The 24th Symposium on Information Theory
and Its Applications (SITA2001)
Kobe, Hyogo, Japan, Dec. 4 - 7, 2001

単語単位で系列を出力する情報源に対するLZ78符号のユニバーサル性について
On universality of LZ78 codes for sources which emit data sequence with word unit

石田 崇 * 　　後藤 正幸 * 　　平澤 茂一 †
Takashi ISHIDA 　　Masayuki GOTO 　　Shigeichi HIRASAWA

Abstract— The Lempel-Ziv code is conventional method for universal coding. It was shown that the LZ78 code is asymptotically optimal for the class of stationry sources, and the improved algorithms are used in practice.

In the theory of universal codes, the class for which a code is asymptotically optimal is important. The LZ78 code is proved to be universal for stationary sources.

In this paper, we consider coding for sources which emit data sequence with word unit. This source is called a word-valued source and is generally nonstationary. We show that the LZ78 code is universal for this source.

Keywords— Lempel-Ziv codes, Universal coding, word-valued source

1 はじめに

確率構造が未知のある情報源クラスに対して漸近最良性をもつ情報源符号化をユニバーサル符号という．代表的なユニバーサル符号は Lempel-Ziv(LZ) 符号 [1],[2] であり，その改良型アルゴリズムである compress, gzip, LHA などは，実際の圧縮用ソフトウェアとして広く用いられている．ユニバーサル符号の議論ではユニバーサル性を保証する確率分布のクラスが重要となる．LZ 符号の 1 つである LZ78 符号 [2] は現在までに開発された符号の中で理論的にも実用的にも最も重要な符号である．LZ78 符号は定常情報源クラスに対して漸近最良性が保証されている [8].

また近年，単語単位の情報源 [3],[4],[5] に対するユニバーサル符号の議論がなされている [5],[6],[7]. テキストデータなど実際のデータは単語単位で系列が出力されると仮定できる場合が多いと考えられる事から，このような情報源に対して LZ 符号化アルゴリズムの有効性を議論することが重要である．単語単位の定常情報源は，一般にはシンボル単位で見た時に定常になるとは限らない [5]. すなわち，この情報源は (シンボル単位の) 非定常情報源クラスに位置する情報源であり，従来から考察されてきたよりも広いクラスに属している．

現在，LZ77 符号 [1] については単語単位の定常エルゴード情報源に対してユニバーサル性を有していることが示されている [5]. 一方，LZ78 符号については単語の長さが一定長という制限のある，ブロック単位の定常情報源に対してユニバーサル性が示されている [7].

そこで本研究では，“単語の長さが一定長”という制限をなくした，“任意の長さの単語単位”で系列を出力する情報源を考え，LZ78 符号のユニバーサル性を議論する．その結果，単語単位の定常エルゴード情報源クラスに対して，LZ78 符号がユニバーサルとなることを示す．

2 従来研究

2.1 LZ78 符号

ここでは LZ78 符号について簡単に説明する [2],[8]. なお，本稿を通じて $j \geq i$ に対して x_i^j は部分系列 $x_i x_{i+1} \cdots x_j$ を表すものとする．

情報源アルファベットを $\mathcal{X} = \{0, 1, \ldots, J-1\}$ とする．第 n 番目のシンボルまでの十分長いシンボル系列 $x_1^n = x_1 x_2 \cdots x_n$, $x_i \in \mathcal{X}$ を $p+1$ 個の空でない部分系列

$$x_1^n = x_{n(0)+1}^{n(1)} x_{n(1)+1}^{n(2)} \cdots x_{n(p)+1}^{n(p+1)}$$
$$(n(0) = 0, n(p+1) = n)$$

に次の規則で分解する．ここで，$n(j)$ は j 番目の部分系列の末尾にあるシンボルの番号を表している．

(1) $n(1) = 1$ とし，$x_{n(0)+1}^{n(1)} = x_1$ を第 1 番目の部分系列とする．

(2) 第 j 番目の部分系列 $x_{n(j-1)+1}^{n(j)}$ は最後の部分系列 $x_{n(p)+1}^{n(p+1)}$ を除きそれ以前のどの部分系列とも一致しない．すなわち，$j \leq p$ に対して
$$x_{n(j-1)+1}^{n(j)} \notin \{x_{n(0)+1}^{n(1)}, x_{n(1)+1}^{n(2)}, \cdots, x_{n(j-2)+1}^{n(j-1)}\}.$$

(3) $x_{n(j-1)+1}^{n(j)}$ はその最後 1 シンボルを除くと空系列かあるいは過去のある部分系列に一致する．すなわち
$$x_{n(j-1)+1}^{n(j)} = x_{n(r_j-1)+1}^{n(r_j)} x_{n(j)}, \ x_{n(j)} \in \mathcal{X},$$
となる r_j が存在する．ただし，$r_j \in \{0, 1, \cdots, j-1\}$ であり，$x_{n(-1)+1}^{n(0)}$ は空系列とする．

このような系列の分解を増分分解と呼ぶ．この分解アルゴリズムで得られた正整数の組 $(r_j, n(j))$ によって，j 番目の部分系列 $x_{n(j-1)+1}^{n(j)}$ に対する符号語を

$$\varphi\left(x_{n(j-1)+1}^{n(j)}\right) = r_j J + x_{n(j)} \tag{1}$$

なる整数として定める．ただし，$x_{n(j)} \in \{0, 1, \cdots, J-1\}$,

---

* 東京大学大学院 工学系研究科 環境海洋工学専攻，〒 113-8656 東京都文京区本郷 7-3-1. Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan.

† 早稲田大学 理工学部 経営システム工学科，〒 169-8555 新宿区大久保 3-4-1. School of Science and Engineering, Waseda University, 3-4-1 Okubo Shinjyuku-ku, Tokyo 169-8555, Japan.

$r_j < j$ より

$$0 \leq \varphi(x_{n(j-1)+1}^{n(j)}) \leq (j-1)J + (J-1) = jJ - 1 \quad (2)$$

である．いま，$L_j = \lceil \log_2(jJ) \rceil$ と定義する．ここで，$\lceil x \rceil$ は x 以上の最小の整数である．このとき，$\varphi(n_{n(j-1)+1}^{n(j)})$ は L_j という部分系列の番号 j だけに依存する桁数の 2 進数で表現できる．そして系列 x_1^n 全体に対する符号語は，これらの 2 進列を連接したものとする．系列 x_1^n を LZ78 符号で符号化したときの符号長を $L_{LZ}(x_1^n)$ で表す．このとき，$L_{LZ}(x_1^n) = \sum_{j=1}^{p+1} L_j$ である．

2.2 単語集合上に確率分布を持つ情報源 [5]

対象となる情報源アルファベットを有限とし，$\mathcal{X} = \{0, 1, \cdots, J-1\}$ とする．また，$\mathcal{X}$ 上を値域とする確率変数を X とし，その実現値を x とする．さらに，情報源から出力される長さ n の系列を $x_1^n = x_1 x_2 \cdots x_n$，$x_i \in \mathcal{X}$ とし，その確率分布を $P(x_1^n) = \Pr(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$ と記述する．

単語単位で系列を出力する情報源として “単語集合上に確率分布を持つ情報源 [5]” を仮定する．まず単語を定義する．

定義 1 [5] $\mathcal{X}$ の要素を有限個連結したものを単語と呼び，w で表す．単語の集合を $\mathcal{W}$ で表し，$\mathcal{W}$ を値域とする確率変数を W で表す．また，$\forall w \in \mathcal{W}$ が $\mathcal{W}$ 内の他の全ての単語の語頭になっていないとき，$\mathcal{W}$ を語頭条件をみたす単語集合 (語頭単語集合) と呼ぶ． □

情報源から出力される長さ n_w の単語系列を $w_1^{n_w} = w_1 w_2 \cdots w_{n_w}$ とし，その確率分布を $P(w_1^{n_w}) = \Pr(W_1 = w_1, W_2 = w_2, \cdots, W_{n_w} = w_{n_w})$ と記述する．

情報源から出力される単語系列 $w_1^{n_w} = w_1 w_2 \cdots w_{n_w}$ は，$n = |w_1| + |w_2| + \cdots + |w_{n_w}|$ のとき $\mathcal{X}$ のシンボルで記述すると $x_1^n = w_1^{n_w}$ となる．ただし，$|w|$ は単語 w の長さを表す．

例 1 $\mathcal{X} = \{0, 1\}$ において，$\mathcal{W} = \{0, 10, 110, 111\}$ は語頭条件をみたす単語集合となっている．また，出力系列 110010 は

$$110\ 0\ 10 = \begin{cases} w_1 w_2 w_3 & = \quad w_1^3 \\ x_1 x_2 x_3 x_4 x_5 x_6 & = \quad x_1^6 \end{cases}$$

とそれぞれ記述される．ここで，

$$\begin{aligned} n_w &= 3 \\ n &= |w_1| + |w_2| + |w_3| \\ &= 3 + 1 + 2 \; = \; 6 \end{aligned}$$

である． □

情報源アルファベット $\mathcal{X}$ を有する任意の情報源を $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ で表す．このとき，もし

$$\begin{aligned} H(\boldsymbol{X}) &= \lim_{n \to \infty} \frac{1}{n} H_n(X_1^n) \\ &= \lim_{n \to \infty} \left[-\frac{1}{n} \sum_{x_1^n \in \mathcal{X}_1^n} P(x_1^n) \log P(x_1^n) \right] \quad (3) \end{aligned}$$

が存在すれば，これを情報源 $\boldsymbol{X}$ のエントロピー・レートと呼ぶ．同様に，単語集合 $\mathcal{W}$ を情報源アルファベットとする情報源 $\boldsymbol{W} = \{W_i\}_{i=1}^{\infty}$ のエントロピー・レートを

$$\begin{aligned} H(\boldsymbol{W}) &= \lim_{n_w \to \infty} \frac{1}{n_w} H_{n_w}(W_1^{n_w}) \\ &= \lim_{n_w \to \infty} \left[-\frac{1}{n_w} \sum_{w_1^{n_w} \in \mathcal{W}_1^{n_w}} P(w_1^{n_w}) \log P(w_1^{n_w}) \right] \end{aligned} \quad (4)$$

とする．

いま，単語単位で系列を出力する情報源に対して，情報源アルファベットを $\mathcal{W}$，$\mathcal{X}$ とみたときのエントロピー・レートをそれぞれ $H(\boldsymbol{W})$，$H(\boldsymbol{X})$ とし，以降これを “単語単位のエントロピー・レート”，“シンボル単位のエントロピー・レート” とよぶ．ここで $H(\boldsymbol{W}) \geq H(\boldsymbol{X})$ である．本稿では特に断らない限り対数の底は 2 とする．

補題 1 [5] 有限語頭単語集合 $\mathcal{W}$ に対して，確率分布 $P(W_1^{n_w})$ が n_w に関して定常エルゴードであれば，シンボル単位のエントロピー・レート $H(\boldsymbol{X})$ は常に存在し，

$$H(\boldsymbol{X}) = \frac{H(\boldsymbol{W})}{E[|W|]} \quad (5)$$

で与えられる．ただし，$E[|W|]$ は平均単語長レートで

$$E[|W|] = \lim_{n_w \to \infty} \frac{1}{n_w} E\left[\sum_{i=1}^{n_w} |W_i| \right] \quad (6)$$

で与えられる． □

3 LZ78 符号のユニバーサル性

ここでは，LZ78 符号を単語単位の情報源クラスに適用した場合の漸近的性能について議論する．ただし，情報源の単語集合 $\mathcal{W}$ は有限語頭単語集合であるとする．ここで，情報源は系列を単語単位で出力するが，符号化アルゴリズムは系列をシンボル単位で符号化する点がポイントとなる．すなわち，シンボル単位の定常情報源クラスに対してユニバーサル性が保証されている LZ78 符号化法が，そのままのアルゴリズムで単語単位の定常エルゴード (シンボル単位では一般に非定常) 情報源に対してもユニバーサル性を有することを以下で示す．

まず，情報源から出力される系列 $x_1^n (= w_1^{n_w})$ の複雑量 $c(x_1^n)$ を以下のように定義する．

定義 2 (系列の複雑量 [8]) 系列 $x_1^n = x_1 x_2 \cdots x_n$ を空でない相異なる部分系列に分解した時の最大個数を系列 x_1^n の複雑量 $c(x_1^n)$ とする． □

系列 x_1^n の複雑量 $c(x_1^n)$ の上界は次のように与えられることが証明されている．

補題 2 [8] すべての $n = 1, 2, \cdots$ とすべての x_1^n に対して，

$$c(x_1^n) < \frac{n}{(1-\varepsilon) \log_J n} \quad (7)$$

が成立する．ここで，ε は $n \to \infty$ のとき $\varepsilon \to 0$ を満たす． □

いま，系列 $w_{-k_w+1}^{n_w} = x_{-k+1}^{n}$ を固定して考える．ただし，$n = |w_1| + |w_2| + \cdots + |w_{n_w}|, k = |w_0| + |w_{-1}| + \cdots + |w_{-k_w+1}|$ である．

シンボル系列 x_1^n が c 個の相異なる部分系列に $x_1^n = x_{n(0)+1}^{n(1)} x_{n(1)+1}^{n(2)} \cdots x_{n(c-1)+1}^{n(c)}$ $(n(0) = 0, n(c) = n)$ と分解されたとする．このとき，部分系列の区切れが単語単位による並び $w_1 w_2 \cdots w_{n_w}$ の区切れと一致しているとは限らない．そのような場合を考慮しながら，各部分系列 $x_{n(j-1)+1}^{n(j)}$ に対して状態 $s_{n(j-1)+1}$ を以下のように与える．

単語 w_t を $w_t = x_{t,1} x_{t,2} \cdots x_{t,|w_t|}$ と表すこととし，いま，単語 w_t の中に部分系列の区切れがある場合を考える．部分系列の先頭のシンボル $x_{n(j-1)+1}$ が $x_{t,i}$ となっているとき，$\alpha = i - 1$ $(\alpha \in \{0, 1, \ldots, |w_t| - 1\})$ とする．α は単語 w_t の頭から数えて $\alpha + 1$ 番目のシンボルが部分系列 $x_{n(j-1)+1}^{n(j)}$ の一番先頭のシンボル $(x_{n(j-1)+1})$ となっていることを表している．すなわち，単語 w_t は部分系列の区切れによって以下のように 2 つに分けられることになる．

$$[x_{t,1}\ x_{t,2} \cdots x_{t,\alpha}],\ [x_{t,\alpha+1} \cdots x_{t,|w_t|}].$$

また，$\beta = |w_{t-1}| + |w_{t-2}| + \cdots + |w_{t-k_w}|$ とする．

この α と β を用いて，各部分系列 $x_{n(j-1)+1}^{n(j)}$ の状態 $s_{n(j-1)+1}$ を $s_{n(j-1)+1} = x_{n(j-1)-\alpha-\beta+1}^{n(j-1)}$ と定義する．これは，部分系列 $x_{n(j-1)+1}^{n(j)}$ の直前 $\alpha + \beta$ シンボルの系列を表している．1 つの単語が部分系列の区切れを複数含んでいる場合も同様に状態を定義できる．

ここで，w が k_w 次のマルコフ情報源から出力されるものとすると，状態 $s_{n(j-1)+1}$ により，$x_{n(j-1)}^{n(j)}$ の条件付確率は完全に規定される．

単語 $w_t = x_{t,1} x_{t,2} \cdots x_{t,|w_t|}$ の途中に部分系列の区切れがある場合，単語 w_t の生起確率 $P(w_t | w_{t-k_w} \cdots w_{t-1})$ を以下のように分解する．シンボル $x_{t,i-1}$ と $x_{t,i}$ の間に区切れがあるとき，

$$\begin{aligned} &P(w_t | w_{t-k_w} \cdots w_{t-1}) \\ &\quad = P(x_{t,1} \cdots x_{t,i-1} | x_{t-k_w,1} \cdots x_{t-1,|w_{t-1}|}) \\ &\qquad \cdot P(x_{t,i} \cdots x_{t,|w_t|} | x_{t-k_w,1} \cdots x_{t,i-1}) \end{aligned} \tag{8}$$

とできる．ただし

$$\begin{aligned} &P(x_{t,1} \cdots x_{t,i-1} | x_{t-k_w,1} \cdots x_{t-1,|w_{t-1}|}) \\ &\quad = \sum_{x_{t,i} \cdots x_{t,|w_t|}} P(w_t | x_{t-k_w,1} \cdots x_{t-1,|w_{t-1}|}) \end{aligned} \tag{9}$$

を周辺分布とし，

$$\begin{aligned} &P(x_{t,i} \cdots x_{t,|w_t|} | x_{t-k_w,1} \cdots x_{t,i-1}) \\ &\quad = \frac{P(x_{t,1} \cdots x_{t,|w_t|} | x_{t-k_w,1} \cdots x_{t-1,|w_{t-1}|})}{P(x_{t,1} \cdots x_{t,i} | x_{t-k_w,1} \cdots x_{t-1,|w_{t-1}|})} \end{aligned} \tag{10}$$

と定義する．

そして，$l = 1, 2, \ldots$ と，s に対して，$c_{l,s}$ を c 個の部分系列のうちで，長さが l でそれに対する状態が $s_{n(j-1)+1} = s$ であるような部分系列の個数であるとする．すると，

$$\sum_{l,s} c_{l,s} = c \tag{11}$$

$$\sum_{l,s} l\ c_{l,s} = n \tag{12}$$

が成り立っている．

ここで，次の補題が成り立つ．

補題 3 (Ziv の不等式の一般化) 単語単位で k_w 次マルコフ情報源の確率分布 P に対して，系列 $x_1^n (= w_1^{n_w})$ の相異なる部分系列への任意の分解は

$$\log P(x_1 x_2 \cdots x_n | s_1) \le -\sum_{l,s} c_{l,s} \log c_{l,s} \tag{13}$$

を満たす．ただし $s_1 = x_{-kh+1}^0$ は任意である．

(証明) [8] と同様の議論による． □

以上の用意から複雑量とエントロピー・レートの関係について次のことがいえる

補題 4 $\boldsymbol{W} = \{W_i\}_{i=1}^{\infty}$ は，エントロピー・レート $H(\boldsymbol{W})$ をもつ単語単位の定常エルゴード情報源とし，情報源からの出力を $W_1^{n_w} = W_1 W_2 \cdots W_{n_w} = X_1 X_2 \cdots X_{N(W_1^{n_w})} = X_1^{N(W_1^{n_w})}$ とする．ただし，$N(W_1^{n_w}) = |W_1| + |W_2| + \cdots + |W_{n_w}|$ であるが，これは $W_1^{n_w}$ により決まる確率変数である．$n_w \to \infty$ のとき $N(W_1^{n_w}) \to \infty$ となる．このときこの系列の複雑量 $c\left(X_1^{N(W_1^{n_w})}\right)$ は，

$$\limsup_{n_w \to \infty} \frac{c\left(X_1^{N(W_1^{n_w})}\right) \log c\left(X_1^{N(W_1^{n_w})}\right)}{N(W_1^{n_w})} \le H(\boldsymbol{X}),\ a.s. \tag{14}$$

を満たす．

(証明) まず単語単位の k_w 次マルコフ情報源の確率分布 P と，そこから出力される系列 $W_1^{n_w} = W_1 W_2 \cdots W_{n_w}$ を考えると，[8] の議論と同様に

$$\begin{aligned} &\frac{c\left(X_1^{N(W_1^{n_w})}\right) \log c\left(X_1^{N(W_1^{n_w})}\right)}{N(W_1^{n_w})} \\ &\quad \le -\frac{1}{N(W_1^{n_w})} \log P(W_1^{n_w} | W_{-k_w+1}^0) + \delta(n_w) \end{aligned} \tag{15}$$

が成立する．ただし，$\delta(n_w) \to 0$ $(n_w \to \infty)$.

ところで，W が n_w に対して定常エルゴードであることから，

$$\lim_{n_w \to \infty} \frac{n_w}{N(W_1^{n_w})} = \frac{1}{E[|W|]},\ a.s. \tag{16}$$

また

$$\begin{aligned} &\lim_{n_w \to \infty} \left\{ -\frac{1}{n_w} \log P(W_1^{n_w} | W_{-k_w+1}^0) \right\} \\ &\quad = H(W_1 | W_{-k_w+1}^0),\quad a.s. \end{aligned} \tag{17}$$

が成り立つから，式 (15) の右辺は $n_w \to \infty$ のとき

$$-\frac{n_w}{N(W_1^{n_w})} \cdot \frac{1}{n_w} \log P(W_1^{n_w} | W_{-k_w+1}^0) + \delta(n_w)$$

$$\to \frac{H(W_1|W^0_{-k_w+1})}{E[|W|]}, \quad a.s. \tag{18}$$

となる．ここで $k_w \to \infty$ とすると，

$$\frac{H(W_1|W_{-k_w+1})}{E[|W|]} \to \frac{H(\boldsymbol{W})}{E[|W|]} = H(\boldsymbol{X}) \tag{19}$$

となり，帰結する． □

補題4よりLZ78符号は単語単位の情報源クラスに対してユニバーサルであることが示される．

定理 1 (LZ78 符号のユニバーサル性)

LZ78符号は，単語単位の定常エルゴード情報源のクラスに対して，漸近最良性を有するユニバーサル符号である．すなわち，系列 $W_1^{n_w}$ をLZ78符号化したときの，情報源出力1シンボルあたりの符号長 $L_{LZ}(W_1^{n_w})/N(W_1^{n_w})$ は $n_w \to \infty$ のとき

$$\frac{L_{LZ}(W_1^{n_w})}{N(W_1^{n_w})} \to H(\boldsymbol{X}), \quad a.s. \tag{20}$$

を満たす．

(証明) 文献 [8] の議論と同様．式 (14) から

$$\begin{aligned}\limsup_{n_w\to\infty} & \frac{L_{LZ}(W_1^{n_w})}{N(W_1^{n_w})} \\ &= \limsup_{n_w\to\infty} \frac{c\left(X_1^{N(W_1^{n_w})}\right)\log c\left(X_1^{N(W_1^{n_w})}\right)}{N(W_1^{n_w})} \\ &\le H(\boldsymbol{X}), \quad a.s. \end{aligned} \tag{21}$$

が成立することがわかる．

逆に，

$$\liminf_{n_w\to\infty} \frac{c\left(X_1^{N(W_1^{n_w})}\right)\log c\left(X_1^{N(W_1^{n_w})}\right)}{N(W_1^{n_w})} < H(\boldsymbol{X}), \quad a.s. \tag{22}$$

であるとすると，平均符号長は $H(\boldsymbol{X})$ よりも小さくなる．しかし，$H(\boldsymbol{X})$ は平均符号長の下界を与えるので矛盾する．よって，

$$\liminf_{n_w\to\infty} \frac{c\left(X_1^{N(W_1^{n_w})}\right)\log c\left(X_1^{N(W_1^{n_w})}\right)}{N(W_1^{n_w})} \ge H(\boldsymbol{X}), \quad a.s. \tag{23}$$

となり，式 (21)，(23) より，式 (20) を得る． □

定理1より，符号長がエントロピー・レートに概収束することが示されたが，平均符号長についても有界収束定理より，エントロピー・レートに収束することもわかる．

4 考察

3章の議論から，単語単位で系列を出力する情報源に対して，LZ78符号がシンボル単位のそのままのアルゴリズムでユニバーサルであることが示された．すなわち，LZ78符号のアルゴリズムはシンボル単位で処理を行っているが，本来このアルゴリズムがユニバーサル性を保証しているシンボル単位の定常情報源クラスよりも広い情報源クラスである，シンボル単位で非定常のクラス(ただし，単語単位では定常エルゴード)に対してもユニバーサル性を保証する情報源クラスが存在することが示されたことになる．

この結果から，テキストデータなどの実際のデータ系列に対してLZ78符号化を適用してもよい圧縮性能が示されることについて一応の理論的な解釈を与えることができると考える．

5 むすび

本稿では，実際のデータ系列を出力する情報源モデルをより適切に表現できるモデルとして，系列が単語単位で出力される情報源を考慮し，この単語単位の定常エルゴード情報源に対して，LZ78符号化法の符号化性能の解析を行った．これにより，実データに対するLZ78符号化の圧縮性能につて理論的な解釈を与えることができた．今後は本稿の考察をもとに，さらに実データに近い情報源モデルを考察し，LZ78符号のユニバーサル性の議論をする必要がある．また，LZ78符号がユニバーサル性を保証する非定常情報源クラスとその性質を明らかにすることも今後の課題である．

謝辞

著者の1人石田は，日頃よりご指導いただいております東京大学 松島克守教授に心より感謝いたします．また，本研究を行なうにあたり，ご検討ご助言をいただきました早稲田大学 小林学氏，及び平澤研究室の皆様に深く感謝いたします．なお，本研究の一部は文部科学省平成12・13年度科学研究費補助金(課題番号12875072)の助成による．

参考文献

[1] J.Ziv and A.Lempel, "A universal algorithm for data compression," *IEEE Trans. Inform. Theory*, vol.IT-23, no.3, pp.337-343, May 1977.

[2] J.Ziv and A.Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol.IT-24, no.5, pp.530-536, Sept. 1978.

[3] 西新幹彦，森田啓義，"言語アルファベット情報源の漸近等分割性について，" 第20回情報理論とその応用シンポジウム予稿集 (SITA97), pp.345-348. 1997.

[4] M.Nishiara and H.Morita, "On the AEP of word-valued sources," *IEEE Trans. Inform. Theory*, vol.IT-46, no.3, pp.1116-1120, May 2000.

[5] 後藤 正幸，松嶋 敏泰，平澤 茂一，"単語単位で系列を出力する情報源，" 第22回情報理論とその応用シンポジウム予稿集 (SITA99), pp.359-362, 1999.

[6] 石田 崇，後藤 正幸，平澤 茂一，"単語単位で出現する系列に対するベイズ符号について，" 電子情報通信学会技術研究報告,IT99-27,pp73-78,1999.

[7] 石田 崇，後藤 正幸，平澤 茂一，"ブロック単位で系列を出力する情報源に対するベイズ符号とZiv-Lempel符号のユニバーサル性について，" 電子情報通信学会論文誌, vol.J84-A, no.9,pp.1167-1178, 2001.

[8] 韓 太舜，小林 欣吾，情報の符号化の数理，岩波講座応用数学，岩波書店,1994.