On universal partial words

Herman Z.Q. Chen[*1], Sergey Kitaev[†2], Torsten Mütze[‡3] and Brian Y. Sun[§4]

[1]School of Science, Tianjin Chengjian University, P.R. China
[2]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, UK
[3]Institut für Mathematik, TU Berlin, Germany
[4]College of Mathematics and System Science, Xinjiang University, Urumqi, P.R.China

Abstract. A *universal word* for a finite alphabet A and some integer $n \geq 1$ is a word over A such that every word of length n appears exactly once as a (consecutive) subword. It is well-known and easy to prove that universal words exist for any A and n. In this work we initiate the systematic study of universal *partial* words. These are words that in addition to the letters from A may contain an arbitrary number of occurrences of a special 'joker' symbol $\diamond \notin A$, which can be substituted by any symbol from A. For example, $u = 0\diamond011100$ is a universal partial word for the binary alphabet $A = \{0,1\}$ and for $n = 3$ (e.g., the first three letters of u yield the subwords 000 and 010). We present several result on the existence and non-existence of universal partial words in different situations (depending on the number of $\diamond$s and their positions), including various explicit constructions. We also provide numerous examples of universal partial words that we found with the help of a computer. The full version of the paper is available at [4].

1. Introduction

For a finite alphabet A, we say that a word u is *universal for* A^n if u contains every word of length $n \geq 1$ over A exactly once as a (consecutive) subword. For example, for the binary alphabet $A = \{0,1\}$ and for $n = 3$, $u = 0001011100$ is a universal word for A^3. The cyclic version of universal words is known as *De Bruijn sequences*, which are a centuries-old and well-studied topic in combinatorics, and over the years they found widespread use in real-world applications (see [4]). The following classical result is the starting point for our work.

Theorem 1. *For any finite alphabet A and any $n \geq 1$, there exists a universal word for A^n.*

In this paper we consider universality of so-called *partial words*, words that in addition to letters from A may contain any number of occurrences of an additional special symbol $\diamond \notin A$. The idea is that every occurrence of $\diamond$ can be substituted by any symbol from A, so we can think of $\diamond$ as a 'joker' or 'wildcard' symbol. Formally, we define $A_\diamond := A \cup \{\diamond\}$ and we say that a word $v = v_1 v_2 \cdots v_m \in A^m$ appears as a *factor* in a word $u = u_1 u_2 \cdots u_n \in A_\diamond^n$ if

*E-Mail: zqchern@163.com
†E-Mail: sergey.kitaev@cis.strath.ac.uk
‡E-Mail: muetze@math.tu-berlin.de
§E-Mail: brian@mail.nankai.edu.cn

there is an integer i such that $u_{i+j} = \diamond$ or $u_{i+j} = v_j$ for all $j = 1, 2, \ldots, m$. For example, for the alphabet $A = \{0, 1, 2\}$, the word $v = 021$ occurs twice as a factor in $u = 120\diamond120021$ because of the subwords $0\diamond1$ and 021 of u, whereas v does not appear as a factor in $u' = 12\diamond11\diamond$. Partial words were introduced in [2], and they too have real-world applications (see [3]) and appear in various contexts in combinatorics (see [4]).

The notion of universality given above extends straightforwardly to partial words, and we refer to a universal partial word as an *upword* for short. The simplest example for an upword for A^n is $\diamond^n := \diamond\diamond\cdots\diamond$, the word consisting of n many $\diamond$s, which we call *trivial*. For another example, $\diamond\diamond0111$ is an upword for A^3, whereas $\diamond\diamond01110$ is *not* an upword for A^3, because replacing the first two letters $\diamond\diamond$ by 11 yields the same factor 110 as the last three letters. Similarly, $0\diamond1$ is *not* an upword for A^2 because $10 \in A^2$ does not appear as a factor (while $01 \in A^2$ appears twice as a factor).

2. Our results

In this work we initiate the systematic study of universal partial words. In stark contrast to Theorem 1, there are very few general existence results on upwords, but many non-existence results. The borderline between these two cases seems rather complicated, which makes the subject even more interesting. This is also reflected in our proofs, which become more technical than the straightforward proof of Theorem 1. In addition to the size of the alphabet A and the length n of the factors, we also consider the number of $\diamond$s and their positions in an upword as problem parameters. The following lemma was useful in obtaining some of our results.

Lemma 2. *Let $u = u_1u_2\cdots u_N$ be an upword for A^n, $A = \{0, 1, \ldots, \alpha-1\}$, $n \geq 2$, such that $u_k = \diamond$ and $u_{k+n} \neq \diamond$ (we require $k + n \leq N$). Then for all $i = 1, 2, \ldots, n-1$ we have that if $u_i \neq \diamond$, then $u_{k+i} = u_i$. Moreover, we have that if $u_n \neq \diamond$, then $\alpha = 2$ and $u_{k+n} = \overline{u_n}$ (the complement of u_n).*

For upwords containing a *single* $\diamond$, we have the following results.

Theorem 3. *For $A = \{0, 1, \ldots, \alpha-1\}$, $\alpha \geq 3$, and any $n \geq 2$, there is no upword for A^n with a single $\diamond$.*

For the binary alphabet, the situation is more interesting (see Table 1).

Theorem 4. *For $A = \{0, 1\}$, any $n \geq 3$ and any $k \in \{1, 2, \ldots, n-1\}$, there is an upword for A^n with a single $\diamond$ at position k.*

Theorem 5. *For $A = \{0, 1\}$ and any $n \geq 2$, there is no upword for A^n with a single $\diamond$ at position n.*

Theorem 6. *For $A = \{0, 1\}$, there is no upword for A^n with a single $\diamond$ at position k for $n = 3$ and $k = 4$, and $n = 4$ and $k \in \{5, 7\}$.*

Conjecture 7. *For $A = \{0, 1\}$ and any $n \geq 1$, there is an upword for A^n with a single $\diamond$ at position k except the cases covered in Theorems 5 and 6.*

To support Conjecture 7, we performed a computer-assisted search and indeed found upwords for all values of $2 \leq n \leq 13$ and all possible values of k other than the ones excluded by the beforementioned results. Some of

n	k	
1	1	◇ (Thm. 4)
2	1	◇011 (Thm. 4, Thm. 12)
	2	— (Thm. 5)
3	1	◇00111010 (Thm. 4)
	2	0◇011100 (Thm. 4)
	3	— (Thm. 5)
	4	— (Thm. 6)
4	1	◇00011110100101100 (Thm. 4)
	2	0◇010011011110000 (Thm. 4)
	3	01◇0111100001010 (Thm. 4)
	4	— (Thm. 5)
	5	— (Thm. 6)
	6	01100◇011110100
	7	— (Thm. 6)
	8	0011110◇0010110
5	1	◇00001111101110011000101101010010000 (Thm. 4)
	2	0◇01011000001101001110111110010001 (Thm. 4)
	3	01◇0110000010001110010101111110100 (Thm. 4)
	4	011◇011111000001010010001101 0110 (Thm. 4)
	5	— (Thm. 5)
	6	00101◇0010011101111100000110101
	7	010011◇0100000101011011111000011
	8	0100110◇0100000111001011111 0110
	9	01110010◇0111110110100110000010
	10	010011011◇01000111111000000101011
	11	0101000001◇0101111100111 0110001
	12	01010000011◇0101101111100010011
	13	001001101011◇001010000011111011
	14	0011101111100◇00110100010101100
	15	01010000010011◇0101101111100011
	16	001000001101011◇001010011111011

Table 1. Examples of upwords for A^n, $A = \{0,1\}$, with a single ◇ at position k.

these examples are listed in Table 1, and the remaining ones are available on the third authors' website [1].

For upwords containing *two* ◇s we have the following results (see Table 2).

Theorem 8. *For $A = \{0,1\}$ and any $n \geq 5$, there is no upword for A^n with two ◇s of the form $u = x◇y◇z$ if $|x|, |y|, |z| \geq n$ or $|x| = n-1$ or $|z| = n-1$ or $|y| \leq n-2$.*

Corollary 9. *For $A = \{0,1\}$ and any $n \geq 2$, ◇◇ and ◇◇0111 are the only upwords for A^n containing two ◇s that are adjacent (up to reversal and letter permutations).*

We also construct an infinite family of binary upwords with two ◇s.

$n = 2$	$\Diamond\Diamond$
$n = 3$	$\Diamond\Diamond 0111$ (Thm. 12)
	$\Diamond 001011 \Diamond$
$n = 4$	$\Diamond 00011 \Diamond 1001011$ (Thm. 10)
	$\Diamond 0001011 \Diamond 10011$
	$001 \Diamond 110 \Diamond 001$
$n = 5$	$\Diamond 0100 \Diamond 1010110000011101111110010$
	$\Diamond 0000111 \Diamond 1000100101011001101111$ (Thm. 10)
	$\Diamond 00001001 \Diamond 10001101011111011001$
	$0 \Diamond 0011 \Diamond 0100010101101111100000$
	$0 \Diamond 010110 \Diamond 00011101111100100110$
	$0 \Diamond 0101110 \Diamond 0001101100100111110$
	$00 \Diamond 0011 \Diamond 0010101101111101000 0$
	$01 \Diamond 01100101110 \Diamond 0100000111110$
	$01 \Diamond 0110010111110 \Diamond 01000001110$
	$001 \Diamond 0101 \Diamond 0011101111110000010$
	$011 \Diamond 011010010 \Diamond 0111110000010$
	$01001 \Diamond 1110 \Diamond 010000011011001$

TABLE 2. Examples of upwords for A^n, $A = \{0, 1\}$, with two $\Diamond$s.

Theorem 10. *For $A = \{0, 1\}$ and any $n \geq 4$, there is an upword for A^n with two $\Diamond$s that begins with $\Diamond 0^{n-1} 1^{n-2} \Diamond 1 0^{n-2} 1$.*

Cyclic upwords, where factors are taken cyclically across the word boundaries, are also of our interest. Note that the trivial solution $\Diamond^n$ is a cyclic upword only for $n = 1$. For the cyclic setting we have the following rather general non-existence result.

Theorem 11. *Let $A = \{0, 1, \ldots, \alpha - 1\}$ and $n \geq 2$. If $\gcd(\alpha, n) = 1$, then there is no cyclic upword for A^n. In particular, for $\alpha = 2$ and odd n, there is no cyclic upword for A^n.*

In fact, we know only a single cyclic upword, namely $\Diamond 001 \Diamond 110$ for $n = 4$ (up to cyclic shifts, reversal and letter permutations).

3. Directions of further research

Concerning the binary alphabet, it would be interesting to achieve complete classification of upwords containing a single $\Diamond$ (see Conjecture 7). For two $\Diamond$s such a task seems somewhat more challenging (recall Table 2, Theorem 8 and see the data from [1]). Some examples of binary upwords with three $\Diamond$s are listed in Table 3, and deriving some general existence and non-existence results for this setting would certainly be of interest.

The next step would be to consider the situation of more than three $\Diamond$s present in an upword. The following easy-to-verify example in this direction was communicated to us by Rachel Kirsch [5].

Theorem 12. *For $A = \{0, 1\}$ and any $n \geq 2$, $\Diamond^{n-1} 0 1^n$ is an upword for A^n with $n - 1$ many $\Diamond$s.*

$n = 3$	◇◇◇
$n = 4$	◇◇◇01111 (Thm. 12)
	◇◇001◇11010
	0◇001◇110◇0
$n = 5$	◇0010◇0111◇10011011000001
	◇0000111◇10001001101100101◇1
	◇0000101110◇1000110101001111◇
	◇00001111101◇10001011001◇01
	◇00001101010011110◇1000101111◇
	◇00001101100100111◇1000101◇1
	0◇1100◇0011111011010001O1◇1

Table 3. Examples of upwords for A^n, $A = \{0, 1\}$, with three ◇s for $n = 3, 4, 5$.

Complementing Theorem 12, we can prove (applying Lemma 2 to the first and second ◇s) the following non-existence result in this direction.

Theorem 13. *For $A = \{0, 1\}$, any $n \geq 4$ and any $2 \leq d \leq n - 2$, there is no upword for A^n that begins with $\diamond^d x_{d+1} x_{d+2} \ldots x_{n+2}$ with $x_i \in A$ for all $i = d + 1, \ldots, n + 2$.*

It would also be interesting to find examples of *cyclic* upwords other than ◇001◇110 for $n = 4$ mentioned before.

Finally, a natural direction would be to search for (cyclic) upwords for *non-binary* alphabets, but we anticipate that no nontrivial upwords exist in most cases (recall Theorem 3), if they exist at all (we do not know any). As evidence for this we have the following general non-existence result.

Theorem 14. *For $A = \{0, 1, \ldots, \alpha - 1\}$, $\alpha \geq 3$, and any $d \geq 2$, for large enough n there is no upword for A^n with exactly d many ◇s.*

4. Acknowledgements

The authors thank Martin Gerlach for his assistance in our computer searches. This work was supported by the 973 Project, the PCSIRT Project of the Ministry of Education and the National Science Foundation of China. The last author was supported by the Scientific Research Program of the Higher Education Institution of Xinjiang Uygur Autonomous Region (No. XJEDU2016S032).

References

[1] Currently `http://www.math.tu-berlin.de/~muetze`.

[2] J. Berstel and L. Boasson. Partial words and a theorem of Fine and Wilf. *Theoret. Comput. Sci.*, 218(1):135–141, 1999. WORDS (Rouen, 1997).

[3] F. Blanchet-Sadri. Open problems on partial words. In G. Bel-Enguix, M. D. Jiménez-López, and C. Martín-Vide, editors, *New Developments in Formal Languages and Applications*, pages 11–58. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[4] H. Z.Q. Chen, S. Kitaev, T. Mütze, and B. Y. Sun. On universal partial words. *arXiv:1601.06456*, 2016.

[5] B. Goeckner, N. Graber, C. Groothuis, C. Hettle, B. Kell, R. Kirsch, P. Kirkpatrick, and R. Solava. Personal communication, 2016.