The Defect of Language Tables

Ulrike Brandt*

and

Hermann K.-G. Walter†

15th November 2005

Technical Report TUD-CS-2005-5

*brandt@informatik.tu-darmstadt.de

†walter@informatik.tu-darmstadt.de

Contents

List of Tables

1 Introduction

Language tables are two dimensional data structures build up in a “crossword puzzle” manner. The concept is introduced for arbitrary formal languages. By this approach the global structure of the pattern given by a table is controlled vertically and horizontally by the syntactical structure of the given language.

Moreover with respect to the row-column-grid the set of all alphabetic entries has to be connected.

We study at length the “density”of packing the alphabetic entries. There the minimal number of empty places we are forced to use, the defect of the language, is our main measure. It indicates how compact a pattern can be designed for a given language.

Going to densities we are concerned with asymptotic behaviour.

We show that there are four main cases:

- infinite languages over arbitrary alphabets
- infinite languages over one-letter alphabets
- finite languages over arbitrary alphabets
- finite languages over one-letter alphabets

Every case needs different constructions and methods.

Infinite languages over arbitrary alphabets allow in the worst case a density of 75% in the limit, the bound is sharp.

Infinite languages over one-letter alphabets allow always a density of 100% in the limit. For finite languages over arbitrary alphabets the bounds depend on the largest word in the language. Again we show sharp bounds, especially with respect to one-letter alphabets. For decidability results on this topic we refer the reader to [1].

2 Basic notations and preliminary results

Consider an alphabet X, then X^* is the free monoid over X. Elements $w \in X^*$ are called *words*. If $w = x_1 \dots x_n$ is a word and $1 \leq i \leq n$, we denote by $(w)_i = x_i$ the i-th character of w. Especially $\text{first}(w) = (w)_1$ and $\text{last}(w) = (w)_n$. $|w| = n$ is the *length* of w. For any $x \in X$ we denote by $|w|_x$ the number of occurrences of x in w.

The special symbol $0 \notin X$ denotes empty "entries", therefore we consider (n, m)-matrices A over $X \cup 0$. For all $1 \leq i \leq n$ denote by RA_i the word $RA_i = A[i, 1] \dots A[i, m]$. In the same way we can define CA_i by $CA_i = (RA^T)_i$ $(1 \leq i \leq n)$.

To any (n, m)-matrix A we can associate the graph G_A with nodes $\{(i, j) | A[i, j] \neq 0, 1 \leq i \leq n, 1 \leq j \leq m\}$ and vertices $\{(i, j) \to (k, l) |\ |i-k|+|j-l| = 1, A[i, j] \neq 0, A[k, l] \neq 0\}$. We call A *connected* if and only if G_A is connected.

Definition 2.1 Let $L \subseteq X^*$ be a language, A a matrix over $X \cup 0$ of size (n, m). Then A is called a *R-L*-table iff $\{RA_i | 1 \leq i \leq n\} \subseteq 0^* \cdot ((L \cup X) \cdot 0^+)^* \cdot ((L \cup X) \cup 0^*)$.

Definition 2.2 We call A a L-table iff A and A^T are R-L-tables and A is connected.

Before dealing with examples two further definitions are necessary.

Definition 2.3 If A is a (n, m)-matrix then $\text{def}\,(A) = \sharp \{(i, j)\,|A[i, j] = 0\}$ is the *defect* of A.

In the quadratic case we extend this function to L in the following way:

Definition 2.4 $\text{def}\,(n, L) = \min\{\text{def}\,(A)\,|A \text{ is a } L\text{-table of size}\,(n, n)\}$.

Since $\text{def}(n, L)$ may be hard to compute, we introduce asymptotic measures, too.

Definition 2.5 $\overline{\delta}(L) = \limsup\ \text{def}\,(n, L)\,/n^2$ and $\underline{\delta}(L) = \liminf\ \text{def}\,(n, L)\,/n^2$.

Note that $0 \leq \underline{\delta}(L) \leq \overline{\delta}(L) \leq 1$.

Example 2.1 If $X = \{a\}$ is an one-letter alphabet and $L \subseteq X^*$ is an infinite language, then obviously, for any word $w \in L$ there exists a L-table A of size $|w|$ with $\text{def}(A) = 0$, hence $\underline{\delta}(L) = 0$.

Example 2.2 Consider $X = \{a, b\}$ and $L = \{ab\}$. Obviously one of two best L-tables of size n has $\text{def}(A_n) = n^2 - 2n + 1$, hence $\underline{\delta}(L) = \overline{\delta}(L) = 1$.

$$\begin{bmatrix} & & & & & & \mathbf{a} \\ & & & & & \mathbf{a} & \mathbf{b} \\ & \mathbf{0} & & & \mathbf{a} & \mathbf{b} & \\ & & & \mathbf{a} & \mathbf{b} & & \\ & & & \cdot & & & \\ & & \cdot & & & & \\ & \cdot & & & & \mathbf{0} & \\ & \mathbf{a} & \mathbf{b} & & & & \\ \mathbf{a} & \mathbf{b} & & & & & \\ \mathbf{a}\ \mathbf{b} & & & & & & \end{bmatrix}$$

Example 2.3 Consider $X = \{a, b\}$ and $L = (ab)^*$. Consider a L-table A of size $n \geq 4$. To estimate $\operatorname{def}(A)$ we consider the first two columns. Obviously

$$\begin{aligned} n = |CA_1|_a + |CA_1|_b \quad &+ \quad |CA_1|_0 \quad \textit{and} \\ |CA_2|_0 \quad &\geq \quad |CA_1|_b \quad \textit{and} \\ \frac{n}{2} \quad &\geq \quad |CA_1|_a \end{aligned}$$

By this

$$|CA_1|_0 + |CA_2|_0 \quad \geq \quad n - (|CA_1|_a + |CA_1|_b) + |CA_1|_b \geq \frac{n}{2}.$$

The same argument is true for the last two columns, the first and the last two rows. Since 16 places may be counted twice we get

$$\operatorname{def}(A) \geq 2n - 16.$$

On the other hand consider the following construction for $n = 2m + 2$ $(m \geq 0)$.

$$\mathbf{A_n} = \begin{bmatrix} 0 & 0 & a & 0 & a & & 0 & a & b \\ 0 & a & b & a & b & \dots & a & b & 0 \\ a & b & a & b & a & \dots & b & a & b \\ \vdots & \vdots & & & & & & & \vdots \\ 0 & a & b & a & b & \dots & a & b & 0 \\ a & b & a & b & a & \dots & b & a & b \\ b & 0 & b & 0 & b & \dots & 0 & b & 0 \end{bmatrix}.$$

Obviously,

$$\operatorname{def}(A_n) = 2n - 2.$$

By this we get

$$\underline{\delta}(L) = \overline{\delta}(L) = 0.$$

□

We now exhibit a language L with $\underline{\delta}(L) \geq \frac{1}{4}$.

Theorem 2.1

$$\underline{\delta}((a^2b^2)^*) \geq \frac{1}{4}$$

Proof Consider a $(a^2b^2)^*$-table A of size n. Without loss of generality we assume n is even and $n \geq 2$. We want to show, that any (2,2)-submatrix of A contains at least one empty entry. By this we get immediately $\operatorname{def}(A) \geq n^2/4$, dividing A into $n^2/4$ (2,2)-submatrices.

Assume, to the contrary, that there exist a (2,2)-submatrix A' of A with no "0"-entry. Let

$$A' = \begin{bmatrix} A[i, j] & A[i, j+1] \\ A[i+1, j] & A[i+1, j+1] \end{bmatrix}.$$

Choose i to be minimal. Studying all cases asserts that we can find a (2,2)-submatrix A'' of A with no "0"-entry, which is positioned in row i-1, contradicting the minimality of i. Table 1 shows all cases. The "forced" A'' is represented by dotted line-boxes together with intermediate steps. □

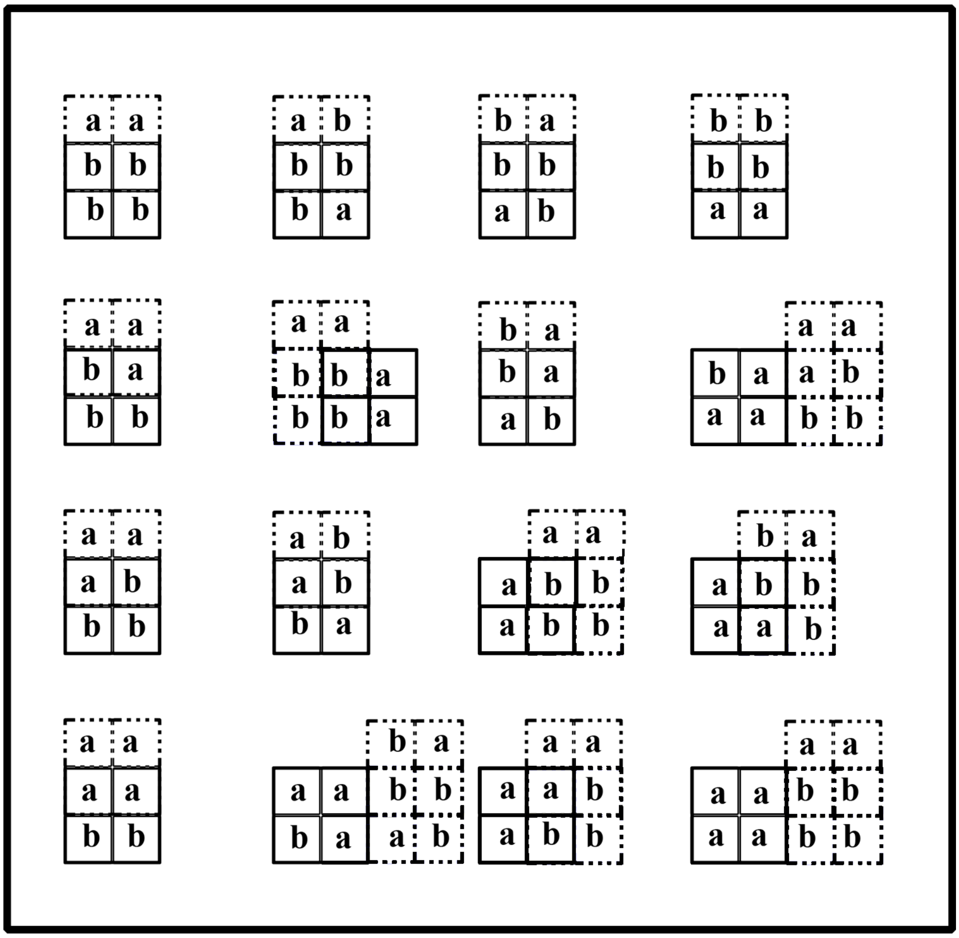
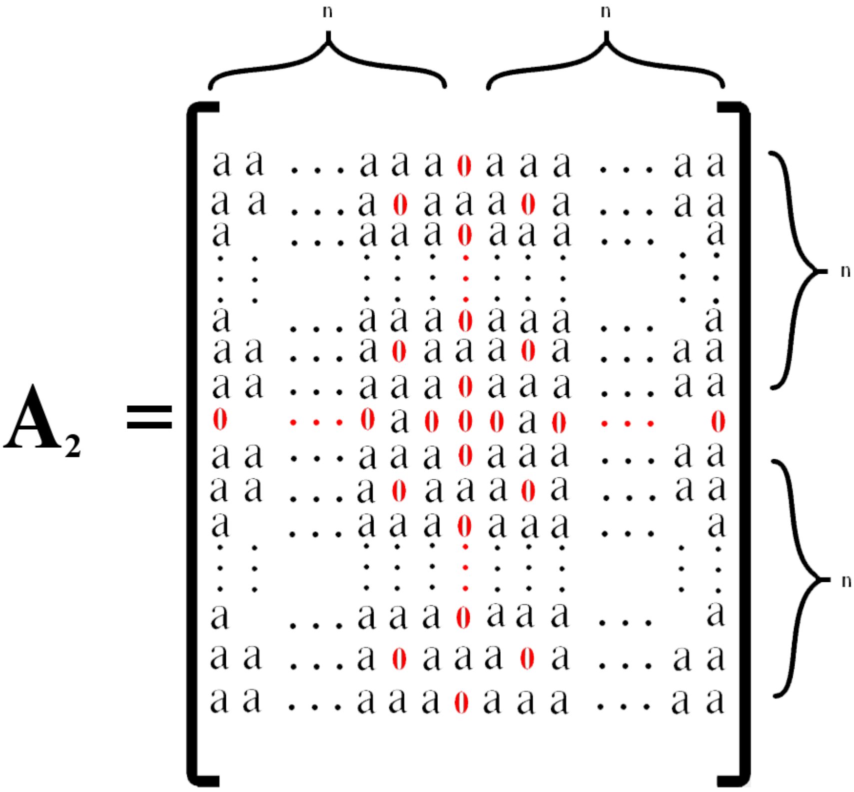


Table 1: Forced completion with intermediate steps.

Example 2.4 Consider $L_n = \{a^3, a^n, a^{n-1}, a^{n-2}\}$. We assume $n > 3$. Suppose n is odd (if n is even, a slightly different construction is possible). Construct the matrix



which is an L-table of size $k \cdot n + (k-1), k \cdot n + (k-1)$, if the construction is done k-times.

$$\begin{aligned}\operatorname{def}(A_k) = 4 \cdot k^2 \cdot \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{n}{2} \right\rfloor \cdot (k-1)^2 + (k-1)^2 \\ - 4(k-1) \cdot 2 \left\lfloor \frac{n}{2} \right\rfloor + 4 \left\lfloor \frac{n}{2} \right\rfloor\end{aligned}$$

This means

$$\overline{\delta}(L) \leq \frac{2n}{(n+1)^2} + \frac{\frac{n}{2}+1}{(n+1)^2} \leq \frac{5}{2n}$$

3 Infinite languages over arbitrary alphabets

We want to show that theorem 2.1 is the best possible result for infinite languages.

Theorem 3.1 If $L \subseteq X^*$ is an infinite language, then

$$\overline{\delta}(L) \leq \frac{1}{4}.$$

Before going into the details of the proof the following observation is necessary.

Observation If L is infinite, then there exist $x, y \in X$ such that $\{w \in L |\ \mathrm{first}\,(w) = x \text{ and } \mathrm{last}\,(w) = y\}$ is infinite.

Proof Obviously:

$$L = \bigcup_{x,y \in X} \{xvy |\ xvy \in L, v \in X^*\}$$

Hence, one of these components must be infinite. □

Now we can assume without loss of generality that for all words w, $w' \in L$: $\mathrm{first}\,(w) = \mathrm{first}\,(w')$ and $\mathrm{last}\,(w) = \mathrm{last}\,(w')$ holds.

For the proof we need a special shift-operation. If $u \in L$ and $k \geq 1$, then for each $w \in 0^*(u0^k)^*u0^*$ with $|w| = m$ we define:

$$\mathrm{shift}\,(w) \quad = \quad p(w)_{|u|}(w)_{|u|+1} \dots (w)_{m-(|u|+k)+1}q$$

with

$$p = \begin{cases} (w)_2 \dots (w)_{|u|-1} & , \text{ if } (w)_1 = 0 \\ 0^{|u|-1} & , \text{ otherwise} \end{cases}$$

and

$$q = \begin{cases} (w)_{m-(|u|+k)+2} \cdots (w)_m 0 & , \text{ if } w \notin (X \cup 0)^* \cdot 0^{|u|+k-1} \\ 0^k u & , \text{ otherwise.} \end{cases}$$

The result of this operation is to shift each letter of w one step to the left, and introduce u from the right if there is enough space. Proper suffixes or prefixes of u on the right or left are not allowed and are replaced by 0's. For example for $u = aba$ and $k = 2$

$$\mathrm{shift}\,(aba00aba00aba0000) \quad = \quad 0000aba00aba00aba.$$

We now proceed in the following way

- First construct a R-L-table using u and k.
- Connect the parts of this R-L-table with the help of special connection-pieces.

The idea is to use u and $k \geq 1$ in such a way that the $u\acute{}s$ separated by k $0's$ are inserted as much as possible. Do it as much as possible in the first row. Then the other rows with odd numbers are generated with the help of the defined shift-operation. To get the third row we apply it two times to the first row. Next we apply it two times to the third row yielding the fifth row and so on.

Rows with even numbers remain empty (filled with 0's) more precisely :
For any odd m with $|u| \leq m$ we define a (m,m)-R-L-table A by

- $RA_{2i} = 0^m \qquad \left(1 \leq i \leq \left\lfloor \frac{m}{2} \right\rfloor\right)$
- $RA_1 = 0^r \left(u0^k\right)^s u$

$$\begin{array}{lcl} with\ r & = & (m - |u|) \mod (|u| + k) \\ and\ s & = & (m - |u|) \operatorname{div} (|u| + k) \end{array}$$

- $RA_{2i+1} = \text{shift}^2 \left(RA_{2i}\right) \qquad \left(1 \leq i \leq \left\lfloor \frac{m}{2} \right\rfloor\right)$

For the sake of adding connection-pieces any u in A which starts in the second column will be shifted one row to the bottom and one column to the left. Analogous on the right side, every u ending in column $m-1$ is shifted one row to the top and one column to the right. Finally in the case that m is even every u in row $m-1$ is shifted one row to the bottom and one column to the left.

The resulting (m,m)-R-L-table A' is shown as part of the table in table 2, represented by horizontal and dotted lines.

Obviously

$$\operatorname{def}\left(A'\right) \leq m^2 - |u| \cdot p \cdot \frac{m}{2},$$

because u occurs in every second row at least $p = \left(\frac{m}{|u|+k} - 1\right)$ times.

Next, we build up the vertical structure:

(i) Fix the first letter of every u in A' and go vertically down with the word $(u)_2(u)_3 \ldots (u)_{|u|}$, if u starts in the upper left triangle of A'.

(ii) Fix the last letter of u and go on vertically up with the word $(u)_{|u|-1}(u)_{|u|-2} \ldots (u)_1$, if u finishes in the lower right triangle of A'.

More precisely A'' is the (m,m)-R-L-table constructed from A' with the following properties:

(i) $\forall\ 1 \leq i, j \leq m, \qquad i + j \leq m$

$$A'[i,j]\, A'[i,j+1] \ldots A'[i,j+|u|-1] = u \Leftrightarrow A''[i,j]\, A''[i+1,j] \ldots A''[i+|u|-1,j] = u$$

and

(ii) $\forall\ 1 \leq i \leq m - |u| + 1\ ,\ |u| \leq j \leq m\ ,\ i + j + |u| - 1 > m$

$$\begin{array}{rl} & A'[i,j]\, A'[i,j+1] \ldots A'[i,j+|u|-1] = u \\ \Leftrightarrow & A''[i,j+|u|-1]\, A''[i-1,j+|u|-1] \ldots A''[i-|u|-1,j+|u|-1] = u \end{array}$$

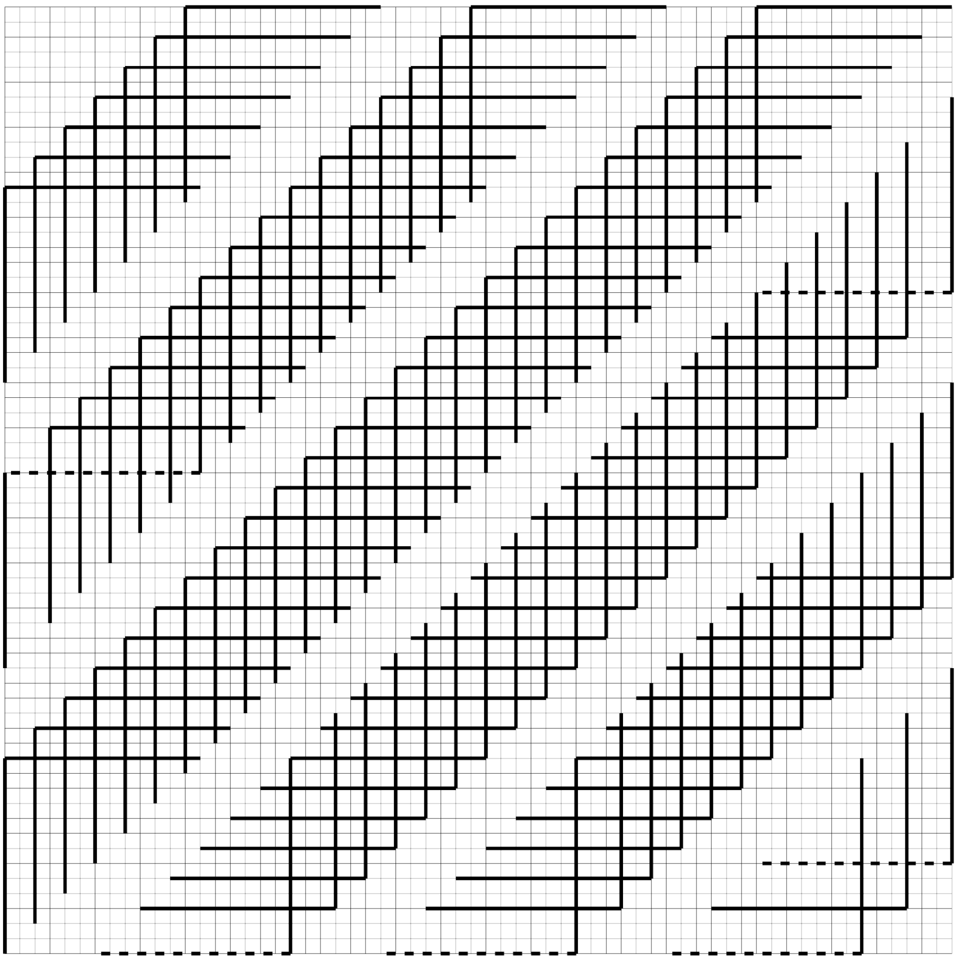


Table 2: $(m,m)-R-L-table\ A''$ with $m = 64, |u| = 14, k = 5$.

(iii) For all other entries : $A''[i,j] = A'[i,j]$

Remember that first (u) and last (u) are fixed letters.

The new (m,m)-R-L-table is A''. Table 2 shows the result of all these transformations. The additional words are indicated by vertical lines. A simple calculation shows

$$\operatorname{def}(A'') \leq \operatorname{def}(A') - \left\lfloor \frac{|u|}{2} \right\rfloor \cdot p \cdot \frac{m}{2}.$$

The reason is, that there are at least $p \cdot \frac{m}{2}$ horizontal $u's$ in A'. Every of these $u's$ is completed by a vertical u. Every time a new vertical u is added at least $\left\lfloor \frac{|u|}{2} \right\rfloor$ 0's are replaced by letters of u.

Therefore we get

$$\begin{aligned}
\text{def}\,(A'') &\leq m^2 - \left(|u| + \left\lfloor \frac{|u|}{2} \right\rfloor\right) \cdot p \cdot \frac{m}{2} \\
&\leq m^2 - \frac{3|u|-1}{4} \left(\frac{m}{|u|+k} - 1\right) \cdot m \\
&= m^2 \cdot \left(1 - \frac{3}{4}\frac{|u|}{|u|+k} + \frac{1}{4(|u|+k)} + \frac{3|u|-1}{4m}\right) \\
&= m^2 \cdot \left(\frac{1}{4} + \frac{3}{4}\frac{k}{|u|+k} + \frac{1}{4(|u|+k)} + \frac{3|u|-1}{4m}\right) \\
&\leq m^2 \left[\frac{1}{4} + \frac{k}{|u|} + \frac{|u|}{m}\right]
\end{aligned}$$

It remains to show that we can connect all these trapezoidal areas, which are indeed connected by the use of the single word u. The failing connections are achieved adjusting the distance of the areas given by the number k.

Choose $v \in L$ with $|v| \geq 3$ as short as possible. Let $\#(X) = t$. Then every $w \in L$ with $|w| \geq |v| \cdot t$ has a decomposition $w = w_0 a_0 \ldots w_t a_t w'$ where $a_0, \ldots, a_t \in X, w_0, \ldots, w_t \in X^{|v|-1}$ and $w' \in X^*$. Clearly, $a_i = a_j$ for some $0 \leq i < j \leq t$, since there are only t different letters in X. Hence our infinite language L has a representation

$$\begin{aligned}
L = &\{w \in L \mid |w| < |v| \cdot t\} \\
\cup &\bigcup_{0 \leq i < j \leq t} \{w_0 a_0 \ldots w_t a_t w' \mid \quad \begin{array}{l} w_0 a_0 \ldots w_t a_t w' \in L, a_0, \ldots, a_t \in X, \\ w_0, \ldots, w_t \in X^{|v|-1}, w' \in X \;\text{ and }\; a_i = a_j\} \end{array}
\end{aligned}$$

which means that at least one of the components of the union must be infinite.

In conclusion there is a letter $a \in X$ and a number $l \in \mathbb{N}$ such that

$$L' = \{w_1 a w_2 a w_3 \mid \quad \begin{array}{l} w_1 a w_2 a w_3 \in L,\; w_1, w_2, w_3 \in X^*, \\ |w_1| \geq |v| - 1, |w_3| \geq |v| - 1 \;\text{ and }\; |w_2| = l(|v|-1) - 1\} \end{array}$$

is infinite. Consider a word $u \in L'$. Then for sufficiently large n a final (n,n)-L-table A_{comp} can be constructed in the following way:
Construct the (m,m)-R-L-table A'' for $m = n - 4|u|$ and $k = (2l-1)(|v|-1) - 1$ according to the rules described above. Why the number k for the distance of the trapeziodal stripes of A'' is chosen in this way is explained later.
To connect the trapeziodal stripes of A'' we create a frame around A'' of width $2|u|$.

First, we number this stripes from the top left corner to the bottom right corner

$$T_1, T_2, \ldots, T_{2r}$$

where

$$r = \lfloor \frac{m - |u|}{|u| + k} \rfloor + 1.$$

Now we connect every stripe with odd number with the next stripe by a special piece on the top respectively on the right of the frame and every stripe with even number with the next stripe on the left respectively on the bottom of the frame. To simplify the description of the process we add a further vertical u ending at the last letter of the last u in row 1 of A'', in the case when r is uneven. In the other case add a horizontal u ending at the last letter of the last u in column 1 of A''.

For all $\frac{r}{2} > z \geq 1$ we use the following connection pieces:

- T_{r-2z+1} and T_{r-2z}, top border: see table 3

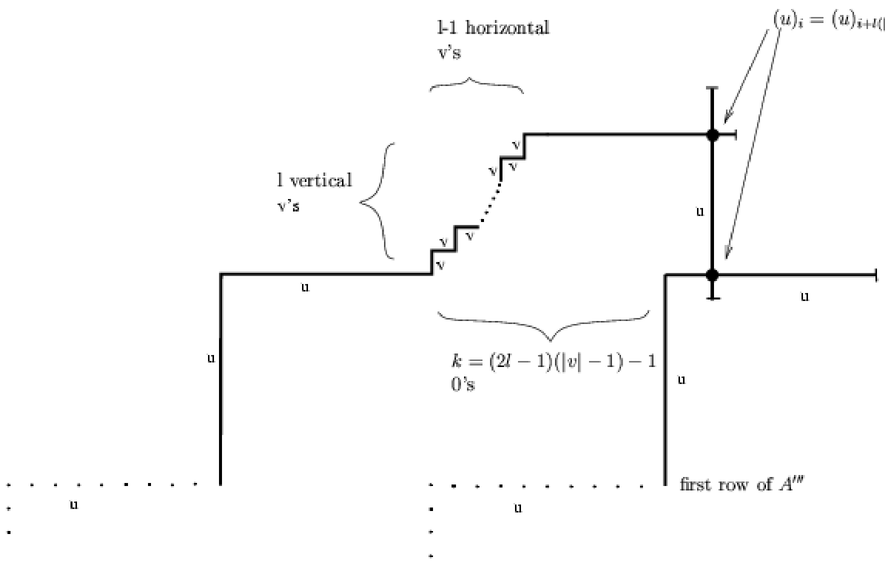


Table 3: T_{r-2z+1} and T_{r-2z}, top border

- $T_{r-2(z-1)}$ and T_{r-2z+1}, left border:
  the corresponding connection piece is obtained from table 1 reflecting it on the y-axis and then turning it 90° to the left.

For all $1 \leq z < \frac{r}{2}$:

- T_{r+2z-1} and T_{r+2z}, bottom border: see table 4

- T_{r+2z} and T_{r+2z+1},right border :
  like in the last case the corresponding connection piece is obtained reflecting table 2 on the y-axis and then turning it 90° to the left.

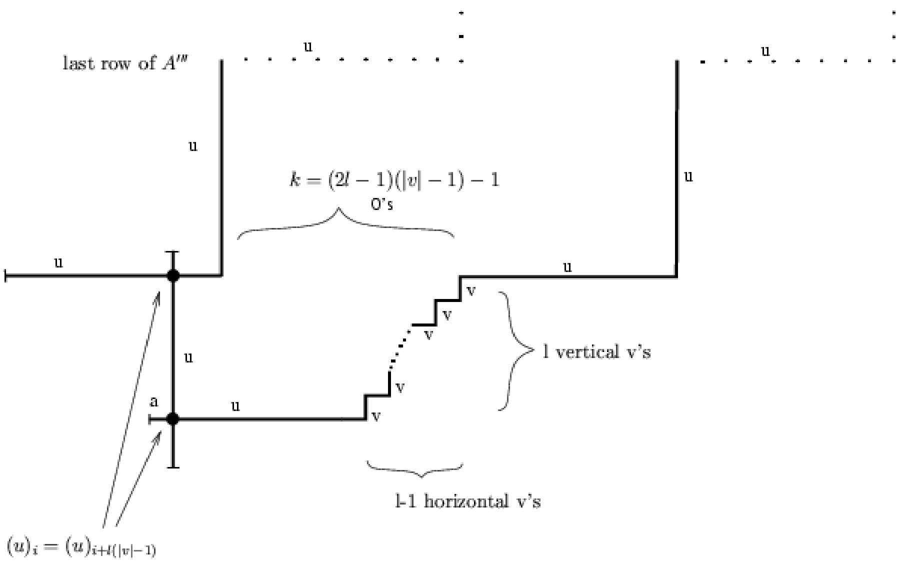


Table 4: T_{r+2z-1} and T_{r+2z}, bottom border

Table 5 shows the resulting (n, n)-L-table A_{comp} for

$$\begin{array}{rcl} n &=& 64 + 4|u|, \\ |u| &=& 14, \\ l &=& 1, \\ |v| &=& 7 \quad (i.e.\ k = (2l-1)(|v|-1)-1 = 5) \\ and\ \ (u)_4 &=& (u)_{10} \quad . \end{array}$$

Estimating the defect of the final A_{comp} we get for $m = n - 4|u|$

$$\operatorname{def}(A_{\text{comp}}) \quad \leq \quad 4 \cdot 2|u| \cdot n + 2 \cdot 2|u| \cdot m + \operatorname{def}(A'') \, .$$

$\Longrightarrow$

$$\begin{aligned} \frac{\operatorname{def}(n, L)}{n^2} &\leq \frac{1}{n^2}\left[12|u| + m^2\left(\frac{1}{4} + \frac{k}{|u|} + \frac{2|u|}{n}\right)\right] \\ &\leq \frac{1}{n^2}\left[12|u| \cdot n + n^2\left(\frac{1}{4} + \frac{k}{|u|} + \frac{2|u|}{n}\right)\right] \\ &\leq \frac{1}{4} + \frac{k}{|u|} + \frac{14|u|}{n} \end{aligned}$$

$\Longrightarrow$

$$\forall \varepsilon > 0 : \frac{\operatorname{def}(n, L)}{n^2} \leq \frac{1}{4} + \varepsilon$$

$$for\ u \in L\ with\ |u| \geq \frac{2k}{\varepsilon}$$
$$and\ all\ n \geq \frac{1}{4}|u|^2$$

$\Longrightarrow$

$$\limsup \frac{\operatorname{def}(n, L)}{n^2} = \bar{\delta}(L) \leq \frac{1}{4}$$

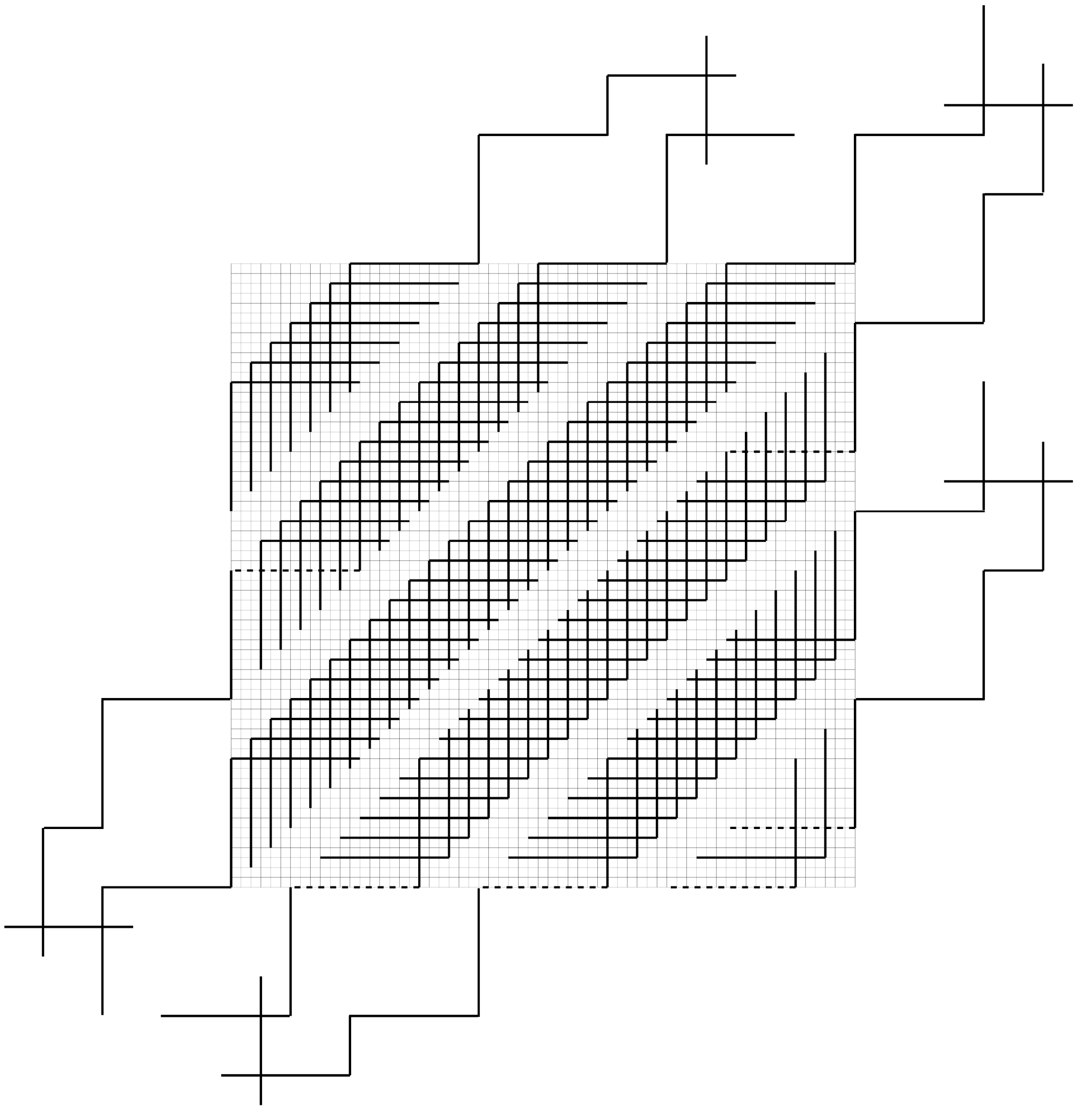


Table 5: A_{comp}

4 Languages over one-letter alphabets

The whole situation looks quite different, if we consider languages over one-letter alphabets. In this case we get the following result.

Theorem 4.1 If $L \subseteq \{a\}^*$ is infinite, then $\bar{\delta}(L) = 0$.

Proof If A is a (l,n)-matrix we denote by A^{rev} the matrix obtained from A by converting A at the "middle"column, that means

$$A^{rev}[i,j] = A[i, n-j+1] \,(\, 1 \leq i \leq l, 1 \leq j \leq n)\,.$$

We now construct "puzzle"-pieces which are the main blocks of our construction.

Consider $u, w \in \{a\}^*$ with $|u|, |w| \geq 4$, then $M'(u,w)$ is a $(|u|+4, |w|+4)$-matrix defined by

$$M'(u,w)[i,j] = \begin{cases} a & 2 < i \leq |u|+2 \text{ and } 2 < j < |w|+2 \\ 0 & \text{otherwise,} \end{cases}$$

Using $M'(u,w)$ we define $M(u,w)$ by

$$M(u,w)[i,j] = \begin{cases} a & \text{if } (i,j) \in \{(1,|w|+2), (2,|w|+2), \\ & (4,|w|+3), (|u|+1,2), \\ & (|u|+3,3), (|u|+4,3)\} \\ 0 & \text{if } (i,j) \in \{(4,3), (|u|+1,|w|+2)\} \\ M'(u,w)[i,j] & \text{otherwise.} \end{cases}$$

For $|u| = 5$ and $|w| = 6$ we get the example

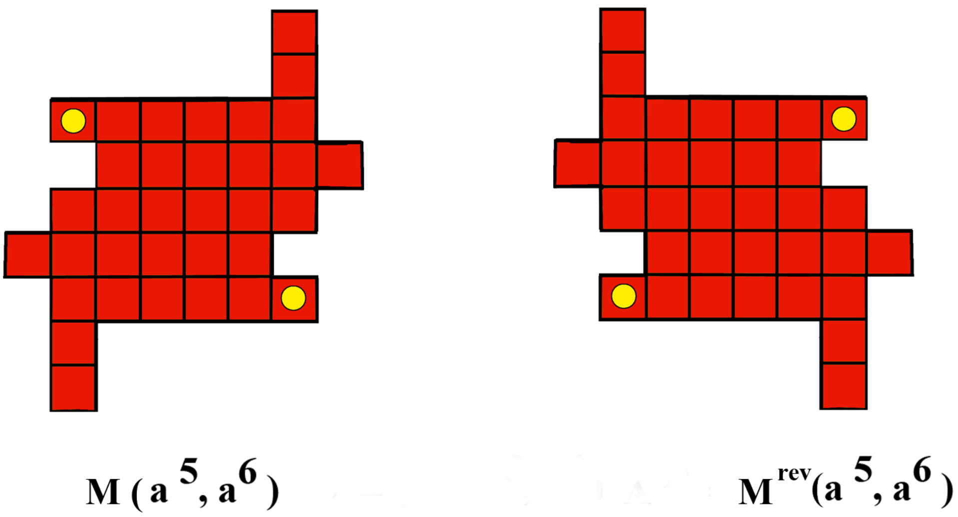


Table 6: "puzzle"-pieces

If $u, w \in L$, then $M(u,w)$ is an L-table with

$$\operatorname{def}(M(u,w)) = (|u|+4)(|w|+4) - (|u| \cdot |w| + 4).$$

Consider a wordsequence $\sigma = (u_1, \ldots, u_k)$ which is ordered (i.e. $|u_i| \le |u_{i+1}|$) and $u_i \in \{a\}^*$ for $1 \le i \le k$. Choose a number $m \ge 1$. Define the R-L-table $A(\sigma, m)$ in the following way:

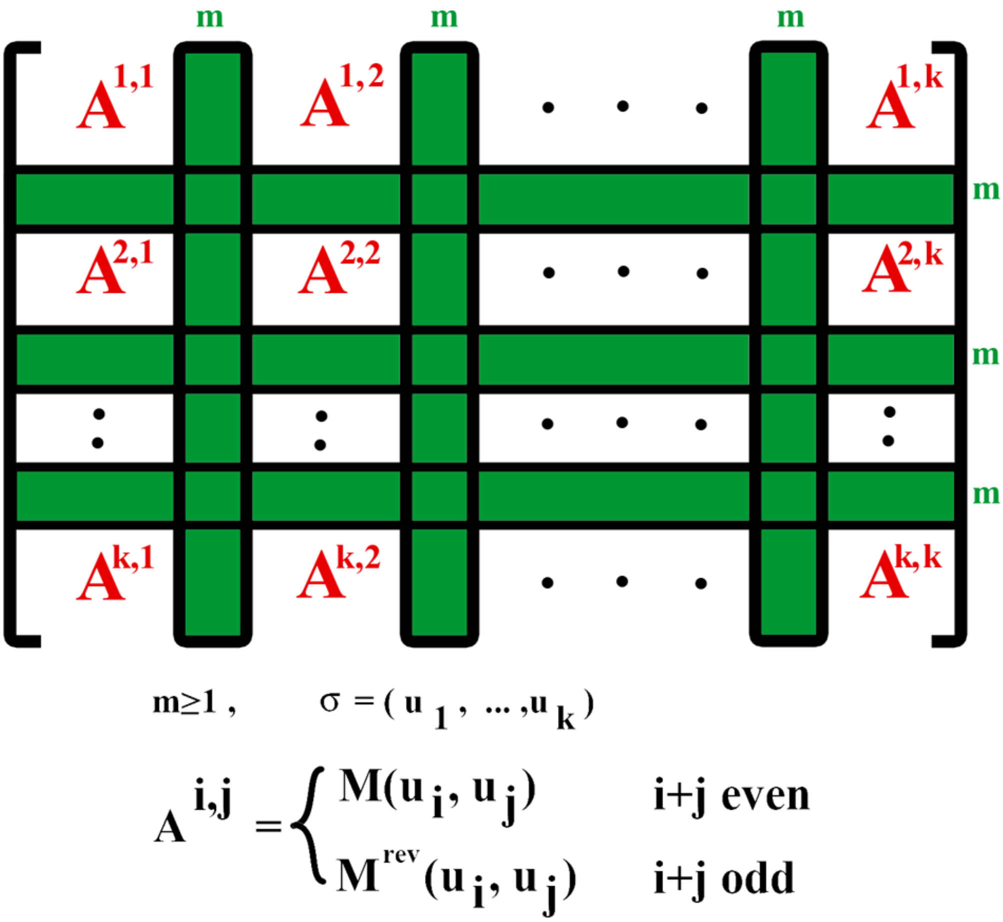


Table 7: $A(\sigma, m)$

Now, if l is the size of $A(\sigma, m)$ then

$$\operatorname{def}(A(\sigma,m)) = l^2 - \sum_{1 \le i,j \le k} (|u_i| \cdot |u_j| + 4)$$

In the special case when there is a word $u \in L$ such that $u_i = u (1 \le i \le k)$

$$\operatorname{def}(A(\sigma,m)) = l^2 - k^2(|u|^2 + 4) \le l^2 - k^2|u|^2.$$

To get an L-Table we choose a word v with $|v| \geq 7$. Then for $m = |v| - 4$ we connect the parts of $A(\sigma, m)$, by an **H**-type table of size $(|v|, |v|)$ with respect to connection points shown in table 5. To the left and the right we only use an **I**-type table of size $(1, |v|)$.

To define $B(\sigma, m)$ we proceed as it shown in the following table:

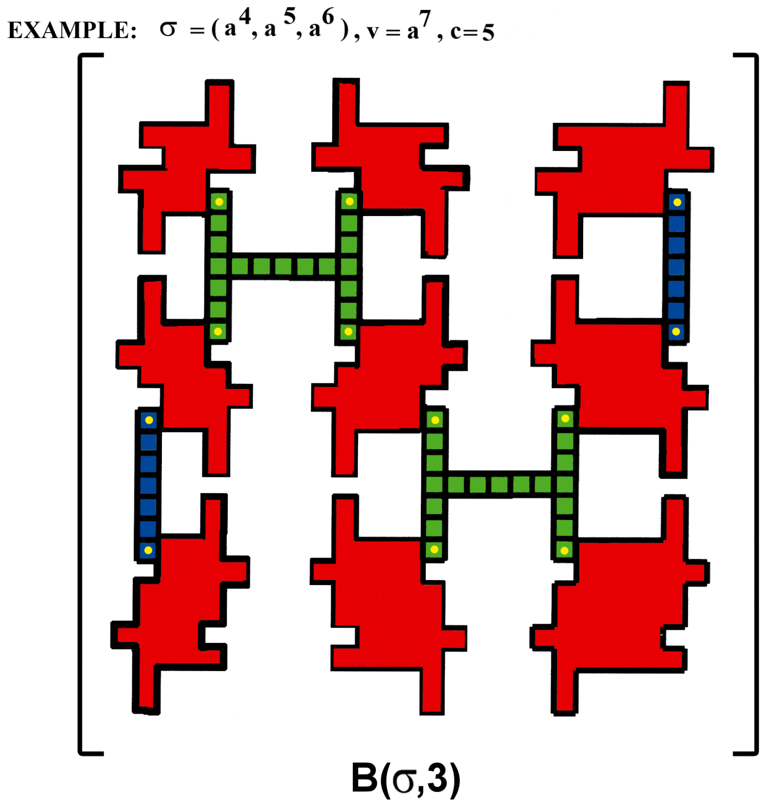


Table 8: Connection of "puzzle"-pieces with connection points

By construction

$$\operatorname{def}(B(\sigma, m)) \quad \leq \quad \operatorname{def}(A(\sigma, m))$$

Now let L be given according to the assumption of the theorem, i.e. there exist $u, v \in L$ with $|u| \geq 4$ and $|v| \geq 7$.

Let $n \geq |u| + 4$. Construct the (n, n)-L-table:

$$\mathbf{C} = \begin{array}{ll} \left[\begin{array}{c|c} \mathbf{B(\sigma,m)} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] & \left.\vphantom{\mathbf{0}}\right\}x \\ \hspace{6em}\underbrace{\hphantom{\mathbf{00}}}_{x} & \end{array}$$

Table 9: Padding

In the construction

$$\begin{aligned} m &= |v| - 4, \\ k &= \left\lfloor \frac{n - (|u| + 4)}{|u| + 4 + m} \right\rfloor + 1, \\ \sigma &= (u_1, \ldots, u_k) \textit{ with } u_i = u (1 \le i \le k), \\ x &= n - l, \end{aligned}$$

where $l = (k-1)(|u| + 4 + m) + |v| + 4$ is the size of $B(\sigma, m)$.

Since

$$k \le \frac{n-1}{|u| + 5}$$

we get

$$\begin{aligned} \mathrm{def}\,(C) &\le n^2 - k^2 |u|^2 \\ &\le n^2 - \left(\frac{n-1}{|u|+5} \right)^2 |u|^2 \\ &= n^2 - (n-1)^2 \cdot \frac{(|u|+5)^2 - (10|u| + 25)}{(|u|+5)^2} \\ &\le 2n - 1 + (n-1)^2 \cdot \frac{10}{|u|} \end{aligned}$$

Therefore

$$\frac{\mathrm{def}\,((n, L))}{n^2} \le \frac{2}{n} + \frac{10}{|u|}$$

yielding

$$\frac{\mathrm{def}\,((n, L))}{n^2} \le \epsilon$$

for every $n \geq |u_0|$
where u_0 is the smallest word in L with

$$|u_0| \geq \frac{12}{\epsilon}$$

In conclusion $\bar{\delta}(L) = 0$. $\square$

5 Finite languages

Finite languages need completely different constructions to obtain “maximal”density. In the case that there exist a $w \in L$ such that w can be decomposed into

$$w = uxtxv$$

with

$$x \in X, t \neq \Box, |uv| < |t| \text{ and } u, v, t \in X^*,$$

we can use a construction which resembles the “rook-tour”in chess:

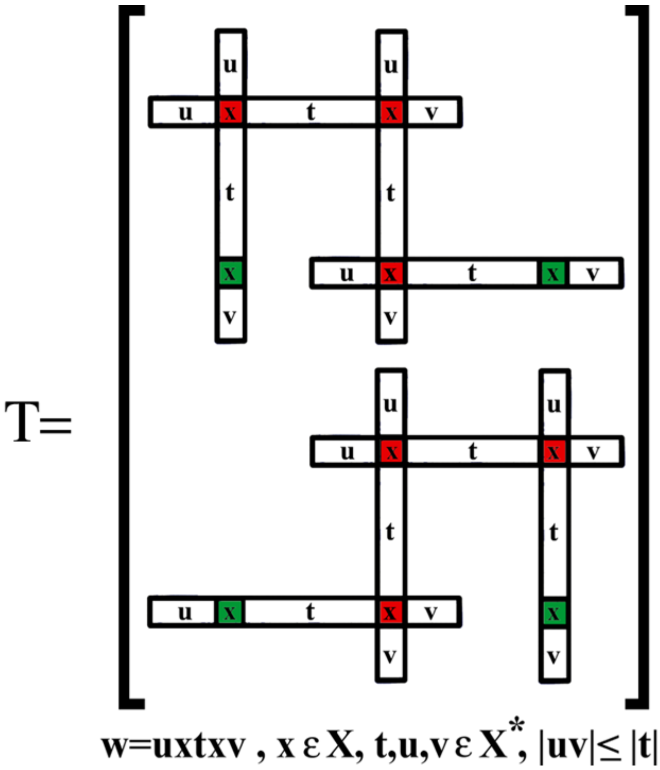
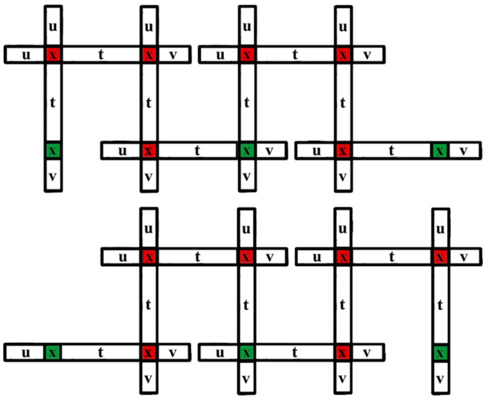


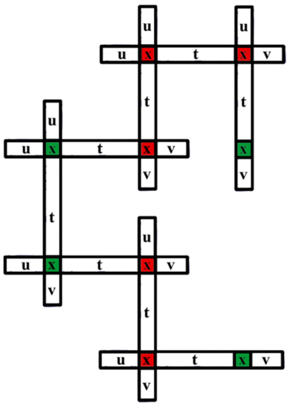
Table 10: Stripe component

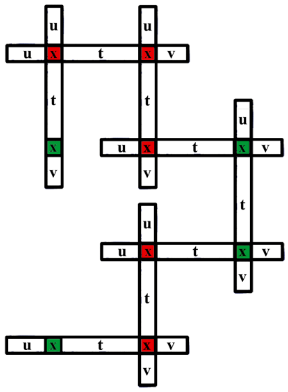
T is a $(2|w| + |t| - |uv|, |w| + |t| + 1)$-matrix.

We can build the “rook-tour”out of T with the help of these constructors:



T & T





LEFT CONNECTION **RIGHT CONNECTION**

Table 11: “Rook-tour”constructors

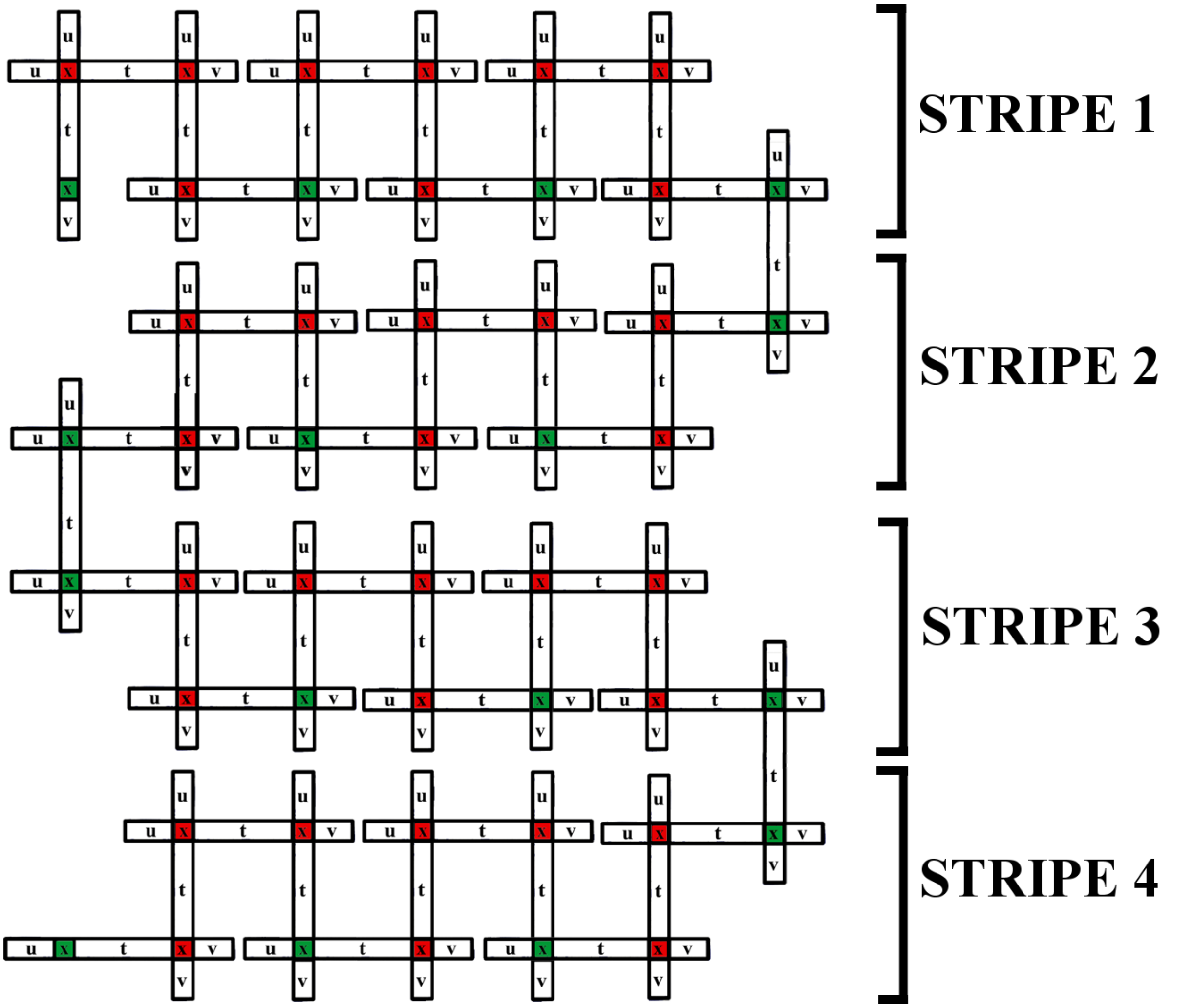


Table 12: Complete “rook-tour”with stripes

Now, the defect of a (n,n)-L-table A constructed with intent to maximize the number and length of the stripes can be estimated by

$$\operatorname{def}(A) \leq n^2 - \left\lfloor \frac{n - |uxv|}{2|tx|} \right\rfloor \cdot 4 \cdot \left\lfloor \frac{n - |uxv|}{2|tx|} \right\rfloor (|w| - 1)$$

Since

(i) The word w occurs in every stripe

$$4 \cdot \left\lfloor \frac{n - |uxv|}{2|tx|} \right\rfloor -\text{times.}$$

(ii) x is used twice for two "different" w's.

(iii) Within a (n,n) −matrix A

$$\left\lfloor \frac{n-|uxv|}{2|tx|} \right\rfloor$$

stripes can be allocated.

By this

$$\begin{aligned} \operatorname{def}(n,L) &\leq n^2 - \left\lfloor \frac{n-|uxv|}{2|tx|} \right\rfloor \cdot 4 \cdot \left\lfloor \frac{n-|uxv|}{2|tx|} \right\rfloor (|w|-1) \\ &\leq n^2 - \left(\frac{n-|uxv|-2|tx|}{|tx|} \right)^2 (|w|-1) \end{aligned}$$

hence

$$\begin{aligned} \frac{\operatorname{def}(n,L)}{n^2} &\leq 1 - \frac{|w|-1}{|tx|^2} + \mathcal{O}\left(\frac{1}{n}\right) \\ &\leq 1 - \frac{|tx|+|uxv|-1}{|tx|^2} + \mathcal{O}\left(\frac{1}{n}\right) \\ &\leq 1 - \frac{1}{|tx|} + \mathcal{O}\left(\frac{1}{n}\right). \end{aligned}$$

Therefore

Theorem 5.1 If $w \in L$ is a word with a decomposition $w = uxtxv \quad (x \in X, |t| \geq 1, |uv| < |t|,\ u,t,v \in X^*)$ then

$$\overline{\delta}(L) \leq 1 - \frac{1}{|tx|}$$

Observe, that we used only one single word to construct the whole table. Thus, the result $\overline{\delta}(L) < 1$ holds even for languages containing only one word, if this word has the desired property.
This is not always the case. Consider a language $L = \{w\}$, where every letter occurs only once in w.
In this case the only possibilities to annex a vertical w to a horizontal w respectively vice versa a horizontal w to a vertical w are shown in table 13. That means: if w starts at position (i,j) of an arbitrary (n,n)-L-table A then

$$i' + j' = i + j$$

for every starting point (i',j') of every w in A.
Since there is at most one j' for every i' fullfilling the equation, there are at most n possibilities w starts in A. Hence

$$\operatorname{def}(A) \geq n^2 - n \cdot |w|$$

and we conclude

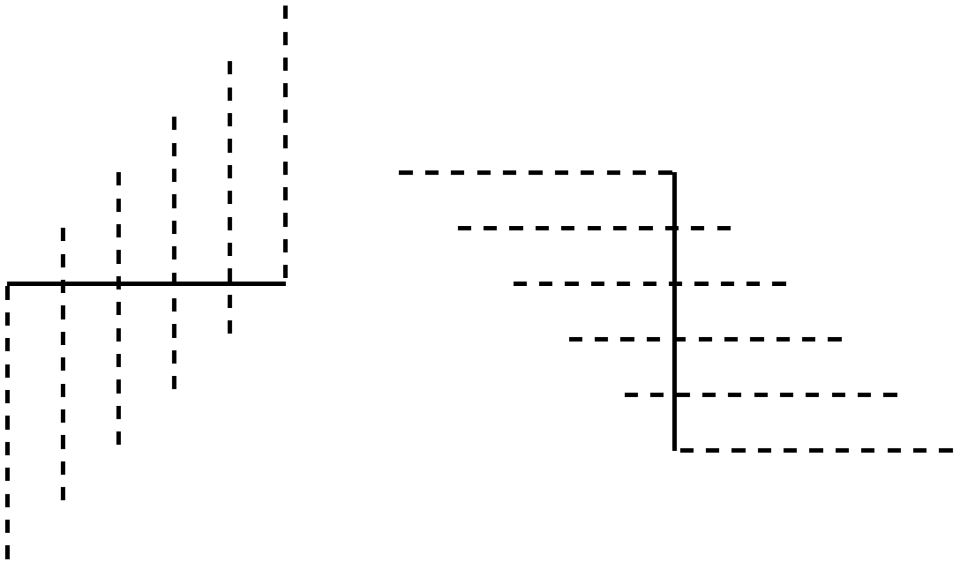


Table 13: Possibilities (represented by dotted lines) to annex a vertical (horizontal) w to a horizontal (vertical) w.

Lemma 5.1 :
For every language $L \subseteq X^*$ with $L = \{w\}$

$$\overline{\delta}\,(L) = 1,$$

if $|w|_z \leq 1$ for every $z \in X$.

A lower bound can be obtained quite easily. Let L be finite and $k = \max\{|w| \big| w \in L\}$. Consider an $(n, n) - L$ table A, n much larger than k.

Consider a row. This row can in the best case contain n/k words from L, but all these words must be separated from each other, hence at least $n/k - 1$ empty entries must exist in the row. Since there are n rows we get

$$\operatorname{def}(A) \geq \frac{n^2}{k} - n$$

Which means

$$\underline{\delta}(L) \geq \frac{1}{k} \qquad .$$

Lemma 5.2 If L is a finite language, then

$$\underline{\delta}(L) \geq \frac{1}{\max\{|w| \big| w \in L\}}$$

6 Concluding remarks

The main open research area is to introduce more combinatorial features of languages to get more refined results.

As an example we look at commutative languages L, which are invariant under wordwise permutations of letters. Obviously, for infinite commutative languages L $\underline{\delta}\,(L) = 0$. It seems to be possible to transfer the result for infinite languages over an one-letter alphabet to infinite commutative languages.

Another topic is the density of languages. At the moment we do not see how to use this concept to get more refined results.

A third aspect would be pumping lemmata, but until now we didn't get better results in the general case [2].

References

[1] U.Brandt, H.K.-G. Walter, Complete Language Tables, Papers on Automata and Languages VIII. Dep. of Mathematics, Karl Marx University of Economics, Budapest, 1986 - 3, pp. 13-32.

[2] V. Obermeit, Defekt von Tableaus für kontextfreie Sprachen. Diplomarbeit am Institut für Theoretische Informatik, Fachbereich Informatik, Technische Hochschule Darmstadt 1983.