

последовательность из K_J двоичных символов, можно образовать 2^{K_J} возможных кодовых слов, причем $K_J \geq J \log_2(L)$. Следовательно, требуется минимальное целое значение для K_J :

$$K_J = \lfloor J \log_2(L) \rfloor + 1.$$

Теперь среднее число символов кода на один символ источника $K = \frac{K_J}{J}$. При эффективности кодирования увеличивается в J раз: $\frac{H(X)}{K} = \frac{H(X)J}{K_J}$. Взяв J достаточно большим, можно эффективность приблизить к 1.

Такие методы кодирования не приводят к искажениям, т.к. кодирование символов источника или блоков символов в кодовые слова выполняется однозначно (уникально). Эти коды называются **бесшумными**.

Теперь рассмотрим ситуацию, когда только часть L^J блоков символов источника кодируется однозначно. Например, $2^{K_J} - 1$ наиболее вероятных J символьных блоков кодируется однозначно. Остальные $L^J - (2^{K_J} - 1)$ блоков длины J представляются одним оставшимся кодовым словом. Такая процедура кодирования вызывает ошибку декодирования каждый раз, когда источник выдает маловероятный блок. Обозначим через p_e вероятность ошибки декодирования. Шеннон в 1948 г. доказал теорему кодирования источника.

Теорема Шеннона кодирования ДИБП. Пусть X - ансамбль символов ДИБП с конечной энтропией $H(X)$. Блоки из J символов источника кодируются в двоичные кодовые слова длины K_J . Тогда для любого $\varepsilon > 0$ p_e можно сделать сколь угодно малой, если выполняется неравенство

$$K = \frac{K_J}{J} \geq H(X) + \varepsilon \quad (4.12)$$

и J достаточно велико.

2. Кодовые слова переменной длины.

Если символы источника не равновероятны, то более эффективно использовать кодовые слова переменной длины. Пример: код Морзе (19 век). Символам, возникающим более часто, ставятся в соответствие более короткие кодовые слова, а символам, возникающим менее часто, сопоставляются более длинные кодовые слова. Такой метод кодирования, который требует знания вероятностей появления символов источника, называется **энтропийным**.