitted, are left to one side for the sake of quantitative analysis.

Information theory is nevertheless pertinent to semiotics because it suggests ways of thinking about efficiency in communication. How can a given amount of data be most efficiently sent over a physically limited channel such as a telephone wire or a fiber optic cable or the sensory neurons in the inner ear? According to information theory, maximal quantitative efficiency in transmission over a channel frequently involves complicated coding schemes to compress data and to add checks to correct errors in transmission. The efficiency in transmission is obtained at the cost of complexity in calculation at the two ends of the channel. In ordinary human communication, by contrast, "efficiency" is a broader, not merely quantitative concept. Complicated human coding and decoding by hand at the two ends of a channel would be "inefficient" by human standards. We want quick, easy communication. At the same time, we want checks against error, when we mishear or misread a word. So the practical concerns in human communication do overlap with those in information theory.

2 Aspects of signs

We will start with semiotic principles, and then show how they are confirmed by information theory. Communication involves signs with three interlocking characteristics: (1) different signs must be identifiable and distinguishable from one another; (2) the signs must be reusable, and identifiable in spite of variations in details; and (3) the signs occur within a larger system to which they are related and in terms of which they have significance.

These three kinds of characteristics are closely related to Kenneth L. Pike's theory of signs (1967, 1982). Kenneth Pike has discussed three aspects of signs, namely, contrast, variation, and distribution (Pike 1982: 42–65).

We may illustrate these aspects with a simple sample from language. Consider the word *sing*. First, it is word distinct in sound and in meaning from other words like *ring* and *clang*. It has a *contrastive* relation to these other words. Second, *sing* has several forms: *sang* is the past tense of *sing*, and *sung* is the past participle. Each of these forms of *sing* is a *variation*, a variant form of the same word, which can be considered "the same" for all of its forms taken together. Third, *sing* has a *distributional* relation to the patterns of other verbs, as displayed in Table 1.

present tense	present participle	past tense	past participle
sing	singing	sang	sung
help	helping	helped	helped
come	coming	came	come
bring	bringing	brought	brought
...	...	...	...

Table 1: Verb forms

3 Contrast, variation, and distribution

We could analyze other words in a similar manner. We could also analyze units of other kinds. Contrast, variation, and distribution occur not only with words but with smaller units (e.g., graphical letters) and with larger units (e.g., sentences, paragraphs, and human behavior; Pike 1967: 84–91).

For example, the graphical letter *a* contrasts with letters *b*, *c*, and so on. The letter *a* shows *variation* in the different forms that it can take, such as a lower case *a*, an upper case *A*, an italic *a*, and so on. Finally, the letter *a* occurs in contexts of other letters before and after, and these contexts are its *distribution* in sequence (Pike 1982: 62–64).

Pike's tagmemic approach is especially relevant to our task because he explicitly sets his discussion in the larger semiotic context of human behavior. Hence, he implies that they may be applicable outside of the specialized concerns with language (Pike 1967: 297–98, 511).

Let us consider an example outside of the system of natural language. We can apply the triad of contrast, variation, and distribution to the realm of music. Each note in music *contrasts* with notes of different pitch: middle C contrasts with C#, D, and E above middle C. Each note in music can show *variation* in length, loudness, and timbre, and still be the same note. Each note has a *distribution* in the context of melodies and chords with which it is linked.

4 Justification of semiotic relations from information theory

All three types of relations – contrast, variation, and distribution – are common in semiotic systems. We now propose to show that all three of these relations can be seen as implications of information theory.

According to information theory, information sent in a sequence of signals from a source is maximized when the signal at any one point is *independent* of the signals sent before and after (Yeung 2002: 25, Theorem 2.39).

Let us illustrate. Suppose that a transmitter can send any one of four symbols, *a*, *b*, *c*, *d* at a particular time. One second later, it again sends any one of the four symbols *a*, *b*, *c*, *d*. The two signals that are sent at the two times are called independent if the probability of sending *a*, *b*, *c*, or *d* at the later time is unaffected by the probability at the earlier time (Yeung 2002: 5–6). Intuitively, it can be seen that this independence maximizes the information flow. Suppose that sending the letter *a* makes it almost certain that the next signal to be sent will be *b*. Then the actual sending of the next signal, namely, a *b*, gives very little extra information to the recipient, because he expected it. If, on the other hand, sending an *a* does not influence the next signal sent, the next signal has its full informational content.

If we plot all the possible combinations of two signals in a table, we get Table 2.

first signal:	**a**	**b**	**c**	**d**
second: a	aa	ba	ca	da
b	ab	bb	cb	db
c	ac	bc	cc	dc
d	ad	bd	cd	dd

Table 2: Possible sequences of signals

The principle of maximization through probabilistic independence holds when the signals are simultaneous as well as when they are successive. That is, if we send two signals through two different channels X and Y, rather than sending them successively through the same physical channel X at two different times, the information-theoretic properties of the system are the same.

In natural language and in other semiotic systems, probabilistic independence does not occur exactly, because meanings carry over from one time to another, and interact with one another. Meaningful communication links the micromeanings of the parts into a macromeaning of the whole, in such a way that the whole makes sense. Thus a nonsense sequence of words “median rector surf pose” is not as statistically likely an utterance as a meaningful sequence: “the boy fed the dog.” Yet the words in a normal sentence are still *partially* independent of one another. The freedom to construct many sentences using the same vocabulary stock depends on freedom in combinations. This freedom in combinations is the focus of *distribution*. In fact, Table 2 as an analysis of sequences is a simple case of both distribution *as part of a structural sequence* (a *followed by* b) and *distribution as a point in a system* (Pike 1982: 62–65). Information theory applied to semiotic systems predicts that distributional relations will regularly occur, and that these distributional relations are an implication of the pressure to maximize communicative content while using a relatively small number of signs.

Why then does human communication not match perfectly the theoretical ideal of information theory? For human beings, communicative meaning fits the context, and that fit means that probabilistic independence is not perfect. Communication is constrained by needs that press in other directions. And that leads us to the next insight from information theory.

5 Information loss

Elementary information theory studies situations where data transmission may be imperfect. Data may be garbled or incorrectly reproduced after transmission through a telephone line. Noise on the line may interfere.

Let us imagine a situation where Abbie transmits information to Barbara using an "alphabet" of characters x_1, x_2, . . . These may take the form of a written alphabet, a system of sounds, digital data (*0*'s and *1*'s), a system of hand signs (as in sign language), or some other system. After transmittal and possible garbling, Barbara detects information from an "alphabet" of characters y_1, y_2, . . . , which may or may not be identical with the x's. According to information theory, the conditional entropy (or information loss) of the transmission is

$$H(x \mid y) = - \Sigma_{\mathrm{i}} \Sigma_{\mathrm{j}} \, p(x_i,y_j) \log p(x_i \mid y_j).$$

where Σ_{i} symbolizes the summation over all the values of i, $p(x_i,y_j)$ is the probability that both x_i and y_j occur, and $p(x_i \mid y_j)$ is the conditional probability of x_i occurring, given that y_j occurs: $p(x_i \mid y_j) \equiv p(x_i,y_j)/p(y_j)$ (Pierce 1980:153: Yeung 2002:12; Cover and Thomas 1991:16). Since the probability $p(x_i \mid y_j)$ is always 1 or less, the logarithm ("log") is always zero or negative. (And by convention, $0 \log 0 = 0$.) So each term inside the double summation is zero or negative, and the total expression is always zero or positive.

Information loss can be minimized if each term in the summation is minimized. But there are some constraints, due to the fact that the probabilities

$p(x_i, y_j)$ must add up correctly. For example, the sum over all possible y_j values of the probabilities $p(x_i, y_j)$ must be the total probability $p(x_i)$ that x_i will occur. But if the number of letters x_i is less than or equal to the number of y's, there is a way of satisfying all the constraints and reducing the information loss to zero.

Any one term $p(x_i,y_j) \log p(x_i \mid y_j)$ is 0 if either $p(x_i,y_j) = 0$ or $p(x_i \mid y_j) = 1$ (Yeung 2002: 24). Let us consider these two alternatives separately. The condition $p(x_i,y_j) = 0$ says that x_i and y_j never occur together. One excludes the other. This condition is equivalent to a contrastive relation between x_i and y_j. Next, $p(x_i \mid y_j) = 1$ means that x_i occurs whenever y_j occurs. This situation corresponds to a variational condition, in which x_i is implied by y_j, and so y_j is one "variant" of x_i. Barbara's detection of y_j guarantees that x_i must have been transmitted by Abbie.

Intuitively, both of these conditions make sense when they operate properly. The contrastive relation is important so that we can distinguish the different x's once they are transmitted. The variational relation is important, because, if we are going to avoid loss in transmission, each x must lead to one result, and not more than one, at the end of the transmission.

We can also make sense of these constraints within a natural semiotic system. In a noisy auditory environment, meaningful sounds (phonemes) must be identifiably distinct (contrastive relation) in spite of variations in sound due to noise (variation). And they must be capable of being built up into series of sounds forming syllables and words (distribution), if a small total number of phonemes is able to produce a large number of distinct words. If the total number of phonemes is not kept reasonably small, it will be difficult to distinguish them against background noise.

6 Deviations from an information-theoretic ideal

Contrary to the information-theoretic ideal of complete probabilistic independence, not all sequences of phonemes are equally probable. Why not? There are at least two obvious reasons. First, consonant clusters with too many consonants in a row become difficult or almost impossible to pronounce. And they become difficult to identify without error (the reception problem). Consonants separated by vowels are easier both to pronounce and to identify. Similarly, distinct vowels are easier to identify consistently if they are separated by consonants. We can see that practical constraints in the skills for human pronunciation and for human hearing enter the process. These constraints make it less "costly" and therefore more "efficient" from a human point of view to use certain kinds of patterns of syllables. The familiarity of these syllabic patterns, as emic patterns (Pike 1967: 37–72), also aids in correcting mistakes in a noisy environment. Just as error correction has a role in information theory, a qualitative analog to such correction has a role in natural semiotic systems, that is, semiotic systems that are adapted to human beings who are not using special technological aids such as signal-processing calculations.

Similar observations hold with respect to words and longer compositions. The meanings of words within sentences, as we have observed, are not independent of one another in a probabilistic sense. On the other hand, the meanings of words are relatively stable as context varies, and that stability is a lesser kind of "independence."

7 Quantitative approximation for a semiotic system

The input from information theory suggests that we can produce a quantitative mathematical model that might represent the interplay of contrast, variation, and distribution within natural semiotic systems. The goal of maximum information content leads to probabilistic independence in the data, which we saw is associated with the distributional condition. The goal of transmitting information without loss leads to contrastive and variational conditions. We need to combine the two goals. Both have a role if we seek maximal information delivery.

Let us suppose that information sources can be represented by random variables $A_1, A_2, \ldots, A_n$, each having an innate information capacity or "entropy" $\mathrm{H}(A_i)$. A random variable like A_1 can take on values from an "alphabet" $a, b, \ldots, k$. The variable is an abstraction that can represent any of a whole spectrum of cases in natural semiotic systems. For example, A_1 may represent the result of rolling a die, and the alphabet would be the distinct numbers 1, 2, . . . , 6 that can come up on the die. Or A_1 could represent the Roman letter written by Abbie at a particular time and place, in which case it could be any of the letters in the Roman alphabet plus punctuation signs. Or it could represent whichever word Abbie is uttering. Let us suppose that Abbie is transmitting information. If information is delivered to Barbara in random variables $B_1, \ldots, B_m$, maximizing Barbara's information about the A's means maximizing the mutual information between the A's and B's, which in information theory is designated $\mathrm{I}(A_1, A_2, \ldots, A_n; B_1, \ldots, B_m)$. According to a simple theorem of information theory, this mutual information can be decomposed into two terms:

$$\mathrm{I}(A_1, A_2, \ldots, A_n; B_1, \ldots, B_m) = \mathrm{H}(A_1, \ldots, A_n) - \mathrm{H}(A_1, A_2, \ldots, A_n \mid B_1, \ldots, B_m)$$

(Yeung 2002:14; Cover and Thomas 1991:18).[1]

The information content or entropy $\mathrm{H}(A_1, A_2, \ldots, A_n)$ concerning the values of all the A's is maximal if the A's are independent (Yeung 2002: 25–26; Cover and Thomas 1991: 28). This is the distributional condition. The second term, the conditional entropy $\mathrm{H}(A_1, A_2, \ldots, A_n, \mid B_1, \ldots, B_m)$, is 0 (its minimal value) only if contrastive and variational relations hold in the relation of the A values to the B values. Hence, the requirement that we maximize mutual information $\mathrm{I}(A;B)$ implies that appropriate contrastive, variational, and distributional relations hold.

This situation within information theory suggests an analogy within natural semiotic systems. The variables A represent in probabilistic fashion different thoughts and meanings of a person Abbie who is the source of communication. Typically, rather than having A_i represent merely physical quantities, it will represent an element of meaning, an emic unit such as a word or a musical note. This unit may or may not be present at a particular time and place for Abbie as a particular participant. Typically, A_i can then have two values, "yes" or "no," depending on whether Abbie thinks that the emic unit is present. Similarly, the variables B represent thoughts and meanings of a person Barbara who is the recipient. The mutual information $\mathrm{I}(A_1, A_2, \ldots, A_n; B_1, \ldots, B_m)$ is analogous to sharability between Abbie and Barbara. Maximizing $\mathrm{I}(A_1, A_2, \ldots, A_n; B_1, \ldots, B_m)$ means making sure that the semiotic system enables good communication. Since the sharability of ideas and meanings takes place through some kind of physical channel(s), the mutual information shared must go through the channels, such as motor nervous system, articulation by the vocal system, sound waves, reception by the ears, and processing by the sensory nervous system.

1 Suppose that a random variable X can take on values from an "alphabet" $x_1, x_2, \ldots, x_n$. Let $\mathrm{p}(x_i)$ be the probability that the value x_i occurs. Then information theory defines the information content or entropy of X as

$$\mathrm{H}(X) = -\Sigma_i\, \mathrm{p}(x_i) \log \mathrm{p}(x_i)$$

where Σ_i indicates that we are to take the sum over all values of i.

If Y is a second random variables with values $y_1, \ldots, y_m$, the mutual information between X and Y is defined as

$$\mathrm{I}(X;Y) = \Sigma_i\, \Sigma_j\, \mathrm{p}(x_i,y_j) \log \mathrm{p}(x_i,y_j)/[\mathrm{p}(x_i)\mathrm{p}(y_j)]$$

Entropies and mutual information for several random variables are defined similarly.

There is no direct route from ideas in one person's mind to ideas in another's. For the most part, communication has to go successively from ideas to grammatical expression to phonological realization to motor neuron signals to sound and so on. Access between two distinct expressions takes place exclusively through specific intermediate levels. From an information-theoretic point of view, this presence of intermediates is equivalent to the criteria for a Markov chain (which is defined precisely to represent the transmission of information through distinct steps, in which information lost at one step cannot be recovered at a later step). Consider a Markov chain with two steps, from A to X and from X to B. The mutual information $\mathrm{I}(A;B)$ between A and B can be no greater than the mutual information $\mathrm{I}(A;X)$ between A and X, nor can it be greater than the mutual information $\mathrm{I}(X;B)$ between X and B. Similar conditions hold for a Markov chain involving more than two steps (Yeung 2002: 7–8, 27). Hence, for efficient communication, the mutual information has to be maximized at each stage.

Maximizing several different stages of mutual information implies a cluster of constraints. And these constraints interact with one another. For example, the information capacity of the channel for human speech sound is increased if we increase the number of distinct phonemes. But if we allow too many phonemes, the error rate will increase because the transition from abstract, "mental" phoneme into an actual sound (a phone) will produce errors in production (due to lack of perfect muscular control and the possibility of hasty or sloppy speech), in transmission (noises in the environment of a speech), and in reception (due to lack of perfect hearing and the possibility of misidentifying one sound that is very similar to another). The physical difficulty of pronouncing some consonant clusters or in making a transition between a particular consonant and a particular vowel may also generate errors in production and reception. So natural languages tend to avoid combinations that are physically difficult, even though the use of such combinations would theoretically increase the information capacity of the channel.

8 Examples

We can illustrate how these constraints work on small pieces of natural language. We first consider some examples from phonemics. We use the standard notation in which slashes "//" enclose a *phonemic* representation, that is, a representation that expresses the "emic" identification of meaningful sound units by native speakers. /p/ is the English phoneme *p*. We use brackets "[ ]" to enclose *phonetic* representation, that is, representation using a language-universal common code applicable to vocal articulation or to sound waves. [p] is the *phone p*, a bilabial voiceless stop (plosive).

8.1 Simple contrast

In English the words *seal* and *zeal* contrast in meaning. Suppose Abbie wants to communicate one of the two meanings to Barbara. The transmission goes from word meaning A to word W to phonemic realization P to sound S to phonemic interpretation Q to word V to meaning B for Barbara. If English had only one generic phoneme /s/ covering both voiced sibilant [z] and voiceless sibilant [s], the two sounds [sil] and [zil] would both receive a single phonemic interpretation /sil/, and the mutual information $\mathrm{I}(S;Q)$ would be dramatically less than $\mathrm{H}(S)$. Information is lost. On the other hand, if English has two distinct phonemes /s/ and /z/, the uncertainty $\mathrm{H}(S|Q)$ is low and the mutual information $\mathrm{I}(Q;S) = \mathrm{H}(S) - \mathrm{H}(S|Q)$ is high. Hence, the contrast in meaning is evidence for the existence of two distinct English phonemes /s/ and /z/.

8.2 Simple complementation (conditioned variation)

Consider now the situation with the words *black* and *plaque*. Transmission from Abbie to Barbara goes from the word *black* to phonemic realization /blæk/ to sound [blæk] or [plæk] (the voicing being weak in [b]). The word *plaque* has phonemic realization /plæk/ and sound [pɫæk], where "ɫ" is a voiceless lateral. Thus there are two distinct lateral sounds, a voiced sound [l] (in *black*) and a voiceless lateral [ɫ] (in *plaque*). Are these two distinct phonemes?

If they were distinct phonemes, we would expect an approximate probabilistic independence between the occurrence of [l] or [ɫ] (voiced or voiceless) on the one hand, and various phonemes coming before and after on the other. Such probabilistic independence maximizes the information capacity $\mathrm{H}(P)$. But that is not what we find. Instead, the occurrence of the voiced or voiceless alternate is predictable on the basis of whether the lateral phone is preceded by a voiceless stop. /p/ is followed by a voiceless lateral. A similar phenomenon occurs with [k]: *cleat* (/klit/) has phonetic realization [kɫit].

We can become more specific as to what the difficulty is. Consonants in English are produced by the vocal apparatus through a simultaneous manifestation of a point of articulation (the lips, "bilabial," for /p/), and voicing (voiced or voiceless). But the exact timing of the beginning or ending of voice is subject to some variation, in relation to the timing for the cutting off of the air stream with a stop like [p]. The timing involves variation both at the level of articulation (muscular production) and at the level of interpretation of sound by the ear and the auditory nerves. If the requirements for timing are too tight, errors are generated in transmission and/or interpretation. The mutual information I(*A*;*B*) shared by Abbie and Barbara decreases from its ideal value. In addition, on the level of articulation, it is too difficult to produce a voiced stop [b] followed by a voiceless lateral [ɬ] followed by a voiced vocoid (e.g., [i]). The vocal chords cannot easily be turned off and on again with this kind of rapidity, and even if they could the ear could not easily detect the difference consistently. Hence in practice a single switch from no voicing to voicing is all that is feasible. And transmission must allow for the variation in the exact time when voicing begins. The phonemic sequence /bl/ is mapped into the phonetic sequence [bl] or [pl], with some leeway given as to whether the voicing starts during the [b] or during the [l]. The phonemic sequence /pl/ is mapped into [pɬ], with voicing starting more or less with the beginning of the following vocoid. This compromise allows for fairly confident identification by Barbara of the phonemic difference between /p/ and /b/.

The identification takes place by means of an allowance for variation. The sounds [bl] and [pl] both represent the phonemic consonant cluster /bl/. Notationally, in probabilities P(B/bl/ | [bl]) ≈ 1 and P(B/bl/ | [pl]) ≈ 1, where "B/bl/" is the event where Barbara recognizes the presence of the cluster /bl/. (The notation "≈ 1" denotes that the probability is approximately 1.) H(*Q*|*S*) ≈ 0, as it should be. That is, no information is lost in going from sound ("*S*") to Barbara's phonemic interpretation ("*Q*"). That is, Barbara is able correctly to identify the phoneme sequences on the basis of the physical sound. In transmission, we cannot predict whether Abbie will say [bl] or [pl]. P(A[bl] | A/bl/) = *a* and P(A[pl] | A/bl/) = 1 – *a*. But that results in no diminution of the mutual information between Abbie and Barbara, since the original information from Abbie is still recoverable.

In sum, we rightly analyze the variation between voiced and voiceless laterals as two variational manifestations of the same phoneme /l/. If we postulate instead two phonemes, we do not increase the information-carrying capacity.

8.3 Free variation

In English, when the word *soup* is pronounced in isolation (as, for example, a complete utterance in answer to a question, "What did you have for lunch?"), the final consonant /p/ can be either released ([pʰ]) or unreleased ([p̚]). Various factors in the circumstances may have some influence on the probability of the occurrence of the released or unreleased variant. But in many circumstances the variation is difficult to predict. It is "free" variation; either form may occur.

Similarly phenomena occur with other word-final stops: the word *toot* can come out [tutʰ] or [tut̚] (with released or unreleased *t*); the word *sack* can come out [sækʰ] or [sæk̚] (with released or unreleased *k*).

The word *soup* in Abbie's mind gets realized phonemically as /sup/, unambiguously. So far, there is no loss in information. The phonemic syllable /sup/ gets realized phonetically, both in the articulation of the air stream and in the sound waves, as either [supʰ] or [sup̚]. In either case, the information about the phoneme /p/ can be unambiguously reconstructed. In terms of probabilities, P(/p/ | [pʰ]) ≈ 1 and P(/p/ | [p̚]) ≈ 1. When, however, Barbara reconstructs the word soup, we cannot reliably infer whether [pʰ] or [p̚] occurred based on Barbara's interpretation. Probabilistically, 0 << P([pʰ] | Barbara's /p/) << 1 and 0 << P([p̚] | Barbara's /p/) << 1. If the difference between the released and unreleased consonants is an emic difference, that is, if the language uses it to convey meaning, that aspect of meaning is irrevocably lost at the interpretation stage. On the basis of minimizing information loss, we conclude that [pʰ] and [p̚] are *etic* variants of the same phoneme /p/.

9 Particle, wave, and field views

We can deepen our understanding of natural semiotic systems if we introduce alternate perspectives on the emic structure of the systems. Kenneth Pike proposed that we can fruitfully look at language not only as composed of "pieces" like phonemes, but as composed of waves (phonological *processes*, as well as grammatical or semantic processes) and as composed of field relationships (multidimensional relationships, as in a phonetic chart or grammatical chart of verb forms). He introduced three views of language, which he dubbed the particle, wave, and field views (1959, 1982: 19–38; Poythress 1982a, 1982b). An entire area within language or semiotics can be analyzed from any one of these three points of view.

So far our discussion of phonemics in English has mostly used the particle view. A phoneme is a distinct stable "piece"; it is a *particle* in Pike's terminology. But the analysis of conditioned variation in the consonant sequences [bl] and [pɬ] tacitly depends on *wave* phenomena. /b/ and /p/ contrast as voiced and voiceless (or aspirated) alternatives. The voicing or lack of voicing "bleeds over" from the time of stoppage of air in /b/ or /p/ to the time during which /l/ is being articulated. We could say that the boundary between the opening stop /b/ or /p/ and the lateral phoneme /l/ is not firmly fixed or air-tight. Or we could say that /b/ and /p/ and /l/ are articulated in "waves" of motion that involve the *gradual* rise or fall of motion in the vocal chords, the gradual movement of the lips, and the gradual movement of the tongue tip up to the alveolar ridge, so that during the /l/ air is passing around the two sides of the tongue. The conditioned variation in /l/ becomes explicable against the background of "wave" characteristics in articulation and sound waves.

The particle view of semiotics has an affinity to contrastive relations. Each phoneme is distinct from every other, and the probability of joint occurrence should be near zero. The wave view of semiotics has an affinity with variation. The peak or "nucleus" of a wave entrains with it the margins. The margins are implied by the peak, and so are like variational manifestations of the peak. If /p/ represents the peak of a voiceless stop [p], then it is surrounded by marginal phenomena of closing and opening of lips and a time in which the vocal chords are silent. If we let *m* designate these marginal phenomena, then in probabilities, $P(m \mid /p/) \approx 1$. That represents a variational relation. But – and this is important – we are not dealing with variation within a single particle phoneme /p/. Rather, we are *viewing* the whole pattern of communication as a wave, as a process with peaks and margins.

The field view on phonology is also important. In fact, we have implicitly touched on it when we observed that patterns with respect to released or unreleased final /p/ are analogous to patterns for /t/ and /k/. The free variation between final released [pʰ] and unreleased [p̚] stops is analogous to other released and unreleased forms: [tʰ] and [t̚], [kʰ] and [k̚]. The analogical relation between patterns is a *field* phenomenon.

The field view has an affinity with distributional relations. Distributional relations frequently manifest probabilistic independence. So does the field view. The patterns for released and unreleased stops are approximately independent of the point of articulation, that is, whether we are dealing with /p/ or /t/ or /k/. The field view analyzes particle-like phonemes by breaking them down, as it were, into several dimensions of different types of features, in this case point of articulation, voicing, and manner of articulation. Each one of these features exhibits a certain independence relative to others. We can see from an information-theoretic point of view how this contributes to efficiency. A comparatively small number of distinctive features, distinguished from one another, generate a larger number of distinct phonemes when we combine the features in a simultaneous manifestation of all of them.

Thus the field view can look at phonemes not as individual particles, but as intersections of simultaneous features, features existing in several dimensions. The probabilistic independence of these dimensions contributes to maximum information capacity in the channel – just as it does in information theory proper.

9.1 Contrastive field features

If, in a language such as English, we find that /t/, /d/, /k/, and /g/ are contrastive phonemes, and if [p] and [b] occur, they are likely to be contrastive. This phenomenon is basically an instance of analogy, involving the "field" of phonemes in their interaction with one another. Each unit within the chart of phonemic units influences the others in terms of the dimensions of voicing and point of articulation.

We can explain these phenomena within the framework that we have provided if we assume that phonological features such as voicing, point of articulation, or manner of articulation can themselves function in either an emic or an etic manner, depending on the language. The requirement that emes be largely definable in terms of contrast, variation, and distribution of other emes leads us to expect that sometimes voicing or point of articulation may be a contrastive feature. The fact that, from the field view, a key feature such as voicing can itself be "emic," that is, contrastive within the phonological system, leads naturally to predicting emic contrasts for voicing at several different points of articulation.

9.2 Noncontrastive field features

The opposite type of situation occurs when voicing or some other field feature is noncontrastive. In Campa from Pajonal, Peru (Robinson 1970: 71), we find free variation between [k] and [g] and free variation between [t] and [d]. We then expect free variation between [p] and [b] as well. We can make this prediction because voicing as a feature in the field view is not emically contrastive. So the lack of contrast will hold for the bilabial point of articulation, i.e., for [p] and [b].

10 Examples from grammar

Note that our formal reasoning from information theory does not explicitly distinguish between grammar and phonology. The same formal properties apply to each. Hence the same framework automatically cover all the grammatical phenomena that are analogous to phonological phenomena. Grammatical phenomena at the level of morphology and at the level of syntax are both included automatically.

For example, the formalism automatically has room for *simple contrast* between two morphemes (such as verbal endings *-ing* and *-ed* in English) or between two syntactical constructions (interrogative versus declarative form). Likewise, we automatically have room for *simple complementation or conditioned variation*, both in morphology (plurals formed with *-s* or *-es*) and syntax (clause word-order in independent and relative clauses). We have room for *free variation* in morphology (*dived*, *dove*) and syntax (if, for example, the relative order of direct object and indirect object after *give* is seen as an instance of free variation at least in some cases).

Some more interesting phenomena turn up as illustrations of *contrastive field features* in grammar. In the area of morphology, the regularity of verb paradigms in inflecting languages is a large-scale instance of field phenomena: the classification of verbs in terms of person, number, tense, or voice represents a multidimensional classification. The dimension of variation in the verb roots is still another, final dimension. When taken together these dimensions form a multidimensional array or network that influences the expectations of language speakers. Hence, for example, we confidently distinguish in Latin two homophonous verbal forms *venit* "he comes," and *venit* "he came," because we know that elsewhere in the verbal system there is a clear distinction between present tense and perfect tense forms. Similarly in English we can distinguish from context occurrences of singular and plural forms of *sheep*. The two homophonic forms *sheep* are confidently distinguished as singular and plural on the basis of context and a regular pattern of forming plurals. The field effect arises from the interaction of two dimensions, one the dimension in which we distinguish singular and plural, and the other the dimension including the lexical root forms of all count nouns.

In the arena of syntax, field-like features are most evident in the area of grammatical transformations of classic transformational grammar. We plot the distinction between active and passive sentence constructions in one dimension and let the detailed contents of the clauses vary along a second dimension. To say that there is a regular pattern relating active and passive constructions is equivalent to saying that there is a field regularity, such that the relation between active and passive forms is independent (in the sense of distributional condition) of the particular contents of any one particular sentence. Similarly, if we plot the distinction between independent and relative clauses in one dimension, and the contents of the clauses in a second dimension, the existence of a regular transformation is equivalent to the assertion of the probabilistic independence of the two dimensions.

The information-theoretical framework applies automatically to semiotic systems like music and dance, because the framework starts with the generic issue of meaningful communication rather than with specific assumptions unique to natural languages.

References

Cover, Thomas M. & Joy A. Thomas. 1991. *Elements of information theory*. New York: Wiley-Interscience.

Cropley, David H. 1998a. Towards formulating a semiotic theory of measurement information: Part 1 Fundamental concepts and measurement theory. *Measurement* 24(4). 237–248.

Cropley, David H. 1998b. Towards formulating a semiotic theory of measurement information: Part 2 Semiotics and related concepts. *Measurement* 24(4). 249–262.

Nauta, Doede. 1972. *The meaning of information*. The Hague: Mouton.

Pierce, John R. 1980. *An introduction to information theory: Symbols, signals, and noise*, 2nd edn. New York: Dover.

Pike, Kenneth L. 1959. Language as particle, wave, and field. *Texas Quarterly* 2(2), 37–54.

Pike, Kenneth L. 1967. *Language in relation to a unified theory of the structure of human behavior*, 2nd edn. The Hague & Paris: Mouton.

Pike, Kenneth L. 1982. *Linguistic concepts: An introduction to tagmemics*. Lincoln, NB & London: University of Nebraska Press.

Poythress, Vern S. 1982a. A framework for discourse analysis: The components of a discourse, from a tagmemic viewpoint. *Semiotica* 38(3/4). 277–298.

Poythress, Vern S. 1982b. Hierarchy in discourse analysis: A revision of tagmemics. *Semiotica* 40(1/2). 107–137.

Robinson, Dow F. 1970. *Workbook for phonological analysis*. Santa Ana, CA: Summer Institute of Linguistics.

Scarrott, G. G. 1993. Some functions and properties of information. *Engineering Science and Education Journal*. 177–186.

Yeung, Raymond W. 2002. *A first course in information theory*. New York: Kluwer/Plenum.

Bionote

Vern S. Poythress (b. 1946) is a professor at Westminster Theological Seminary ⟨vpoythress@wts.edu⟩. His research interests include hermeneutics, mathematical linguistics, and theology. His publications include *The gender-neutral Bible controversy* (2000); *Redeeming science* (2006); *In the beginning was the word: Language – a God-centered approach* (2009); and *Redeeming sociology* (2011).