# MORE ON INFORMATIONAL ENTROPY, REDUNDANCY AND SOUND CHANGE[1]

EVANGELOS A. AFENDRAS

*International Center for Research on Bilingualism*
*Laval University, Quebec*

NICOLAOS S. TZANNES

*University of Oklahoma, Norman, Okla*

## 0. Introduction — the linguistic problem

Several practical problems concerning phonology could be approached much more effectively by using a rigorous mathematical model. Quantification of bilingual interference and eventual phonological convergence necessitated the search for such models and some findings were reported in the first author's earlier papers (1968, 1969a, 1969b).

The present model is based on a more powerful mathematical theory and should be applicable to a much wider range of phonological problems. The following test case from diachronic phonology was singled out: the commonness as to geographical distribution and diachronic stability of the five vowel "classical" triangle, consisting of /i, e, a, o, u/ as in Spanish and Modern Greek. The striking frequency of this vocalic type has been pointed out by many linguists, including Trubetzkoy (1929; 1949: 106, 116, 117), Hockett (1958), Viggo Brøndal (1936) and others.

The stability of this type can be attested too by Modern Greek in which it has remained unaltered since, roughly, 900 A. D. Actually, it is most likely extreme stability which accounts for the frequent occurrence of this pattern among the most diverse genetic groups. If one considers a phonological pattern as a system, there must be systemic characteristics responsible for stability, which in turn are discoverable by systems analytic methods. Doubtless, these characteristics are due to the nature of the human speech production-perception mechanism in which lies the final answer to our problem. But first we must find the systemic constraints in the sound pattern and then turn to psychology or neurophysiology for a natural explanation of the constraints.

[1] An original version of this paper was read at the Second International Congress of Applied Linguistics, Cambridge, England, September 1969.

## 1. The random process

The model employed was the following:

First, the five vowel pattern was rendered into Jakobsonian distinctive features. (Actually, any other distinctive features can be adopted, binary or not and this has little effect on choice of model). Next, vowel pattern, individual vowels and distinctive features were expressed as a stochastic process.

A random (stochastic) process is a collection of numerically valued functions of a parameter $f$, each of which is assigned to points $\omega_i$ of a probability model $(\Omega, p, \tilde{B})$.

When the parameter $f$ assumes only a finite number of points, the process is called a finite discrete parameter process. At each point $f_i$ of the parameter $f$ we have a collection of numbers, each number corresponding to each function of the process and thus each number coming from some $\omega_i \varepsilon \Omega$. We thus have a random variable at each $f_i$ of the parameter $f$. If the range of all the random variables is a finite set of points, the process is called a finite range random process.

Here we consider only finite range, discrete parameter random processes. Such random processes are completely specified probabilistically by the joint probability mass function of all the random variables at each $f_i$ of the parameter $f$.

## 2. Adaptation of the model

Next, the adaptation of the stochastic process to the phonological system is outlined: We consider a vowel pattern consisting of individuals $v_1, v_2, \ldots v_n$ (at first approximation equiprobable), which are completely specified by giving their values (+, −, or 0) on a set of distinctive features $f_1, f_2, \ldots f_n$.

In other words the vocalic pattern is the abstract space $\Omega$ and to each vowel is assigned a realization, which is a function of a "distinctive feature parameter". The following diagram illustrates the idea:

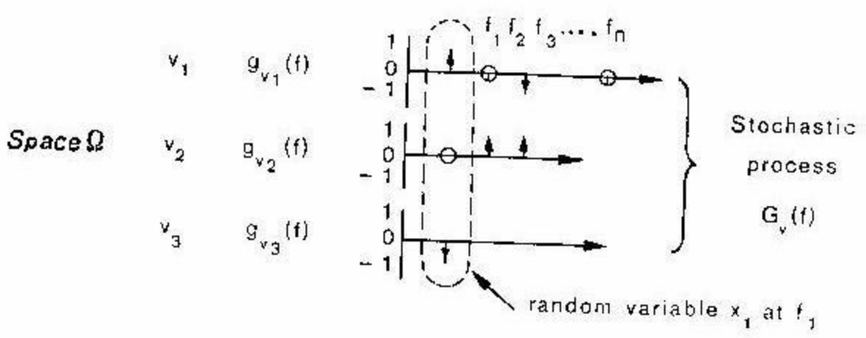


Fig. 1. Illustration of a stochastic process representing a phonological system (hypothetical)



This says that vowels:

$v_1 = \{[+\text{dist. feat. 1}], [0 \text{ dist. feat. 2}], [-\text{dist. feat. 3}]\}$

$v_2 = \{[0 \text{ dist. feat. 1}], [+\text{dist. feat. 2}] \ldots\}$

etc.

The process is now a collection of $n$ random variables $(x_1, x_2, \ldots x_n)$ each one defined at $f_1, f_2, \ldots f_n$. Each random variable represents a distinctive feature.

A *complete description* of the process is given when the joint probability distribution function of these random variables is specified:

$$p(x_1, x_2, \ldots x_n) \tag{1}$$

There are many partial descriptions, all interesting, some more interesting than others.

1) Marginal p.d.f.'s of each random variable:

$$p(x_1),\ p(x_2) \ldots p(x_n) \tag{2}$$

2) Joint probabilities of random variables taken two at a time:

$$p(x_1, x_2),\ p(x_1, x_3) \ldots p(x_1, x_n)$$
$$p(x_2, x_3) \ldots p(x_2, x_n) \tag{3}$$
$$p(x_{n-1}, x_n)$$

3) Joint probabilities taken three at a time:

$$\text{e.g. } p(x_1, x_2, x_3), \ldots \tag{4}$$

4) Conditional probability distribution functions of many types:

a) $p(x_i/x_j)$ i. e. given $x_i$, what is the probability that it will co-occur with $x_j$.

b) $p(x_i/x_j, x_k)$ etc. (5)

All of the above are specifiable by knowing (1).

Furthermore, moments can also be obtained in the established procedure.

## 3. Some possible uses of the model

Having now described sound patterns in such terms we can perform many operations which in traditional phonology are difficult or impossible. For instance, we can compare two, or many more patterns and measure their similarity (conversely distance) very precisely. We can thus find the mathematical (systems) properties of various phonological types. In a previous paper (1969a), a type of conditional probabilities for distinctive features was calculated for several Balkan vocalic systems and the resulting matrices cross correlated, then factor analysed to establish grouping. We can study the informational properties of a system diachronically and see how it changes and how exactly its various stages relate to each other. Or, as in the present model we can look into the mathematical properties which characterize diachronically stable systems.

Here the *entropy* of the random process was examined. This measures roughly informational efficiency of a given system. It was assumed that stable systems should display extremal value entropies. As a test case, the Latin

pattern (without the length feature) and the Modern Greek pattern were analysed by eventually taking into consideration frequencies of occurrence for each vowel.

## 4. Definition of entropy

Consider a discrete random variable X with range $x_1, x_2, \ldots x_n$ with probability density (or mass) function (p.d.f.) $p(x_i)$. The entropy of X given by:

$$H(X) = \sum_{i=1}^{n} p(x_i) \ln \frac{1}{p(x_i)} = -\sum_{=1}^{n} p(x_i) \cdot \ln p(x_i) \tag{6}$$

where ln x is logarithm to base 2[2].

The interpretation of H(X) is average information (if outcomes of X, $x_i$ are observed) or average uncertainty about X (if outcomes are not observed). However, some other interpretations can also be valid, based on some of the properties of this functional H(X), and these will be discussed later.

Consider now the situation when we deal with an n-dimensional random variable model $(x_1, x_2, \ldots x_n)$ which is what our phonological process is. In this case we have various kinds of entropies which can be defined depending on which p.d.f.'s one uses ((1), (2) or (3) etc.), all of which have interesting interpretations for our model.

The most general one is of course the entropy of the entire model i.e.

$$H(X_1, X_2, \ldots X_n) = \sum_{ijk\ldots n} p(x_i, x_j, \ldots x_n) \cdot \ln p(x_i, \ldots x_n) \tag{7}$$

$$\text{where} \sum_{ijk\ldots n} \text{means} \sum_{i=1}^{m} \cdot \sum_{j=1}^{m} \cdot \sum_{k=1}^{m} \ldots \sum_{n=1}^{m}$$

in the above notation each random variable is discrete with values (0, 1, 2, ...m) and there are *n* random variables.

One may also define higher order of Entropies using the other p.d.f.'s defined by (2), (3) etc. All of these yield interpretations for our model, in fact they probably have more than the overall entropy (5) as we will discuss below, when we apply it to our language model.

## 5. Application to our model (zero order Entropies)

We now apply the above concepts to our phonological model. The entropies obtained are called "zero order" entropies since the model is the simplest possible, i.e. since it does not take into account other features such as frequency of vowels, universality of the vowel pattern, etc.

In order to illustrate the situation we also present an example, simultaneously with the development.

Consider the language called L with 3 vowels $v_1, v_2, v_3$ and three features $f_1, f_2, f_3$.

[2] p (x) is P ($X_i$) now given so as to show discreteness of the model.



Assume further that each vowel has the values on $f_1, f_2, f_3$ as shown on Figure 2:

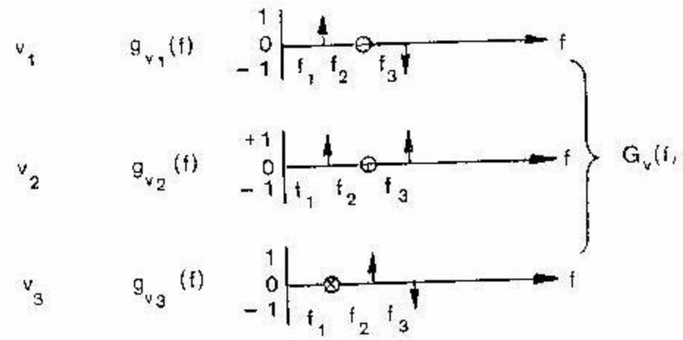


Fig. 2. A random process representing a sound pattern

The joint p.d.f. of the random variables $x_1, x_2, x_3$ is specified as follows (vowels are assumed equiprobable):

$$\begin{aligned} p(x_i, x_j, x_k) = p(1, 0, -1) &= 1/3 \\ p(1, 0, 1) &= 1/3 \\ p(0, 1, -1) &= 1/3 \end{aligned} \tag{8}$$

all other combinations of values have probability zero

where p (1, 0, −1) means probability that $x_1=1$, $x_2=0$ and $x_3=-1$.

The above specification is sufficient in order to find the overall entropy of the process by applying expression (7). Here it is very simply:

$$\begin{aligned} H(x_1, x_2, x_3) &= \frac{1}{3} \ln 3 + \frac{1}{3} \ln 3 + \frac{1}{3} \ln 3) \\ &= \ln 3 = 1.585 \end{aligned} \tag{9}$$

since the summation is very simple. In this case the total entropy per random variable is:

$$\frac{H(x_1, x_2, x_3)}{3} = \frac{1.585}{3} = 0.5283 \tag{10}$$

It should be noted that the total entropy (9) is equal to the entropy of the vowel structure of our system since there is a one to one mapping between the vowels and the realization of our joint model. Thus, the interpretation of (9) is both total entropy of the feature process and entropy of the vowels. However, the value (10) is total entropy per feature, that is, a different concept, since it gives an idea of the average uncertainty about features not vowels.

Since it is our intent to make statements about features and not vowels, the partial descriptions and their entropies take on a useful form.

Let us define and compute the marginal entropies of each feature. To this purpose we need the marginal p.d.f.'s of each feature. They are:

$$p(x_1) = \begin{cases} p(1) = 2/3 \\ p(-1) = 0 \\ p(0) = 1/3 \end{cases}$$
$$p(x_2) = \begin{cases} p(1) = 1/3 \\ p(-1) = 0 \\ p(0) = 2/3 \end{cases} \quad (11)$$
$$p(x_3) = \begin{cases} p(1) = 1/3 \\ p(-1) = 2/3 \\ p(0) = 0 \end{cases}$$

The entropies of each feature are:

$$H(x_1) = -\left(\frac{2}{3}\ln\frac{2}{3} + \frac{1}{3}\ln\frac{1}{3}\right) \quad (12)$$
$$= \frac{2}{3}\ln\frac{3}{2} + \frac{1}{3}\ln 3 = 0.918$$

and the same value is also for $H(x_1)$ and $H(x_3)$ in language L. The interpretation of these entropies is as average uncertainty of each feature. It is not the same as total entropy per feature [expression (10)], since in that case the features were taken jointly whereas here, they are taken individually.

Finally, we define entropies of random variables taken two at a time, i.e. joint partial entropies.

For example $H(x_1, x_2)$ for Fig. 2 is obtained as follows. First $p(x_1, x_j)$ or $p(x_1, x_2)$:

$$p(x_1, x_2) = p(1, 0) = 2/3$$
$$p(0, 1) = 1/3 \quad (13)$$
$$\text{all others zero.}$$

$$p(x_2, x_3) = p(0, -1) = 1/3$$
$$p(0, 1) = 1/3 \quad (14)$$
$$p(1, -1) = 1/3$$
$$\text{all others zero.}$$

$$p(x_1, x_3) = p(1, -1) = 1/3$$
$$p(1, 1) = 1/3$$
$$p(0, -1) = 1/3 \quad (15)$$
$$\text{all others zero.}$$

Then we calculate:

$$H(x_1, x_2) = 0.918 \quad (16)$$
$$H(x_2, x_3) = H(x_1, x_3) = 1.585$$

We now form the following table for our sound pattern:

| $\frac{H(x_1, x_2, x_3)}{n}$ | $H(x_1)$ | $H(x_2)$ | $H(x_3)$ | $H(x_1, x_2)$ | $H(x_2, x_3)$ | $H(x_1, x_3)$ |
|---|---|---|---|---|---|---|
| 0.5283 | 0.918 | 0.918 | 0.918 | 0.918 | 1.585 | 1.585 |



The hypothesis which we formulate is that stability of a structure is reflected in stationary values of these entropies (either minimum or maximum values). If a language uses one vowel or even none, all the entropies are zero and this represents a stable structure. However, if a language uses more than one vowel, it seems natural to expect that the entropies reaching large values represent stable structures since such a tendency tends to eliminate redundancy.

Thus, one must gradually begin to consider other effects, for a more adequate study of stability.

Nevertheless, on the basis of the above primitive model comparisons of languages as regards their stability can indeed be made, even though they are rather superficial. But when they are made, all other effects are really assumed equally likely (zero order entropies). Of two languages, for instance, with the same number of vowels and features, the one with the higher entropies must be the more stable.

## 6. Higher order entropies

Here, account is taken of varying vowel frequencies, in computing entropies. Let us assume that the vowels have probabilities, i.e. we have a random variable called $v$ (vowel) with,

$$p(v_1) = 2/3$$
$$p(v_2) = 1/6 \quad (17)$$
$$p(v_3) = 1/6$$

The total vowel entropy is:

$$H(v) = -\left(\frac{2}{3}\ln\frac{2}{3} + \frac{2}{6}\ln\frac{1}{6}\right) \quad (18)$$

To find now all the feature entropies (say first-order type) we simply have to recalculate the p.d.f.'s based on these values.

To find $p(x_1, x_2, x_3)$ we use:

$$p(x_1, x_2, x_3) = \sum_v p(x_1, x_2, x_3 / v) \cdot p(v)$$

In this case, it trivially becomes:

$$p(1, 0, -1) = 1 \cdot p(v_1) = 2/3$$
$$p(1, 0, 1) = 1 \cdot p(v_2) = 1/6$$
$$p(0, 1, -1) = 1 \cdot p(v_3) = 1/6$$

and the total entropy per feature is:

$$\boxed{\frac{H(v)}{3}}$$

Next we compute $p(x_1)$, $p(x_2)$, $p(x_3)$. First $p(x_1)$:

$$p(1) = p(1/v_1) \cdot p(v_1) + p(1/v_2) \cdot p(v_2) + p(1/v_3) \cdot p(v_3) =$$
$$= 1 \cdot \frac{2}{3} + 1 \cdot \frac{1}{6} + 0 \cdot \frac{1}{6} = \frac{2}{3} + \frac{1}{6} = \frac{5}{6}$$

$$p(0) = 1 \cdot \frac{1}{6} = \frac{1}{6}$$

$$p(-1) = 0.$$

Likewise we compute all the others.

## 7. Latin and Modern Greek

As an application we shall consider the two languages Greek and Latin. Since both use the same number of vowels it would be interesting to compare them. Furthermore since we have data on the frequency of the vowels for both, we shall compute the entropies which take vowel frequency into account.

(1) *Greek*:

Greek uses five vowels i, e, a, o, u. Rename them $v_1$, $v_2$, $v_3$, $v_4$, $v_5$. The p.m.f. of these vowels estimated on the basis of the relative frequency interpretation of probability is as follows:

$$p(v_1) = 0.268 \qquad p(v_2) = 0.06 \qquad p(v_3) = 0.2053$$
$$p(v_4) = 0.1916 \quad \text{and} \quad p(v_5) = 0.2751$$

The stochastic process of interest is given below. The features diffuse, compact and grave are denoted by $f_1$, $f_2$, $f_3$.

| | i | e | a | o | u |
|---|---|---|---|---|---|
| diffuse | + | − | − | − | + |
| compact | 0 | − | + | − | 0 |
| grave | − | − | 0 | + | + |

Fig. 3. Distinctive feature matrix for Modern Greek and Latin

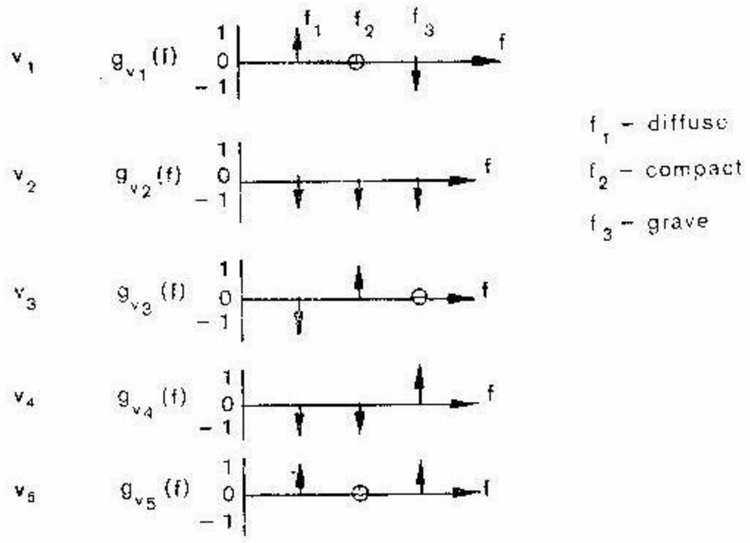


Fig. 4. Stochastic process for the vocalic system of fig. 3



The total entropy of Greek (logarithm to the base 2) is:

$$\begin{aligned} H(\text{vowel}) &= -[0.268 \ln 0.268 + 0.06 \ln 0.06 + 0.2053 \ln(0.2053) + \\ &\quad + 0.191 \ln(0.191) + 0.2751 \ln 0.2751] \\ &= 2.179 \end{aligned}$$

The entropy per feature is then:

$$\boxed{\frac{H(x_1, x_2, x_3)}{3} = .7293}$$

Next, the entropies of each feature i.e. $(H\, x_1)$, $H(x_2)$, $H(x_3)$, are computed.

We first need $p(x_1)$ using, of course, the vowel frequencies.

$$p(x_1) = \begin{cases} p(1) = 0.268 + 0.2751 = .5431 \\ p(0) = 0 \\ p(-1) = .06 + 0.2053 + 0.1916 = .4569 \end{cases}$$

Thus, the entropy for the first feature $f_1$ is:

$$\begin{aligned} H(x_1) &= -0.5431\ (\ln 543 - \ln 1000) - 0.4569\ (\ln 457 - \ln 1000) \\ &= 0.543\ (9.96 - 9.08) + 0.4569\ (9.96 - 8.84) \\ &= 0.478 + 0.512 = 0.990 \end{aligned}$$

$$\boxed{H(x_1) = 0.990}$$

Next feature $f_2$:

$$p(x_2) = \begin{cases} p(1) = 0.2053 = 0.2053 \\ p(0) = 0.268 + .2751 = 0.5431 \\ p(-1) = 0.06 + 0.1916 = 0.2516 \end{cases}$$

$$\begin{aligned} \text{Thus, } H(x_2) &= 0.2053\ (\ln 1000 - \ln 205) + 0.5431\ (\ln 1000 - \ln 543) + \\ &\quad + 0.2516\ (\ln 1000 - \ln 252) \\ &= 0.468 + 0.478 + 0.498 \end{aligned}$$

$$\boxed{H(x_2) = 1.444}$$

Finally feature $f_3$:

$$p(x_3) = \begin{cases} p(1) = 0.1916 + 0.2751 = .4667 \\ p(0) = 0.2053 = 0.2053 \\ p(-1) = 0.268 + 0.06 = 0.328 \end{cases}$$

$$\begin{aligned} H(x_3) &= .4667\ (9.96 - \ln 468) + 0.2053\ (9.96 - \ln 205) + 0.328\ (9.96 - \ln 328) \\ &= 0.51 + 0.468 + 0.525 = 1.503 \end{aligned}$$

$$\boxed{H(x_3) = 1.503}\ .$$

Final table for Greek:

| $\frac{H(x_1, x_2, x_3)}{n}$ | $H(x_1)$ | $H(x_2)$ | $H(x_3)$ |
|---|---|---|---|
| 0.7293 | 0.990 | 1.444 | 1.503 |

(2) *Latin*:

Latin has presumably the same stochastic phonological process as concerns realizations, but different probability distribution. The vowel p.m.f.'s are (from Table VIII)[3]:

$$\begin{aligned} p(v_1) &= 0.2527 \\ p(v_2) &= 0.1499 \\ p(v_3) &= 0.2719 \\ p(v_4) &= 0.1232 \\ p(v_5) &= 0.2023 \end{aligned}$$

$$\begin{aligned} H(x_1, x_2, x_3) = {} & 0.2527\ (9.96 - \ln 253) = 0.502 \\ & +0.145\ (9.96 - \ln 150) = 0.396 \\ & +0.2719\ (9.96 - \ln 271) = 0.508 \\ & +0.1232\ (9.96 - \ln 123) = 0.372 \\ & +0.2023\ (9.96 - \ln 202) = 0.465 \\ & \overline{2.243} \end{aligned}$$

Thus total entropy is:

$$\boxed{\frac{H(x_1, x_2, x_3)}{3} = 0.747}$$

a value slightly higher than Greek.

Next feature $f_1$:

$$p(x_1) = \begin{cases} p(1) = 0.2527 + 0.2023 = .4550 \\ p(0) = 0 \\ p(-1) = 0.1499 + 0.2719 + 0.1232 = .5450 \end{cases}$$

$$\begin{aligned} H(x_1) = {} & 0.455\ (9.96 - 8.83) = 0.455\ (1.13) = 0.514 \\ & +0.545\ (9.96 - 9.09) = 0.545\ (0.87) = 0.474 \\ & \overline{.988} \end{aligned}$$

$$\boxed{H(x_1) = 0.988}$$

Feature $f_2$:

$$p(x_2) = \begin{cases} p(1) = 0.2719 = 0.2719 \\ p(0) = 0.2527 + 0.2023 = 0.4550 \\ p(-1) = 0.1499 + 0.1232 = 0.2731 \end{cases}$$

$$\begin{aligned} H(x_2) = {} & 0.2719\ (9.96 - 8.09) = 0.509 \\ & +(0.4550)\ (9.96 - 8.83) = 0.514 \\ & +\ 0.2731\ (9.96 - 8.09) = 0.510 \\ & \overline{1.533} \end{aligned}$$

$$\boxed{H(x_2) = 1.533}$$

[3] Krámský (1966).



Feature $f_3$:

$$p(x_3) = \begin{cases} p(1) = 0.3255 \\ p(0) = 0.2719 \\ p(-1) = 0.4026 \end{cases}$$

$$\begin{aligned} H(x_3) = {} & 0.3255\ (9.96 - 8.35) = 0.524 \\ & +0.2719\ (9.96 - 8.09) = 0.509 \\ & +0.4026\ (9.96 - 8.65) = 0.527 \\ & \overline{1.560} \end{aligned}$$

$$\boxed{H(x_3) = 1.560}$$

Final table for Latin:

| $\frac{H(x_1, x_2, x_3)}{3}$ | $H(x_1)$ | $H(x_2)$ | $H(x_3)$ |
|---|---|---|---|
| 0.747 | 0.988 | 1.533 | 1.560 |

## 8. Conclusion

The results are inconclusive, as Latin displayed consistently higher entropies, therefore, lower redundancy, therefore according to our assumption higher stability.

Perhaps, as has been often assumed, average value for Entropy is the most stable, as it incorporates aspects of both speaker and listener.

But linguistic diachrony does not proceed in a sociolinguistic vacuum. It is most likely that the particular pattern of prosodic features (stress and length) made Latin vowels more unstable, but also that the sociolinguistic situation (wide-spread bilingualism, often only incipient or subordinate) would have required higher systematic redundancy.

A more final answer will have to wait. However with as rigorous and flexible a model, solutions to different problems can be actively, and successfully sought, and long standing riddles in diachronic phonology finally answered

## REFERENCES

Afendras, Evangelos A. 1968. *The Balkans as a Linguistic Area: A Study in Phonological Convergence.* Unpublished Ph. D. Dissertation. The Johns Hopkins University, Baltimore.

— 1969a. "Mathematical Models for Balkan Phonological Convergence". Paper read at the *Third International Conference on Computational Linguistics,* Sänga-Säby, Sweden, September 1–4, 1969. (Preprint No. 56 of papers).

— 1969b. "Quantification of Balkan Phonological Distribution. Paper read at the Third Annual Meeting of the *Societas Linguistica Europaea,* Vienna, September 6, 1969.

Afendras, E. and N. S. Tzannes. 1969c. "Stochastic Models of Sound Pattern Stability". Paper read at the *Second International Congress of Applied Linguistics*, Cambridge, England, September 9, 1969.

Avram, A. 1964. "Sur la typologie phonologique quantitative". *Revue Roumaine de Linguistique. IX* 131–134.

Borko, H. (Ed.) 1967. *Automated Language Processing: The State of the Art.* New York: Wiley.

Brøndal, V. 1936. "La structure des systèmes vocaliques". Paper read at the *Cercle Linguistique de Prague.*

Cramer, H. and M. R. Leadbetter 1967. *Stationary and Related Stochastic Processes.* New York: Wiley.

Edmundson, H. P. 1967. "Mathematical Models in Linguistics and Language Processing". In Borko, (ed.).

Hockett, C. F. 1955. *A Manual of Phonology.* Baltimore: Waverley Press, Inc.

— 1967. *Language, Mathematics and Linguistics.* The Hague: Mouton.

Jaynes, E. T. 1968. "Prior Probabilities". *IEEE Transactions on Systems Science and Cybernetics.* Vol. SSC-4, No. 3 (September).

Krámský, J. 1966. "The Frequency of Occurrence of Vowel Phonemes in Languages Possessing Vowel Systems of Identical Structure". *Prague Studies in Mathematical Linguistics 1.* 13-31.

Kučera, H. and G. K. Monroe. 1968. *A Comparative Quantitative Phonology of Russian, Czech and German.* New York: American Elsevier (A rather extensive bibliography with references to Kučera's other works is appended).

Papoulis, A. 1965. *Probability Theory, Random Variables and Stochastic Processes.* New York: McGraw Hill (Ch. 2, 4, 9).

Trubetzkoy, N. S. 1929. "Zur allgemeinen Theorie der phonologischen Vokalsysteme". *TCLP 1,* 39–67.

— 1949. *Principes de Phonologie.* (French translation of J. Cantineau). Paris: Klincksieck.

Voegelin, C. F. 1956. "The Scope of whole System (Distinctive Feature) and Subsystem Typologies". *Word 12.* 444 ff.

—, F. M. Voegelin, S. Wurm, O'Grady and T. Matsuda. 1963. "Obtaining an index of phonological differentiation from the construction of non-existent Minimax systems". *IJAL 29.* 5–28.