Relaxation of Square-Freeness

Hiroki Shibata[1], Takuya Mieno[2], Dominik Köppl[3], and Shunsuke Inenaga[4]

[1] Joint Graduate School of Mathematics for Innovation, Kyushu University
shibata.hiroki.753@s.kyushu-u.ac.jp
[2] Graduate School of Informatics and Engineering, University of Electro-Communications
tmieno@uec.ac.jp
[3] Department of Computer Science and Engineering, University of Yamanashi
dkppl@yamanashi.ac.jp
[4] Department of Informatics, Kyushu University
inenaga.shunsuke.380@m.kyushu-u.ac.jp

Abstract. We extend the analysis of nonrepetitive sequences of Entringer et al. [Journal of Combinatorial Theory, 1974] to relaxations of equality testing under nonstandard equivalence relations, in particular parameterized equivalence and order-preserving equivalence. For this setting, we introduce ℓ^+-squares, defined as squares whose total length is at least 2ℓ. We obtain an infinite 3^+-parameterized-square-free ternary word and an infinite 3^+-order-preserving-square-free binary word through an approach based on combinatorics on words. In addition, we report the longest ℓ^+-square-free words across several equivalence relations.

Keywords: squarefree words, parameterized equivalence, order-preserving equivalence, morphic word

1 Introduction

A *square* is a non-empty word of the form uu. Squares are one of the simplest and most studied repetitions in words [4,9]. A word is *square-free* if it contains no square as a subword. More than a century ago, Thue [21] constructed an infinite square-free word over a ternary alphabet. A ternary alphabet is the smallest possible alphabet for an infinite square-free word. Indeed, every infinite binary sequence contains at least one of the squares aa, bb, and $abab$.

The notion of a square can be generalized by replacing equality between the two halves with a weaker equivalence relation on words. In this paper, we use two equivalence relations from pattern matching: *parameterized equivalence* [2] and *order-preserving equivalence* [12]. We say that two words u and v of the same length ℓ are *parameterized equivalent* if there is a bijection f with $f(u[i]) = v[i]$ for each text position i. We say that they are *order-preserving equivalent* if $u[i] \preceq u[j] \Leftrightarrow v[i] \preceq v[j]$ holds for every pair of text positions i and j. Order-preserving equivalence can also be viewed through the *increasing bijection* on the alphabet. Equivalently, u and v are parameterized equivalent under an increasing bijection. A non-empty word uv is called a *parameterized* (resp. *order-preserving*) *square* if u and v are parameterized (resp. order-preserving) equivalent [13].

Some equivalence relations admit no infinite square-free words. For example, every word of length two is a parameterized square, so no parameterized-square-free infinite word can exist. Even if we consider only *non-trivial* squares (those of length greater than two), no infinite word avoids every non-trivial parameterized square [13]. Consequently, weaker notions of square-freeness are required to give positive results on infinite words that avoid a weaker type of square. We study ℓ^+*-squares*, squares of

Hiroki Shibata, Takuya Mieno, Dominik Köppl, Shunsuke Inenaga: Relaxation of Square-Freeness, pp. 58–70.
Proceedings of PSC 2026, Jan Holub and Jan Žďárek (Eds.), ISBN 978-80-01-07576-0 © Czech Technical University in Prague, Czech Republic

Table 1. Lengths of the longest ℓ^+-square-free words under various equivalence relations and alphabet sizes. An entry marked ∞ indicates the existence of an infinite ℓ^+-square-free word. A question mark (?) denotes a case that remains open. A dash ($-$) indicates that infiniteness follows from another case (e.g., a smaller ℓ or σ).

Strict Equality

$\sigma\backslash\ell$	1	2	3	4
2	3	18	∞	$-$
3	∞	$-$	$-$	$-$
4	$-$	$-$	$-$	$-$
5	$-$	$-$	$-$	$-$

Parameterized Equivalence

$\sigma\backslash\ell$	1	2	3	4
2	1	7	∞	$-$
3	1	9	$-$	$-$
4	1	9	$-$	$-$
5	1	9	$-$	$-$

Order-Preserving Equivalence

$\sigma\backslash\ell$	1	2	3	4
2	1	7	∞	$-$
3	1	∞	$-$	$-$
4	1	$-$	$-$	$-$
5	1	$-$	$-$	$-$

Cartesian-Tree Equivalence [17]

$\sigma\backslash\ell$	1	2	3	4
2	1	5	29	?
3	1	9	?	?
4	1	9	?	?
5	1	9	?	?

length at least 2ℓ, and investigate *ℓ^+-square-free words* under parameterized equivalence.

A complete characterization of the avoidability of length-restricted parameterized and order-preserving squares can be derived by combining existing results [13,16]. However, the proofs by Ng et al. [16] do not rely solely on combinatorics on words because their approach uses large morphisms and automated theorem proving tools. In this paper, we construct square-free words using an approach based on combinatorics on words.

Our main results are summarized as follows. We first prove the existence of an infinite 3^+-parameterized-square-free ternary word. We then show the existence of an infinite 3^+-order-preserving-square-free binary word. Both infinite words are constructed using small morphisms, and their square-freeness is proved through an approach based on combinatorics on words. Finally, through computational experiments, we establish the maximum lengths of words that avoid length-restricted squares under several nonstandard equivalence relations and give explicit longest words.

Table 1 summarizes the maximum lengths of ℓ^+-square-free words under several equivalence relations.

Related Work. Square-free words were first studied by Thue [21] over a century ago and have since remained an active topic of research. Variants involving length-restricted squares have also been explored. For example, it is known that every binary word of length greater than eighteen contains a 2^+-square [8]. In contrast, infinite binary 3^+-square-free words exist [8,20]; they are constructed by applying a morphism to a ternary square-free word. This kind of freeness has also been studied for Gray codes [18] and circular words [5].

Squares and other repetitions under generalized equivalence relations on words have also been studied. In the abelian model, where equivalence is defined by the multiset of characters, Keränen [11] showed the existence of infinite abelian-square-free words over a four-character alphabet, but it is known that none exist over a ternary alphabet. In the parameterized setting, the number of parameterized squares

in a word has also been studied [13,10]. Cubic and unary patterns under permutations, which are closely related to parameterized matching, have also been studied [14,6].

Words avoiding parameterized and order-preserving squares have also been studied in several works. Kociumaka et al. [13] showed that an infinite 2^+-order-preserving-square-free word exists over a ternary alphabet, whereas no such word exists in the parameterized model even without restrictions on the alphabet size. Ng et al. [16] constructed an infinite binary word that avoids every ordinary square and *antisquare* of length at least 6, where an antisquare is a binary word of the form $x\overline{x}$ and $\overline{x}$ is obtained by changing every character of x. Since parameterized squares over a binary alphabet are precisely ordinary squares and antisquares, this infinite binary word is 3^+-parameterized-square-free. Moreover, every order-preserving square is a parameterized square, and hence the word is 3^+-order-preserving-square-free. Their results were proved using large morphisms found through computer search and `Walnut` [15]. Combining these results gives a complete characterization of the values of ℓ and the alphabet sizes for which infinite ℓ^+-square-free words exist in the parameterized and order-preserving models.

2 Preliminaries

Let Σ be an *alphabet*. We denote the size of Σ by σ. An element of Σ is referred to as a *character*, and an element of Σ^* is referred to as a *word*. We denote a word x of length n as $x = x[0] \cdots x[n-1]$, where $x[i]$ denotes the i-th character of x. For $0 \leq i < j \leq |x|$, we denote by $x[i..j)$ the subword of x consisting of the characters $x[i] \cdots x[j-1]$, where the notation $[i..j)$ follows the half-open interval convention. If a word x does not contain another word w as a subword, we say that x *avoids* w.

Two words $x, y \in \Sigma^n$ are *parameterized equivalent* if there exists a bijective function $f : \Sigma \to \Sigma$ such that $f(x[i]) = y[i]$ for all $0 \leq i < n$. Two words $x, y \in \Sigma^n$ on a totally ordered alphabet Σ are *order-preserving equivalent* if $x[i] \preceq x[j] \Leftrightarrow y[i] \preceq y[j]$ holds for every pair $i, j \in [0..n-1]$. We note that order-preserving equivalence is stronger than parameterized equivalence; in particular, if two words x and y are order-preserving equivalent, then they are also parameterized equivalent.

A word of the form ww is called a (standard) *square*. A word that contains no standard square as a subword is called *square-free*. For an integer ℓ, we call such a square an *ℓ^+-square* if its total length is at least 2ℓ. A word of the form uv, where u and v are *parameterized* (resp. *order-preserving*) *equivalent*, is called a *parameterized* (resp. *order-preserving*) *square*. We use the term *ℓ^+-parameterized* (resp. *ℓ^+-order-preserving*) *square* to refer to such a square when its length is at least 2ℓ. A word is said to be *ℓ^+-square-free*, with respect to an equivalence model (exact, parameterized, order-preserving, etc.), if it contains no square of the corresponding type with length at least 2ℓ.

A *morphism* on two alphabets Σ and Σ' is a function $h : \Sigma^* \to \Sigma'^*$ such that $h(xy) = h(x)h(y)$ for all words $x, y \in \Sigma^*$. Hence, to define a morphism h, it suffices to specify $h(c)$ for each character $c \in \Sigma$. When $\Sigma' = \Sigma$ (i.e., $h : \Sigma^* \to \Sigma^*$), we define $h^k(x)$ for $x \in \Sigma^*$ as the word obtained by applying h to x repeatedly k times. If x is a prefix of $h(x)$, then $h^k(x)$ is a prefix of $h^{k+1}(x)$ for every $k \geq 0$. If the lengths $|h^k(x)|$ are unbounded, these finite words define an infinite word. Let $\Sigma_2 = \{0, 1\}$ and $\Sigma_3 = \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$ denote the binary and ternary alphabets, respectively. In this paper, we put special emphasis on the morphism $\phi : \Sigma_3^* \to \Sigma_3^*$ defined by

$$\begin{aligned} \mathrm{a} &\mapsto \mathrm{abc}, \\ \mathrm{b} &\mapsto \mathrm{ac}, \\ \mathrm{c} &\mapsto \mathrm{b}. \end{aligned}$$

It is known that $\phi^k(\mathrm{a})$ is square-free for any $k \geq 0$ [3]. For instance, $\phi(\mathrm{a}) = \mathrm{abc}$, $\phi^2(\mathrm{a}) = \mathrm{abcacb}$, and $\phi^3(\mathrm{a}) = \mathrm{abcacbabcbac}$, none of which contains a square.

3 3^+-Parameterized-Square-Free Ternary Words

In this section, we aim to construct an infinite 3^+-parameterized-square-free ternary word. We define $C = \mathrm{cbbbc}$ and let $\mu : \Sigma_3^* \to \Sigma_3^*$ be the morphism that maps c to C while leaving a and b unchanged. For each $k \geq 0$, we define $x'_k = \phi^k(\mathrm{a})$ and $x_k = \mu(x'_k)$. Recall that x'_k is square-free [3].

We first show that x_k contains no 3^+-parameterized squares except for the standard squares.

Lemma 1. *For any $k \geq 0$, if a 3^+-parameterized square $w = uv$ ($|u| = |v| \geq 3$) exists as a subword of x_k, then $u = v$.*

Proof. Let $w = uv$ be a 3^+-parameterized square subword of x_k of length 2ℓ ($\ell \geq 3$), where u and v are parameterized equivalent under some bijection f.

We first determine the nonempty subwords of x'_k that avoid c. Since every subword of the square-free word x'_k is square-free, such a subword cannot contain any of the binary squares aa, bb, abab, and baba. Indeed, avoiding aa and bb forces a binary word to alternate, and every alternating word of length at least four contains abab or baba. Thus, any nonempty subword of x'_k that avoids c belongs to $\{\mathrm{a}, \mathrm{b}, \mathrm{ab}, \mathrm{ba}, \mathrm{aba}, \mathrm{bab}\}$. We next exclude aba. This is clear for $x'_0 = \mathrm{a}$. For $k \geq 1$, write $x'_k = \phi(z)$ with $z = x'_{k-1}$. In any occurrence of aba, the middle b is preceded and followed by a. Consider an occurrence of b in $\phi(z)$. It is either the middle character of $\phi(\mathrm{a}) = \mathrm{abc}$ or the whole image $\phi(\mathrm{c}) = \mathrm{b}$, since $\phi(\mathrm{b}) = \mathrm{ac}$ contains no b. In the first case, this b is followed by c, and hence it cannot be the middle character of aba. In the second case, if this c is the first character of z, then this b has no preceding character and cannot be the middle character of aba. Otherwise, the character preceding this b is the last character of $\phi(d)$ for the character d preceding this c in z. Since $\phi(\mathrm{a}) = \mathrm{abc}$, $\phi(\mathrm{b}) = \mathrm{ac}$, and $\phi(\mathrm{c}) = \mathrm{b}$ end with c, c, and b, respectively, the character preceding this b is not a. Hence aba does not occur in x'_k. Consequently, every nonempty subword of x'_k that avoids c belongs to $S = \{\mathrm{a}, \mathrm{b}, \mathrm{ab}, \mathrm{ba}, \mathrm{bab}\}$.

We next show that bb occurs in w. Since x_k is obtained from x'_k by replacing each occurrence of c with C, any subword of x_k that avoids bb cannot contain an occurrence of C, because $C = \mathrm{cbbbc}$ contains bb. Moreover, no suffix or prefix of C of length at least 3 avoids bb. Suppose, for the sake of contradiction, that w avoids bb. Then, it must be of the form $w = \alpha s \beta$, where α (resp. β) is a suffix (resp. prefix) of C of length at most 2 and $s \in S$. This gives $|s| = 2\ell - |\alpha| - |\beta| \geq 2$, so the only possibilities for s are ab, ba, and bab. Consequently, the only possible candidates for w, whose length is even and at least six, are bcbabc, cbabcb, bcabcb, and bcbacb. None of them is a parameterized square. Hence, bb occurs in w.

Since $|w| \geq 6$ and every occurrence of bb in x_k must be a part of an image of c (i.e., $C = \mathrm{cbbbc}$), either u or v must contain the subword cbb or bbc. Since x'_k is

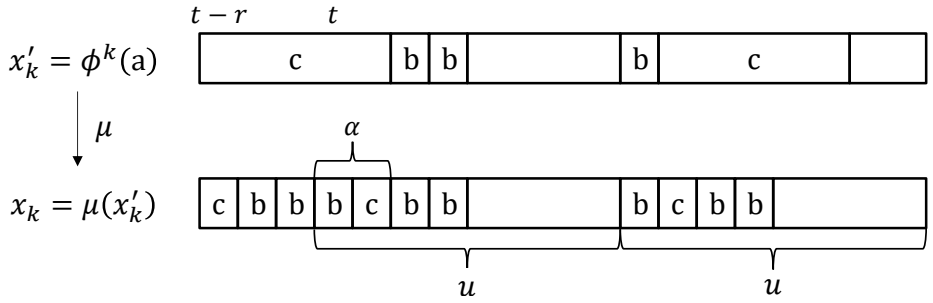


Figure 1. Illustration for Lemma 2. In this figure, the two occurrences of u in $w = u^2$ have different offsets. The left one has offset $r = 3$, and the right one starts at a boundary, that is, with offset 0. In this case, u starts with αbb = bcbb, where α = bc. The factorization starting at position $t - r$ is C|b|b. This implies that x'_k contains the square bb.

square-free, it contains no two equal consecutive characters. The morphism μ only replaces c with C = cbbbc. Thus, every occurrence of two equal consecutive characters in x_k is contained in the block bbb of an occurrence of C. Hence, if a length-three subword of x_k has its last two (resp. first two) characters equal and the remaining character different, then it must be cbb (resp. bbc). Thus, the only subword of x_k parameterized equivalent to cbb (resp. bbc) is cbb itself (resp. bbc itself). Therefore, the corresponding subword in the other half of w is identical to cbb or bbc, and hence f(b) = b and f(c) = c. Consequently, f must be the identity function, and w is a standard square $w = uu$. □

We also show the 3^+-square-freeness of x_k.

Lemma 2. *For any $k \geq 0$, the word x_k is 3^+-square-free.*

Proof. Let $\mathcal{S} = \{\mathrm{a}, \mathrm{b}, C\}$. In this proof, we define a *factorization* of a word z as a decomposition $z = f_1 \cdots f_m$ with $f_i \in \mathcal{S}$ for every i. The words f_i are called *factors*. Since $\mathcal{S}$ is prefix-free, such a factorization is unique whenever it exists. Moreover, since μ maps c to C while leaving a and b unchanged, the word $x_k = \mu(x'_k)$ has a factorization, and hence it has a unique factorization. We call a position a *boundary* if a factor starts there. If a position is not a boundary, then it lies inside an occurrence of C. In this case, we denote its offset from the beginning of that occurrence to be an element of $\{1, 2, 3, 4\}$. Boundary positions are said to have offset 0. If both endpoints of a subword z of x_k are boundaries, then its factorization determines a unique word v such that $\mu(v) = z$.

For the sake of contradiction, assume that there exists a square subword $w = uu = x_k[p - \ell..p + \ell)$ of x_k with $\ell \geq 3$ centered at text position p. Let $i = p - \ell$ be the starting position of the square.

We first show that the two occurrences $x_k[i..p)$ and $x_k[p..p + \ell)$ of u start at the same offset. Assume that one occurrence starts at offset $r \in \{1, 2\}$. Then it begins with bbb if $r = 1$ and with bbc if $r = 2$. These prefixes occur only inside an occurrence of C, and the offset is uniquely determined. Hence the other occurrence also starts at the same offset r.

Next assume that one occurrence starts at offset $r \in \{3, 4\}$. Let $\alpha = C[r..5)$, so α = bc for $r = 3$ and α = c for $r = 4$. If an occurrence starts away from a boundary, then the prefix α determines its offset uniquely. Suppose, for the sake of contradiction, that the other occurrence starts at a different offset. Since the prefix α determines every non-boundary offset, the other occurrence must start at a boundary. In this

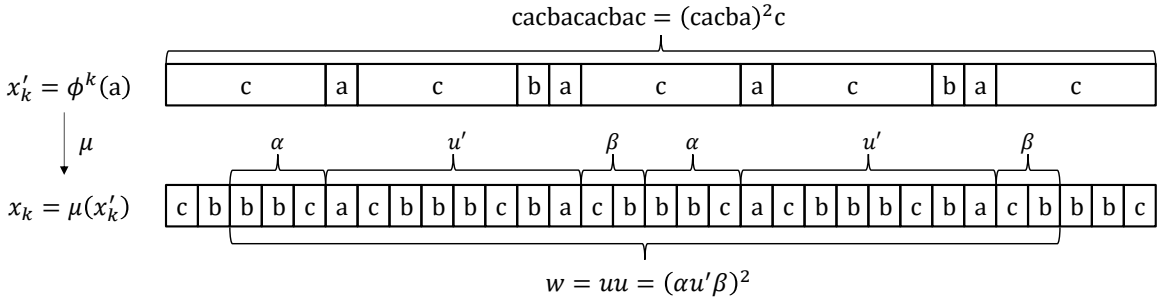


Figure 2. Illustration for Lemma 2. In the figure, x_k has a square $w = uu = (\mathrm{bbcacbbbcbacb})^2$ and u is written as $u = \alpha z \beta$ where $\alpha = \mathrm{bbc}, \beta = \mathrm{cb}$, and $z = \mathrm{acbbbcba}$. In this case, $w = uu$ is generated from $(\mathrm{cacba})^2 c$ and thus x'_k has a square $(\mathrm{cacba})^2$.

case, the other occurrence of u starts at offset 0 with prefix α. If $r = 4$ ($\alpha = \mathrm{c}$), then the first factor in the factorization beginning at this occurrence must be C. If $r = 3$ ($\alpha = \mathrm{bc}$), then the first and second factors in this factorization must be b and C, respectively. In both cases, u starts with $\alpha\mathrm{bb}$. Consider the occurrence of u with offset r, and let its starting position be t. Since α appears as a suffix of the occurrence of C starting at position $t - r$, the factorization of x_k at position $t - r$ begins with $C|\mathrm{b}|\mathrm{b}$. This factorization implies that x'_k contains the square bb, contradicting the square-freeness of x'_k (see Figure 1). Thus, whenever one occurrence of u starts away from a boundary, the other occurrence starts at the same offset in $\{1, 2, 3, 4\}$. Equivalently, if one occurrence starts at a boundary, then the other occurrence also starts at a boundary.

If both occurrences of u start at boundaries, then both endpoints of the first occurrence $x_k[i..p)$ are boundaries. The equality of the two occurrences and the uniqueness of the factorization imply that $x_k[p..p + \ell)$ has the same factorization. Thus, both endpoints of $x_k[p..p + \ell)$ are also boundaries. Let v be the unique word such that $\mu(v) = x_k[i..p)$. Then both endpoints of w are boundaries, and $w = \mu(v^2)$. Hence v^2 occurs in x'_k. This contradicts the square-freeness of x'_k. Otherwise, let $r \in \{1, 2, 3, 4\}$ be the common offset of the two occurrences of u. Let α be the suffix of C of length $5 - r$ and β be the prefix of C of length r. Since $\beta\alpha = C$ and the center position p of the square $w = uu$ lies on an occurrence of C with offset $r = |\beta|$, u can be written as $u = \alpha z \beta$, where $z = x_k[i + |\alpha|..p - |\beta|)$ (see Figure 2). Since the positions $i + |\alpha|$, $p - |\beta|$, $p + |\alpha|$, and $p + \ell - |\beta|$ are boundaries, the two subwords $x_k[i + |\alpha|..p - |\beta|)$ and $x_k[p + |\alpha|..p + \ell - |\beta|)$ have exactly the same factorization. Let v be the unique word such that $\mu(v) = z$. Hence, $w = \alpha\mu(v)\beta\alpha\mu(v)\beta$. Since both the starting and ending position of w lie inside occurrences of C, by extending this subword to the left by r positions and to the right by $5 - r$ positions, we obtain $C\mu(v)C\mu(v)C$ as a subword of x_k. This subword is equal to $\mu(\mathrm{cvcvc})$. Thus, x'_k contains the square $(\mathrm{cv})^2$. This contradicts the square-freeness of x'_k. □

By combining the above results, we can obtain the main result.

Theorem 3. *For any $k \geq 0$, the ternary word $x_k = \mu(\phi^k(\mathrm{a}))$ is 3^+-parameterized-square-free.*

Proof. By Lemma 1, if a subword w is a 3^+-parameterized square, then w must be a standard square. However, by Lemma 2, such square does not exist. Therefore x_k does not contain any 3^+-parameterized-square. □

4 3^+-Order-Preserving-Square-Free Binary Words

In this section, we show that there exists an infinite 3^+-order-preserving-square-free binary word. The proof is similar to [19], where it is observed that applying the morphism $\psi : \Sigma_3^* \to \Sigma_2^*$ with $\psi(\mathrm{a}) = 0000, \psi(\mathrm{b}) = 0101$, and $\psi(\mathrm{c}) = 1111$ to a square-free ternary word x yields a 3^+-square-free binary word (in the strict sense). We show that the same property holds when we slightly modify the morphism ψ to another morphism $\tau : \Sigma_3^* \to \Sigma_2^*$ (Lemma 4). Then, we construct an infinite binary word by combining the two morphisms τ and ϕ (Theorem 5).

Let $\tau : \Sigma_3^* \to \Sigma_2^*$ be the morphism defined as

$$\begin{aligned} \mathrm{a} &\mapsto 00, \\ \mathrm{b} &\mapsto 0101, \\ \mathrm{c} &\mapsto 11. \end{aligned}$$

The following lemma will be proved in Section 4.1.

Lemma 4. *For any square-free ternary word x, the word $\tau(x)$ is 3^+-square-free.*

Let $x'_k = \phi^k(\mathrm{a})$ and $y_k = \tau(x'_k)$ for every $k \geq 0$. Recall that each x'_k is square-free, and thus, each y_k is 3^+-square-free by Lemma 4. We prove the main theorem of this section.

Theorem 5. *For any $k \geq 0$, the binary word $y_k = \tau(\phi^k(\mathrm{a}))$ is 3^+-order-preserving-square-free.*

Proof. Assume, for the sake of contradiction, that there exists a subword $w = uv$ ($|u| = |v|$) of y_k that is a 3^+-order-preserving square of length 2ℓ ($\ell \geq 3$). Since y_k is 3^+-square-free and $\ell \geq 3$, we have $u \neq v$. If either u or v contains both 0 and 1, then the only order-preserving bijection between their alphabets is the identity. Therefore, since $u \neq v$, the only possible case is when both u and v consist of a single repeated character but differ from each other (i.e., $\{u, v\} = \{0^\ell, 1^\ell\}$). To prove that such a square cannot occur in y_k, it suffices to show that y_k contains neither $0^3 1^3$ nor $1^3 0^3$ as a subword.

Since $\tau(\mathrm{a}) = 00$, $\tau(\mathrm{b}) = 0101$, and $\tau(\mathrm{c}) = 11$, any occurrence of 0^3 (resp. 1^3) in $y_k = \tau(x'_k)$ is contained in $\tau(\mathrm{aa})$ or $\tau(\mathrm{ab})$ (resp. $\tau(\mathrm{cc})$ or $\tau(\mathrm{bc})$), if it exists. Since x'_k is square-free, it contains neither aa nor cc. Thus, 0^3 can appear only as the prefix of $\tau(\mathrm{ab}) = 000101$, and 1^3 can appear only as the suffix of $\tau(\mathrm{bc}) = 010111$. Therefore, a subword of the form $0^3 1^3$ or $1^3 0^3$ can occur only inside $\tau(\mathrm{bcab})$, so that

$$\tau(\mathrm{bcab}) = \tau(\mathrm{b})\tau(\mathrm{c})\tau(\mathrm{a})\tau(\mathrm{b}) = 010\,1^3 0^3\,101.$$

However, bcab cannot occur in x'_k. If $k = 0$, then this is immediate from $x'_0 = \mathrm{a}$. If $k \geq 1$, then both bc and ab can only be generated by applying ϕ to the character a. Thus, an occurrence of bcab in $x'_k = \phi(x'_{k-1})$ would imply an occurrence of aa in x'_{k-1}, contradicting the square-freeness of x'_{k-1}. Therefore, y_k contains no 3^+-order-preserving square. □

In the rest of this section, we give a proof of Lemma 4.

4.1 Proof of Lemma 4

Before starting the proof, we introduce some notation. Let x be a square-free ternary word. We denote $w = \tau(x)$. A subword $w[i..j)$ of w is called a *block* if it is exactly generated from the subword $x[s..e)$, i.e., there exists a pair of integers s and e such that $\tau(x[0..s)) = w[0..i)$ and $\tau(x[s..e)) = w[i..j)$ hold. We denote $\mathcal{F} = \{\tau(\mathrm{a}), \tau(\mathrm{b}), \tau(\mathrm{c})\} = \{00, 0101, 11\}$. In this subsection, we define a *factorization* of a block z as a decomposition $z = f_1 \cdots f_m$ with $f_i \in \mathcal{F}$ for every i. The words f_i are called *factors*. Note that this convention differs from the factorization used in Lemma 2. Since $\mathcal{F}$ is prefix-free, such a factorization is unique whenever it exists. Every block $w[i..j)$ has such a factorization by definition, and therefore its factorization is uniquely defined. Moreover, the factorization of a block z determines a unique word v such that $\tau(v) = z$. A position i with $0 \le i \le |w|$ is called a *boundary* if $i = |w|$ or a factor starts at position i in the factorization of w. For any position i, let o_i and o'_i be the smallest non-negative integers such that $i + o_i$ and $i - o'_i$ are boundaries, respectively, where $0 \le o_i, o'_i \le 3$. In particular, $o_i = 0$ if and only if $o'_i = 0$. Since the length of a factor is either two (for 00, 11) or four (for 0101), $o_i \ge 2$ implies that the factor ending at position $i + o_i$ is 0101. Similarly, $o'_i \ge 2$ implies that the factor starting at position $i - o'_i$ is 0101.

Next, we introduce Fact 4.1, which states that w contains no short squares of lengths between 6 and 12. Let $w[i..j)$ be a subword of w of length 2ℓ. By construction of w, the subword $w[i - o'_i..j + o_j)$ is a block generated from a square-free subword of x of length at most $\ell + 1$. Indeed, each factor has length at least two, so a subword of length 2ℓ can intersect at most $\ell + 1$ factors. Thus, we can verify Fact 4.1 by enumerating all words in $\{\tau(v) \mid v \in \Sigma_3^*, 1 \le |v| \le 7, v \text{ is square-free}\}$ and checking 3^+-square-freeness[1].

Fact: For every $3 \le \ell \le 6$, the word w contains no square of length 2ℓ.

We next establish two lemmas, which rule out long equal subwords with different boundary offsets.

Lemma 6. *There is no pair of subwords $w[i..i + \ell)$ and $w[j..j + \ell)$ such that $\ell \ge 7$, $w[i..i + \ell) = w[j..j + \ell)$, and $o_i \not\equiv o_j \pmod 2$.*

Proof. Assume on the contrary that there are two subwords $w[i..i + \ell) = w[j..j + \ell)$ of length $\ell \ge 7$ with $o_i \not\equiv o_j \pmod 2$. Let $u = w[i..i + \ell) = w[j..j + \ell)$. Assume $o_i < o_j$ without loss of generality. We first consider the case $o_i = 0, o_j = 3$. In this case, $w[j - 1..j + 3)$ must be a factor, specifically $w[j - 1..j + 3) = 0101$. Then, $w[j..j + 3) = w[i..i + 3) = 101$, which contradicts $o_i = 0$ since no factor starts with 101.

The remaining case is $o_j = o_i + 1$. Let f_i and f_j be the factors starting at positions $i + o_i$ and $j + o_j$, respectively. Since $o_j = o_i + 1$, we have $0 \le o_i \le 2$. Together with $\ell \ge 7$ and the fact that every factor has length at most four, this ensures that the factor immediately following f_i exists. Let g_i be this factor. Since $\ell \ge 7$, both f_i and f_j are contained in $u = w[i..i + \ell) = w[j..j + \ell)$. Further, since $o_j = o_i + 1$, the starting positions of the occurrences of f_i and f_j inside u are different by one position. These facts imply that $(f_i g_i)[1..|f_j| + 1) = f_j$ holds.

We consider the following three cases according to the value of f_i.

Case 1: $f_i = 0101$: Then f_j must start with 10 because $f_j[0..2) = u[o_j..o_j + 2) = u[o_i + 1..o_i + 3) = f_i[1..3) = 10$. However, no factor starts with 10, a contradiction.

[1] The source code is available at `https://github.com/koeppl/squarechecker/check_short_string_3plus_square_freeness.py`

Case 2: $f_i = 11$: Then f_j must start with 1, hence $f_j = 11 \in \mathcal{F}$. From $(f_i g_i)[1..|f_j|+1) = f_j = 11$, g_i also starts with 1, so $g_i = 11$. Therefore, $f_i g_i = 1111$, which is the image of cc under τ. This implies that x contains the square cc, contradicting the square-freeness of x.
Case 3: $f_i = 00$: Then f_j must start with 0, so f_j is either 00 or 0101. If $f_j = 00$, as in the previous case, we obtain $g_i = 00$ from $(f_i g_i)[1..|f_j| + 1) = f_j = 00$. Then, $f_i g_i = 0000 = \tau(\mathrm{aa})$, contradicting the square-freeness of x. Otherwise, if $f_j = 0101$, then g_i must start with 10, which contradicts that $g_i \in \mathcal{F}$. □

Lemma 7. *There is no pair of subwords $w[i..i+\ell)$ and $w[j..j+\ell)$ such that $\ell \geq 7$, $w[i..i+\ell) = w[j..j+\ell)$, and $o_i - o_j \equiv 2 \pmod 4$.*

Proof. Assume on the contrary that there are two subwords $w[i..i+\ell) = w[j..j+\ell)$ of length $\ell \geq 7$ with $o_i - o_j \equiv 2 \pmod 4$. Let $u = w[i..i+\ell) = w[j..j+\ell)$. Assume $o_i < o_j$ without loss of generality. Then, the only possible pairs for (o_i, o_j) are $(0, 2)$ and $(1, 3)$, so $o_j = o_i + 2$. Let f_i and f_j be the factors starting at positions $i + o_i$ and $j + o_j$, respectively. Also, let e_j be the factor immediately preceding f_j. Since $o_j \geq 2$, we have $e_j = 0101$.

From $u = w[i..i+\ell) = w[j..j+\ell)$ and $o_j = o_i + 2$, it follows that $u[o_i..o_i+2) = u[o_j - 2..o_j) = e_j[|e_j| - 2..|e_j|) = 01$. Hence $f_i = 0101$, because it is the only factor that starts with 01. Also, $u[o_j..o_j+2) = u[o_i+2..o_i+4) = f_i[2..4) = 01$, so $f_j = 0101$. Therefore, the two consecutive factors around position $j + o_j$ satisfy $e_j = f_j = 0101$. Consequently, $e_j f_j = 01010101 = \tau(\mathrm{bb})$, which implies that x contains the square bb, contradicting the square-freeness of x. □

Now, we are ready to prove Lemma 4.

Proof (of Lemma 4). For the sake of contradiction, assume w contains a square $w[i..k)$ of length $2\ell \geq 6$. Let $j = i + \ell$ where $\ell = (k - i)/2$, so $w[i..k) = uu$ with $u = w[i..j) = w[j..k)$. By Fact 4.1, w has no square with $3 \leq \ell \leq 6$, thus $\ell \geq 7$. Let $d = (o_i - o_j) \bmod 4$. We classify squares in w by the values of o_i, o'_j, o_j, o'_k and d.
Case 1: $d = 2$: This case contradicts Lemma 7.
Case 2: $d \in \{1, 3\}$: This case is equivalent to $o_i \not\equiv o_j \pmod 2$. Thus, the existence of square $w[i..k) = uu$ contradicts Lemma 6.
Case 3: $d = 0$: In this case, $o_i = o_j$ holds. Assume $w[i..j)$ is factorized as $w[i..j) = \alpha f_1 \cdots f_t \beta$, where $\alpha = w[i..i+o_i)$ and $\beta = w[j-o'_j..j)$ may be empty, and each f_p is a factor. Then, $w[j..k)$ is also factorized as $w[j..k) = \alpha f_1 \cdots f_t \beta$, because $w[i+o_i..j) = w[j+o_j..k)$, both $i + o_i$ and $j + o_j$ are boundaries, and the factorization of a block is unique. More precisely, $o'_j = o'_k$ and $w[i..k) = \alpha f_1 \cdots f_t \beta \alpha f_1 \cdots f_t \beta$ hold. We denote $f = f_1 \cdots f_t$, and let g be the word such that $\tau(g) = f$. Note that such g is uniquely defined.

Case 3-(a): $o_i = o_j = 0$: In this case, both i and j are boundaries. Hence, both α and β are the empty word, and $w[i..j)$ is fully factorized as $w[i..j) = f_1 \cdots f_t = f$. Likewise, $w[j..k) = f$, and thus $w[i..k) = w[i..j)w[j..k) = f^2$. Since $\tau(g^2) = f^2$, the block $w[i..k)$ is generated from the word g^2. Hence g^2 occurs in x, a contradiction.

Case 3-(b): $o_i = o_j \in \{2, 3\}$: In this case, the factors crossing positions i and j must be 0101. In particular, $w[i-o'_i..i+o_i) = w[j-o'_j..j+o_j) = 0101$. Thus, the block $w[i - o'_i..k - o'_k)$ is factorized as $|0101|f|0101|f|$. Therefore, this block is generated from the word $(\mathrm{b}g)^2$, implying that x contains a square (see the top of Figure 3), a contradiction.

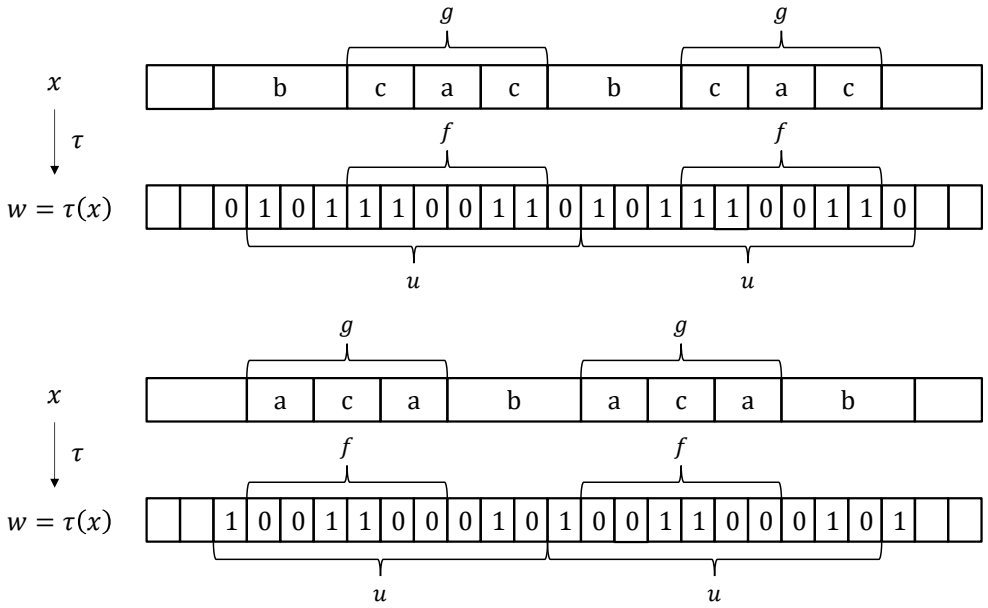


Figure 3. Illustrations of examples for Case 3-(b) (top) and Case 3-(c)-(ii) (bottom) of Lemma 4. In the first figure, the word $w = \tau(x)$ contains a square $uu = (1011100110)^2$ and x contains a square $(\mathrm{b}g)^2 = (\mathrm{bcac})^2$. In the second figure, the word $w = \tau(x)$ contains a square $uu = (1001100010)^2$ and x contains a square $(g\mathrm{b})^2 = (\mathrm{acab})^2$.

Case 3-(c)-(i): $o_i = o_j = 1$ and $o'_j = o'_k = 1$: In this case, the factor crossing position j has length two, i.e., $w[j-o'_j..j+o_j)$ is either 00 or 11. Let c be the character such that $w[j-o'_j..j+o_j) = cc$. Since $w[j-o'_j]$ and $w[j]$ correspond to the last and first characters of u, respectively, we have $\alpha = \beta = c$. Therefore, $w[i..k) = uu$ is of the form $cfccfc$. If $c = 0$, the factor crossing position i must be 00, because it is the only factor that ends with 0. Then, the block $w[i-o'_i..k-o'_k)$ is factorized as $|00|f|00|f|$. Therefore, this block is generated from the word $(\mathrm{a}g)^2$, implying that x contains a square (see the top of Figure 4), a contradiction. If $c = 1$, the factor crossing position k must be 11, because it is the only factor that starts with 1. Then, the block $w[i+o_i..k+o_k)$ is factorized as $|f|11|f|11|$. Therefore, this block is generated from the word $(g\mathrm{c})^2$, implying that x contains a square (see the bottom of Figure 4), a contradiction.

Case 3-(c)-(ii): $o_i = o_j = 1$ and $o'_j = o'_k \neq 1$: In this case, $o'_j = o'_k = 3$. This is because $o_j = 1$, so j is not a boundary and the factor crossing position j has length $o_j + o'_j = 1 + o'_j$. Every factor has length either 2 or 4. The proof of this case is analogous to that of Case 3-(b). In this case, the factors crossing positions j and k are 0101, so the block $w[i+o_i..k+o_k)$ is factorized as $|f|0101|f|0101|$. Therefore, this block is generated from the word $(g\mathrm{b})^2$, implying that x contains a square (see the bottom of Figure 3), a contradiction.

The preceding cases exhaust all possible scenarios for a square uu occurring in $w = \tau(x)$. Since each of these cases leads to a contradiction with the square-freeness of x, we conclude that $w = \tau(x)$ is 3^+-square-free. □

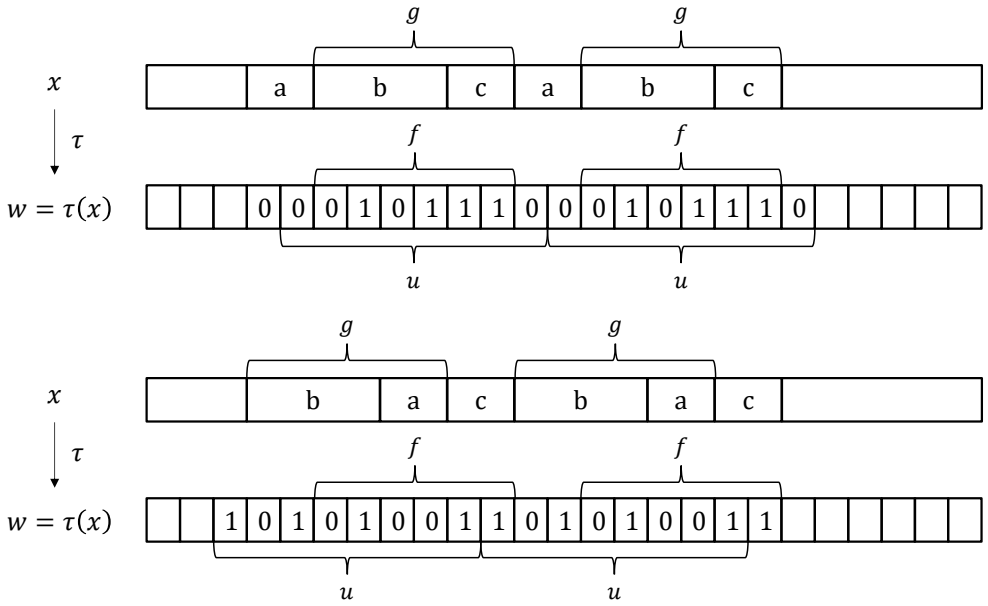


Figure 4. Illustrations of examples for Case 3-(c)-(i) of Lemma 4. In the first figure, the word $w = \tau(x)$ contains a square $uu = (00101110)^2$ and x contains a square $(\mathrm{a}g)^2 = (\mathrm{abc})^2$. In the second figure, the word $w = \tau(x)$ contains a square $uu = (10101001)^2$ and x contains a square $(g\mathrm{c})^2 = (\mathrm{bac})^2$.

5 Conclusions

In this paper, we studied ℓ^+-square-free words under several equivalence relations. We gave combinatorial proofs of the existence of an infinite 3^+-parameterized-square-free ternary word and an infinite 3^+-order-preserving-square-free binary word. Both infinite words were constructed using small morphisms.

Through computational experiments, we also determined the maximum lengths in the finite cases reported in Table 1. Table 2 gives an explicit longest word for each such case. These values and words were obtained with software tools[2]. We list alphabet sizes only up to $\sigma = 5$. For the finite entries in the tables, our experiments also gave the same values when $\sigma \geq 6$.

As a direction for future work, it would be interesting to study repetition thresholds [7,1] under these equivalence relations. A repetition threshold is the smallest real number r such that some infinite word avoids every factor of exponent greater than r. This can be viewed as another direction in generalizing square-freeness: while ℓ^+-square-freeness relaxes classical square-freeness by considering only long squares, repetition thresholds relax it by bounding the exponent of repetitions. In the strict-equality setting, repetition thresholds have been studied extensively, including questions about the maximum exponent and the number of occurrences of factors attaining it. Whether analogous results hold under other equivalence relations is an interesting question.

Acknowledgement. We thank Pascal Ochem for pointing out relevant prior work. This work was supported by JSPS KAKENHI Grant Numbers JP24K20734 (TM),

[2] `https://github.com/koeppl/squarechecker`

Table 2. Examples of longest ℓ^+-square-free words under various equivalence relations and alphabet sizes. Each finite entry gives one longest word. An entry marked ∞ indicates the existence of an infinite ℓ^+-square-free word. A waved line indicates a conjectured result based on experimental results. A dash (−) indicates that infiniteness follows from another case (e.g., a smaller ℓ or σ).

Strict Equality

$\sigma\backslash\ell$	1	2	3	4
2	aba	abaabbaaabbbaabbab	∞	−
3	∞	−	−	−
4	−	−	−	−
5	−	−	−	−

Parameterized Equivalence

$\sigma\backslash\ell$	1	2	3	4
2	a	aaabaaa	∞	−
3	a	aabaaabcc	−	−
4	a	aabaaabcc	−	−
5	a	aabaaabcc	−	−

Order-Preserving Equivalence

$\sigma\backslash\ell$	1	2	3	4
2	a	aaabaaa	∞	−
3	a	∞	−	−
4	a	−	−	−
5	a	−	−	−

Cartesian-Tree Equivalence

$\sigma\backslash\ell$	1	2	3	4
2	a	baaba	bbaaabbabaabbbababaaabbababaa	$\underset{\sim}{\infty}$
3	a	aacbaccba	$\underset{\sim}{\infty}$	−
4	a	aacbaccba	−	−
5	a	aacbaccba	−	−

JP25K21150 (DK), JP23K24808, JP23K18466 (SI), and JST BOOST, Japan Grant Number JPMJBS2406 (HS).

References

1. G. Badkobeh, M. Crochemore, and M. Rao: *Finite repetition threshold for large alphabets.* RAIRO Theor. Informatics Appl., 48(4) 2014, pp. 419–430.
2. B. S. Baker: *Parameterized pattern matching: Algorithms and applications.* J. Comput. Syst. Sci., 52(1) 1996, pp. 28–42.
3. R. Cori and M. R. Formisano: *Partially abelian squarefree words.* RAIRO Theor. Informatics Appl., 24 1990, pp. 509–520.
4. M. Crochemore and W. Rytter: *Squares, cubes, and time-space efficient string searching.* Algorithmica, 13(5) 1995, pp. 405–425.
5. J. D. Currie and J. T. Johnson: *Characterization of the lengths of binary circular words containing no squares other than 00, 11, and 0101.* Theor. Comput. Sci., 850 2021, pp. 30–39.
6. J. D. Currie, F. Manea, D. Nowotka, and K. Reshadi: *Unary patterns under permutations.* Theor. Comput. Sci., 743 2018, pp. 72–82.
7. F. Dejean: *Sur un théorème de thue.* J. Comb. Theory A, 13(1) 1972, pp. 90–99.
8. R. C. Entringer, D. E. Jackson, and J. A. Schatz: *On nonrepetitive sequences.* J. Comb. Theory A, 16(2) 1974, pp. 159–164.
9. A. S. Fraenkel and J. Simpson: *How many squares can a string contain?* J. Comb. Theory, Ser. A, 82(1) 1998, pp. 112–120.
10. R. Hamai, K. Taketsugu, Y. Nakashima, S. Inenaga, and H. Bannai: *On the number of non-equivalent parameterized squares in a string*, in String Processing and Information Retrieval - 31st International Symposium, SPIRE 2024, Puerto Vallarta, Mexico, September 23-25, 2024, Proceedings, Z. Lipták, E. S. de Moura, K. Figueroa, and R. Baeza-Yates, eds., vol. 14899 of Lecture Notes in Computer Science, Springer, 2024, pp. 174–183.
11. V. Keränen: *Abelian squares are avoidable on 4 letters*, in Automata, Languages and Programming, 19th International Colloquium, ICALP92, Vienna, Austria, July 13-17, 1992, Proceedings, vol. 623 of Lecture Notes in Computer Science, Springer, 1992, pp. 41–52.
12. J. Kim, P. Eades, R. Fleischer, S. Hong, C. S. Iliopoulos, K. Park, S. J. Puglisi, and T. Tokuyama: *Order-preserving matching.* Theor. Comput. Sci., 525 2014, pp. 68–79.
13. T. Kociumaka, J. Radoszewski, W. Rytter, and T. Walen: *Maximum number of distinct and nonequivalent nonstandard squares in a word.* Theor. Comput. Sci., 648 2016, pp. 84–95.
14. F. Manea, M. Müller, and D. Nowotka: *Cubic patterns with permutations.* J. Comput. Syst. Sci., 81(7) 2015, pp. 1298–1310.
15. H. Mousavi: *Automatic theorem proving in walnut.* CoRR, abs/1603.06017 2016.
16. T. Ng, P. Ochem, N. Rampersad, and J. O. Shallit: *New results on pseudosquare avoidance*, in Combinatorics on Words - 12th International Conference, WORDS 2019, Loughborough, UK, September 9-13, 2019, Proceedings, R. Mercas and D. Reidenbach, eds., vol. 11682 of Lecture Notes in Computer Science, Springer, 2019, pp. 264–274.
17. S. G. Park, A. Amir, G. M. Landau, and K. Park: *Cartesian tree matching and indexing*, in Proc. CPM, vol. 128 of LIPIcs, 2019, pp. 16:1–16:14.
18. H. Prodinger: *Non-repetitive sequences and Gray code.* Discret. Math., 43(1) 1983, pp. 113–116.
19. H. Prodinger and F. J. Urbanek: *Infinite 0-1-sequences without long adjacent identical blocks.* Discret. Math., 28(3) 1979, pp. 277–289.
20. N. Rampersad, J. O. Shallit, and M. Wang: *Avoiding large squares in infinite binary words.* Theor. Comput. Sci., 339(1) 2005, pp. 19–34.
21. A. Thue: *Über unendliche Zeichenreihen.* Norske Vid. Selsk. Skr. Mat. Nat. Kl., 7 1906, pp. 1–22, Reprinted in *Selected mathematical papers of Axel Thue*, T. Nagell (ed.), Universitetsforlaget, Oslo, 1977, pp. 139–158.