(16 July 1996) I am indebted to Prof. Valdemar C. da Rocha, Jr., for pointing out a blunder in the proof of Proposition 4–the expression for $H(V|U = u)$ does not increase monotonically with $P(U = u)$ as claimed. There follows a correct proof of this proposition.

Let $p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}$ be the probability distribution $P(V = .|U = u)$. From the manner in which the homophones are generated, it follows that

$$p_{i+1}^{(1)} \leq p_i^{(1)}/2 \quad \text{for} \quad 1 \leq i < L^{(1)}$$

which implies further that

$$p_1^{(1)} > 1/2.$$

But, by property 3 of entropy in Section 6 of Shannon's 1948 paper (or, equivalently, by the "leaf-entropy theorem"),

$$H(p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}) = H(p_1^{(1)}, 1 - p_1^{(1)}) + (1 - p_1^{(1)}) H(p_1^{(2)}, p_2^{(2)}, \ldots, p_{L^{(2)}}^{(2)})$$

where

$$p_i^{(2)} = p_{i+1}^{(1)}/(1 - p_1^{(1)}) \quad \text{and} \quad L^{(2)} = L^{(1)} - 1.$$

Thus

$$H(p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}) < 1 + \frac{1}{2} H(p_1^{(2)}, p_2^{(2)}, \ldots, p_{L^{(2)}}^{(2)}) \quad \text{bits}.$$

But the probability distribution $p_1^{(2)}, p_2^{(2)}, \ldots, p_{L^{(2)}}^{(2)}$ inherits the above properties of $p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}$, namely

$$p_{i+1}^{(2)} \leq p_i^{(2)}/2 \quad \text{for} \quad 1 \leq i < L^{(2)} \quad \text{and} \quad p_1^{(2)} > 1/2.$$

Thus, we may apply the same argument again to obtain (in obvious notation)

$$H(p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}) < 1 + \frac{1}{2}[1 + \frac{1}{2} H(p_1^{(3)}, p_2^{(3)}, \ldots, p_{L^{(3)}}^{(3)})].$$

Repeating this argument a total of $L^{(1)} - 2$ times gives

$$H(p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}) < 1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^{L^{(1)}-3}} + \frac{1}{2^{L^{(1)}-2}} H(p_1^{(L^{(1)}-1)}, p_2^{(L^{(1)}-1)})$$

and hence

$$H(p_1^{(1)}, p_2^{(1)}, \ldots, p_{L^{(1)}}^{(1)}) < 1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^{L^{(1)}-3}} + \frac{1}{2^{L^{(1)}-2}} < 2 \quad \text{bits}.$$