INFORMATIQUE THÉORIQUE ET APPLICATIONS

RAINER KEMP

On the average minimal prefix-length of the generalized semi-Dycklanguage

Informatique théorique et applications, tome 30, n^{o} 6 (1996), p. 545-561

<http://www.numdam.org/item?id=ITA_1996__30_6_545_0>

© AFCET, 1996, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 30, n° 6, 1996, pp. 545-561)

ON THE AVERAGE MINIMAL PREFIX-LENGTH OF THE GENERALIZED SEMI-DYCKLANGUAGE (*)

by Rainer KEMP [1]

Communicated by J. BERSTEL

Abstract. – *Given two disjoint alphabets* $T_\sqsubset$ *and* $T_\sqsupset$ *and a relation* $\mathfrak{R} \subseteq T_\sqsubset \times T_\sqsupset$, *the "generalized semi-Dycklanguage"* $D^{\mathfrak{R}}$ *over* $T_\sqsubset \cup T_\sqsupset$ *consists of all words* $w \in (T_\sqsubset \cup T_\sqsupset)^*$ *which are equivalent to the empty word under the congruence* δ *defined by* $xy \equiv \varepsilon \bmod \delta$ *for all* $(x, y) \in \mathfrak{R}$. *For arbitrary* $\mathfrak{R}$, *we compute the average length of the shortest prefix which has to be read in order to decide whether or not a given word of length* n *over* $(T_\sqsubset \cup T_\sqsupset)^*$ *belongs to* $D^{\mathfrak{R}}$.

Résumé. – *Étant donnés deux alphabets disjoints* $T_\sqsubset$ *et* $T_\sqsupset$ *et une relation* $\mathfrak{R} \subseteq T_\sqsubset \times T_\sqsupset$, *le « langage de semi-Dyck généralisé »* $D^{\mathfrak{R}}$ *sur* $T_\sqsubset \cup T_\sqsupset$ *est composé des mots* $w \in (T_\sqsubset \cup T_\sqsupset)^*$ *qui sont équivalents au mot vide pour la congruence* δ *définie par* $xy \equiv \varepsilon \bmod \delta$ *pour tout* $(x, y) \in \mathfrak{R}$. *Pour tout* $\mathfrak{R}$, *nous calculons la longueur moyenne du plus court préfixe d'un mot de longueur* n *sur* $(T_\sqsubset \cup T_\sqsupset)^*$ *qu'il faut lire pour décider si ce mot appartient ou non au langage* $D^{\mathfrak{R}}$.

1. INTRODUCTION AND BASIC DEFINITIONS

The *membership problem*, *i.e.* the question whether or not a given word w belongs to a given language $\mathcal{L}$, is a fundamental problem in formal language theory. A simple strategy to solve this problem is as follows: Let $\ell(w)$ be the length of the word w. Scan w from left to right letter by letter until the last symbol of the shortest prefix v which has no extension rightwards to any word of length $\ell(w)$ of the language $\mathcal{L}$. If $w \in \mathcal{L}$, then we have to read $\ell(w)$ symbols; but, if $w \notin \mathcal{L}$, then we only have to read $\ell(v) \leq \ell(w)$ symbols. Naturally, such a recognition procedure presupposes information about the words which have an extension rightwards to a word of length $\ell(w)$ belonging to $\mathcal{L}$ and those ones not having such a continuation.

(*) Received July 1996.

([1]) Johann Wolfgang Goethe-Universität, Fachbereich Informatik, D-60054 Frankfurt am Main, Germany.

0988-3754/96/06/$ 7.00/© AFCET-Gauthier-Villars

Given a formal language $\mathcal{L}$ over an alphabet T furnished with a probability distribution, a general approach to the computation of the average length of the shortest prefix v has been presented in [8]. This approach covers a complete average-case analysis of that parameter, including higher moments about the origin and the cumulative distribution function. In this note, we shall deal with the class of generalized semi-Dycklanguages which is defined as follows:

DEFINITION 1: Let $k_1, k_2 \in \mathbb{N}$, $T_{\sqsubset} := \{\sqsubset_i \mid 1 \leq i \leq k_1\}$,

$$T_{\sqsupset} := \{\sqsubset_i \mid 1 \leq i \leq k_2\}$$

and $T := T_{\sqsubset} \cup T_{\sqsupset}$. Given a relation $\mathfrak{R} \subseteq T_{\sqsubset} \times T_{\sqsupset}$, the *generalized semi-Dycklanguage* $D^{\mathfrak{R}}$ *associated with* $\mathfrak{R}$ is defined by

$$D^{\mathfrak{R}} := \{w \in T^* \mid w \equiv \varepsilon \bmod \delta\},$$

where ε denotes the empty word and δ is the congruence over T given by $(\forall (x, y) \in \mathfrak{R})$ $(xy \equiv \varepsilon \bmod \delta)$. The elements of $T_{\sqsubset}$ (resp. $T_{\sqsupset}$) are called *opening* (resp. *closing*) *brackets*. The sets R_1 and R_2 are defined by

$$R_1 := \{x \in T_{\sqsubset} \mid (\exists y \in T_{\sqsupset}) ((x, y) \in \mathfrak{R})\}$$

and

$$R_2 := \{y \in T_{\sqsupset} \mid (\exists x \in T_{\sqsubset}) ((x, y) \in \mathfrak{R})\},$$

respectively. $\diamond$

Choosing $k_1 := k_2 := k \in \mathbb{N}$ and $\mathfrak{R} := \{(\sqsubset_i, \sqsupset_i) \mid 1 \leq i \leq k\}$ in the preceding definition, $D^{\mathfrak{R}}$ coincides with the usual semi-Dycklanguage D_k with k types of brackets (e.g. [6], pp. 312). Applying the general approach presented in [8], we are able to determine the exact asymptotical behaviour of the average length of the shortest prefix which has to be read in order to decide whether or not a word $w \in T^n$ belongs to $D^{\mathfrak{R}}(n) := D^{\mathfrak{R}} \cap T^n$ provided that all words w are equally likely. The presented analysis includes the computation of the higher moments about the origin, too (*cf.* Theorem 1, Corollary 1). Informally, we shall show that the growth of the average length of the shortest prefix is of order

– $\Theta(1)$ if and only if the alphabet T contains brackets not appearing in $\mathfrak{R}$,

or

all brackets in T appear in $\mathfrak{R}$, but $\mathfrak{R}$ is a proper subset of $T_{\sqsubset} \times T_{\sqsupset}$, or

$\mathfrak{R}$ is equal to $T_\sqsubset \times T_\sqsupset$, but there are less opening brackets in T than closing brackets in T;

- $\Theta(n^{\frac{1}{2}})$ if and only if $\mathfrak{R}$ is equal to $T_\sqsubset \times T_\sqsupset$ and there are as much opening brackets in T as closing brackets in T;
- $\Theta(n)$ if and only if $\mathfrak{R}$ is equal to $T_\sqsubset \times T \sqsupset$ and there are more opening brackets in T than closing brackets in T.

Let us conclude this introductory section by some further definitions and notations used in the rest of the paper.

Given a formal language $\mathcal{L} \subseteq T^*$, the set

$$\text{INIT}(\mathcal{L}) := \{u \in T^* \mid (\exists\, v \in T^*)\,(uv \in \mathcal{L})\}$$

denotes the set of all prefixes appearing in words belonging to $\mathcal{L}$. The set $\text{INIT}_r(\mathcal{L})$ is defined by $\text{INIT}_r(\mathcal{L}) := \text{INIT}(\mathcal{L}) \cap T^r$. We say that T is the *smallest alphabet for* $\mathcal{L}$ if $(\forall\, \hat{T} \subset T)\,(\mathcal{L} \nsubseteq \hat{T}^*)$. A prefix $u \in \text{INIT}(\{w\})$, $w \in T^n$, is called a *minimal prefix of* w *with respect to* $\mathcal{L}(n) := \mathcal{L} \cap T^n$ iff $u \in \text{INIT}_{k-1}(\mathcal{L}(n)) \cdot T \backslash \text{INIT}_k(\mathcal{L}(n))$, where $k \in [1:n]$ is minimal. Obviously, the minimal prefix of a word $w \in T^n$ with respect to $\mathcal{L}(n)$ cannot be extended rightwards to any word of length n in $\mathcal{L}$; after reading such a prefix, the given input word w has to be rejected by the recognition procedure described above.

Next, let us consider the generalized semi-Dycklanguage $D^{\mathfrak{R}}$. Since all words in $D^{\mathfrak{R}}$ have an even length, a minimal prefix of $w \in T^n$ with respect to $D^{\mathfrak{R}}(n)$, $n \equiv 1 \bmod 2$, does not exist. If $n \equiv 0 \bmod 2$, a minimal prefix u with respect to $D^{\mathfrak{R}}(n)$ satisfies the following properties:

(i) $\#_\sqsubset(u) < \#_\sqsupset(u)$

or

(ii) $\#_\sqsubset(u) > \frac{1}{2}n$

or

(iii) $u \in T^* \cdot (\{x\} \cap T_\sqsubset) \cdot D^{\mathfrak{R}} \cdot (\{y\} \cap T_\sqsupset)$ with $(x,\, y) \notin \mathfrak{R}$.

Here, $\#_\sqsubset(u)$ (resp. $\#_\sqsupset(u)$) denotes the number of opening (resp. closing) brackets appearing in u. The condition $\#_\sqsubset(u) < \#_\sqsupset(u)$ means that a word u consisting of more closing brackets than opening brackets cannot be a prefix of $D^{\mathfrak{R}}$. The second property $\#_\sqsubset(u) > \frac{1}{2}n$ takes the fact into account that a Dyckword $w \in D^{\mathfrak{R}}(2n)$ cannot have more than n opening brackets. The last condition reflects the property that the opening bracket $x \in T_\sqsubset$ and the closing bracket $y \in T_\sqsupset$ do not clash because $(x,\, y) \notin \mathfrak{R}$.

Given a prefix $u \in T^* \cdot (\{x\} \cap T_{\sqsubset}) \cdot D^{\mathfrak{R}} \cdot (\{y\} \cap T_{\sqsupset}) \cap \mathrm{INIT}\,(D^{\mathfrak{R}}(n))$, $n \equiv 0 \bmod 2$, with $(x, y) \in \mathfrak{R}$, the tuple (x, y) of brackets is said to be a *correct pair of brackets* of $u \in \mathrm{INIT}\,(D^{\mathfrak{R}}(n))$. All opening brackets not appearing in a correct pair of brackets of u are called *free opening brackets* of u. The *structure of a prefix* $u \in \mathrm{INIT}\,(D^{\mathfrak{R}})$ is the word $\varphi(u)$, where $\varphi : T^* \to \{\sqsubset_1, \sqsupset_1\}$ is the monoidhomomorphism defined by $\varphi(x) := \sqsubset_1$ if $x \in T_{\sqsubset}$, and by $\varphi(x) := \sqsupset_1$ if $x \in T_{\sqsupset}$. Note that $\varphi(D^{\mathfrak{R}}) = D_1$ for all $\mathfrak{R} \subseteq T_{\sqsubset} \times T_{\sqsupset}$.

2. THE AVERAGE MINIMAL PREFIX-LENGTH OF $D^{\mathfrak{R}}$

Let $Y_{\mathrm{pref}}\,(D^{\mathfrak{R}}(n))$ be the random variable describing the length of the minimal prefix which has to be read in order to decide whether or not an input word $w \in T^n$ belongs to the language $D^{\mathfrak{R}}(n)$. Assuming that all words $w \in T^n$ are equally likely, the general considerations presented in [8] imply that the s-th moment about the origin of $Y_{\mathrm{pref}}\,(D^{\mathfrak{R}}(n))$ is equal to

$$\mathbb{E}\,[Y_{\mathrm{pref}}^s\,(D^{\mathfrak{R}}(n))] = \sum_{0 \le k \le n} [(k+1)^s - k^s]\,|\,\mathrm{INIT}_k\,(D^{\mathfrak{R}}(n))\,|\,|T|^{-k}. \quad (1)$$

Thus, we first have to compute an expression for $|\,\mathrm{INIT}_k\,(D^{\mathfrak{R}}(n))\,|$. Since $\mathrm{INIT}_k\,(D^{\mathfrak{R}}(n)) = \emptyset$ for $n \equiv 1 \bmod 2$, we have $\mathbb{E}\,[Y_{\mathrm{pref}}^s\,(D^{\mathfrak{R}}(n))] = 0$ for odd n. In the sequel, we shall only deal with $\mathbb{E}\,[Y_{\mathrm{pref}}^s\,(D^{\mathfrak{R}}(n))]$ for even n.

Consider the well-known one-to-one correspondence (e.g. [7], p. 173) between the Dyckwords $w \in D_1(2n)$ and the labelled paths from $(0, 0)$ to $(2n, 0)$ in the diagram drawn in figure 1.

Obviously, the number of prefixes of length k in $D_1(2n)$ is equal to the number of all paths from $(0, 0)$ to all points (k, i), $0 \le i \le \min\,\{k, 2n-k\}$ with $(k+i) \equiv 0 \bmod 2$. It is well-known that the number $w(k, i)$ of paths from $(0, 0)$ to (k, i) is equal to the ballot number $a_{\frac{1}{2}(k+i),\, \frac{1}{2}(k-i)}$ ([3], p. 259), *i.e.*

$$w\,(k, i) = \binom{k}{\frac{1}{2}(k-i)} - \binom{k}{\frac{1}{2}(k-i)-1}. \quad (2)$$

Such a path corresponds to a word $u \in \mathrm{INIT}_k(D_1(2n))$ with $\frac{1}{2}(k+i)$ opening brackets and $\frac{1}{2}(k-i)$ closing brackets. Hence, the number of correct pairs of u is equal to $\frac{1}{2}(k-i)$, and the number of free opening brackets in u is equal to $\frac{1}{2}(k+i) - \frac{1}{2}(k-i)$. Now, we obtain all prefixes

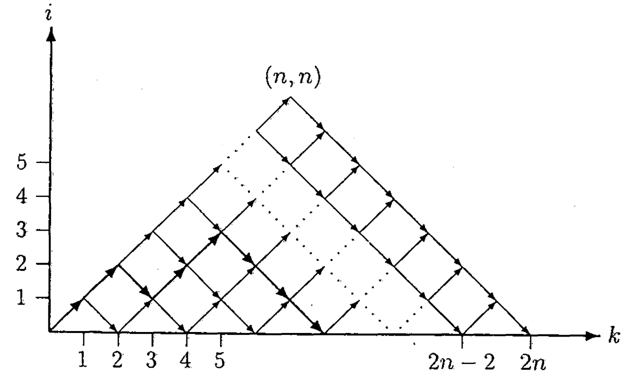

Figure 1. – The one-to-one correspondence between Dyckwords in D_1 of length $2n$ and the paths from $(0, 0)$ to $(2n, 0)$. Each segment $\nearrow$ and $\searrow$ is labelled by "$\sqsubset_1$" and "$\sqsupset_1$", respectively. A successive concatenation of the labels of the segments appearing on a path from $(0, 0)$ to $(2n, 0)$ yields a Dyckword of length $2n$. For example, the marked path corresponds to $\sqsubset_1 \sqsubset_1 \sqsupset_1 \sqsubset_1 \sqsubset_1 \sqsupset_1 \sqsupset_1 \sqsupset_1 \in D_1$ (8).

$u' \in \mathrm{INIT}_k(D^{\mathfrak{R}}(2n))$ with the same structure as u by the following consideration:

– Replace each correct pair of brackets appearing in u by a $(x, y) \in \mathfrak{R}$ (giving $|\mathfrak{R}|^{\frac{1}{2}(k-i)}$ possibilities);

– Replace each free opening bracket in u by a $x \in R_1$ (giving $|R_1|^{\frac{1}{2}(k+i)-\frac{1}{2}(k-i)}$ possibilities).

Hence,

$$|\mathrm{INIT}_k(D^{\mathfrak{R}}(2n))| = \sum_{\substack{0 \le i \le \min\{k, 2n-k\} \\ k+i \equiv 0 \bmod 2}} |\mathfrak{R}|^{\frac{1}{2}(k-i)} \, |R_1|^{\frac{1}{2}(k+i)-\frac{1}{2}(k-i)} \, w(k, i).$$

Inserting this expression into (1), we obtain by a straightforward computation

$$\mathbb{E}[Y_{\mathrm{pref}}^{s}(D^{\mathfrak{R}}(2n))] = \sum_{0 \le k < 2n} [(k+1)^s - k^s] \, p^{\lfloor \frac{k+1}{2} \rfloor} q^{\lfloor \frac{k}{2} \rfloor} \times \sum_{0 \le i \le \min\{\lfloor \frac{k}{2} \rfloor, n - \lfloor \frac{k+1}{2} \rfloor\}} p^{i} q^{-i} w\left(k, 2i + k - 2\left\lfloor \frac{k}{2} \right\rfloor\right),$$

where $p := |R_1|\,|T|^{-1}$ and $q := |\mathfrak{R}|(|R_1|\,|T|)^{-1}$. Since

$$|R_1| \leq |T_{\sqsubset}| \qquad \text{and} \qquad |\mathfrak{R}| \leq |R_1|\,|T_{\sqsupset}|,$$

we have $0 \leq p+q \leq |T_{\sqsubset}|\,|T|^{-1} + |T_{\sqsupset}|\,|T|^{-1} = 1$. In order to simplify the last expression for $\mathbb{E}\,[Y^s_{\mathrm{pref}}(D^{\mathfrak{R}}(2n))]$, we split the sum over k into two sums, the first one over $k \in [0:n]$ and the second one over $k \in]n:2n[$. Then, both sums are divided again into two parts, one for even k and for odd k. Finally, gathering the terms for q^λ and using (2), the described procedure ends in the following explicit result.

LEMMA 1: *Let $D^{\mathfrak{R}} \subseteq T^*$ be the generalized semi-Dycklanguage associated with $\mathfrak{R}$, $p := |R_1|\,|T|^{-1}$ and $q := |\mathfrak{R}|(|R_1|\,|T|)^{-1}$. Assuming that all words in $w \in T^{2n}$ are equally likely, the s-th moment $\mathbb{E}\,[Y^s_{\mathrm{pref}}(D^{\mathfrak{R}}(2n))]$ about the origin is given by*

$$\mathbb{E}\,[Y^s_{\mathrm{pref}}(D^{\mathfrak{R}}(2n))] = \sum_{0 \leq \lambda < n} q^\lambda \sum_{\lambda \leq k \leq n} [(k+\lambda+1)^s - (k+\lambda)^s]\,p^k \times \left[\binom{k+\lambda}{\lambda} - \binom{k+\lambda}{\lambda-1}\right]. \quad \square$$

Remark: Assume that all brackets in T are equally likely and that the brackets appearing in a word $w \in T^*$ are independently chosen from T. Obviously, a word $w \in D^{\mathfrak{R}}(m)$ has the probability $\Pr[w] = |\mathfrak{R}|^{\frac{1}{2}m}\,|T|^{-m} = p^{\frac{1}{2}m}\,q^{\frac{1}{2}m}$. Thus, $p = |R_1|\,|T|^{-1}$ (resp. $q = |\mathfrak{R}|(|R_1|\,|T|)^{-1}$) is the probability that an opening bracket $x \in R_1$ (resp. closing bracket $y \in R_2$ with $(x, y) \in \mathfrak{R}$) has been selected. Note that $|\mathfrak{R}|\,|R_1|^{-1}$ is the average quota of closing brackets per opening bracket in $\mathfrak{R}$.

Now, the prefixes in $\mathrm{INIT}_k\,(D^{\mathfrak{R}}(2n))$ can be partitioned according to their structure. There are $w(k, i)$ different structures consisting of $\frac{1}{2}(k+i)$ opening and $\frac{1}{2}(k-i)$ closing brackets. As each prefix in $\mathrm{INIT}_k\,(D^{\mathfrak{R}}(2n))$ with $\frac{1}{2}(k+i)$ opening and $\frac{1}{2}(k-i)$ closing brackets has the probability $p^{\frac{1}{2}(k+i)}\,q^{\frac{1}{2}(k-i)} = |R_1|^i\,|\mathfrak{R}|^{\frac{1}{2}(k-i)}\,|T|^{-k}$, we rediscover the above expression for $|\mathrm{INIT}_k\,(D^{\mathfrak{R}}(2n))|$. Furthermore, these considerations show that $|\mathrm{INIT}_k\,(D^{\mathfrak{R}}(2n))|\,|T|^{-k}$ is equal to the sum of the "weights" of all paths from $(0, 0)$ to the points (k, i), $0 \leq i \leq \min\{k, 2n-k\}$ with $(k+i) \equiv 0 \bmod 2$, in the grid presented in Figure 1 provided that each segment $\nearrow$ (resp. $\searrow$) is additionally labelled by p (resp. q). Here, the weight of a path is the product of the additional labels p and q taken over all segments appearing on that path.

Next, we shall compute an asymptotic equivalent to $\mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n))]$ for large n.

THEOREM 1: *Let $D^{\mathfrak{R}} \subseteq T^*$ be the generalized semi-Dycklanguage associated with $\mathfrak{R}$, $p := |R_1|\,|T|^{-1}$ and $q := |\mathfrak{R}|\,(|R_1|\,|T|)^{-1}$. Assuming that all words in $w \in T^{2n}$ are equally likely, the s-th moment $\mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n))]$ has for $n \to \infty$ the asymptotic equivalent*

$$\mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n))] \sim \begin{cases} \dfrac{1}{2p}\left[(-1)^s + (1-2p)\,P_s\left(\dfrac{p(1-p)}{(1-2p)^2}\right)\right] \\ \quad if \quad p = 1-q < \dfrac{1}{2} \\ \pi^{-\frac{1}{2}}\,\dfrac{s}{2s-1}\,2^{s+1}\,n^{s-\frac{1}{2}} \\ \quad if \quad p = 1-q = \dfrac{1}{2} \\ \dfrac{2p-1}{p^{s+1}}\,n^s \\ \quad if \quad p = 1-q > \dfrac{1}{2} \\ \dfrac{1-p}{2pq}\left[(-1)^s + (1-4pq)^{\frac{1}{2}}\,P_s\left(\dfrac{pq}{1-4pq}\right)\right] \\ \quad + \dfrac{1}{p(1-p-q)}\,F_{p,\,q,\,s}\,(1) \\ \quad if \quad p+q < 1 \end{cases}$$

Here, $P_s(x)$ denotes the polynomial

$$P_s(x) := \sum_{0\le k\le i\le j\le s} (-1)^{i+k}\,(2k-2)^j\,\frac{1}{2i-1}\binom{s}{j}\binom{2i}{i}\binom{i}{k}\,x^i.$$

The function $F_{p,\,q,\,s}(z)$ is given by

$$F_{p,\,q,\,s}(z) = \sum_{0\le k\le i\le j\le s} \binom{s}{j}\,2^{j-k}\,q^{-k}\,\mathcal{S}_j^{(i)}\binom{i}{k}\,(-1)^k\,z^{i-1} \times \frac{d^{i-k}}{dz^{i-k}}\,(h^{[k]}_{p,\,q}(z)\,g^{[j,\,k]}_{p,\,q,\,s}(z)),$$

where

$$h^{[k]}_{p,\,q}(z) := z^{-k}[1-(1-4pqz)^{\frac{1}{2}}]^k$$

and

$$g_{p,q,s}^{[j,k]}(z) := \frac{d^k}{dx^k}\left(\frac{\mathrm{A}_{s-j}(x)}{(1-x)^{s-j+1}} - x - \delta_{s,j}\right)\Bigg|_{x=\frac{1}{2q}[1-(1-4pq)^{\frac{1}{2}}]}$$

Here, $S_m^{(i)}$ *is a Stirling number of the second kind and* $\mathrm{A}_s(x)$ *denotes the s-th Eulerian polynomial* ([4], p. 245).

Proof: Starting with the expression in the preceding Lemma, we immediately find

$$\begin{aligned}
&\mathbb{E}\,[Y_{\text{pref}}^s\,(D^{\mathfrak{R}}(2n+2))] \\
&\quad = \mathbb{E}\,[Y_{\text{pref}}^s\,(D^{\mathfrak{R}}(2n))] \\
&\qquad + p^n\,q^n\,[(2n+1)^s - (2n)^s]\left[\binom{2n}{n} - \binom{2n}{n-1}\right] \\
&\qquad + p^{n+1}\,q^n\,[(2n+2)^s - (2n+1)^s]\left[\binom{2n+1}{n} - \binom{2n+1}{n-1}\right] \\
&\qquad + S_2(n) - S_1(n)
\end{aligned} \tag{3}$$

where

$$S_j(n) := p^{n+1} \sum_{0 \le \lambda < n} q^\lambda (n+\lambda+j)^s \times \left[\binom{n+\lambda+1}{\lambda} - \binom{n+\lambda+1}{\lambda-1}\right], \qquad j \in \{1, 2\}.$$

Introducing the numbers

$$X_{a,b,s}(n) := a^{n+1} \sum_{0 \le \lambda < n} (n+\lambda+1)^s\, b^\lambda \left[\binom{n+\lambda}{n} - \binom{n+\lambda}{n+1}\right], \quad (a,\,b,\,s) \in \mathbb{R} \times \mathbb{R} \times \mathbb{N}, \tag{4}$$

the sums $S_j(n)$, $j \in \{1, 2\}$, can easily be transformed into

$$\begin{aligned}
S_1(n) &= p^{n+1} \sum_{0 \le \lambda < n} q^\lambda\,(n+\lambda+1)^s \left[\binom{n+\lambda}{n} - \binom{n+\lambda}{n+1}\right] \\
&\quad + p^{n+1} \sum_{0 \le \lambda < n} q^\lambda\,(n+\lambda+1)^s \left[\binom{n+\lambda}{n+1} - \binom{n+\lambda}{n+2}\right] \\
&= X_{p,q,s}(n) - p^{-1}\,q X_{p,q,s}(n+1) \\
&\quad + p^{n+1}\,q^n\,(2n+1)^s \left[\binom{2n}{n+1} - \binom{2n}{n+2}\right] \\
&\quad + p^{n+1}\,q^{n+1}\,(2n+2)^s \left[\binom{2n+1}{n+1} - \binom{2n+1}{n+2}\right]
\end{aligned}$$

and

$$S_2(n) = p^{-1}\, X_{p,q,s}(n+1) - p^{n+1}\, q^n\, (2n+2)^s \left[\binom{2n+1}{n} - \binom{2n+1}{n-1}\right].$$

Inserting these alternative expressions for $S_j(n)$, $j \in \{1, 2\}$, into (3) and using the relation $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$, we obtain the recurrence

$$\begin{aligned} &\mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n+2))] \\ &\quad = \mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n))] + (1-q)\, p^{-1}\, X_{p,q,s}(n+1) - X_{p,q,s}(n) \\ &\qquad + (1-p)\, p^n\, q^n\, (2n+1)^s \left[\binom{2n}{n} - \binom{2n}{n+1}\right] \\ &\qquad - p^n q^n (2n)^s \left[\binom{2n}{n} - \binom{2n}{n+1}\right] \\ &\qquad + p^{n+1}\, q^{n+1}\, (2n+2)^s \left[\binom{2n+2}{n+1} - \binom{2n+2}{n+2}\right] \end{aligned}$$

with the initial condition $\mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(0))] = 0$.

Next, we introduce the generating function

$$\mathcal{E}_{p,q,s}(z) := \sum_{n\geq 0} \mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n))]\, z^n.$$

Translating the derived recurrence for $\mathbb{E}\,[Y^s_{\text{pref}}\,(D^{\mathfrak{R}}(2n))]$ into terms of $\mathcal{E}_{p,q,s}(z)$, we find

$$\begin{aligned} \mathcal{E}_{p,q,s}(z) &= \sum_{n\geq 0} (pq)^n\, (2n)^s\, \frac{1}{n+1} \binom{2n}{n} z^n + (1-p)\, G_{pq,s}(z) \\ &\quad + \frac{(1-q)\, p^{-1} - z}{1-z}\, F_{p,q,s}(z), \end{aligned}$$

where $F_{a,b,s}(z) := \sum_{n\geq 0} X_{a,b,s}(n)\, z^n$ is the generating function of the numbers $X_{a,b,s}(n)$ defined in (4) ([2]) and

$$G_{a,s}(z) := \frac{z}{1-z} \sum_{n\geq 0} (2n+1)^s\, (az)^n\, \frac{1}{n+1} \binom{2n}{n}, \qquad (a, s) \in \mathbb{R} \times \mathbb{N}.$$

([2]) Note that the function $F_{a,b,s}(z)$ has already been studied in [8], Lemma 1; it can be represented by the intricate expression given in the theorem.

Hence,

$$\mathbb{E}\left[Y_{\text{pref}}^{s}\left(D^{\mathfrak{R}}(2n)\right)\right] = 2^{s}\, p^{n}\, q^{n}\, \frac{n^{s}}{n+1}\binom{2n}{n} + (1-p)\,\langle z^{n}; G_{pq,\,s}(z)\rangle + \left\langle z^{n};\ \frac{(1-q)\,p^{-1}-z}{1-z}\, F_{p,\,q,\,s}(z)\right\rangle. \tag{5}$$

Here, the abbreviation $\langle z^n; f(z)\rangle$ denotes the coefficient of z^n in the expansion of $f(z)$ at $z = 0$.

Now, we have to find asymptotic equivalents to the three quantities appearing on the right-hand side of equation (5). Computing these equivalents, we can assume that $0 \leq p + q \leq 1$.

(a) By Stirling's approximation (e.g. [9], p. 111) we immediately obtain

$$2^{s}\, p^{n}\, q^{n}\, \frac{n^{s}}{n+1}\binom{2n}{n} \sim \pi^{-\frac{1}{2}}\, 2^{s}\, (4pq)^{n}\, n^{s-\frac{3}{2}}, \qquad n \to \infty. \tag{6}$$

(b) The coefficient $\langle z^n; G_{a,\,s}(z)\rangle$ is the number

$$Z_{a,\,s}(n) := \sum_{0 \leq k < n} (2k+1)^{s}\, a^{k}\, \frac{1}{k+1}\binom{2k}{k}$$

with an already computed asymptotic equivalent in [8], Lemma 2. We have

$$Z_{a,\,s}(n) \sim \begin{cases} (2a)^{-1}\left[(-1)^{s} + (1-4a)^{\frac{1}{2}}\, P_{s}\left(a(1-4a)^{-1}\right)\right] & \text{if } a < \frac{1}{4} \\ \pi^{-\frac{1}{2}}(2s-1)^{-1}\, 2^{s+1}\, n^{s-\frac{1}{2}} & \text{if } a = \frac{1}{4}, \\ \pi^{-\frac{1}{2}}(4a-1)^{-1}\, 2^{s}\, (4a)^{n}\, n^{s-\frac{3}{2}} & \text{if } a > \frac{1}{4} \end{cases} \qquad n \to \infty, \tag{7}$$

where $P_s(x)$ denotes the polynomial introduced in our theorem.

Since $4pq \leq (p+q)^2 \leq 1$, only the first two alternatives in (7) must be considered to obtain an asymptotic equivalent to $(1-p) < z^n; G_{pq,s}(z) >= (1-p)\, Z_{pq,s}(n)$. We find for $n \to \infty$

$$(1-p)\langle z^n; G_{pq,s}(z)\rangle \sim \begin{cases} \dfrac{1-p}{2pq}\left[(-1)^s + (1-4pq)^{\frac{1}{2}}\, P_s\left(pq(1-4pq)^{-1}\right)\right] \\ \quad \text{if} \quad p+q<1 \qquad \vee\, p = 1-q \neq \dfrac{1}{2} \\ \dfrac{1-p}{2s-1}\,\pi^{-\frac{1}{2}}\, 2^{s+1}\, n^{s-\frac{1}{2}} \\ \quad \text{if} \quad p = q = \dfrac{1}{2} \end{cases} \tag{8}$$

(c) To compute an asymptotic equivalent to the coefficient $\langle z^n; \frac{(1-q)p^{-1}-z}{1-z} F_{p,q,s}(z)\rangle$, we have to consider the function $F_{p,q,s}(z)$. As mentioned above, this function has already been investigated in [8], Lemma 1. Its singularity z_0 nearest to the origin and therefore its expansion around z_0 depends on the choice of p and q. The corresponding results are summarized in the following table:

q	z_0	expansion $F_{p,q,s}(z) = \sum_{\lambda\geq 0} \gamma_{p,q,s}(\lambda)\left(1-\frac{z}{z_0}\right)^{-\omega_\lambda}$	
		ω_λ	$\gamma_{p,q,s}(\lambda)$
$< \frac{1}{2}$	$(1-q)\,p^{-1}$	$s-\lambda+1$	$p(1-2q)\,(1-q)^{-(s+2)}\, s!$
$= \frac{1}{2}$	$(2p)^{-1}$	$\frac{1}{2}(2s-\lambda+1)$	$p\,2^{-(s-1)}\,(2s)!\, s!^{-1}$
$> \frac{1}{2}$	$(4pq)^{-1}$	$\frac{1}{2}(2s-\lambda-1)$	$2p\,(4q-1)\,(2q-1)^{-2}\, 2^{-(s-1)}\,(2s-2)!\,(s-1)!^{-1}$

If $p+q=1$, we have $\frac{(1-q)p^{-1}-z}{1-z} F_{p,q,s}(z) = F_{p,q,s}(z)$ and we can apply the theorem of Darboux (e.g. [2, 5, 10]) to the expansions presented in the above table. For $n \to \infty$, we obtain by means of the relation $\Gamma\left(s+\frac{1}{2}\right) = \pi^{\frac{1}{2}}\,(2s)!\,4^{-s}\, s!^{-1}$ satisfied by the complete gamma function (e.g. [1])

$$\left\langle z^n;\ \frac{(1-q)\,p^{-1}-z}{1-z}\,F_{p,q,s}(z)\right\rangle \sim \begin{cases} p(1-2q)\,(1-q)^{-s-2}n^s \\ \quad \text{if} \quad q=1-p<\dfrac{1}{2} \\ \pi^{-\frac{1}{2}}\,2^s\,n^{s-\frac{1}{2}} \\ \quad \text{if} \quad q=1-p=\dfrac{1}{2} \\ \pi^{-\frac{1}{2}}\,p(4q-1)\,(2q-1)^{-2} \\ \quad \times\, 2^s\,(4pq)^n\,n^{s-\frac{3}{2}} \\ \quad \text{if} \quad q=1-p>\dfrac{1}{2} \end{cases}, \tag{9}$$

If $p+q<1$, we have to consider the cases whether the singularity at $\hat{z}:=1$ induced by the factor $\frac{(1-q)\,p^{-1}-z}{1-z}$ is less or greater than z_0, or equal to z_0. For $q<\frac{1}{2}$, the assumption $z_0\leq\hat{z}$ implies $p+q\geq 1$ which is a contradiction to $p+q<1$. For $q=\frac{1}{2}$, the same assumption leads to $p\geq\frac{1}{2}$ and therefore to the same contradiction. Finally, for $q>\frac{1}{2}$, the assumption $z_0\leq\hat{z}$ implies $4pq\geq 1$ which again is a contradiction because generally the relation $4pq\leq(p+q)^2$ holds. Thus, in all cases, the singularity nearest to the origin of the function $\frac{(1-q)\,p^{-1}-z}{1-z}\,F_{p,q,s}(z)$ is at $\hat{z}:=1$. Therefore,

$$\left\langle z^n;\ \frac{(1-q)\,p^{-1}-z}{1-z}\,F_{p,q,s}(z)\right\rangle \sim (1-p-q)\,p^{-1}\,F_{p,q,s}(1), \qquad p+q<1, \qquad n\to\infty. \tag{10}$$

Now, inserting the asymptotic equivalents presented in (6), (8), (9) and (10) into (5), we obtain the result stated in the theorem. This completes the proof. □

Choosing $s\in\{1,2\}$, the representation of the function $F_{p,q,s}(z)$ established in the preceding theorem implies the following explicit expressions:

$$F_{p,q,1}(z)=\frac{(1-\sqrt{1-4pqz})^2\,(\sqrt{1-4pqz}+4q-1)}{2qz\,\sqrt{1-4pqz}\,(\sqrt{1-4pqz}+2q-1)^2}$$

and

$$F_{p,q,2}(z)=\frac{1-3q+4q^2-zpa_1-z^2p^2a_2-z^3p^3}{2qz(pz+q-1)^3} - \frac{1-3q+4q^2-zpa_3+z^2p^2a_4+z^3p^3a_5+8z^4p^4q}{2qz(pz+q-1)^3\,(1-4pqz)\,\sqrt{1-4pqz}},$$

where $a_1 := 3q^2 - 5q + 3$, $a_2 := 4q - 3$, $a_3 := 24q^3 - 15q^2 + q + 3$, $a_4 := 16q^4 + 24q^3 - 48q^2 + 22q + 3$ and $a_5 := 16q^3 + 16q^2 - 24q - 1$. Using these expressions together with the relations $P_1(x) = 1 + 4x$ and $P_2(x) = -1 + 16x^2$, the following result is implied by Theorem 1.

COROLLARY 1: *Let $D^{\mathfrak{R}} \subseteq T^*$ be the generalized semi-Dycklanguage associated with $\mathfrak{R}$, $p := |R_1|\,|T|^{-1}$ and $q := |\mathfrak{R}|\,(|R_1|\,|T|)^{-1}$. Assuming that all words in $w \in T^{2n}$ are equally likely, the average minimal prefix-length is asymptotically given by*

$$\mathbb{E}[Y_{\text{pref}}(D^{\mathfrak{R}}(2n))] \sim \begin{cases} (1-2p)^{-1} & \text{if } \; p = 1-q < \frac{1}{2} \\ 4\pi^{-\frac{1}{2}}\, n^{\frac{1}{2}} & \text{if } \; p = 1-q = \frac{1}{2} \\ (2p-1)\,p^{-2}\,n & \text{if } \; p = 1-q > \frac{1}{2} \\ \dfrac{\sqrt{1-4pq}+2p-1}{2p\,(1-p-q)} & \text{if } \; p+q < 1 \end{cases}, \quad n \to \infty.$$

The asymptotical behaviour of the variance

$$\sigma^2\,(Y_{\text{pref}}(D^{\mathfrak{R}}(2n))) := \mathbb{E}\,[Y^2_{\text{pref}}(D^{\mathfrak{R}}(2n))] - (\mathbb{E}\,[Y_{\text{pref}}(D^{\mathfrak{R}}(2n))])^2$$

is described by

$$\sigma^2(Y_{\text{pref}}(D^{\mathfrak{R}}(2n))) \sim \begin{cases} 4p(1-p)\,(1-2p)^{-3} \\ \quad \text{if } \; p = 1-q < \frac{1}{2} \\ 16\,(9\,\pi)^{-\frac{1}{2}}\, n^{\frac{3}{2}} \\ \quad \text{if } \; p = 1-q = \frac{1}{2} \\ (2p-1)\,(1-p)\,p^{-4}\,n^2 \\ \quad \text{if } \; p = 1-q > \frac{1}{2} \\ v(p,\,q) \quad \text{if } \; p+q < 1 \end{cases}, \quad n \to \infty,$$

where

$$v(p,\,q) := \frac{(2p-1)\left(1-p+p^2+pq\right)}{2p^2\,(1-p-q)^2} + \frac{1-3p+3p^2-pq+4p^2q-4p^3q-4p^2q^2}{2p^2(1-p-q)^2\,\sqrt{1-4pq}}. \qquad \square$$

The first few numerical values for

$$\mathbb{E}\left[Y_{\text{pref}}\left(D^{\mathfrak{R}}(2n)\right)\right] \qquad \text{and} \qquad \sigma^2\left(Y_{\text{pref}}\left(D^{\mathfrak{R}}(2n)\right)\right)$$

are summarized in Table 1. For example, consider the generalized semi-Dycklanguage $D^{\mathfrak{R}}$ with

$$\begin{aligned}\mathfrak{R} := \{&\sqsubset_1, \sqsubset_2, \sqsubset_3, \sqsubset_4, \sqsubset_6\} \times \{\sqsupset_1, \sqsupset_2\} \cup \{\sqsubset_3, \sqsubset_5, \sqsubset_6\} \\ &\times \{\sqsupset_3, \sqsupset_4\} \cup \{(\sqsubset_1, \sqsupset_3), (\sqsubset_2, \sqsupset_4)\}.\end{aligned}$$

We find $|T| = 10$, $|R_1| = 6$ and $|\mathfrak{R}| = 18$, and therefore $p = 0.6$ and $q = 0.3$. An inspection of table 1 shows that we have to read ≈ 6.07625 $[= \frac{5}{3}(1 + \sqrt{7})]$ brackets in order to decide whether or not a word $w \in T^{2n}$ belongs to $D^{\mathfrak{R}}(2n)$, $n \to \infty$, on the average. The variance is ≈ 63.6976 $[= \frac{5}{63}(329 + 179\sqrt{7})]$.

We conclude this note by discussing some interesting consequences implied by the preceding corollary.

(a) If T is not the smallest alphabet for $D^{\mathfrak{R}}$, we have the inequality $|R_1| + |R_2| < |T|$ and therefore the relation

$$p + q = \frac{|R_1|^2 + |\mathfrak{R}|}{|R_1|\,|T|} < \frac{|R_1|^2 + |\mathfrak{R}|}{|R_1|(|R_1| + |R_2|)} \leq 1$$

because $|\mathfrak{R}| \leq |R_1|\,|R_2|$. In that case, the fourth alternative in the relations presented in Corollary 1 implies that we only have to read a minimal prefix of length $\sim \frac{\sqrt{1-4pq}+2p-1}{2p(1-p-q)} = \Theta(1)$, $n \to \infty$, to decide whether or not an input word $w \in T^{2n}$ belongs to $D^{\mathfrak{R}}(2n)$, on the average; the variance is also bounded by a constant, namely $v(p, q)$.

(b) If T is the smallest alphabet for $D^{\mathfrak{R}}$, we have $R_1 = T_{\sqsubset}$ and $R_2 = T_{\sqsupset}$ and therefore $|T| = |R_1| + |R_2|$. In that case, the relation $p + q \leq 1$ is equivalent to $|\mathfrak{R}| \leq |R_1|\,|R_2| = |T_{\sqsubset}|\,|T_{\sqsupset}|$.

(b1) If $\mathfrak{R}$ is not maximal, *i.e.*, $\mathfrak{R} \subset T_{\sqsubset} \times T_{\sqsupset}$, the fourth alternative appearing in the above corollary implies again that a prefix of minimal length $\sim \frac{\sqrt{1-4pq}+2p-1}{2p(1-p-q)} = \Theta(1)$, $n \to \infty$, has to be read, on the average; the variance is bounded by the constant $v(p, q)$.

(b2) If $\mathfrak{R}$ is maximal, *i.e.*, $\mathfrak{R} = T_{\sqsubset} \times T_{\sqsupset}$, the situation changes completely.

(b2.1) If $p < \frac{1}{2}$, *i.e.*, $|T_{\sqsubset}| < |T_{\sqsupset}|$, the first alternative in the relations of the preceding corollary implies that we only have to read

p	n \ q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	10	1.1237	1.1369	1.1507	1.1652	1.1803	1.1963	1.2132	1.2311	1.2500
		1.1237	*1.1369*	*1.1507*	*1.1652*	*1.1803*	*1.1963*	*1.2132*	*1.2311*	*1.2500*
		0.1651	0.2107	0.2608	0.3161	0.3773	0.4454	0.5214	0.6067	0.7031
		0.1651	*0.2107*	*0.2608*	*0.3161*	*0.3773*	*0.4454*	*0.5214*	*0.6068*	*0.7031*
	100	1.1237	1.1369	1.1507	1.1652	1.1803	1.1963	1.2132	1.2311	1.2500
		1.1237	*1.1369*	*1.1507*	*1.1652*	*1.1803*	*1.1963*	*1.2132*	*1.2311*	*1.2500*
		0.1651	0.2107	0.2608	0.3161	0.3773	0.4454	0.5214	0.6068	0.7031
		0.1651	*0.2107*	*0.2608*	*0.3161*	*0.3773*	*0.4454*	*0.5214*	*0.6068*	*0.7031*
0.2	10	1.2827	1.3188	1.3590	1.4039	1.4550	1.5138	1.5827	1.6652	
		1.2827	*1.3188*	*1.3589*	*1.4039*	*1.4550*	*1.5139*	*1.5831*	*1.6667*	
		0.4327	0.5789	0.7593	0.9860	1.2771	1.6604	2.1788	2.8993	
		0.4327	*0.5789*	*0.7593*	*0.9862*	*1.2780*	*1.6643*	*2.1952*	*2.9630*	
	100	1.2827	1.3188	1.3590	1.4039	1.4550	1.5139	1.5831	1.6667	
		1.2827	*1.3188*	*1.3590*	*1.4039*	*1.4550*	*1.5139*	*1.5831*	*1.6667*	
		0.4327	0.5789	0.7593	0.9862	1.2780	1.6643	2.1952	2.9630	
		0.4327	*0.5789*	*0.7593*	*0.9862*	*1.2780*	*1.6643*	*2.1952*	*2.9630*	
0.3	10	1.4947	1.5726	1.6665	1.7834	1.9343	2.1385	2.4292		
		1.4947	*1.5726*	*1.6667*	*1.7840*	*1.9371*	*2.1525*	*2.5000*		
		0.8866	1.2634	1.8017	2.6093	3.8880	6.0174	9.6917		
		0.8868	*1.2642*	*1.8056*	*2.6293*	*3.9961*	*6.6036*	*13.125*		
	100	1.4947	1.5726	1.6667	1.7840	1.9371	2.1525	2.5000		
		1.4947	*1.5726*	*1.6667*	*1.7840*	*1.9371*	*2.1525*	*2.5000*		
		0.8868	1.2642	1.8056	2.6293	3.9961	6.6036	13.125		
		0.8868	*1.2642*	*1.8056*	*2.6293*	*3.9961*	*6.6036*	*13.125*		
0.4	10	1.7911	1.9510	2.1672	2.4792	2.9656	3.7834			
		1.7913	*1.9519*	*2.1713*	*2.5000*	*3.0902*	*5.0000*			
		1.7041	2.6439	4.2419	7.1784	12.913	24.332			
		1.7097	*2.6685*	*4.3666*	*7.9167*	*18.262*	*120.00*			
	100	1.7913	1.9519	2.1713	2.5000	3.0902	4.9952			
		1.7913	*1.9519*	*2.1713*	*2.5000*	*3.0902*	*5.0000*			
		1.7097	2.6685	4.3666	7.9167	18.262	117.79			
		1.7097	*2.6685*	*4.3666*	*7.9167*	*18.262*	*120.00*			
0.5	10	2.2326	2.5678	3.0915	3.9969	5.7206				
		2.2361	*2.5820*	*3.1623*	*4.4721*	*7.1365*				
		3.2722	5.6383	10.378	20.503	41.825				
		3.3541	*6.0246*	*12.649*	*40.249*	*95.153*				
	100	2.2361	2.5820	3.1623	4.4721	20.764				
		2.2361	*2.5820*	*3.1623*	*4.4721*	*22.568*				
		3.3541	6.0246	12.649	40.249	2251.0				
		3.3541	*6.0246*	*12.649*	*40.249*	*3009.0*				
0.6	10	2.9429	3.6796	5.0257	7.7867					
		2.9772	*3.8380*	*6.0763*	*5.5556*					
		6.3708	12.130	24.619	49.738					
		7.1900	*16.589*	*63.698*	*61.728*					
	100	2.9772	3.8380	6.0763	58.331					
		2.9772	*3.8380*	*6.0763*	*55.556*					
		7.1700	16.589	63.697	5799.2					
		7.1900	*16.589*	*63.698*	*6172.8*					
0.7	10	4.1790	5.8687	9.4235						
		4.4590	*7.5952*	*8.1633*						
		12.104	23.773	43.115						
		18.973	*79.371*	*49.979*						
	100	4.4590	7.5952	82.908						
		4.4590	*7.5952*	*81.633*						
		18.973	79.368	4888.3						
		18.973	*79.371*	*4997.9*						
0.8	10	6.4707	10.416							
		8.9039	*9.3750*							
		19.614	28.234							
		87.236	*29.297*							
	100	8.9038	94.792							
		8.9039	*93.750*							
		87.223	2914.6							
		87.236	*2929.7*							
0.9	10	10.880								
		9.8765								
		12.708								
		12.193								
	100	99.769								
		98.765								
		1224.2								
		1219.3								

Table 1. Exact [in roman] and asymptotical values [in italics] for $\mathbb{E}[Y_{\text{pref}}(D^{\Re}(2n))]$ and $\sigma^2(Y_{\text{pref}}(D^{\Re}(2n)))$, $(n, p, q) := (n, \frac{|R_1|}{|T|}, \frac{|\Re|}{|R_1||T|}) \in \{10, 100\} \times \{0.1, \ldots, 0.9\}^2$. For each (p, n), the first two lines refer to the expected value and the second two lines to the variance.

a minimal prefix of length $\sim \frac{|T_\sqsupset|+|T_\sqsubset|}{|T_\sqsupset|-|T_\sqsubset|} = \Theta(1)$, $n \to \infty$, to decide whether or not an input word $w \in T^{2n}$ belongs to $D^{\mathfrak{R}}(2n)$, on the average. Note that $\frac{|T_\sqsupset|+|T_\sqsubset|}{|T_\sqsupset|-|T_\sqsubset|} \leq |T|$. The variance has the asymptotical behaviour $\sim 4|T_\sqsubset|\,|T_\sqsupset|\frac{|T_\sqsupset|+|T_\sqsubset|}{(|T_\sqsupset|-|T_\sqsubset|)^3} \leq 4|\mathfrak{R}|\,|T|$, $n \to \infty$.

(b2.2) If $p = \frac{1}{2}$, *i.e.*, $|T_\sqsubset| = |T_\sqsupset|$, the second alternative established in Corollary 1 shows that a prefix of minimal length $\sim 4\pi^{-\frac{1}{2}} n^{\frac{1}{2}} = \Theta(n^{\frac{1}{2}})$, $n \to \infty$, has to be read, on the average. In this case, the variance is asymptotically given by $\sim 16(9\pi)^{-\frac{1}{2}} n^{\frac{3}{2}}$, $n \to \infty$.

(b2.3) If $p > \frac{1}{2}$, *i.e.*, $|T_\sqsubset| > |T_\sqsupset|$, the third alternative appearing in the preceding corollary implies that a prefix of minimal length $\sim (1 - \frac{|T_\sqsupset|^2}{|T_\sqsubset|^2})\, n = \Theta(n)$, $n \to \infty$, has to be read, on the average; the variance is asymptotically equal to $\sim \frac{|T_\sqsupset|}{|T_\sqsubset|}\left(1 - \frac{|T_\sqsupset|}{|T_\sqsubset|}\right)\left(1 + \frac{|T_\sqsupset|}{|T_\sqsubset|}\right)^2 n^2$, $n \to \infty$. Note that the factor before n^2 is maximal for $\frac{|T_\sqsupset|}{|T_\sqsubset|} = \frac{1}{8}(1 + \sqrt{17})$, *i.e.*, $\frac{|T_\sqsupset|}{|T_\sqsubset|} \approx 0.640\,388\ldots$ Thus, this factor is less than or equal to $\frac{1}{512}(107 + 51\sqrt{17}) \approx 0.619\,684\ldots$

Considering the semi-Dycklanguage D_k with k types of brackets over its smallest alphabet, we have to read a minimal prefix

– of average length $\sim 4\pi^{-\frac{1}{2}} n^{\frac{1}{2}}$ if $k = 1$ [Case (b2.2)],

and

– of average length $\sim 2\sqrt{\frac{k}{k-1}}$ if $k \geq 2$ [Case (b1)]

to decide whether or not a given word $w \in T^{2n}$ belongs to $D^{\mathfrak{R}}(2n)$, $n \to \infty$. In the former case, the variance is asymptotically given by $\sim 16(9\pi)^{-\frac{1}{2}} n^{\frac{3}{2}}$, $n \to \infty$, and in the latter case by $\sim v\left(\frac{1}{2}, \frac{1}{2k}\right) = 2(k+1)\sqrt{\frac{k}{(k-1)^3}}$, $n \to \infty$. Note that the former result ($k = 1$) has already been proved in [8] ([3]).

Considering the generalized semi-Dycklanguage $D^{\mathfrak{R}}$ with

$$\mathfrak{R} := \{(\sqsubset_i, \sqsupset_1) \mid 1 \leq i \leq k\}$$

over its smallest alphabet, we again have to read a minimal prefix of average length $\sim 4\pi^{-\frac{1}{2}} n^{\frac{1}{2}}$ if $k = 1$, $n \to \infty$ [Case (b2.2)]. For $k \geq 2$, the average minimal prefix-length is asymptotically given by $\sim \left(1 - \frac{1}{k^2}\right) n$, $n \to \infty$

([3]) The result for $k = 1$ answers a question mooted by J. Berstel while he visits the department of computer science at the Johann Wolfgang Goethe-Universität Frankfurt am Main in January, 1992.

[Case (b2.3)]. In the former case, the asymptotical behaviour of the variance is $\sim 16\,(9\,\pi)^{-\frac{1}{2}}\,n^{\frac{3}{2}}$, and in the latter case $\sim \frac{1}{k}\left(1-\frac{1}{k}\right)\left(1+\frac{1}{k}\right)^{2} n^{2}$, $n \to \infty$.

For the generalized semi-Dycklanguage $D^{\mathfrak{R}}$ with

$$\mathfrak{R} := \{(\sqsubset_1,\, \sqsupset_i) \,|\, 1 \leq i \leq k\}$$

over its smallest alphabet, we again have to read a minimal prefix of average length $\sim 4\,\pi^{-\frac{1}{2}}\,n^{\frac{1}{2}}$ if $k = 1$, $n \to \infty$ [Case (b2.2)]. For $k \geq 2$, the minimal prefix-length is asymptotically given by $\frac{k+1}{k-1}$, $n \to \infty$, on the average [Case (b2.1)]. The variance is asymptotically given by $\sim 16\,(9\,\pi)^{-\frac{1}{2}}\,n^{\frac{3}{2}}$ for $k = 1$, and by $\sim 4k\,\frac{k+1}{(k-1)^3}$ for $k \geq 2$, $n \to \infty$.

REFERENCES

1. M. ABRAMOWITZ and A. STEGUN, *Handbook of Mathematical Functions*, Dover, 1970.
2. E. A. BENDER, Asymptotic Methods in Enumeration, *SIAM Review*, 1974, *16* (4), pp. 485-515.
3. L. CARLITZ, D. P. ROSELLE and R. A. SCOVILLE, Some Remarks on Ballot-Type Sequences of Positive Integers, *J. Comb. Theory* (A), 1971, *11*, pp. 258-271.
4. L. COMTET, *Advanced Combinatorics*, D. Reidel, 1974.
5. Ph. FLAJOLET and A. M. ODLYZKO, Singularity Analysis of Generating Functions, *SIAM J. Discrete Math.*, 1990, *3* (2), pp. 216-240.
6. M. A. HARRISON, *Introduction to Formal Languages*, Addison-Wesley, 1978.
7. R. KEMP, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley-Teubner, 1984.
8. R. KEMP, On Prefixes of Formal Languages and Their Relation to the Average-Case Complexity of the Membership Problem, *Journal of Automata, Languages and Combinatorics*, 1996 (to appear).
9. D. E. KNUTH, *The Art of Computer Programming*, Vol. 1, 2nd ed., Addison-Wesley, 1973.
10. A. ODLYZKO, Asymptotic Enumeration Methods, in: *Handbook of Combinatorics*, Chapt. 22, Elsevier, 1995.