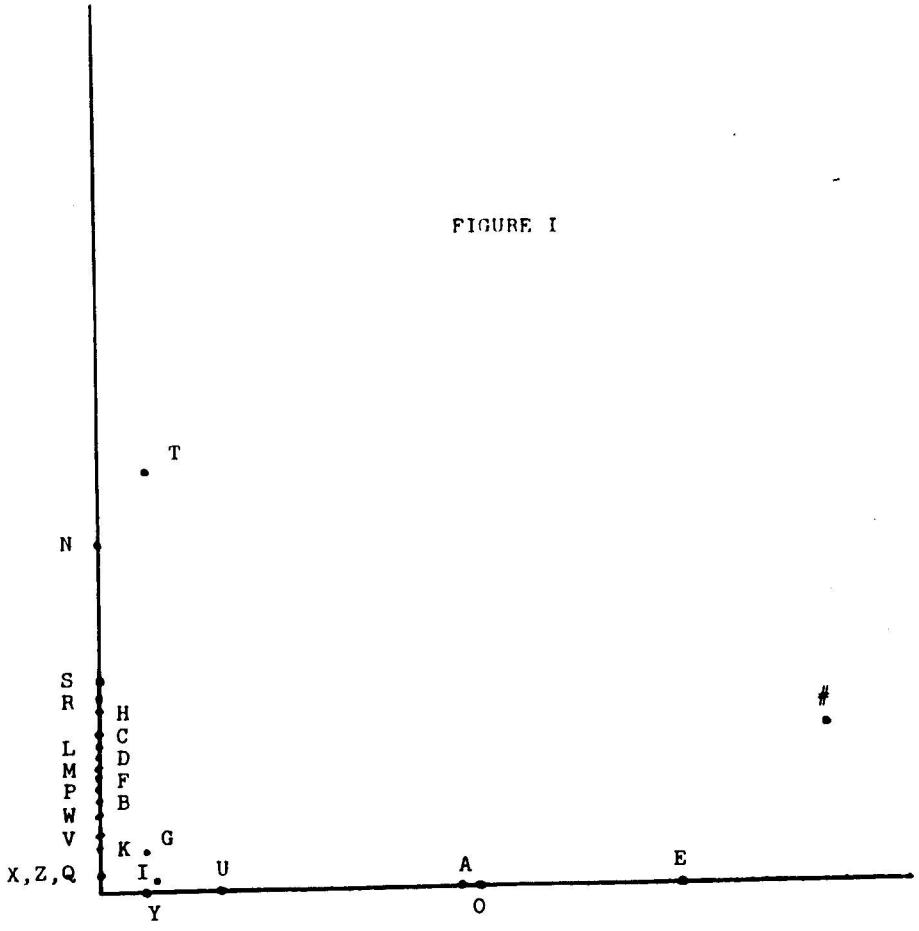
FIGURE I



This may be expressed by saying that, given a long sequence of English text, there are, on average, about two possibilities for the very next letter. Since entropy is usually measured in $\log_2$ the entropy of English is then about 1. It is possible, using the following formulas, to compare the information (or resolution of uncertainty) in our models with Shannon's estimate: We let $H(X)$ denote the uncertainty (entropy) in the Hidden Markov Chain, $H(Y)$ the uncertainty in the output of our model for English, $H(X|Y)$ the uncertainty in the Hidden Markov Chain given a sequence of text, $H(Y|X)$ the uncertainty in the text given the Hidden Markov Chain, and $H(X,Y)$ the uncertainty in the joint process. The following formulas are valid:

$$H(X|Y) = H(X, Y) - H(Y) \tag{1}$$

$$H(Y|X) = H(X, Y) - H(X) . \tag{2}$$

Subtracting gives

$$H(Y) = H(Y|X) + H(X) - H(X|Y) .$$

Now each term on the right may be calculated independently (the first two precisely, and the third estimated). When this is done we may compare the uncertainty in our