Conditional Entropy

- **The conditional entropy of a random variable Y given another X, expresses how much extra information one still needs to supply on average to communicate Y given that the other party knows X**

$$H(Y \mid X) = \sum_{x \in \Omega_X} p(x) H(Y \mid X = x)$$

$$= - \sum_{x \in \Omega_X} p(x) \sum_{y \in \Omega_Y} p(y \mid x) \log p(y \mid x)$$

$$= - \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} p(x, y) \log p(y \mid x) \ = - E\big(\log p(Y \mid X)\big)$$

- **H(Topic| "computer") vs. H(Topic | "the")?**