Subword Complexity and (non)-automaticity of certain completely multiplicative functions

YINING HU
CNRS, Institut de Mathématiques de Jussieu-PRG
Université Pierre et Marie Curie, Case 247
4 Place Jussieu
F-75252 Paris Cedex 05 (France)
yining.hu@imj-prg.fr

Abstract
In this article, we prove that for a completely multiplicative function \( f \) from \( \mathbb{N}^* \) to a field \( K \) such that the set \( \{ p \mid f(p) \neq 1_K \text{ and } p \text{ is prime} \} \) is finite, the asymptotic subword complexity of \( f \) is \( \Theta(n^t) \), where \( t \) is the number of primes \( p \) that \( f(p) \neq 0_K, 1_K \). This proves in particular that sequences like \( ((-1)^{v_2(n)+v_3(n)})_n \) are not \( k \)-automatic for \( k \geq 2 \).

Keywords. Completely multiplicative functions; \( k \)-automatic sequences; subword complexity.

1 Introduction
The subject of this article is the subword complexity and (non-)automaticity of certain completely multiplicative functions. From the definition we see that a completely multiplicative function is completely determined by its value on prime numbers.

We recall the definition and a few results about subword complexity and automaticity. The proof of Theorem 1 can be found for example in Chapter 10 of [2], and the proof of Theorem 2 can be found in [3].

Definition 1. Let \( u \) be a infinite sequence of symbols from an alphabet. We define the subword complexity \( p_u(n) \) of \( u \) to be the number of different factors (consecutive letters) of length \( n \) in \( u \).

Theorem 1 (Morse and Hedlund [4]). A sequence \( u \) is ultimately periodic if and only if there exists a non-negative integer \( k \) such that \( p_u(k) \leq k \).

We recall one of the equivalent definitions of an \( k \)-automatic sequence.

Definition 2. Let \( k \geq 2 \) be an integer. A sequence \( u \) is called \( k \)-automatic if there is a finite set of sequences containing \( u \) and closed under the operations

\[
(v(n))_n \rightarrow (v(kn+r))_n, \quad \text{for } r = 0, ..., k-1.
\]

Lemma 1. Let \( u = (u_n)_{n \geq 0} \) be a \( k \)-automatic sequence, and let \( \rho \) be a coding. Then the sequence \( \rho(u) \) is also \( k \)-automatic.
We recall some asymptotic notations for asymptotic complexity:
\( f(n) = O(g(n)) \): \( f \) is bounded above by \( g \) (up to constant factor) asymptotically;
\( f(n) = \Omega(g(n)) \): \( f \) is bounded below by \( g \) (up to constant factor) asymptotically;
\( f(n) = \Theta(g(n)) \): \( f \) is bounded both above and below (up to constant factors) by \( g \) asymptotically.

Theorem 2. (Cobham) If a sequence \( u \) is \( k \)-automatic, then \( p_u(n) = O(n) \).

Theorem 3 (Cobham). Let \( k,l \geq 2 \) be two multiplicatively independent integers, and suppose the sequence \( u = (u(n))_{n \geq 0} \) is both \( k \)- and \( l \)-automatic, then \( u \) is ultimately periodic.

As a first example we can consider a completely multiplicative sequence \( (u(n))_n \) that takes values in \( \mathbb{R} \) such that \( u(2) = -1 \), and \( u(p) = 1 \) for all \( p \in \mathbb{P} \setminus \{2\} \), where \( \mathbb{P} \) denotes the set of prime numbers. Then we have for all \( n \in \mathbb{N}^* \), \( u(n) = (-1)^{v_2(n)} \), where \( v_p \) is the \( p \)-adic valuation. This sequence is 2-automatic, for we have

\[
\begin{align*}
u(n) &= -u(n), \\
u(n+1) &= 1.
\end{align*}
\]

It is easy to see that \( u \) is not ultimately periodic. Thus by Theorem 1 and Theorem 2 we have \( p_u(n) = \Theta(n) \). In general, for all prime \( p \), the sequence \( ((-1)^{v_p(n)})_{n \geq 1} \) is \( p \)-automatic and has asymptotic subword complexity \( \Theta(n) \).

For a more interesting example, consider the completely multiplicative sequence \( a = (a(n))_{n \geq 1} \) taking values in \( \mathbb{R} \) such that \( a(2) = a(3) = -1 \), and \( a(p) = 1 \) for \( p \in \mathbb{P} \setminus \{2,3\} \). Intuitively this sequence cannot be \( k \)-automatic: if \( a \) is 2-automatic then there is no reason why it should not be 3-automatic, but \( a \) is not ultimately periodic (which we will prove later), so by Theorem 3 \( a \) cannot be at the same time 2- and 3-automatic. It can be shown that \( a \) is not 6-automatic, and there is no reason for \( a \) to be \( k \)-automatic for \( k \) other than 2, 3 or 6 either. Indeed, we will prove in the Section 2 that the subword complexity of \( a \) is \( \omega(n^2) \), which implies by Theorem 2 that it cannot be \( k \)-automatic. In Section 3 we a general result.

2 The example of \((-1)^{v_2(n)+v_3(n)}\)

In this section we prove that the asymptotic subword complexity of the sequence \((u(n))_{n \geq 1} = ((-1)^{v_2(n)+v_3(n)})_{n \geq 1}\) is \( \Theta(n^2) \).

It is easy to see that since \( u \) is the product of two sequences \( ((-1)^{v_2(n)})_{n \geq 1} \) and \( ((-1)^{v_3(n)})_{n \geq 1} \), both of which have asymptotic subword complexity \( \Theta(n) \), the asymptotic subword complexity of \( u \) is \( O(n^2) \). Indeed, we have the following lemma:

Lemma 2. Let the function \( f_0(n) \) be the product of functions \( f_1(n), f_2(n), \ldots, f_k(n) \). Then for all \( n \in \mathbb{N} \), \( p_{f_0}(n) \leq \prod_{i=1}^{k} p_{f_i}(n) \).

Proof. Let \( F_i,n \) be the set of factors of length \( n \) of \( f_i \). There is an surjection from \( \prod_{i=1}^{k} F_i,n \) to \( F_{0,n} \).

We recall Bézout’s Lemma in the form that we need:

Lemma 3 (Bézout). Let \( n \) be an integer and let \( p \) and \( q \) be coprime integers. Then

\[ \{k \cdot p - l \cdot q \mid k, l \in \mathbb{Z} \text{ and } k, l > n \} = \mathbb{Z} \]
The usual form of Bézout’s Lemma says that if $p$ and $q$ are coprime integers, then

$$\{k \cdot p - l \cdot q \mid k, l \in \mathbb{Z}\} = \mathbb{Z}.$$ 

We may assume that $k, l > n$ by replacing, when this is not the case, $(k, l)$ by $(k + m \cdot q, l + m \cdot p)$, for an integer $m$ large enough.

Now we prove that $p_a(n) = \Omega(n)$.

First, in order to isolate $v_2(n)$ and $v_3(n)$, we consider the subsequences $(u(n))_n = (u(3n + 1))_n$ and $(b(n))_n = (u(2n + 1))_n = (-1)^{u_2(3n + 1)}_n$. By Lemma 3 in Section 3 we know that $a$ and $b$ are not ultimately periodic. Therefore, $p_a(n) > n$ and $p_b(n) > n$ by Theorem 1. This mean that there exist $n + 1$ factors of length $n$ in $a$ (resp. $b$), which we denote by $A_0, A_1, \ldots, A_n$ (resp. $B_0, B_1, \ldots, B_n$) and by $a_0, a_1, \ldots, a_n$ (resp. $\beta_0, \beta_1, \ldots, \beta_n$) their starting position in $a$ (resp. $b$). We choose an integer $N$ such that these factors are contained in the initial segment of length $N$ in $a$ or $b$ respectively.

In the next step, for $(i, j) \in \{0, 1, \ldots, n\}^2$, we want the scattered subwords $A_i$ and $B_j$ to occur “in the same place” in the original sequence $u$, such that they would form distinct factors $U_{ij}$ of $u$. While this does not happen for all couples $(i, j)$, we will show in the following lines that it is true for at least $\lceil (n + 1)^2/6 \rceil$ couples in $\{0, 1, \ldots, n\}^2$. First we remark that the factors $A_0, A_1, \ldots, A_n$ (resp. $B_0, B_1, \ldots, B_n$) recur periodically in $a$ (resp. $b$). This is due to the property of the $p$-adic valuation that

$$v_p(m) > v_p(n) \Rightarrow v_p(m + n) = v_p(n).$$

Thus, let $M$ be an integer such that $M > v_2(3n+1)$ and $M > v_3(2n+1)$ for all $n \in \{0, 1, \ldots, N-1\}$. Then for all integer $k$ and all $n \in \{0, 1, \ldots, N-1\}$, we have $a(n+k \cdot 2^M) = a(n)$ and $b(n+k \cdot 3^M) = b(n)$. Therefore the set of starting positions of the factor $A_i$ in $a$ include

$$\{\alpha_i + k \cdot 2^M \mid k \in \mathbb{N}\},$$

and the set of starting positions of the factor $B_j$ in $b$ include

$$\{\beta_j + l \cdot 3^M \mid l \in \mathbb{N}\}.$$

In the original sequence $u$, the set of starting positions of the scattered subword $A_i$ include

$$\{3\alpha_i + 3k \cdot 2^M + 1 \mid k \in \mathbb{N}\},$$

and the set of starting positions of the scattered subword $B_j$ in $u$ include

$$\{2\beta_j + 2l \cdot 3^M + 1 \mid l \in \mathbb{N}\}.$$

By Lemma 3 the set

$$\{3k \cdot 2^M - 2l \cdot 3^M \mid k, l \in \mathbb{N}\} = \{6k \cdot 2^{M-1} - l \cdot 3^{M-1} \mid k, l \in \mathbb{N}\} = \{6z \mid z \in \mathbb{Z}\}$$

(1)

There exist $r \in \{0, 1, \ldots, 5\}$ such that

$$\text{card}((i, j) \in \{0, 1, \ldots, n\}^2 \mid 2\beta_j - 3\alpha \equiv r \mod 6)) \geq \lceil (n + 1)^2/6 \rceil.$$

This, combined with equation (1), means that for at least $\lceil (n + 1)^2/6 \rceil$ couples $(i, j)$, there is an occurrence of $A_i$ and an occurrence of $B_j$ in $u$, such $A_i$ starts $r$ letters before $B_j$. This gives at least $\lceil (n + 1)^2/6 \rceil$ distinct factors of $u$ of length $\max\{3n - 2, r + 2n - 1\}$, which proves that the asymptotic subword complexity of $u$ is $\Omega(n^2)$, which implies by Theorem 2 that $u$ is not automatic.
3 The general case

To study the general case, we need the following generalization of Lemma 3.

**Lemma 4.** Let \( q_0, q_1, \ldots, q_n \) be pairwise coprime integers. Let \( l_1, \ldots, l_n \) be integers. Then exists positive integers \( k_0, k_1, \ldots, k_n \) such that for all \( i \in \{1, \ldots, n\} \), \( k_0q_0 - k_1q_i = l_i \).

**Proof.** We prove this by induction.

By Lemma 3 we know that there exist \( k_0 \) and \( k_1 \in \mathbb{N} \) such that \( k_0q_0 - k_1q_1 = l_1 \).

Suppose that for an integer \( j \in \{1, \ldots, n\} \), we have proven that there exist \( k_0, k_1, \ldots, k_j \) such that for all \( i \in \{1, \ldots, j\} \), \( k_0q_0 - k_iq_i = l_i \). If \( j = n \) we are done. Otherwise by the pairwise coprime assumption and Lemma 3 we know that there exists \( k \in \mathbb{N} \) and \( k_{j+1}' \in \mathbb{N} \) such that
\[
k(k_0q_0 \ldots q_j) - k_{j+1}'q_{j+1} = l_{j+1} - k_0q_0.
\]
That is,
\[
q_0(k_0 + k(q_0 \ldots q_j)) - k_{j+1}'q_{j+1} = l_{j+1}
\]
For \( i = 0, 1, \ldots, j \), we define \( k_i' \) as \( k_i + k \cdot \prod_{m=0}^{j} q_m \). Then \( k_0', k_1', \ldots, k_{j+1}' \) satisfy the condition that for all \( i \in \{1, \ldots, j+1\} \), \( k_0'q_0 - k_i'q_i = l_i \).

We introduce the following lemma before proving our main theorem.

**Lemma 5.** Let \( a \) be an element of finite order different from the identity element in a multiplicative group \( G \). Let \( p \) be a prime number. Let \( q \) and \( b \) be positive integers such that \( p \nmid q \). Then the sequence \( (u(n))_n := (a^{vp(qm+b)})_{n \geq 0} \) is not ultimately periodic.

**Proof.** Suppose that \( (u(n))_n \) were ultimately periodic and let \( T \) be a period of \( (u(n))_n \). We write \( T \) as \( p^k \cdot T' \) where \( k = v_p(T) \) and \( p \nmid T' \). We claim that there exists an integer \( m \) larger than the length of the initial non-periodic segment of \( u \) such that \( v_p(qm + b) = k + 1 \). In fact, since \( q \) and \( p^{k+1} \) are coprime, by Lemma 3 we know that there exist integers \( m, n \) large enough to be in the periodic part of \( u \) such that
\[
n \cdot p^{k+1} - m \cdot q = b.
\]
Furthermore, we can assume that \( p \nmid n \), by eventually replacing \( (m, n) \) by \( (m + p^{k+1}, n + q) \) when this is not the case. Thus we have \( v_p(qm + b) = v_p(n + p^{k+1}) = k + 1 \). We also have \( v_p(q(m + T) + b) = v_p(qm + b + qT) = k \). Therefore \( u(m) = a^{k+1} \neq a^k = u(m + T) \) since \( a \) is not the identity element, which contradicts the definition of \( m \) and \( T \). Therefore \( (u(n))_n \) cannot be ultimately periodic.

**Theorem 4.** Let \( u = (u(n))_{n \geq 1} \) be a completely multiplicative sequence taking values in a field \( K \). We suppose that the number of prime numbers \( p \) such that \( u(p) \neq 1_K \) is finite. Let \( t \) be the number of primes \( p \) such that \( u(p) \neq 1_K, 0_K \). Then the asymptotic subword complexity of \( u \) is \( \Theta(n^t) \).

**Proof.** If \( t = 0 \), then \( u \) is periodic. Then \( u(n) \) is ultimately constant.

If not, we denote by \( P = \{p_1, \ldots, p_t\} \) the set of primes \( p \) such that \( u(p) \neq 1_K, 0_K \) and by \( q \) the product of primes where \( u \) takes the value 0\(_K\). The sequence \( u \) is the product of an ultimately periodic sequence and non-ultimately periodic automatic sequences:
\[
u(n) = \prod_{p \in P} (u(p))^{v_p(n)},
\]
where \((z(n))_n\) is the sequence defined as \(z(n) = 0\) if and only if \(u(n) = 0\), \(z(n) = 1\) otherwise. By Lemma 2 and Theorem 2 we know that \(p_u = O(n^t)\).

We consider the subsequence \((w(n))_n = (u(qn + 1))_n\). To prove that \(p_w(n) = \Omega(n^t)\) we only have to prove that \(p_w(u) = \Omega(n^t)\).

If \(t = 1\), we denote by \(p\) the prime such that \(u(p) \neq 1_K, 0_K\) and by \(a\) the value of \(u(p)\). Then \(w(n) = a^{vu(qn + 1)}\). By Lemma 5 we know that \(w\) is not ultimately periodic. So by Theorem 1 we have \(p_w(u) = \Omega(n)\).

If \(t \geq 2\), for \(i = 1, ..., t\) we define \(q_i\) to be \(\prod_{j=1, j \neq i}^t p_j\), and we consider the sequences \((w_i(n))_n := (w(q_i n))_n = (u(q_i q_n + 1))_n = (u(p_i)^{vu(q_i q_n + 1)})_n\). By Lemma 5 and Theorem 1 \(w_i\) is not ultimately periodic so \(w_i\) at least \(n + 1\) factors of length \(n\), which we denote by \(W_{i,j}\) for \(j = 0, ..., n\), and by \(\alpha_{i,j}\) their staring position in \(w_i\). We choose an integer \(N\) such that for all \(i = 1, ..., t\), and all \(j = 0, ..., n\), \(W_{i,j}\) occurs in the initial segment of \(w_i\) of length \(N\). There exists an integer \(M\) such that for all \(k = 0, 1, ..., N - 1\), all \(i = 0, ..., t\), and all \(m \in \mathbb{N}\), \(w_i(k + m \cdot p_i^M) = w_i(k)\). Thus the starting position of \(W_{i,j}\) in \(w\) include the set

\[
\{q_i \alpha_{i,j} + q_i m p_i^M \mid m \in \mathbb{N}\}.
\]

There exist \(r_2, ..., r_t \in \{0, 1, ..., p_1...p_t\}\) such that the cardinality of the set

\[
\{(j_1, ..., j_t) \in \{0, 1, ..., n\}^t \mid (q_i \alpha_{i,j_1} - q_i \alpha_{2,j_2}, ..., q_i \alpha_{t,j_t} - q_i \alpha_{t,j_t}) \equiv (r_2, ..., r_t) \mod p_1...p_t\}
\]

is at least \(n^t/(p_1^{t-1} p_2...p_t)\). By Lemma 4 we know that

\[
\{(q_1 m_1 p_1^M - q_2 m_2 p_2^M, ..., q_1 m_t p_1^M - q_t m_t p_t^M) \mid m_1, ..., m_t \in \mathbb{N}\} = (p_1...p_t \mathbb{Z})^{t-1}
\]

Therefore \(p_w\) has at least \(n^t/(p_1^{t-1} p_2...p_t)\) factors of length \((n \cdot \max\{q_1, ..., q_t\} + p_1...p_t)\). This means that \(p_w(n) = \Omega(n)\) and therefore \(p_u(n) = \Omega(n)\). \(\Box\)

The following corollary is an immediate consequence of Theorem 2 and 4.

**Corollary 1.** Let \((u(n))_n\) be a completely multiplicative sequence taking values in a field \(K\). We suppose that the number of prime numbers \(p\) such that \(u(p) \neq 1_K\) is finite and that among them, for at least two primes \(p_1\) and \(p_2\) we have \((p_1) \neq 0_K\) and \((p_2) \neq 0_K\). Then \((u(n))_n\) is not \(k\)-automatic for any \(k \geq 2\).

**Corollary 2.** Let \((u(n))_n\) be a completely multiplicative sequence taking values in a field \(K\). We suppose that the number of prime numbers \(p\) such that \(u(p) = 0_K\) is finite and that there exists an integer \(d\) such that the cardinality of the set \(\{p \in \mathbb{P} \mid u(p^d) \neq 0, 1_K\}\) is finite and at least \(2\). Then \(u\) is not \(k\)-automatic for any \(k \geq 2\).

**Proof.** By the assumption, the sequence \((u(n)^d)_n\) satisfies the conditions of Corollary 1 and therefore is not \(k\)-automatic. This implies that by \(u\) is not \(k\)-automatic by Lemma 4. \(\Box\)

In the case where \(K\) is the field of complex numbers, there is a more elegant proof of Corollary 4 using the following result in 1 about automatic sequences and Dirichlet series:

Let \(k \geq 2\) be an integer and let \((u(n))_n\) be a \(k\)-automatic sequence with values in \(\mathcal{C}\), then the Dirichlet series

\[
\sum_{n=1}^{\infty} \frac{u(n)}{n^s}
\]
has a meromorphic continuation to the whole complex plane, whose poles (if any) are located at the points
$$s = \frac{\log \lambda}{\log k} + \frac{2i\text{m} \pi}{\log k} - l + 1,$$
where $\lambda$ is any eigenvalue of a certain matrix defined from the sequence $u$, where $m \in \mathbb{Z}$, $l \in \mathbb{N}$, and $\log$ is a branch of the complex logarithm.

Proof of Corollary for $K = \mathbb{C}$. We denote by $P$ the set of prime numbers. By the assumption of the corollary we know that
$$P = A \cup B \cup \{p_1, p_2\},$$
where $A = \{p \in P | u(p) = 1\}$, $B = P \setminus (A \cup \{p_1, p_2\})$.

For $s$ such that $\Re(s) > 1$, we consider the Dirichlet series
$$\sum_{n=1}^{\infty} \frac{u(n)}{n^s} = \prod_{p \in P} \sum_{i=0}^{\infty} \left( \frac{u(p)}{p^s} \right)^i$$
$$= \prod_{p \in P} \frac{1}{1 - \frac{u(p)}{p^s}}$$
$$= \left( \prod_{p \in A} \frac{1}{1 - \frac{1}{p^s}} \right) \left( \prod_{q \in B} \frac{1}{1 - \frac{u(q)}{q^s}} \right) \left( \frac{1}{1 - \frac{1}{p_1^s}} \right) \left( \frac{1}{1 - \frac{1}{p_2^s}} \right)$$
$$= \zeta(s) \left( \prod_{q \in B} \frac{q^s - 1}{q^s - u(q)} \right) \frac{p_1^s - 1}{p_1^s - a} \frac{p_2^s - 1}{p_2^s - b}$$

Since the Riemann Zeta function $\zeta(s)$ is meromorphic on the whole complex plane and the product over $B$ in the last line is finite, the Dirichlet series $\sum_{n=1}^{\infty} \frac{u(n)}{n^s}$ has meromorphic continuation on the whole complex plane. Now we examine the poles of this function. The assumption of automaticity and multiplicativity implies that $a$ and $b$ are roots of unity, therefore the poles of $\frac{p_1^s - 1}{p_1^s - a} \frac{p_2^s - 1}{p_2^s - b}$ are $\{\frac{\pi \text{m} \cdot (\log(p_1)+2n\pi)}{\log(p_1)} | n \in \mathbb{Z}\}$ or $\{\frac{\pi \text{m} \cdot (\log(p_2)+2n\pi)}{\log(p_2)} | m \in \mathbb{Z}\}$. Since the $\zeta(s)$ has no zeros on the imaginary axis, these are also poles of the series $\sum_{n=1}^{\infty} \frac{u(n)}{n^s}$. Therefore $(u(n))_n$ cannot be $k$–automatic according to the result cited above.

\[\square\]

References

[1] J.-P. Allouche, M. Mendès France, J. Peyrière, “Automatic Dirichlet series”, J. Number Theory 81 (2000), no. 2, 359–373.

[2] J.-P. Allouche, J. Shallit, “Automatic sequences. Theory, applications, generalizations”, Cambridge University Press, Cambridge, 2003.
[3] A. Cobham, “Uniform tag sequences”, Math. Systems Theory 6 (1972), 164–192.

[4] M. Morse, G. A. Hedlund. Symbolic dynamics. Amer. J. Math., 60:815–866, 1938.