- let $x = \alpha_1\alpha_2 ... \alpha_n$ a sequence of length n generated by a iid source and Q(x) = the probability to see such a sequence

- say LempelZiv breaks into c phrases $x = y_1 y_2 ... y_c$ and call $c_l = \#$ of phrases of length l then $-\log Q(x) \geq \sum_l c_l \log c_l$
(proof) $\sum_{|y_i|=l} Q(y_i) < 1$ so $\prod_{|y_i|=l} Q(y_i) < (\frac{1}{c_l})^{c_l}$

- if p_i is the source probab for α_i then by law of large numbers x will have roughly np_i occurrences of α_i and then
$logQ(x) = -\log \prod_i p_i^{np_i} \approx n \sum p_i \log p_i = nH_{source}$

- note that $\sum_l c_l \log c_l$ is roughly the LempelZiv encoding length so th einequality reads
$nH \geq\approx LZencoding$ which is to say $H \approx\geq LZrate$.