A 'Tween 80' water extract of egg-yolk was prepared by extracting 10 gm. of egg-yolk in 100 ml. of 6 per cent aqueous solution of 'Tween 80' with 1 gm. of active charcoal, at 120° C. for 20 min. This egg extract was a yellowish transparent solution and very thermostable as regards its growth-promoting property. This was added to the basal medium. The medium was sterilized by autoclaving at 120° C. for 20 min. without adjustment of *p*H.

Culture results with the medium are shown in Table 1, compared with those with Dubos 'Tween' albumin medium. At the fourteenth day of incubation, the 10^{-7} mgm. inoculum of *H37Rv* strain of *Mycobacterium tuberculosis* had grown equally well in the yolk extract medium and in the 'Tween' albumin medium. Similar results were also obtained when several clinical strains isolated from tuberculosis patients were tested.

Table 1. Mean Relative Grade of Growth at Fourteenth Day of Incubation

Culture medium	Inoculum size (mgm. in 2 ml. medium) 10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Yolk extract medium	+++	+++	+++	++	+
	+++	+++	++	++	+
Medium without yolk extract	+++	+	−	−	−
	+++	−	−	−	−
'Tween' albumin medium	+++	+++	++	++	+
	+++	+++	++	++	+

The culture grown in this medium showed a rather dispersed growth, and could therefore be employed as the inoculum for various growth and animal experiments without grinding up the culture. On the other hand, the resistance of many clinical strains to various chemotherapeutic agents could be determined in five days, employing this medium (Imamura, unpublished work). This medium is very convenient from the point of view of simplicity of preparation and the dispersibility of cultures grown on it.

Isao Yamane
Kunio Imamura

Department of Bacteriology,
Fukushima Medical College,
Fukushima, Japan.
Nov. 18.

[1] Kirchner, *Zbl. f. Bakt.*, **124**, 403 (1932).
[2] Dubos and Davis, *J. Exp. Med.*, **83**, 409 (1946).

Distribution of Word Frequencies

The purpose of this communication is to explain, in terms of the theory of information, the implications of the Zipf distribution of word frequencies[1]. The distribution is formally identical with the Pareto income and Willis taxonomic distributions, but the present discussion is restricted to word frequencies. The discussion resembles that of Mandelbrot[2] but is simpler. The discussion by Parker-Rhodes and Joyce[3] also resembles Mandelbrot's, but is fallacious.

Let the population frequencies of the distinct words of a language, arranged in decreasing order, be $p_1, p_2, p_3, \ldots$, where $\sum_r p_r = 1$. Let $\varphi(x)$ be the density of words at population frequency x, so that $\int_{p_1}^{p_r} \varphi(x)\,dx = r$. The Zipf distribution can be described either by $\varphi(x) \propto x^{-2}$ or by $p_r \propto 1/(r+a)$, where a is independent of r. If we wish, for formal reasons, to allow r to be unbounded, then it is desirable to introduce a convergence factor[4] and to write $p_r = b\exp(-\varepsilon r)/(r+a)$, where $\varepsilon > 0$ is very small and b is chosen so that $\Sigma p_r = 1$.

Let the effort required to extract the rth word from the memory be $f(r)$. The expected amount of information[5] per word is $-\sum_r p_r \log p_r$, so that the expected amount of information per unit of effort is $I = -(\Sigma p_r \log p_r)/(\Sigma f(r) p_r)$. We now assume, not Zipf's distribution, but his principle of least effort for the development of languages. Its natural interpretation is that the population frequencies p_r are such as to minimize I, subject to the restraint $\Sigma p_r = 1$. It follows by Lagrange's method of undetermined multipliers that $p_r = \exp(-cf(r))$, where the positive number c is such that the restraint is satisfied. If the Zipf distribution is true, it follows that $f(r)$ is proportional to $\varepsilon r + \log(r+a)$. This, then, is the interpretation of the Zipf distribution. The term εr is not to be taken too seriously.

If, on the other hand, we make the assumption of Parker-Rhodes and Joyce[3] that $f(r)$ is proportional to r, then p_r would be 2^{-r}, which is inconsistent with the Zipf distribution. It is interesting that the distribution of the number of 'strokes' in connected symbols in Pitman's shorthand is close to the distribution defined by $p_r = 2^{-r}$, according to a frequency count given by Herdan[6]. For this shorthand example, $f(r)$ probably is approximately proportional to r.

I. J. Good

25 Scott House,
Princess Elizabeth Way,
Cheltenham.

[1] Zipf, G. K., "The Psychobiology of Language" (Boston, 1935). "Human Behaviour and the Principle of Least Effort" (Cambridge, Mass., 1949).
[2] Mandelbrot, B., in "Communication Theory", 486, edit. by Willis Jackson (London, 1953); "Information Theory: Third London Symposium", 135 (edit. by E. C. Cherry, London, 1956).
[3] Parker-Rhodes, A. F., and Joyce, T., *Nature*, **178**, 1308 (1956).
[4] Corbet, A. S., Fisher, R. A., and Williams, C. B., *J. Anim. Ecol.*, **12**, 42 (1943). Good, I. J., *Biometrika*, **40**, 237 (1953).
[5] Shannon, C. E., *Bell System Tech. J.*, **27**, 379 and 623 (1948). Wiener, N., "Cybernetics", 75 (New York and Paris, 1948). Good, I. J., "Probability and the Weighing of Evidence", 75 (London and New York, 1950); *Proc. Inst. Elect. Eng.*, C, **103**, 200 (1956).
[6] Herdan, G., "Language as Choice and Chance" (Groningen, 1956).

First, we must apologize for having evidently given the impression in our original communication, especially in our assumption (iv), that we were intending to use information theory in treating the problem. Our aim, however, was to attack the problem without recourse to the assumptions of this theory. These necessarily involve regarding language as consisting of messages, and the words as being, for all statistical purposes, packets of information. This model is unrealistic as applied to real language situations, on the basis of which the faculty of speech is learned and which even in the most sophisticated communities dominate its natural evolution. We accordingly sought for some approach which would avoid the need for the concept of 'information' in the technical sense; our condition (iv) should have been put in the form: "Languages will tend to evolve in such a way that the time required to recall