The source compression problem

- **Example:** "There are no people like show people"

  $\overset{\text{encode}}{\rightarrow} x \in \{0, 1\}^n$

  $\overset{\text{decode}}{\rightarrow}$ "there are no people like show people"

- **Lossless:** Message reconstructed perfectly.

- **Goal:** minimize expected length $E(n)$ of coded message.

- Can we do better than $\lceil \log_2(26) \rceil = 5$ bits per character?

- **Basic idea:** Use short codes for common messages.