# The source compression problem

- **Example:** “There are no people like show people”
  $\overset{\text{encode}}{\longrightarrow} x \in \{0,1\}^n$
  $\overset{\text{decode}}{\longrightarrow}$ “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.
- **Stream compression:**
  - Message revealed one character at a time.
  - Code generated as message is revealed.
  - Decoded message is constructed gradually.
- Easier than block codes when processing long messages.
- A natural way for describing a distribution.