Journal of Integer Sequences, Vol. 23 (2020),
Article 20.3.1

The Number of Threshold Words on n Letters Grows Exponentially for Every $n \geq 27$

James D. Currie[1], Lucas Mol, and Narad Rampersad[2]
Department of Mathematics and Statistics
University of Winnipeg
515 Portage Avenue
Winnipeg, MB R3B 2E9
Canada
j.currie@uwinnipeg.ca
l.mol@uwinnipeg.ca
n.rampersad@uwinnipeg.ca

Abstract

For every $n \geq 27$, we show that the number of $n/(n-1)^+$-free words (i.e., threshold words) of length k on n letters grows exponentially in k. This settles all but finitely many cases of a conjecture of Ochem.

1 Introduction

Throughout, we use standard definitions and notations from combinatorics on words (see [13]). A *square* is a word of the form xx, where x is a nonempty word. A *cube* is a word of the form xxx, where x is a nonempty word. An *overlap* is a word of the form $axaxa$, where a is a letter and x is a (possibly empty) word. The study of words goes back to Thue, who demonstrated the existence of an infinite overlap-free word over a binary alphabet, and an infinite square-free word over a ternary alphabet (see [1]).

[1]The work of James D. Currie is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 2017-03901].

[2]The work of Narad Rampersad is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 2019-04111].

A language is a set of finite words over some alphabet A. The *combinatorial complexity* of a language L is the sequence $C_L : \mathbb{N} \to \mathbb{N}$, where $C_L(k)$ is defined as the number of words in L of length k. We say that a language L *grows* exactly as the sequence $C_L(k)$ grows, be it exponentially, polynomially, etc. Since the work of Brandenburg [2], the study of the growth of languages has been a central theme in combinatorics on words. Given a language L, a key question is whether it grows exponentially (fast), or subexponentially (slow). Brandenburg [2] demonstrated that both the language of cube-free words over a binary alphabet, and the language of square-free words over a ternary alphabet, grow exponentially. On the other hand, Restivo and Salemi [19] demonstrated that the language of overlap-free binary words grows only polynomially.

Squares, cubes, and overlaps are all examples of *repetitions* in words, and can be considered in the same general framework. Let $w = w_1 w_2 \cdots w_k$ be a finite word, where the w_i's are letters. A positive integer p is a *period* of w if $w_{i+p} = w_i$ for all $1 \leq i \leq k - p$. In this case, we say that $|w|/p$ is an *exponent* of w, and the largest such number is called *the* exponent of w. For a real number $r > 1$, a finite or infinite word w is called r-free (r^+-free) if w contains no finite factors of exponent greater than or equal to r (strictly greater than r, respectively).

Throughout, for every positive integer n, let A_n denote the n-letter alphabet $\{\mathtt{1}, \mathtt{2}, \ldots, \mathtt{n}\}$. For every $n \geq 2$, the *repetition threshold* for n letters, denoted $\mathrm{RT}(n)$, is defined by

$$\mathrm{RT}(n) = \inf\{r > 1\colon \text{ there is an infinite } r^+\text{-free word over } A_n\}.$$

Essentially, the repetition threshold describes the border between avoidable and unavoidable repetitions in words over an alphabet of n letters. The repetition threshold was first defined by Dejean [7]. Her 1972 conjecture on the values of $\mathrm{RT}(n)$ has now been confirmed through the work of many authors [3–7, 14, 15, 17, 18]:

$$\mathrm{RT}(n) = \begin{cases} 2, & \text{if } n = 2; \\ 7/4, & \text{if } n = 3; \\ 7/5, & \text{if } n = 4; \\ n/(n-1), & \text{if } n \geq 5. \end{cases}$$

The last cases of Dejean's conjecture were confirmed in 2011 by the first and third authors [6], and independently by Rao [18]. However, probably the most important contribution was made by Carpi [3], who confirmed the conjecture in all but finitely many cases.

In this short note, we are concerned with the growth rate of the language of *threshold words* over A_n. For every $n \geq 2$, let T_n denote the language of all $\mathrm{RT}(n)^+$-free words over A_n. We call T_n the *threshold language* of order n, and we call its members *threshold words* of order n. Threshold words are also called *Dejean words* by some authors. For every $n \geq 2$, the threshold language T_n is the minimally repetitive infinite language over A_n.

The threshold language T_2 is exactly the language of overlap-free words over A_2, which is known to grow only polynomially [19].[3,4] However, Ochem made the following conjecture about the growth of threshold languages of all other orders.

Conjecture 1 (Ochem [16])**.** For every $n \geq 3$, the language T_n of threshold words of order n grows exponentially.

Conjecture 1 has been confirmed for $n \in \{3, 4\}$ by Ochem [16], for $n \in \{5, 6, \ldots, 10\}$ by Kolpakov and Rao [12], and for all odd n less than or equal to 101 by Tunev and Shur [23]. In this note, we confirm Conjecture 1 for every $n \geq 27$.

Theorem 2. *For every $n \geq 27$, the language T_n of threshold words of order n grows exponentially.*

The layout of the remainder of the note is as follows. In Section 2, we summarize the work of Carpi [3] in confirming all but finitely many cases of Dejean's conjecture. In Section 3, we establish Theorem 2 with constructions that rely heavily on the work of Carpi. We conclude with a discussion of problems related to the rate of growth of threshold languages.

2 Carpi's reduction to ψ_n-kernel repetitions

In this section, let $n \geq 2$ be a fixed integer. Pansiot [17] was first to observe that if a word over the alphabet A_n is $(n-1)/(n-2)$-free, then it can be encoded by a word over the binary alphabet $B = \{\mathtt{0}, \mathtt{1}\}$. For consistency, we use the notation of Carpi [3] to describe this encoding. Let $\mathbb{S}_n$ denote the symmetric group on A_n, and define the morphism $\varphi_n : B^* \to \mathbb{S}_n$ by

$$\varphi_n(\mathtt{0}) = \begin{pmatrix} \mathtt{1} & \mathtt{2} & \cdots & \mathtt{n-1} \end{pmatrix}; \text{ and}$$
$$\varphi_n(\mathtt{1}) = \begin{pmatrix} \mathtt{1} & \mathtt{2} & \cdots & \mathtt{n} \end{pmatrix}.$$

Now define the map $\gamma_n : B^* \to A_n^*$ by

$$\gamma_n(b_1 b_2 \cdots b_k) = a_1 a_2 \cdots a_k,$$

where

$$a_i \varphi_n(b_1 b_2 \cdots b_i) = \mathtt{1}$$

for all $1 \leq i \leq k$. To be precise, Pansiot proved that if a word $\alpha \in A_n^*$ is $(n-1)/(n-2)$-free, then α can be obtained from a word of the form $\gamma_n(u)$, where $u \in B^*$, by renaming the letters.

[3] Currently the best known bounds on $C_{T_2}(k)$ are due to Jungers et al. [9].

[4] The threshold between polynomial and exponential growth for repetition-free binary words is known to be 7/3 [10]. That is, the language of 7/3-free words over A_2 grows polynomially, while the language of $7/3^+$-free words over A_2 grows exponentially.

Let $u \in B^*$, and let $\alpha = \gamma_n(u)$. Pansiot showed that if α has a factor of exponent greater than $n/(n-1)$, then either the word α itself contains a *short repetition*, or the binary word u contains a *kernel repetition* (see [17] for details). Carpi reformulated this statement so that both types of forbidden factors appear in the binary word u. Let $k \in \{1, 2, \ldots, n-1\}$, and let $v \in B^+$. Then v is called a *k-stabilizing word* (of order n) if $\varphi_n(v)$ fixes the points $\mathtt{1}, \mathtt{2}, \ldots, \mathtt{k}$. Let $\mathrm{Stab}_n(k)$ denote the set of k-stabilizing words of order n. The word v is called a *kernel repetition* (of order n) if it has period p and a factor v' of length p such that $v' \in \ker(\varphi_n)$ and $|v| > \frac{np}{n-1} - (n-1)$. Carpi's reformulation of Pansiot's result is the following.

Proposition 3 (Carpi [3, Proposition 3.2])**.** *Let $u \in B^*$. If a factor of $\gamma_n(u)$ has exponent larger than $n/(n-1)$, then u has a factor v satisfying one of the following conditions:*

(i) *$v \in \mathrm{Stab}_n(k)$ and $0 < |v| < k(n-1)$ for some $1 \leq k \leq n-1$; or*

(ii) *v is a kernel repetition of order n.*

Now assume that $n \geq 9$, and define $m = \lfloor (n-3)/6 \rfloor$ and $\ell = \lfloor n/2 \rfloor$. Carpi [3] defines an $(n-1)(\ell+1)$-uniform morphism $f_n : A_m^* \to B^*$ with the following extraordinary property.

Proposition 4 (Carpi [3, Proposition 7.3])**.** *Suppose that $n \geq 27$, and let $w \in A_m^*$. Then for every $k \in \{1, 2, \ldots, n-1\}$, the word $f_n(w)$ contains no k-stabilizing word of length smaller than $k(n-1)$.*

We note that Proposition 4 was proven by Carpi [3] in the case that $n \geq 30$ in a computation-free manner. The improvement to $n \geq 27$ stated here was achieved later by the first and third authors [4], using lemmas of Carpi [3] along with a significant computer check.

Proposition 4 says that for *every* word $w \in A_m^*$, no factor of $f_n(w)$ satisfies condition (i) of Proposition 3. Thus, we need only worry about factors satisfying condition (ii) of Proposition 3, i.e., kernel repetitions. To this end, define the morphism $\psi_n : A_m^* \to \mathbb{S}_n$ by $\psi_n(v) = \varphi_n(f_n(v))$ for all $v \in A_m^*$. A word $v \in A_m^*$ is called a *ψ_n-kernel repetition* if it has a period q and a factor v' of length q such that $v' \in \ker(\psi_n)$ and $(n-1)(|v|+1) \geq nq - 3$. Carpi established the following result.

Proposition 5 (Carpi [3, Proposition 8.2])**.** *Let $w \in A_m^*$. If a factor of $f_n(w)$ is a kernel repetition, then a factor of w is a ψ_n-kernel repetition.*

In other words, if $w \in A_m^*$ contains no ψ_n-kernel repetitions, then no factor of $f_n(w)$ satisfies condition (ii) of Proposition 3. Altogether, we have the following theorem, which we state formally for ease of reference.

Theorem 6. *Suppose that $n \geq 27$. If $w \in A_m^*$ contains no ψ_n-kernel repetitions, then $\gamma_n(f_n(w))$ is $\mathrm{RT}(n)^+$-free.*

Finally, we note that the morphism f_n is defined in such a way that the kernel of ψ_n has a very simple structure.

Lemma 7 (Carpi [3, Lemma 9.1])**.** *If $v \in A_m^*$, then $v \in \ker(\psi_n)$ if and only if 4 divides $|v|_a$ for every letter $a \in A_m$.*

3 Constructing exponentially many threshold words

In this section, let $n \geq 27$ be a fixed integer, and let $m = \lfloor (n-3)/6 \rfloor$ and $\ell = \lfloor n/2 \rfloor$, as in the previous section. Since $n \geq 27$, we have $m \geq 4$. In order to prove that the threshold language T_n grows exponentially, we construct an exponentially growing language $Z_m \subseteq A_m^*$ of words that contain no ψ_n-kernel repetitions. If $n \geq 33$ (or equivalently, if $m \geq 5$), then we define Z_m by modifying Carpi's construction of an infinite word α over A_m that contains no ψ_n-kernel repetitions. If $27 \leq n \leq 32$ (or equivalently, if $m = 4$), then we define a 3-uniform substitution $g\colon A_4^* \to 2^{A_4^*}$, and let Z_4 be the set of all factors of words obtained by iterating g on the letter $\mathtt{1}$.

3.1 Case I: $n \geq 33$

We first recall the definition of α, the infinite word over A_m defined by Carpi [3] that contains no ψ_n-kernel repetitions. First of all, define $\beta = (b_i)_{i \geq 1}$, where

$$b_i = \begin{cases} \mathtt{1}, & \text{if } i \equiv 1 \pmod 3; \\ \mathtt{2}, & \text{if } i \equiv 2 \pmod 3; \\ b_{i/3}, & \text{if } i \equiv 0 \pmod 3. \end{cases}$$

Now define $\alpha = (a_i)_{i \geq 1}$, where for all $i \geq 1$, we have

$$a_i = \begin{cases} \max\{a \in A_m \colon\ 4^{a-2} \text{ divides } i\}, & \text{if } i \text{ is even;} \\ b_{(i+1)/2}, & \text{if } i \text{ is odd.} \end{cases}$$

Note that if $i \equiv 2 \pmod 4$, then $a_i = \mathtt{2}$. Let Z_m be the set of all finite words obtained from a prefix of α by exchanging any subset of these $\mathtt{2}$'s for $\mathtt{1}$'s. To be precise, if $z = z_1 z_2 \cdots z_k$, then $z \in Z_m$ if and only if all of the following hold:

- $z_i \in \{\mathtt{1}, \mathtt{2}\}$ if $i \equiv 2 \pmod 4$;
- $z_i = \max\{a \in A_m \colon\ 4^{a-2} \text{ divides } i\}$ if $i \equiv 0 \pmod 4$; and
- $z_i = b_{(i+1)/2}$ if i is odd.

Note in particular that if $z = z_1 z_2 \cdots z_k$ is in Z_m, then $z_i \geq 3$ if and only if $i \equiv 0 \pmod 4$.

We claim that no word $z \in Z_m$ contains a ψ_n-kernel repetition. The proof is essentially analogous to Carpi's proof that α contains no ψ_n-kernel repetitions. We begin with a lemma about the lengths of factors of words in Z_m that lie in $\ker(\psi_n)$.

Lemma 8 (Adapted from Carpi [3, Lemma 9.3])**.** *Let $z \in Z_m$, and let v be a factor of z. If $v \in \ker(\psi_n)$, then 4^{m-1} divides $|v|$.*

Proof. The statement is trivially true if $v = \varepsilon$, so assume $|v| > 0$. Set $|v| = 4^b c$, where 4^b is the maximal power of 4 dividing $|v|$. Suppose, towards a contradiction, that $b \le m - 2$. Since $v \in \ker(\psi_n)$, by Lemma 7, we see that 4 divides $|v|$, meaning $b \ge 1$.

Write $z = z_1 z_2 \cdots z_{|z|}$. Then we have $v = z_i z_{i+1} \cdots z_{i+4^b c-1}$ for some $i \ge 1$. By definition, for any $j \ge 1$, we have $z_j \ge b+2$ if and only if 4^b divides j. (Since $b \ge 1$, we have $b+2 \ge 3$, and hence $z_j \ge b+2$ implies $j \equiv 0 \pmod 4$.) Thus, we have that the sum $\sum_{a=b+2}^{m} |v|_a$ is exactly the number of integers in the set $\{i, i+1, \ldots, i+4^b c-1\}$ that are divisible by 4^b, which is exactly c. Since $v \in \ker(\psi_n)$, by Lemma 7, we conclude that 4 divides c, contradicting the maximality of b. □

Now, using Lemma 8 in place of [3, Lemma 9.3], a proof strictly analogous to that of [3, Proposition 9.4] gives the following. The only tool in the proof that we have not covered here is [3, Lemma 9.2], which is a short technical lemma about the repetitions in the word β, and which can be used without any modification.

Proposition 9. *Suppose that $n \ge 33$. Then no word $z \in Z_m$ contains a ψ_n-kernel repetition.*

3.2 Case II: $27 \le n \le 32$

In this case, we have $m = 4$. For any set X, we let 2^X denote the power set of X. So in particular, the set $2^{A_4^*}$ consists of all sets of finite words over A_4. For alphabets Σ and Δ, a *substitution* is a map $s : \Sigma^* \to 2^{\Delta^*}$ such that $s(xy) = s(x)s(y)$ for all $x, y \in \Sigma^*$ and $s(\varepsilon) = \{\varepsilon\}$.

Define a substitution $g : A_4^* \to 2^{A_4^*}$ by

$$\begin{aligned} g(1) &= \{112\} \\ g(2) &= \{114\} \\ g(3) &= \{113\} \\ g(4) &= \{123, 213\}. \end{aligned}$$

We extend g to $2^{A_4^*}$ by $g(W) = \bigcup_{w \in W} g(w)$, which allows us to iteratively apply g to an initial word in A_4^*. Let $Z_4 = \text{Fact}\{v\colon\ v \in g^n(1) \text{ for some } n \ge 1\}$, i.e., we have that Z_4 is the set of factors of all words obtained by iteratively applying g to the initial word 1.

For a word $w \in A_4^*$, let $\pi(w)$ denote the *Parikh vector* of w, defined by

$$\pi(w) = \begin{bmatrix} |w|_1 & |w|_2 & |w|_3 & |w|_4 \end{bmatrix}^T$$

Note that for every $a \in A_4$ and every $x, y \in g(a)$, we have $\pi(x) = \pi(y)$. Thus, we may let $\pi_g(a)$ denote the common Parikh vector of every word in $g(a)$. The *frequency matrix* M_g of g is then the 4×4 matrix defined by

$$(M_g)_{ij} = (\pi_g(j))_i$$

i.e., the jth column of M_g is the Parikh vector of every word in $g(j)$. Explicitly, we have

$$M_g = \begin{bmatrix} 2 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Evidently, if $v \in g(u)$, then we have $\pi(v) = M_g \pi(u)$. Note also that M_g is invertible modulo 4, with

$$M_g^{-1} = \begin{bmatrix} 3 & 3 & 2 & 2 \\ 0 & 0 & 0 & 1 \\ 3 & 2 & 3 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix}.$$

This fact is the key to the proof of the following lemma.

Lemma 10. *Let $u \in A_4^*$ be a word, and let $v \in g(u)$. If $v \in \ker(\psi_n)$, then $u \in \ker(\psi_n)$.*

Proof. By Lemma 7, for every word $w \in A_4^*$, we have $w \in \ker(\psi_n)$ if and only if $\pi(w) \equiv \mathbf{0} \pmod 4$.

Now suppose that $v \in \ker(\psi_n)$. Then $\pi(v) \equiv \mathbf{0} \pmod 4$. Since $v \in g(u)$, we have $\pi(v) = M_g \pi(u)$. It follows that $\pi(u) \equiv M_g^{-1}\pi(v) \equiv \mathbf{0} \pmod 4$ as well, and thus we have $u \in \ker(\psi_n)$. □

If a word $w \in A_4^*$ has period p and the length p prefix of w is in $\ker(\psi_n)$, then we say that p is a *kernel period* of w.

Proposition 11. *Suppose that $27 \le n \le 32$. Then no word in Z_4 contains a ψ_n-kernel repetition.*

Proof. Suppose otherwise that the word $v_0 \in Z_4$ is a ψ_n-kernel repetition. Write $v_0 = x_0 y_0$, where v_0 has kernel period $|x_0|$. From the definition of ψ_n-kernel repetition, we must have

$$(n-1)(|v_0|+1) \ge n|x_0| - 3,$$

or equivalently,

$$|x_0| \le (n-1)|y_0| + n + 2.$$

Since $n \le 32$, we certainly have

$$|x_0| \le 31|y_0| + 34. \tag{1}$$

If $|y_0| \le 3$, then we have $|x_0| \le 127$, and hence $|v_0| \le 130$. We eliminate this possibility by exhaustive search, so we may assume that $|y_0| \ge 4$.

Let z be the length 3 prefix of y_0. By inspection of all words in Z_4 of length 3, there is a unique factorization $z = z'z''$, where z' is a nonempty suffix of some word in $g(A_4)$, and z''

is a (possibly empty) prefix of some word in $g(A_4)$. Since y_0 is a prefix of x_0y_0, it must be the case that z is a prefix of x_0. Write $x_0z = z'z''x_0'z'z''$. Then we must have $z''x_0'z' \in g(A_4^*)$, and hence $|x_0| = |z''x_0'z'|$ is a multiple of 3.

Now write $v_0 = s_0v_0'p_0$ for some suffix s_0 of a word in $g(A_4)$, some prefix p_0 of a word in $g(A_4)$, and some word $v_0' \in g(v_1)$, where $v_1 \in Z_4$. Since $|s_0| \leq 2$ and $|p_0| \leq 2$, we have $|v_0'| \geq |v_0| - 4 \geq |x_0|$, and hence v_0' has kernel period $|x_0|$. Now write $v_1 = x_1y_1$, where $3|x_1| = |x_0|$. Evidently, we have $3|y_1| + 4 \geq |y_0|$. Note that v_1 has period $|x_1|$. Further, by Lemma 10, we have $x_1 \in \ker(\psi_n)$, and hence $|x_1|$ is a kernel period of v_1.

We may now repeat the process described above. Eventually, for some $r \geq 1$, we reach a word $v_r \in Z_4$ that can be written $v_r = x_ry_r$, where $|x_r|$ is a kernel period of v_r, and $|y_r| \leq 3$. For all $1 \leq i \leq r$, one proves by induction that $|x_0| = 3^i|x_i|$ and $|y_0| \leq 3^i|y_i| + 4\sum_{j=0}^{i-1} 3^j = 3^i|y_i| + 2(3^i - 1)$. Thus, from (1), we obtain

$$3^i|x_i| \leq 31\left[3^i|y_i| + 2(3^i - 1)\right] + 34$$

for all $1 \leq i \leq r$. Dividing through by 3^i, and then simplifying, we obtain

$$|x_i| \leq 31\left[|y_i| + 2\right] - \frac{28}{3^i} \leq 31\left[|y_i| + 2\right] \tag{2}$$

for all $1 \leq i \leq r$.

Since $|y_r| \leq 3$, we obtain $|x_r| \leq 155$ from (2). By Lemma 7, the kernel period $|x_r|$ of v_r is a multiple of 4, so in fact we have $|x_r| \leq 152$, and in turn $|v_r| \leq 155$. By exhaustive search of all words in Z_4 of length at most 155, we find that $v_r \in W$, where W is a set containing exactly 200 words. Indeed, the set W contains

- 160 words with kernel period 76 and length 77,
- 36 words with kernel period 92 and length 93, and
- 4 words with kernel period 112 and length 114.

For every $w \in W$, let

$$E_w = \text{Fact}\left(\{g(awb)\colon\ a, b \in A_4,\ awb \in Z_4\}\right).$$

Evidently, we have $v_{r-1} \in E_{v_r}$. For every word $w \in W$, let p_w denote the kernel period of w, and let q_w denote the maximum length of a repetition with kernel period $3p_w$ across all words in E_w. By exhaustive check, for every $w \in W$, we find $3p_w > 31\left[q_w - 3p_w + 2\right]$. However, the word $v_{r-1} = x_{r-1}y_{r-1}$ must be in E_{v_r}, and by (2), we have

$$3p_{v_r} = |x_{r-1}| \leq 31\left[|y_{r-1}| + 2\right] \leq 31\left[q_{v_r} - 3p_{v_r} + 2\right].$$

This is a contradiction. We conclude that the set Z_4 contains no ψ_n-kernel repetitions. □

We now proceed with the proof of our main result.

Proof of Theorem 2. First suppose that $n \geq 33$. By Proposition 9, no word $z \in Z_m$ contains a ψ_n-kernel repetition. From the definition of Z_m, one easily proves that

$$C_{Z_m}(k) = \Omega\left(2^{k/4}\right).$$

By Theorem 6, for every word $z \in Z_m$, the word $\gamma_n(f_n(z))$ is in the threshold language T_n of order n. Moreover, the maps γ_n and f_n are injective, and $|\gamma_n(f_n(z))|/|z| = (n-1)(\ell+1)$, since f_n is $(n-1)(\ell+1)$-uniform and γ_n preserves length. It follows that

$$C_{T_n}(k) = \Omega\left(2^{k/4(n-1)(\ell+1)}\right).$$

Since n, and hence ℓ, are fixed, the quantity $(n-1)(\ell+1)$ is a constant, and we conclude that the language T_n grows exponentially.

Suppose now that $27 \leq n \leq 32$. By Proposition 11, no word $z \in Z_4$ contains a ψ_n-kernel repetition. Since $|g^4(a)| \geq 4$ for all $a \in A_4$, we have

$$C_{Z_4}(k) = \Omega\left(4^{k/81}\right).$$

By the same argument as above, we see that

$$C_{T_n}(k) = \Omega\left(4^{k/81(n-1)(\ell+1)}\right),$$

and we conclude that the language T_n grows exponentially. □

As pointed out by one of the anonymous referees, the constructions given in this section can also be used to demonstrate the existence of uncountably many infinite threshold words over A_n for every $n \geq 27$.

4 Conclusion

Conjecture 1 has now been established for all $n \notin \{12, 14, \ldots, 26\}$. We remark that different techniques than those presented here will be needed to establish Conjecture 1 in all but one of these remaining cases. (It appears that the techniques presented here could potentially be used for $n = 22$, but we do not pursue this isolated case.) For example, let $n = 26$. Then we have $m = 3$. By computer search, for every letter $a \in A_m$, the word $f_n(a\mathtt{3})$ contains a 15-stabilizing word of length 350, which is less than $15(n-1) = 375$. By another computer search, the longest word on $\{\mathtt{1}, \mathtt{2}\}$ avoiding ψ_n-kernel repetitions has length 15. So there are only finitely many words in A_m^* that avoid both ψ_n-kernel repetitions and the forbidden stabilizing words. Similar arguments lead to the same conclusion for all $n \in \{12, 14, 16, 18, 20, 24\}$.

For a language L, the value $\rho(L) = \limsup_{k\to\infty}(C_L(k))^{1/k}$ is called the *growth rate* of L. If L is factorial (i.e., closed under taking factors), then by an application of Fekete's lemma, we can safely replace lim sup by lim in this definition. If $\rho(L) > 1$, then the language L

grows exponentially, and in this case, $\rho(L)$ is a good description of how quickly the language grows.

For all $n \geq 33$, we have established that $\rho(T_n) \geq 2^{1/4(n-1)(\ell+1)}$. However, this lower bound tends to 1 as n tends to infinity, and this seems far from best possible. Indeed, Shur and Gorbunova proposed the following conjecture concerning the asymptotic behaviour of $\rho(T_n)$.

Conjecture 12 (Shur and Gorbunova [22]). The sequence $\{\rho(T_n)\}$ of the growth rates of threshold languages converges to a limit $\hat{\rho} \approx 1.242$ as n tends to infinity.

A wide variety of evidence supports this conjecture – we refer the reader to [8, 20–22] for details. For a fixed n, there are efficient methods for determining upper bounds on $\rho(T_n)$ which appear to be rather sharp, even for relatively large values of n (see [22], for example). Establishing a sharp lower bound on $\rho(T_n)$ appears to be a more difficult problem. We note that a good lower bound on $\rho(T_3)$ is given by Kolpakov [11] using a method that requires some significant computation. For all $n \in \{5, 6, \ldots, 10\}$, Kolpakov and Rao [12] give lower bounds for $\rho(T_n)$ using a similar method. They were then able to estimate the value of $\rho(T_n)$ with precision 0.005 using upper bounds obtained by the method of Shur and Gorbunova [22].

Thus, in addition to resolving the finitely many remaining cases of Conjecture 1, improving our lower bound for $\rho(T_n)$ when $n \geq 27$ remains a significant open problem.

5 Acknowledgments

We thank the anonymous referees, whose comments and suggestions helped to improve the paper.

References

[1] J. Berstel, Axel Thue's papers on repetitions in words: a translation, in *Publications du LaCIM*, Vol. 20, Université du Québec à Montréal, 1995.

[2] F. J. Brandenburg, Uniformly growing k-th power-free homomorphisms, *Theoret. Comput. Sci.* **23** (1983), 69–82.

[3] A. Carpi, On Dejean's conjecture over large alphabets, *Theoret. Comput. Sci.* **385** (2007), 137–151.

[4] J. D. Currie and N. Rampersad, Dejean's conjecture holds for $n \geq 27$, *RAIRO—Theor. Inform. Appl.* **43** (2009), 775–778.

[5] J. D. Currie and N. Rampersad, Dejean's conjecture holds for $n \geq 30$, *Theoret. Comput. Sci.* **410** (2009), 2885–2888.

[6] J. D. Currie and N. Rampersad, A proof of Dejean's conjecture, *Math. Comp.* **80** (2011), 1063–1070.

[7] F. Dejean, Sur un théorème de Thue, *J. Combin. Theory Ser. A* **13** (1972), 90–99.

[8] I. A. Gorbunova and A. M. Shur, On Pansiot words avoiding 3-repetitions, in *Proc. 8th Internat. Conf. Words 2011 (WORDS 2011)*, Electron. Proc. Theor. Comput. Sci., Vol. 63, 2012, pp. 138–146.

[9] R. M. Jungers, V. Y. Protasov, and V. D. Blondel, Overlap-free words and spectra of matrices, *Theoret. Comput. Sci.* **410** (2009), 3670–3684.

[10] J. Karhumäki and J. Shallit, Polynomial versus exponential growth in repetition-free binary words, *J. Combin. Theory Ser. A* **105** (2004), 335–347.

[11] R. Kolpakov, Efficient lower bounds on the number of repetition-free words, *J. Integer Sequences* **10** (2007), 1–16.

[12] R. Kolpakov and M. Rao, On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters, *Theoret. Comput. Sci.* **412** (2011), 6507–6516.

[13] M Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.

[14] M. Mohammad-Noori and J. D. Currie, Dejean's conjecture and Sturmian words, *European J. Combin.* **28** (2007), 876–890.

[15] J. Moulin-Ollagnier, Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10, and 11 letters, *Theoret. Comput. Sci.* **95** (1992), 187–205.

[16] P. Ochem, A generator of morphisms for infinite words, *RAIRO—Theor. Inform. Appl.* **40** (2006), 427–441.

[17] J. J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Discrete Appl. Math.* **7** (1984), 297–311.

[18] M. Rao, Last cases of Dejean's conjecture, *Theoret. Comput. Sci.* **412** (2011), 3010–3018.

[19] A. Restivo and S. Salemi, Overlap free words on two symbols, in M. Nivat and D. Perrin, eds., *Automata on Infinite Words*, Lect. Notes in Comput. Sci., Vol. 192, Springer-Verlag, 1985, pp. 198–206.

[20] A. M. Shur, Growth properties of power-free languages, *Comput. Sci. Rev.* **6** (2012), 187–208.

[21] A. M. Shur, Growth of power-free languages over large alphabets, *Theory Comput. Syst.* **54** (2014), 224–243.

[22] A. M. Shur and I. A. Gorbunova, On the growth rates of complexity of threshold languages, *RAIRO—Theor. Inform. Appl.* **44** (2010), 175–192.

[23] I. N. Tunev and A. M. Shur, On two stronger versions of Dejean's conjecture, in *Proc. 37th Internat. Conf. on Mathematical Foundations of Computer Science (MFCS 2012)*, Lect. Notes in Comput. Sci., Vol. 7464, Springer, 2012, pp. 800–812.

2010 *Mathematics Subject Classification*: Primary 68R15.
Keywords: threshold word, repetition threshold, Dejean word, exponential growth, Dejean's conjecture, Dejean's theorem.

Received November 13 2019; revised version received February 21 2020. Published in *Journal of Integer Sequences*, February 22 2020.

Return to Journal of Integer Sequences home page.