Задача декодування.

1. Зробити частотний аналіз тексту великого обсягу і побудувати матрицю

$$A = (A_{ij})_{i,j=1,33}$$

$$A_{ij} \approx P(X_{n+1} = j \mid X_n = i), \quad \begin{matrix} а & б & я \\ i = 1, 2, \dots, 33 \\ j = 1, 2, \dots, 33 \end{matrix}$$

де X_n – n-та буква в тексті;

2. Закодувати частину <u>цього ж</u> тексту – так буде більше впевненості у відповідності матриці A закодованому тексту. В процесі навчання матрицю A перекодовувати не потрібно.

3. В якості початкового наближення для початкового розподілу $\mu = (\mu_1, \dots, \mu_{33})$ ланцюга $(X_n)_{n \geq 1}$ можна спробувати рівномірний розподіл

$$\mu^{(0)} = \left(\approx \frac{1}{33}, \dots, \approx \frac{1}{33}\right)$$

або інваріантний розподіл вихідного

$$\mu^{(0)} = \mu^{(0)} \cdot A$$

В процесі навчання μ <u>перекодовувати</u>.

4. В якості початкового значення матриці B можна спробувати

$$B = \begin{pmatrix} \approx \frac{1}{33} & \dots & \approx \frac{1}{33} \\ \approx \frac{1}{33} & \dots & \approx \frac{1}{33} \end{pmatrix}$$

Можна також використати результат задачі про поділ букв алфавіту на голосні/приголосні. А саме:

- розділити укр. алфавіт за текстом, на основі якого робився частотний аналіз, на дві множини - голосних Γ_1 і приголосних Π_1.
- розділити закодований алфавіт за закодованим текстом на дві множини – голосних Γ_2 і приголосних Π_2.
- сформувати $B^{(0)}$ наступним чином

$$B^{(0)}_{ij} = \begin{cases} \approx \frac{1}{|\Gamma_2|}, & \text{якщо } i \in \Gamma_1,\ j \in \Gamma_2 \\ \approx 0, & \text{якщо } i \in \Gamma_1,\ j \in \Pi_2 \\ \approx \frac{1}{|\Pi_2|}, & \text{якщо } i \in \Pi_1,\ j \in \Pi_2 \\ \approx 0, & \text{якщо } i \in \Pi_1,\ j \in \Gamma_2 \end{cases}$$

Тобто формуємо матрицю так, щоб голосні/приголосні незакодованого алфавіту переходили в голосні/приголосні закодованого з імовірностями, близькими до рівномірного розподілу на множини Γ_2 і Π_2 відповідно.