RAIRO-Inf. Theor. Appl. **39** (2005) 621-639
DOI: 10.1051/ita:2005032

THE ENTROPY OF ŁUKASIEWICZ-LANGUAGES *

Ludwig Staiger[1]

Abstract. The paper presents an elementary approach for the calculation of the entropy of a class of languages. This approach is based on the consideration of roots of a real polynomial and is also suitable for calculating the Bernoulli measure. The class of languages we consider here is a generalisation of the Łukasiewicz language.

Mathematics Subject Classification. 68Q30, 68Q45, 94A17.

Introduction

The Łukasiewicz language (see [1]) is the language defined by the grammar $S \to aSS \mid b$. It is a deterministic one-counter language and a prefix-code. In this paper we are going to generalise this concept in two ways: First we admit languages generated by grammars $S \to aS^n \mid b$ with $a \in A_0, b \in A_1$, where A_0 and A_1 are disjoint alphabets. The languages thus specified are also deterministic one-counter prefix-codes. Secondly, we allow substitution of letters of $A := A_0 \cup A_1$ by codewords of a previously given code $\widetilde{C}$ (for more details see Sect. 2). This results in languages which are codes but – depending on the code $\widetilde{C}$ – not necessarily context-free and which will be called, in the sequel, generalised Łukasiewicz languages.

In the paper [11] a remarkable information-theoretic property of Łukasiewicz's comma-free notation was developed. The languages of well-formed formulas of the implicational calculus with one variable and one n-ary operation ($n \geq 2$) in Polish parenthesis-free notation have generative capacity $h_2(\frac{n-1}{n})$ where h_2 is the usual

Keywords and phrases. Entropy of languages, Bernoulli measure of languages, codes, Łukasiewicz language.

* *A preliminary version appeared in: Developments in Language Theory (W. Kuich, G. Rozenberg and A. Salomaa Eds.), Lect. Notes Comput. Sci. No. 2295, Springer-Verlag, Berlin (2002) 155–165.*

[1] Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik, von-Seckendorff-Platz 1, D–06099 Halle (Saale), Germany; staiger@informatik.uni-halle.de

Shannon entropy, or, stated in other terms, the languages generated by grammars $S \to aS^n \mid b$ have generative capacity $h_2(\frac{n-1}{n})$.

The main purpose of our investigations is to study the same information theoretic aspect of languages as in [3, 5, 9, 11, 14], namely the generative capacity of languages. This capacity, in language theory called the *entropy of languages* resembles directly Shannon's channel capacity (*cf.* [8]). It measures the amount of information which must be provided on the average in order to specify a particular symbol of a word in a language. For a connection of the entropy of languages to Algorithmic Information Theory see *e.g.* [12, 15]. In [7] an account of interesting connections between the entropy of languages and data compression was presented.

After having investigated basic properties of generalised Łukasiewicz languages we first calculate their Bernoulli measures in Section 3. Here we derive and investigate in detail a basic real-valued equation closely related to the measure of generalised Łukasiewicz languages.

These investigations turn out to be useful not only for the calculation of the measure but also for estimating the entropy of generalised Łukasiewicz languages which will be carried out in Section 4. In contrast to [11] we do not require the powerful apparatus of the theory of complex functions utilised there for the more general task of calculating the entropy of unambiguous context-free languages. We develop a simpler apparatus based on augmented real functions. As announced above, this approach applies also to languages which are not necessarily context-free where the entropy is, in general, not computable [10]. We give also an exact formula for the entropy of pure Łukasiewicz languages with arbitrary numbers of letters representing variables and n-ary operations (n fixed).

The final section deals with the entropy of the star languages (submonoids) of generalised Łukasiewicz languages.

Next we introduce the notation used throughout the paper. By $\mathbb{N} = \{0, 1, 2, \ldots\}$ we denote the set of natural numbers. Let X be an alphabet of cardinality $\# X = r$. By X^* we denote the set (monoid) of words on X, including the *empty word* e. For $w, v \in X^*$ let $w \cdot v$ be their *concatenation*. This concatenation product extends in an obvious way to subsets $W, V \subseteq X^*$. For a language W let $W^* := \bigcup_{i \in \mathbb{N}} W^i$ be the *submonoid* of X^* generated by W, and by $W^\omega := \{w_1 \cdots w_i \cdots : w_i \in W \setminus \{e\}\}$ we denote the set of infinite strings formed by concatenating words in W. Furthermore $|w|$ is the *length* of the word $w \in X^*$ and $\mathbf{A}(B)$ is the set of all finite prefixes of strings in $B \subseteq X^* \cup X^\omega$. We shall abbreviate $w \in \mathbf{A}(\eta)$ ($\eta \in X^* \cup X^\omega$) by $w \sqsubseteq \eta$.

As usual a language $V \subseteq X^*$ is called a code provided $w_1 \cdots w_l = v_1 \cdots v_k$ for $w_1, \ldots, w_l, v_1, \ldots, v_k \in V$ implies $l = k$ and $w_i = v_i$. A language $V \subseteq X^*$ is referred to as an ω-code provided $w_1 \cdots w_i \cdots = v_1 \cdots v_i \cdots$ where $w_i, v_i \in V$ implies $w_i = v_i$. A code V is said to have a *finite delay of decipherability*, provided for every $w \in V$ there is an $m_w \in \mathbb{N}$ such that $w \cdot v_1 \cdots v_{m_w} \sqsubseteq w' \cdot u$, for $v_1, \ldots, v_{m_w}, w' \in V$ and $u \in V^*$, implies $w = w'$ (*cf.* [4, 13]). As usual, V is called a *prefix code* provided $v \sqsubseteq w$ implies $v = w$ for $v, w \in V$, that is, V has a finite delay of decipherability and $m_w = 0$ for every $w \in V$.

Every code having a finite delay of decipherability is an ω-code (see [4, 13]). A simple example of an ω-code having no finite delay of decipherability is the set $V := \{a, c\} \cup \{ac^i b : i \in \mathbb{N}\} \subseteq \{a, b, c\}^*$. Here, for the codeword $a \in V$, the number m_a is infinite, whereas $m_w = 0$ for every other word $w \in V$.

1. Pure Łukasiewicz-languages

In this section we consider languages over a finite or countably infinite alphabet A. Let $\{A_0, A_1\}$ be a partition of A into two nonempty parts and let $n \geq 2$. The *pure* $\{A_0, A_1\}$*-n-Łukasiewicz-language* is defined as the solution of the equation

$$\widetilde{\text{Ł}} = A_0 \cup A_1 \cdot \widetilde{\text{Ł}}^n. \tag{1}$$

It is a simple deterministic language (*cf.* [1], Sect. 6.7) and can be obtained as $\bigcup_{i\in\mathbb{N}} \widetilde{\text{Ł}}_i$ where $\widetilde{\text{Ł}}_0 := \emptyset$ and $\widetilde{\text{Ł}}_{i+1} := A_0 \cup A_1 \cdot \widetilde{\text{Ł}}_i^{\,n}$.

$\{A_0, A_1\}$-n-Łukasiewicz-languages have the following easily verified properties. For the sake of completeness we give a proof.

Proposition 1.1.

1. $\widetilde{\text{Ł}}$ *is a prefix code.*
2. *If* $w \in A^*$ *and* $a_0 \in A_0$ *then* $w \cdot a_0^{|w|\cdot n} \in \widetilde{\text{Ł}}^*$.
3. $\mathbf{A}(\widetilde{\text{Ł}}^*) = A^*$

Proof.

1. Let $v, w \in \widetilde{\text{Ł}}$ be a pair of words such that $v \sqsubset w$, and $|v| + |w|$ is minimal. Since $A_0 \cap A_1 = \emptyset$, we have $v, w \in A_1 \cdot \widetilde{\text{Ł}}^n$, that is, $v = v_0 \cdot v_1 \cdots v_n$ and $w = w_0 \cdot w_1 \cdots w_n$ where $v_0, w_0 \in A_1$ and $v_i, w_i \in \widetilde{\text{Ł}}$ for $i \geq 1$. $v_0 = w_0$ follows readily. Let i, $1 \leq i \leq n$, be the smallest index such that $v_i \neq w_i$. Hence, either v_i is a prefix of w_i or *vice versa*, a contradiction to the length assumption.
2. We show by induction on i that the assertion holds for every $w \in A^i$.

   If $w \in A^0 = \{e\}$ then $w \in \widetilde{\text{Ł}}^*$. Assume $w \in A^{i+1}$. Then $w = a \cdot u$ for $a \in A$ and $u \in A^i$. By the induction hypothesis, $u \cdot a_0^{|u|\cdot n} \in \widetilde{\text{Ł}}^m$ for suitable $m \in \mathbb{N}$. Consequently, $u \cdot a_0^{|w|\cdot n} = u \cdot a_0^{(|u|+1)\cdot n} \in \widetilde{\text{Ł}}^{m+n}$ has a decomposition $u \cdot a_0^{|w|\cdot n} = v_1 \cdots v_n \cdot u'$ where $v_j \in \widetilde{\text{Ł}}$ and $u' \in \widetilde{\text{Ł}}^m$.

   If $a \in A_0$ then $a \cdot v_1 \cdots v_n \cdot u' = w \cdot a_0^{|w|\cdot n} \in \widetilde{\text{Ł}}^{m+n+1}$ and the assertion is true. If $a \in A_1$ then $a \cdot v_1 \cdots v_n \in \widetilde{\text{Ł}}$ whence $a \cdot v_1 \cdots v_n \cdot u' = w \cdot a_0^{|w|\cdot n} \in \widetilde{\text{Ł}}^{m+1}$ and the assertion is also true.
3. Follows from 2. □

Along with $\widetilde{\text{Ł}}$ it is useful to consider its *derived language* $\widetilde{\text{K}}$ which is defined by the following equation.

$$\widetilde{\text{K}} := A_1 \cdot \bigcup\nolimits_{i=0}^{n-1} \widetilde{\text{Ł}}^i. \tag{2}$$

Proposition 1.2.

1. $\mathbf{A}(\widetilde{\text{Ł}}) \setminus \widetilde{\text{Ł}} = \widetilde{\text{K}}^*$.
2. *Every* $w \in A^*$ *has a unique factorisation* $w = v \cdot u$ *where* $v \in \widetilde{\text{Ł}}^*$ *and* $u \in \widetilde{\text{K}}^*$.

Proof.

1. We have $\mathbf{A}(\widetilde{\text{Ł}}) = \{e\} \cup A_0 \cup A_1 \cdot \bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i \cdot \mathbf{A}(\widetilde{\text{Ł}})$.

   Since $\widetilde{\text{Ł}}$ is a prefix code, $\bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i \cdot (\mathbf{A}(\widetilde{\text{Ł}}) \setminus \widetilde{\text{Ł}})$ is the disjoint union of the sets $\widetilde{\text{Ł}}^i \cdot (\mathbf{A}(\widetilde{\text{Ł}}) \setminus \widetilde{\text{Ł}})$, whence $\bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i \cdot \mathbf{A}(\widetilde{\text{Ł}}) = \bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i \cdot (\mathbf{A}(\widetilde{\text{Ł}}) \setminus \widetilde{\text{Ł}}) \cup \widetilde{\text{Ł}}^n$.

   Consequently, $\mathbf{A}(\widetilde{\text{Ł}}) \setminus \widetilde{\text{Ł}} = \mathbf{A}(\widetilde{\text{Ł}}) \setminus (A_0 \cup A_1 \cdot \widetilde{\text{Ł}}^n) = \{e\} \cup A_1 \cdot \bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i \cdot (\mathbf{A}(\widetilde{\text{Ł}}) \setminus \widetilde{\text{Ł}})$. Since $e \notin A_1 \cdot \bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i$, this equation has the unique solution $(A_1 \cdot \bigcup_{i=0}^{n-1} \widetilde{\text{Ł}}^i)^*$.
2. Follows from 1. because for a prefix code C every word in $w \in \mathbf{A}(C^*)$ has a unique factorisation $w = v \cdot u$ with $v \in C^*$ and $u \in \mathbf{A}(C) \setminus C$.

$\square$

Proposition 1.3. $\widetilde{\text{K}}$ *is an* ω*-code having an infinite delay of decipherability.*

Proof. Assume that there are subfamilies $(w_i)_{i\in\mathbb{N}}, (v_i)_{i\in\mathbb{N}}$ of $\widetilde{\text{K}}$ such that $w_0 \cdot w_1 \cdots = v_0 \cdot v_1 \cdots$ and $w_0 \neq v_0$. W.l.o.g. assume w_0 to be a proper prefix of v_0. Then, since $\widetilde{\text{Ł}}$ and A_1 are prefix codes, $w_0 = x \cdot u_0 \cdots u_j$ and $v_0 = x \cdot u_0 \cdots u_{j'}$ where $x \in A_1$, $u_i \in \widetilde{\text{Ł}}$ and $n > j' > j$. Consequently, u_{j+1} is a prefix of $v_1 \cdots v_m$ for a sufficiently large $m \in \mathbb{N}$ which contradicts Proposition 1.1.

The second assertion follows from Lemma 3.5 in [4], because $\widetilde{\text{K}}^\omega \cap \{\xi : \xi \in A^\omega \wedge \mathbf{A}(\xi) \subseteq \mathbf{A}(\widetilde{\text{K}})\} \supseteq A_1^\omega \neq \emptyset$. $\square$

2. The definition of Łukasiewicz languages

Generalised Łukasiewicz-languages are constructed from pure Łukasiewicz-languages *via* composition of codes (*cf.* Sect. I.6 of [2]) as follows. We start with a code $\tilde{C}$, $\# \tilde{C} \geq 2$, an alphabet A with $\# A = \# \tilde{C}$ and an bijective morphism $\psi : A^* \to \tilde{C}^*$. Let $C := \psi(A_0) \subseteq \tilde{C}$ and $B := \psi(A_1) \subseteq \tilde{C}$. This partitions the code $\tilde{C}$ into nonempty parts C and B.

Let $\widetilde{\text{Ł}}$ be the $\{A_0, A_1\}$-n-Łukasiewicz-language and $\text{Ł} := \psi(\widetilde{\text{Ł}})$. Thus, Ł is the composition of the codes $\widetilde{\text{Ł}}$ and $\tilde{C}$ *via* ψ, $\text{Ł} = \tilde{C} \circ_\psi \widetilde{\text{Ł}}$. Analogously to the previous section Ł is called $\{C, B\}$*-n-Łukasiewicz-language.* For the sake of brevity, we shall omit the prefix "$\{C, B\}$-n-" when there is no danger of confusion. Throughout the rest of the paper we suppose C and B to be disjoint nonempty sets for which $C \cup B$ is a code, and we suppose n to be the composition parameter described in equation (1).

Utilising the properties of the composition of codes (*cf.* [2], Sect. 1.6) from the results of the previous section one can easily derive that Ł has the following properties.

$$\text{Ł} = C \cup B \cdot \text{Ł}^n. \tag{3}$$

Proposition 2.1.

1. *Ł $\subseteq (C \cup B)^* \cdot C \subseteq (C \cup B)^*$.*
2. *Ł is a code, and if $C \cup B$ is a prefix code then Ł is also a prefix code.*
3. *If $w \in (B \cup C)^*$ and $v \in C$ then $w \cdot v^{|w| \cdot n} \in$ Ł*.*
4. *$\mathbf{A}($Ł$^*) = \mathbf{A}((C \cup B)^*)$.*

It should be mentioned that Ł might be a prefix code, even if B and hence $C \cup B$ are codes having no finite delay of decipherability.

Example 2.2. Let $X = \{a, b, c\}$, $C := \{ac^{2i+1}b : i \in \mathbb{N}\}$, $B := \{a, c\} \cup \{ac^{2i}b : i \in \mathbb{N}\}$ and $n \geq 2$. It is easily seen that $C \cup B$ as well as B are codes having no finite delay of decipherability.

Moreover, Ł $\cap \{c\}^* = \emptyset$ and $\mathbf{A}($Ł$^*) \cap \{c\}^* \cdot b = \emptyset$, because each nonempty word $u \in$ Ł* contains a factor of the form $ac^{2i+1}b$.

Assume Ł to be no prefix code. Then there are $w, v \in$ Ł such that $w = a \cdot w_1 \cdots w_n$ with $w_j \in$ Ł and v has a prefix of the form ac^jb. Then $w_1 \in \{c\}^*$ or $c^jb \sqsubseteq w_1$ which is impossible.

In the same way as above we define the *derived language* K as K $:= \tilde{C} \circ_\psi \widetilde{\mathrm{K}}$, and we obtain the following.

$$\mathrm{K} := B \cdot \bigcup\nolimits_{i=0}^{n-1} Ł^i. \tag{4}$$

Propositions 1.2.2 and 1.3 prove that the language K is related to Ł *via* the following properties.

Theorem 2.3.

1. *$\mathbf{A}($Ł$) = \mathrm{K}^* \cdot \mathbf{A}(C \cup B)$.*
2. *Every $w \in (C \cup B)^*$ has a unique factorisation $w = v \cdot u$ where $v \in$ Ł* and $u \in \mathrm{K}^*$.*
3. *K is a code having an infinite delay of decipherability.*

3. The measure of Łukasiewicz languages

In this section we consider the measure of Łukasiewicz languages. Measures of languages were considered in Chapters 1.4 and 2.7 of [2]. In particular, we will consider so-called Bernoulli measures.

3.1. Valuations of languages

As in [6] we call a morphism $\mu : X^* \to (0, \infty)$ of the monoid X^* into the multiplicative monoid of the positive real numbers a *valuation.* A valuation μ such that $\mu(X) = \sum_{x \in X} \mu(x) = 1$ is known as *Bernoulli measure* on X^* (*cf.* [2], Chap. 1.4).

A valuation is usually extended to a mapping $\mu : 2^{X^*} \to [0, \infty]$ *via* $\mu(W) := \sum_{w \in W} \mu(w)$.

Now consider the measure $\mu(Ł)$ for a valuation μ on X^*. Since the decomposition $Ł = C \cup B \cdot Ł^n$ is unambiguous, we obtain

$$\mu(Ł) = \mu(C) + \mu(B) \cdot \mu(Ł)^n.$$

The representation $Ł = \bigcup_{i=1}^{\infty} Ł_i$ where $Ł_1 := C$ and $Ł_{i+1} := C \cup B \cdot Ł_i$ allows us to approximate the measure $\mu(Ł)$ by the sequence

$$\begin{aligned} \mu_1 &:= \mu(C) \\ \mu_{i+1} &:= \mu(C) + \mu(B) \cdot {\mu_i}^n \\ \mu(Ł) &= \lim_{i\to\infty} \mu_i. \end{aligned}$$

We have the following

Theorem 3.1. *If the equation $\lambda = \mu(C) + \mu(B) \cdot \lambda^n$ has a positive solution then $\mu(Ł)$ equals its smallest positive solution, otherwise $\mu(Ł) = \infty$.*

Proof. We have $0 < \mu_1 < \cdots < \mu_i < \mu_{i+1} < \ldots$ Let λ_0 be the minimal positive solution of the equation $\mu(Ł) = \mu(C) + \mu(B) \cdot \mu(Ł)^n$. Then $0 < \lambda_0$ and if $\mu_i < \lambda_0$ then $\mu_{i+1} := \mu(C) + \mu(B) \cdot {\mu_i}^n < \mu(C) + \mu(B) \cdot \lambda_0^n = \lambda_0$. Consequently, $\mu(Ł) \le \lambda_0$.

On the other hand, in view of $\lim_{i\to\infty} \mu_{i+1} = \lim_{i\to\infty}(\mu(C) + \mu(B) \cdot {\mu_i}^n) = \mu(C) + \mu(B) \cdot (\lim_{i\to\infty} \mu_i)^n$, the limit $\lim_{i\to\infty} \mu_i$ is a solution of our equation, and the assertion follows. □

In order to give a more precise evaluation of $\mu(Ł)$, in the subsequent section we take a closer look to our basic equation

$$\lambda = \gamma + \beta \cdot \lambda^n, \tag{5}$$

where $\gamma, \beta > 0$ are positive reals.

In order to estimate $\mu(K)$ we observe that the unambiguous representation of equation (4) yields the formula $\mu(K) = \mu(B) \cdot \sum_{i=0}^{n-1} \mu(Ł)^i$. Then the following connection between the valuations $\mu(Ł)$ and $\mu(K)$ is obvious.

Proposition 3.2. *It holds $\mu(Ł) = \infty$ iff $\mu(K) = \infty$.*

3.2. The basic equation $\lambda = \gamma + \beta \cdot \lambda^n$

This section is devoted to a detailed investigation of the solutions of our basic equation (5). As a result we obtain estimates for the Bernoulli measures of Ł and K as well as a useful tool when we are going to calculate the entropy of Łukasiewicz languages in the subsequent sections.

Let $\bar\lambda$ be an arbitrary positive solution of equation (5). Then we have the following relationship to the value $\gamma + \beta$.

$$\begin{aligned} &\bar\lambda < 1 \Leftrightarrow \bar\lambda < \gamma + \beta\ , \\ &\bar\lambda = 1 \Leftrightarrow \bar\lambda = \gamma + \beta, \text{ and} \\ &\bar\lambda > 1 \Leftrightarrow \bar\lambda > \gamma + \beta. \end{aligned} \tag{6}$$

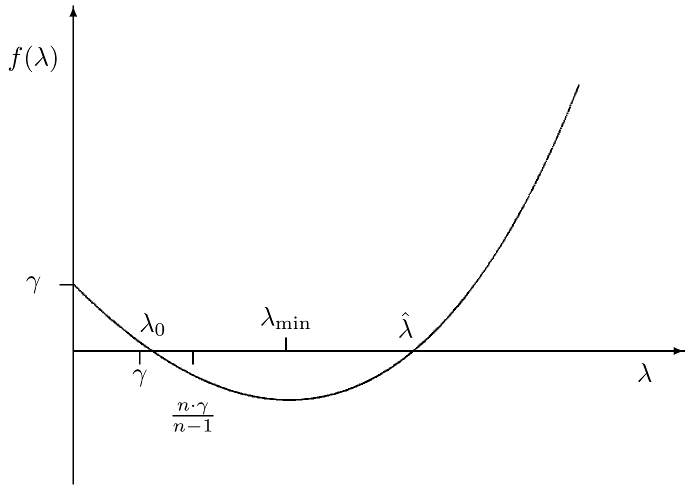

FIGURE 1. Plot of the function $f(\lambda)$ in the case of two positive roots.

Proof. We prove only the last equivalence, the other proofs being similar. From $\bar{\lambda} = \gamma + \beta \cdot \bar{\lambda}^n$, in view of $\bar{\lambda} > 1$, we have immediately $\gamma + \beta < \bar{\lambda}$. Conversely, $\gamma + \beta < \bar{\lambda} = \gamma + \beta \cdot \bar{\lambda}^n$ implies $\bar{\lambda} > 1$. $\square$

In order to study positive solutions it is convenient to consider the positive zeroes of the function.

$$f(\lambda) = \gamma + \beta \cdot \lambda^n - \lambda. \tag{7}$$

The graph of the function f reveals that f has exactly one minimum at $\lambda_{\min}$, $0 < \lambda_{\min} = \frac{1}{\sqrt[n-1]{\beta n}} < \infty$ on the positive real axis, the value of which is $f(\lambda_{\min}) = \gamma - \frac{n-1}{n} \cdot \frac{1}{\sqrt[n-1]{\beta n}} = \frac{n-1}{n}\left(\frac{n}{n-1} \cdot \gamma - \lambda_{\min}\right)$. Thus it has at most two positive roots $\lambda_0, \hat{\lambda}$ which satisfy $0 < \lambda_0 \le \lambda_{\min} \le \hat{\lambda}$.

We obtain the following necessary and sufficient condition for the existence of a positive root λ_0 and its further properties.

Proposition 3.3. *Let $\gamma, \beta > 0$ and let $f(\lambda) = \gamma + \beta \cdot \lambda^n - \lambda$. The function f has a positive root if and only if $\gamma^{n-1} \cdot \beta \le \frac{(n-1)^{n-1}}{n^n}$, and its positive roots satisfy*

$$\gamma < \lambda_0 \le \frac{n \cdot \gamma}{n-1} \le \lambda_{\min} \le \hat{\lambda}\,. \tag{8}$$

Moreover, f has a positive root provided $\gamma + \beta \le 1$, and in this case for its positive roots λ_0 and $\hat{\lambda}$ the following equivalences are valid

$$\begin{array}{l} \gamma + \beta < 1 \Leftrightarrow \lambda_0 < \gamma + \beta < 1 < \hat{\lambda}\ ,\ \textit{and} \\ \gamma + \beta = 1 \Leftrightarrow \lambda_0 = 1 \ \vee\ \hat{\lambda} = 1. \end{array} \tag{9}$$

Proof. As it was mentioned above, the function f has a positive root if and only if $f(\lambda_{\min}) = \gamma - \frac{n-1}{n} \cdot \frac{1}{\sqrt[n-1]{\beta n}} \leq 0$, that is, if $\gamma^{n-1} \cdot \beta \leq \frac{(n-1)^{n-1}}{n^n}$.

Since $f(\lambda) > 0$ for $0 \leq \lambda \leq \gamma$ and $f(\frac{n \cdot \gamma}{n-1}) = \gamma \cdot (\frac{n}{n-1})^n \cdot \left(\gamma^{n-1} \cdot \beta - \frac{(n-1)^{n-1}}{n^n}\right)$, we have $\gamma < \lambda_0 \leq \frac{n}{n-1} \cdot \gamma$ provided f has a positive root.

Now, $f(1) = \gamma + \beta - 1$ and $f(\gamma + \beta) = \beta((\gamma + \beta)^n - 1)$ imply that in case $\gamma + \beta \leq 1$ the function f has positive roots satisfying $\lambda_0 \leq \gamma + \beta \leq 1 \leq \hat{\lambda}$, and in view of equation (6) its positive roots satisfy equation (9). □

Next, we consider the case $f(\lambda_{\min}) = 0$, that is, when f has a positive root of multiplicity two. It turns out that in this case we have some additional restrictions.

Lemma 3.4. *Let $\gamma, \beta > 0$. Then the following conditions are equivalent.*

(1) $\gamma + \beta \cdot \lambda^n - \lambda$ *has a positive root of multiplicity two.*
(2) $\gamma^{n-1} \cdot \beta = \frac{(n-1)^{n-1}}{n^n}$.
(3) $\lambda_{\min} = \frac{n}{n-1} \cdot \gamma$.

Moreover, each one of the conditions implies $\gamma + \beta \geq 1$, and then $\gamma + \beta = 1$ if and only if $\beta = \frac{1}{n}$ or $\gamma = \frac{n-1}{n}$.

Proof. It is obvious that $\lambda_{\min}$ is a root of f iff f has a multiple positive root. The condition $\gamma^{n-1} \cdot \beta = \frac{(n-1)^{n-1}}{n^n}$ is equivalent to $f(\lambda_{\min}) = 0$.

Now, $\gamma^{n-1} \cdot \beta = \frac{(n-1)^{n-1}}{n^n}$ iff $\lambda_{\min} = \frac{1}{\sqrt[n-1]{\beta n}} = \frac{n}{n-1} \cdot \gamma$.

From the arithmetic geometric mean inequality we know that $\sqrt[n]{(\frac{\gamma}{n-1})^{n-1} \cdot \beta} \leq \frac{\gamma+\beta}{n}$ with equality if and only if $\frac{\gamma}{n-1} = \beta$. Thus, Condition 2 implies $\gamma + \beta \geq 1$, and if $\gamma + \beta = 1$ the identity $\gamma^{n-1} \cdot \beta = \frac{(n-1)^{n-1}}{n^n}$ holds iff $\beta = \frac{1}{n}$ or $\gamma = \frac{n-1}{n}$. □

In connection with λ_0 we consider the value $\kappa_0 := \beta \cdot \sum_{i-0}^{n-1} \lambda_0^i$. This value is related to $\mu(K)$ in the same way as λ_0 to $\mu(Ł)$. In view of $\beta(\lambda_0^n - 1) = \beta \cdot \lambda_0^n + \gamma - \gamma - \beta = \lambda_0 - (\gamma + \beta)$, we have

$$\kappa_0 = \begin{cases} n \cdot \beta, & \text{if } \lambda_0 = 1 \text{ and} \\ \dfrac{\lambda_0 - (\gamma + \beta)}{\lambda_0 - 1}, & \text{otherwise.} \end{cases} \tag{10}$$

As a corollary to equation (10) we obtain the following.

Corollary 3.5. $(\kappa_0 - 1) \cdot (\lambda_0 - 1) = 1 - (\gamma + \beta)$.

We obtain our main result on the dependencies between the coefficients of our basic equation (5) and the values of λ_0 and κ_0.

Theorem 3.6. *Let $f(\lambda) = \gamma + \beta \cdot \lambda^n - \lambda$. Then $f(\lambda)$ has a positive root if and only if one of the right hand side conditions in equation (11) to (16) holds. Moreover, the values of its minimum positive root λ_0 and the value $\kappa_0 := \beta \cdot \sum_{i-0}^{n-1} \lambda_0^i$ depend* *in the following way from the coefficients γ and β.*

$$\lambda_0 < 1 \wedge \kappa_0 < 1 \Leftrightarrow \gamma + \beta < 1 \tag{11}$$

$$\lambda_0 < 1 \wedge \kappa_0 = 1 \Leftrightarrow \gamma + \beta = 1 \wedge \beta > \frac{1}{n} \tag{12}$$

$$\lambda_0 < 1 \wedge \kappa_0 > 1 \Leftrightarrow \gamma + \beta > 1 \wedge \beta > \frac{1}{n} \wedge f(\lambda_{\min}) \leq 0 \tag{13}$$

$$\lambda_0 = 1 \wedge \kappa_0 < 1 \Leftrightarrow \gamma + \beta = 1 \wedge \beta < \frac{1}{n} \tag{14}$$

$$\lambda_0 = 1 \wedge \kappa_0 = 1 \Leftrightarrow \gamma + \beta = 1 \wedge \beta = \frac{1}{n} \tag{15}$$

$$\lambda_0 > 1 \wedge \kappa_0 < 1 \Leftrightarrow \gamma + \beta > 1 \wedge \beta < \frac{1}{n} \wedge f(\lambda_{\min}) \leq 0. \tag{16}$$

The remaining cases $\lambda_0 = 1 \wedge \kappa_0 > 1$ and $\lambda_0 > 1 \wedge \kappa_0 \geq 1$ are impossible.

Proof. The function f has a positive root if and only if $f(\lambda_{\min}) \leq 0$. Observe that, in view of equation (9), $\gamma + \beta \leq 1$ implies $f(\lambda_{\min}) \leq 0$. Moreover, if $\gamma + \beta > 1$ and $\beta = \frac{1}{n}$ the function f has no positive root.

Consequently, the six cases on the right hand sides of our equivalences cover the whole range when f has a positive root and, additionally, are mutually excluding each other. Thus it suffices to prove the implications from right to left.

Equation (11). If $\gamma + \beta < 1$ then $\lambda_0 < 1$ (*cf.* Eq. (9)). Hence, Corollary 3.5 implies $\kappa_0 < 1$.

Equations (12) and (13). $\beta > \frac{1}{n}$ is equivalent to $\lambda_{\min} < 1$ whence $\lambda_0 < 1$ if $f(\lambda_{\min}) \leq 0$. Then $\kappa_0 = 1$, in case of equation (12), and $\kappa_0 > 1$, in case of equation (13), follow from Corollary 3.5.

Equation (14). If $\beta < \frac{1}{n}$ we have $\lambda_{\min} > 1$. Thus equation (9) and shows $\lambda_0 = 1$. Now, $\kappa_0 < 1$ follows from equation (10).

Equation (15). This implication is straightforward.

Equation (16). The right hand side is equivalent to $f(1) > 0$, $\lambda_{\min} > 1$ and $f(\lambda_{\min}) \leq 0$, whence $\lambda_{\min} > \lambda_0 > 1$. Again from Corollary 3.5 we obtain $\kappa_0 < 1$. □

From Theorem 3.6 and equation (6) we obtain the following.

Corollary 3.7. *If $\lambda_0 > 1$ then $\lambda_0 > \gamma + \beta > 1$.*

Comparing with the equivalences of Lemma 3.4 we observe that multiple positive roots are possible only in the cases of equations (13), (15) and (16), and that in the case of equation (15) we have necessarily multiple positive roots.

3.3. The Bernoulli measure of Łukasiewicz languages

The last part of Section 3 is an application of the results of the previous subsections to Bernoulli measures. As is well known, a code of Bernoulli measure 1 is maximal (*cf.* [2]). The results of the previous subsection show the following necessary and sufficient conditions.

Theorem 3.8. *Let* $Ł = C \cup B \cdot Ł$*,* $C, B \subseteq X^*$ *be a Łukasiewicz language,* K *its derived language and* $\mu : X^* \to (0,1)$ *a Bernoulli measure. Then* $\mu(Ł) = 1$ *iff* $\mu(C \cup B) = 1$ *and* $\mu(B) \le \frac{1}{n}$*, and* $\mu(\mathrm{K}) = 1$ *iff* $\mu(C \cup B) = 1$ *and* $\mu(B) \ge \frac{1}{n}$*.*

Thus Theorem 3.8 proves that pure Łukasiewicz languages $\widetilde{Ł}$ and their derived languages $\widetilde{\mathrm{K}}$ are maximal codes.

Resuming the results of Section 3 one can say that in order to achieve maximum measure for both Łukasiewicz languages Ł and K it is necessary and sufficient to distribute the measures $\mu(C)$ and $\mu(B)$ as $\mu(C) = \frac{n-1}{n}$ and $\mu(B) = \frac{1}{n}$, thus respecting the composition parameter n in the defining equation (1). A bias in the measure distribution results in a measure loss for at least one of the codes Ł or K.

4. The entropy of Łukasiewicz languages

In [11] Kuich introduced a powerful apparatus in terms of the theory of complex functions to calculate the entropy of unambiguous context-free languages.

For our purposes it is sufficient to consider real functions admitting the value ∞. The coincidence of Kuich's and our approach for Łukasiewicz languages is established by Pringsheim's theorem which states that a power series $\mathfrak{s}(t) = \sum_{i=0}^{\infty} s_i t^i$, $s_i \ge 0$, with finite radius of convergence $\mathsf{rad}\,\mathfrak{s}$ has a singular point at $\mathsf{rad}\,\mathfrak{s}$ and no singular point with modulus less than $\mathsf{rad}\,\mathfrak{s}$. For a more detailed account see [11], Section 2.

Here and in the subsequent section we show that our apparatus establishes a general treatise of the entropies of Ł, K and their star closures $Ł^*$ and K^* provided sufficient information is known about the structure generating functions of the codes C and B.

4.1. Definition and simple properties

The notion of entropy of languages is based on counting words of equal length. Therefore, from now on we assume our alphabet X to be finite of cardinality $\# X = r$, $r \ge 2$.

For a language $W \subseteq X^*$ let $\mathsf{s}_W : \mathbb{N} \to \mathbb{N}$ where $\mathsf{s}_W(n) := \# W \cap X^n$ be its *structure function*, and let

$$\mathsf{H}_W = \limsup_{n \to \infty} \frac{\log_r(1 + \mathsf{s}_W(n))}{n}$$

be its *entropy* (*cf.* [11]).

Informally, this concept measures the amount of information which must be provided on the average in order to specify a particular symbol of a word in a language.

The *structure generating function* corresponding to s_W is

$$\mathfrak{s}_W(t) := \sum_{i\in\mathbb{N}} \mathsf{s}_W(i)\cdot t^i. \tag{17}$$

$\mathfrak{s}_W$ is a power series with convergence radius

$$\operatorname{\mathsf{rad}} W := \liminf_{n\to\infty} \frac{1}{\sqrt[n]{\mathsf{s}_W(n)}},$$

and, considered as a real function on $[0, \operatorname{\mathsf{rad}} W)$, it is nondecreasing.

As it was explained above, it is convenient to consider $\mathfrak{s}_W$ also as a function mapping $[0,\infty)$ to $[0,\infty)\cup\{\infty\}$, where we set

$$\mathfrak{s}_W(\operatorname{\mathsf{rad}} W) := \sup\{\mathfrak{s}_W(\alpha) : \alpha < \operatorname{\mathsf{rad}} W\}\text{ , and} \tag{18}$$

$$\mathfrak{s}_W(\alpha) := \infty,\text{ if } \alpha > \operatorname{\mathsf{rad}} W. \tag{19}$$

Equation (18) is in accordance with Abel's Limit Theorem which states that for $a_i \geq 0$ and $r \geq 0$ one has $\lim\limits_{t\to r,\ t<r} \sum_{i\in\mathbb{N}} a_i\cdot t^i = \sum_{i\in\mathbb{N}} a_i \cdot r^i$ provided $\sum_{i\in\mathbb{N}} a_i\cdot t^i$ converges at $t = r$.

Having in mind this variant of $\mathfrak{s}_W$, we observe that $\mathfrak{s}_W$ is a nondecreasing function which is continuous in the interval $(0, \operatorname{\mathsf{rad}} W)$ and continuous from the left in the point $\operatorname{\mathsf{rad}} W$. Moreover, $\mathfrak{s}_W$ is increasing whenever $W \not\subseteq \{e\}$.

If $W \not\subseteq \{e\}$ we say that $\mathfrak{s}_W$ *reaches* the value s at $t \in [0, \operatorname{\mathsf{rad}} W]$ iff $\mathfrak{s}_W(t) \leq s$ and $\mathfrak{s}_W(t') > s$ for all $t' > t$, that is, for $s \in [0, \mathfrak{s}_W(\operatorname{\mathsf{rad}} W))$ there is a t such that $\mathfrak{s}_W(t) = s$ and, if $\mathfrak{s}_W(\operatorname{\mathsf{rad}} W) < \infty$ any value $s, \mathfrak{s}_W(\operatorname{\mathsf{rad}} W) \leq s < \infty$, is reached at $\operatorname{\mathsf{rad}} W$.

Then the entropy of languages satisfies the following property.

Proposition 4.1.

$$\mathsf{H}_W := \begin{cases} 0, & \textit{if } W \textit{ is finite, and} \\ -\log_r \operatorname{\mathsf{rad}} W, & \textit{otherwise.} \end{cases}$$

Before we proceed to the calculation of the entropy of Łukasiewicz languages we mention still some properties of the entropy of languages which are easily derived from the fact that $\mathfrak{s}_W$ is a positive series (*cf.* [5], Prop. VIII.5.5).

Proposition 4.2. *Let $V, W \subseteq X^*$. Then $0 \leq \mathsf{H}_W \leq 1$ and, if W and V are nonempty, we have $\mathsf{H}_{W\cup V} = \mathsf{H}_{W\cdot V} = \max\{\mathsf{H}_W, \mathsf{H}_V\}$.*

Proof. In fact, $0 \leq \mathsf{s}_W(n) \leq r^n$, and if W and V are nonempty languages then

$$\max\{\mathfrak{s}_V(t), \mathfrak{s}_W(t)\} \leq \mathfrak{s}_{V\cup W}(t) \leq 2\cdot\max\{\mathfrak{s}_V(t), \mathfrak{s}_W(t)\}\text{ and}$$
$$\max\{\mathfrak{s}_{w\cdot V}(t), \mathfrak{s}_{W\cdot v}(t)\} \leq \mathfrak{s}_{V\cdot W}(t) \leq \mathfrak{s}_V(t)\cdot\mathfrak{s}_W(t)$$

when $v \in V$ and $w \in W$.

Consequently, $\operatorname{\mathsf{rad}} V\cup W = \operatorname{\mathsf{rad}} V\cdot W = \min\{\operatorname{\mathsf{rad}} V, \operatorname{\mathsf{rad}} W\}$. □

For the entropy of the star of a language we have the following (*cf.* [5,11]).

Proposition 4.3. *If $V \subseteq X^*$ is a code then*

$$\mathfrak{s}_{V^*}(t) = \textstyle\sum_{i\in\mathbb{N}}(\mathfrak{s}_V(t))^i = \frac{1}{1-\mathfrak{s}_V(t)}, \quad \textit{and}$$
$$\mathsf{H}_{V^*} = \begin{cases} \mathsf{H}_V, & \textit{if } \mathfrak{s}_V(\mathsf{rad}\,V) \le 1, \textit{ and} \\ -\log_r \inf\{\gamma : \mathfrak{s}_V(\gamma) = 1\}, & \textit{otherwise.}\end{cases}$$

In general $\sum_{i\in\mathbb{N}}(\mathfrak{s}_W(t))^i$ is only an upper bound to $\mathfrak{s}_{W^*}(t)$. Hence only in case $\mathfrak{s}_W(t) < 1$ one can conclude that $\mathfrak{s}_{W^*}(t) \le \sum_{i\in\mathbb{N}}(\mathfrak{s}_W(t))^i < \infty$ and, consequently, $t \le \mathsf{rad}\,W^*$. Thus we obtain a sufficient condition for the equality $\mathsf{H}_W = \mathsf{H}_{W^*}$ depending on the value of $\mathfrak{s}_W(\mathsf{rad}\,W)$.

Corollary 4.4. *Let $W \subseteq X^*$. We have $\mathsf{H}_W = \mathsf{H}_{W^*}$ if $\mathfrak{s}_W(\mathsf{rad}\,W) \le 1$, and if W is a code $W \subseteq X^*$ it holds $\mathsf{H}_W < \mathsf{H}_{W^*}$ if and only if $\mathfrak{s}_W(\mathsf{rad}\,W) > 1$.*

Property 4.3 and Corollary 4.4 show that the value t_1 for which $\mathfrak{s}_V(t_1) = 1$ is crucial for the calculation of H_{V^*} and yields an exact estimate of H_{V^*} if V is a code.

4.2. The calculation of the convergence radius

Property 4.1 showed the close relationship between H_W and $\mathsf{rad}\,W$, and Corollary 4.4 proved that the value of $\mathfrak{s}_W$ at the point $\mathsf{rad}\,W$ is of importance for the calculation of the entropy of the star language of W, H_{W^*}.

Therefore, in this section we are going to estimate the convergence radius of the power series $\mathfrak{s}_{\mathrm{L}}(t)$ and simultaneously, the values $\mathfrak{s}_{\mathrm{L}}(\mathsf{rad}\,\mathrm{L})$ and $\mathfrak{s}_{\mathrm{K}}(\mathsf{rad}\,\mathrm{L})$ (observe that $\mathsf{rad}\,\mathrm{L} = \mathsf{rad}\,\mathrm{K}$ in view of (4) and Prop. 4.2). We start with the equation

$$\mathfrak{s}_{\mathrm{L}}(t) = \mathfrak{s}_C(t) + \mathfrak{s}_B(t)\cdot\mathfrak{s}_{\mathrm{L}}(t)^n \tag{20}$$

which follows from the unambiguous representation in equation (3) and the observation that $\mathsf{rad}\,\mathrm{L} = \sup\{t : \mathfrak{s}_{\mathrm{L}}(t) < \infty\} = \inf\{t : \mathfrak{s}_{\mathrm{L}}(t) = \infty\}$, because the function $\mathfrak{s}_{\mathrm{L}}(t)$ is nondecreasing (even increasing on $[0, \mathsf{rad}\,\mathrm{L}]$).

From Section 3.2 we know that, for fixed t, $t < \mathsf{rad}\,\mathrm{L}$, the value $\mathfrak{s}_{\mathrm{L}}(t)$ is one of the solutions of equation (5) with $\gamma = \mathfrak{s}_C(t)$ and $\beta = \mathfrak{s}_B(t)$. Similarly to Theorem 3.1 one can prove the following.

Theorem 4.5. *Let $t > 0$. If equation (5) has a positive solution for $\gamma = \mathfrak{s}_C(t)$ and $\beta = \mathfrak{s}_B(t)$ then $\mathfrak{s}_{\mathrm{L}}(t) = \lambda_0$, and if equation (5) has no positive solution then $\mathfrak{s}_{\mathrm{L}}(t)$ diverges, that is, $\mathfrak{s}_{\mathrm{L}}(t) = \infty$.*

This yields an estimate for the convergence radius of $\mathfrak{s}_{\mathrm{L}}(t)$ as the point at which the product $\mathfrak{s}_C(t)^{n-1}\cdot\mathfrak{s}_B(t)$ reaches the value $\frac{(n-1)^{n-1}}{n^n}$.

$$\begin{aligned}\mathsf{rad}\,\mathrm{L} &= \inf\left\{\mathsf{rad}\,(C\cup B)\right\} \cup \left\{t : \mathfrak{s}_C(t)^{n-1}\cdot\mathfrak{s}_B(t) > \frac{(n-1)^{n-1}}{n^n}\right\} \\ &= \sup\left\{t : \mathfrak{s}_C(t)^{n-1}\cdot\mathfrak{s}_B(t) \le \frac{(n-1)^{n-1}}{n^n}\right\}. \end{aligned}\tag{21}$$

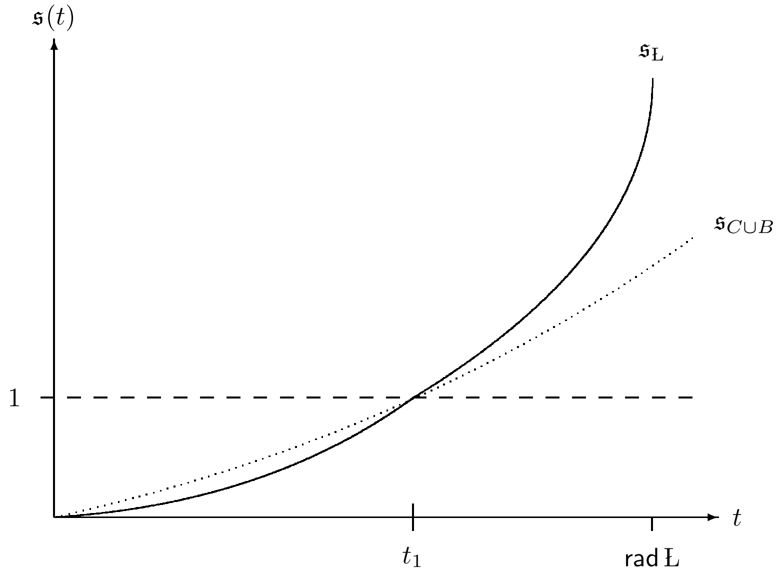

FIGURE 2. A typical plot of the structure generating functions of Ł and $C \cup B$.

Proof. Clearly, $\mathfrak{s}_Ł(t)$ converges only if $\mathfrak{s}_C(t)$ and $\mathfrak{s}_B(t)$ converge. If $t \le \mathsf{rad}\,(C \cup B)$ then $\mathfrak{s}_Ł(t) < \infty$ if and only if our basic equation has a solution. This is the case when $f(\lambda_{\min}) \le 0$, that is, if $\mathfrak{s}_C(t)^{n-1} \cdot \mathfrak{s}_B(t) \le \frac{(n-1)^{n-1}}{n^n}$. □

As $\mathfrak{s}_Ł(t) < \infty$ whenever $\mathfrak{s}_C(t)^{n-1} \cdot \mathfrak{s}_B(t) \le \frac{(n-1)^{n-1}}{n^n}$ we obtain

$$\mathfrak{s}_Ł(\mathsf{rad}\,Ł) < \infty. \tag{22}$$

Using Theorem 4.5 in connection with the results of Section 3.2 we can describe the behaviour of $\mathfrak{s}_Ł$ on $[0, \mathsf{rad}\,Ł]$ as follows (see Fig. 2). Observe that $\mathfrak{s}_{C\cup B}$ and $\mathfrak{s}_Ł$ are increasing on $[0, \mathsf{rad}\,C \cup B)$ and $[0, \mathsf{rad}\,Ł]$, respectively. First equation (8) shows

$$\mathfrak{s}_C(t) < \mathfrak{s}_Ł(t) < \frac{n}{n-1}\mathfrak{s}_C(t) \text{ for } 0 < t < \mathsf{rad}\,Ł\,.$$

Moreover, from equation (9) we obtain that $\mathfrak{s}_Ł(t) < \mathfrak{s}_C(t) + \mathfrak{s}_B(t)$ as long as $\mathfrak{s}_C(t) + \mathfrak{s}_B(t) < 1$. The value t_1 for which $\mathfrak{s}_C(t_1) + \mathfrak{s}_B(t_1) = 1$ is crucial for the behaviour of $\mathfrak{s}_Ł$:

If $\mathfrak{s}_Ł(t_1) = 1$ then $\mathfrak{s}_Ł(t) > 1$ for $t_1 < t \le \mathsf{rad}\,Ł$ and Corollary 3.5 implies that then $\mathfrak{s}_Ł(t) > \mathfrak{s}_C(t) + \mathfrak{s}_B(t)$. On the other hand, if $\mathfrak{s}_Ł(t_1) < 1$ then $\mathfrak{s}_Ł(t) < 1$ in the whole range $0 \le t \le \mathsf{rad}\,Ł$, because $\mathfrak{s}_Ł(t) = 1$ implies $\mathfrak{s}_C(t) + \mathfrak{s}_B(t) = 1$ which is impossible for $t > t_1$.

We obtain two corollaries to Theorem 4.5 and equation (21) which allow us to estimate $\mathsf{rad}\,Ł$. The first one follows from Lemma 3.4 and covers also the case when $\mathsf{rad}\,(C \cup B) = \infty$.

Corollary 4.6. *If* $\frac{(n-1)^{n-1}}{n^n} \leq \mathfrak{s}_C(\mathsf{rad}\,(C \cup B))^{n-1} \cdot \mathfrak{s}_B(\mathsf{rad}\,(C \cup B))$ *then* $\mathsf{rad}\,Ł$ *is the solution of the equation* $\mathfrak{s}_C(t)^{n-1} \cdot \mathfrak{s}_B(t) = \frac{(n-1)^{n-1}}{n^n}$.

In this case $\mathfrak{s}_{C\cup B}(\mathsf{rad}\,Ł) \geq 1$ *and* $\mathfrak{s}_Ł(\mathsf{rad}\,Ł) = \frac{n}{n-1} \cdot \mathfrak{s}_C(\mathsf{rad}\,Ł)$. *Moreover, then, the following conditions are equivalent:*

(1) $\mathfrak{s}_{C\cup B}(\mathsf{rad}\,Ł) = 1$;
(2) $\mathfrak{s}_B(\mathsf{rad}\,Ł) = \frac{1}{n}$;
(3) $\mathfrak{s}_C(\mathsf{rad}\,Ł) = \frac{n-1}{n}$; *and*
(4) $\mathfrak{s}_Ł(\mathsf{rad}\,Ł) = 1$.

The second corollary covers the case when $\frac{(n-1)^{n-1}}{n^n} > \mathfrak{s}_C(t)^{n-1} \cdot \mathfrak{s}_B(t)$ for all $t \leq \mathsf{rad}\,(C \cup B)$. Here $\mathsf{rad}\,(C \cup B) < \infty$.

Corollary 4.7. *We have* $\mathsf{rad}\,Ł = \mathsf{rad}\,(C \cup B)$ *if and only if* $\mathsf{rad}\,(C \cup B) < \infty$ *and* $\frac{(n-1)^{n-1}}{n^n} \geq \mathfrak{s}_C(\mathsf{rad}\,(C \cup B))^{n-1} \cdot \mathfrak{s}_B(\mathsf{rad}\,(C \cup B))$.

If $C \cup B$ is a finite prefix code then $\mathsf{rad}\,(C \cup B) = \infty$. In this case $\mathsf{rad}\,Ł$ is defined *via* $\mathfrak{s}_C(\mathsf{rad}\,Ł)^{n-1} \cdot \mathfrak{s}_B(\mathsf{rad}\,Ł) = \frac{(n-1)^{n-1}}{n^n}$. Hence Corollary 4.6 applies. We give an example that, depending on C and B, all three cases $\mathfrak{s}_Ł(\mathsf{rad}\,Ł) < 1$, $\mathfrak{s}_Ł(\mathsf{rad}\,Ł) = 1$ and $\mathfrak{s}_Ł(\mathsf{rad}\,Ł) > 1$ are possible.

Example 4.8. Fix $m \geq 1$ and $C \cup B \subseteq X^m$. Then $\mathfrak{s}_C(t)^{n-1} \cdot \mathfrak{s}_B(t) = (\#\, C \cdot t)^{n-1} \cdot \#\, B \cdot t = \frac{(n-1)^{n-1}}{n^n}$ has the minimum positive solution

$$\mathsf{rad}\,Ł = \sqrt[m\cdot n]{\left(\frac{n-1}{n \cdot \#\, C}\right)^{n-1} \cdot \frac{1}{n \cdot \#\, B}},$$

and, utilising Corollary 4.6, we obtain

$$\mathfrak{s}_Ł(\mathsf{rad}\,Ł) = \frac{n}{n-1} \cdot \mathfrak{s}_C(\mathsf{rad}\,Ł) = \frac{n}{n-1} \cdot \#\, C \cdot (\mathsf{rad}\,Ł)^m = \sqrt[n]{\frac{\#\, C}{(n-1) \cdot \#\, B}}.$$

Choosing $m := 1$, $X := \{a, b, d\}$, $C_0 := \{a, d\}$ and $B_0 := \{b\}$ and n appropriately we obtain the above mentioned three cases:

Define the Łukasiewicz languages $Ł_i$ $(i = 1, 2, 3)$ *via* the equation

$$Ł_i = \{a, d\} \cup \{b\} \cdot Ł_i^{i+1}.$$

Then we have

	$i = 1\ (n = 2)$	$i = 2\ (n = 3)$	$i = 3\ (n = 4)$
$\mathsf{rad}\,Ł_i$	$\frac{1}{2\sqrt{2}}$	$\frac{1}{3}$	$\frac{3}{8} \cdot \sqrt[4]{\frac{2}{3}}$
$\mathfrak{s}_{Ł_i}(\mathsf{rad}\,Ł_i)$	$\sqrt{2} > 1$	1	$\sqrt[4]{\frac{2}{3}} < 1$
$\mathsf{H}_{Ł_i}$	$h_3(\frac{1}{2}) = \frac{3}{2}\log_3 2$	$h_3(\frac{2}{3}) = 1$	$h_3(\frac{3}{4}) = \frac{1}{4}\log_3 \frac{2048}{27}$

Here $h_r(p) = -(1-p)\cdot\log_r(1-p) - p\cdot\log_r \frac{p}{r-1}$ is the r-ary entropy function well-known from information theory (*cf.* [8, Sect. 2.3]). This function satisfies $0 \le h_r(p) \le 1$ for $0 \le p \le 1$ and $h_r(p) = 1$ iff $p = \frac{r-1}{r}$.

Remark 1. If we set $m := 1$, $\# B := 1$, and $C := X \setminus B$, whence $\# C = r-1$, we obtain a slight generalisation of Kuich's example [11, Example 1] (see also [9], Ex. 4.1) to alphabets of cardinality $\# X = r \ge 2$, yielding $\mathsf{H}_{Ł} = h_r(\frac{n-1}{n})$.

In the case of Corollary 4.7 when $\mathfrak{s}_C(\mathsf{rad}\,Ł)^{n-1}\cdot\mathfrak{s}_B(\mathsf{rad}\,Ł) < \frac{(n-1)^{n-1}}{n^n}$ the value $\mathfrak{s}_Ł(\mathsf{rad}\,Ł)$ is a single root of equation (20). Then the results of Section 3.2 show that $\mathfrak{s}_B(t) = \frac{1}{n}$ and simultaneously $\mathfrak{s}_C(t) = \frac{n-1}{n}$ is impossible for $t \le \mathsf{rad}\,Ł$. The other cases, except for equation (15), listed in Theorem 3.6 are possible. This can be shown using the Łukasiewicz languages $Ł_i$ $(i = 1,2,3)$ constructed in Example 4.8 as basic codes $Ł_i = C \cup B$ and splitting them appropriately.

Example 4.9. We let, generally, $n := 2$ and define our Łukasiewicz languages $Ł_i$ $(i = 4,\dots,8)$ by

$$Ł_i = C_i \cup B_i \cdot Ł_i^2 \text{ where}$$

	C_i	B_i	$r_i := \mathsf{rad}\,C_i \cup B_i$
$i=4$	$b\cdot\{a,d\}^2 \subseteq Ł_1$	$Ł_1 \setminus C_4$	$\mathsf{rad}\,Ł_1 = \frac{1}{2\sqrt{2}}$
$i=5$	$B_4 = Ł_1 \setminus C_4$	$C_4 = b\cdot\{a,d\}^2$	$\mathsf{rad}\,Ł_1 = \frac{1}{2\sqrt{2}}$
$i=6$	$\{a,d\} \subseteq Ł_2$	$Ł_2 \setminus C_6$	$\mathsf{rad}\,Ł_2 = \frac{1}{3}$
$i=7$	$B_6 = Ł_2 \setminus C_6$	$C_6 = \{a,d\}$	$\mathsf{rad}\,Ł_2 = \frac{1}{3}$
$i=8$	$\{a,d\} \subseteq Ł_3$	$Ł_3 \setminus C_8$	$\mathsf{rad}\,Ł_3 = \frac{3}{8}\cdot\sqrt[4]{\frac{2}{3}}$

This yields the following values of $\mathfrak{s}_{C_i}(r_i), \mathfrak{s}_{B_i}(r_i)$, $\mathfrak{s}_{C_i}(r_i)\cdot\mathfrak{s}_{B_i}(r_i)$ and $\mathfrak{s}_{C_i\cup B_i}(r_i)$, where the latter three are compared with the values of $\frac{1}{n}$, $\frac{(n-1)^{n-1}}{n^n}$, and 1, respectively.

	$\mathfrak{s}_{C_i}(r_i)$	$\mathfrak{s}_{B_i}(r_i)$	$\mathfrak{s}_{C_i}(r_i)\cdot\mathfrak{s}_{B_i}(r_i)$	$\mathfrak{s}_{C_i\cup B_i}(r_i)$
$i=4$	$\frac{1}{4\sqrt{2}}$	$\sqrt{2}-\frac{1}{4\sqrt{2}} > 1/2$	$\frac{1}{4}-\frac{1}{32} < 1/4$	$\sqrt{2} > 1$
$i=5$	$\sqrt{2}-\frac{1}{4\sqrt{2}}$	$\frac{1}{4\sqrt{2}} < 1/2$	$\frac{1}{4}-\frac{1}{32} < 1/4$	$\sqrt{2} > 1$
$i=6$	$\frac{2}{3}$	$\frac{1}{3} < 1/2$	$\frac{2}{9} < 1/4$	1
$i=7$	$\frac{1}{3}$	$\frac{2}{3} > 1/2$	$\frac{2}{9} < 1/4$	1
$i=8$	$\frac{3}{4}\cdot\sqrt[4]{\frac{2}{3}}$	$\frac{1}{4}\cdot\sqrt[4]{\frac{2}{3}}$	$\frac{3}{16}\cdot\sqrt{\frac{2}{3}} < 1/4$	$\sqrt[4]{\frac{2}{3}} < 1$

By Corollary 4.7, $\mathsf{rad}\,\text{Ł}_i = \mathsf{rad}\,C_i \cup B_i$ for $i = 4, \ldots, 8$, and we obtain

$$\begin{array}{ll} \mathfrak{s}_{\text{Ł}_4}(\mathsf{rad}\,\text{Ł}_4) < 1 & \text{according to equation (13),} \\ \mathfrak{s}_{\text{Ł}_5}(\mathsf{rad}\,\text{Ł}_5) > 1 & \text{according to equation (16),} \\ \mathfrak{s}_{\text{Ł}_6}(\mathsf{rad}\,\text{Ł}_6) = 1 & \text{according to equation (14),} \\ \mathfrak{s}_{\text{Ł}_7}(\mathsf{rad}\,\text{Ł}_7) < 1 & \text{according to equation (12), and} \\ \mathfrak{s}_{\text{Ł}_8}(\mathsf{rad}\,\text{Ł}_8) < 1 & \text{according to equation (11).} \end{array}$$

4.3. The entropies of Ł* and K*

The previous part of Section 4 was mainly devoted to explain how to give estimates on the entropy of Ł on the basis of the structure generating functions of the basic codes C and B. As a byproduct we could sometimes achieve some knowledge about $\mathfrak{s}_{\text{Ł}}(\mathsf{rad}\,\text{Ł})$.

We are going to explore this situation in more detail in this section.

In particular, we derive estimates for the entropies $\mathsf{H}_{\text{Ł}}, \mathsf{H}_{\text{K}}, \mathsf{H}_{\text{Ł}^*}$ and H_{K^*} relative to the entropies of the basic code $C \cup B$ and its star language $(C \cup B)^*$. Using elementary properties of the entropy established in Property 4.2 we obtain

$$\mathsf{H}_{C\cup B} \leq \mathsf{H}_{\text{Ł}} = \mathsf{H}_{\text{K}} \leq \min\{\mathsf{H}_{\text{Ł}^*}, \mathsf{H}_{\text{K}^*}\} \leq \max\{\mathsf{H}_{\text{Ł}^*}, \mathsf{H}_{\text{K}^*}\} = \mathsf{H}_{(C\cup B)^*}\,. \tag{23}$$

Proof. As $C \cup C \cdot B^n \subseteq \text{Ł}$, $B \cdot \text{Ł} \subseteq \text{K}$ and $\text{Ł}^* \cup \text{K}^* \subseteq (C \cup B)^*$, we have $\mathsf{H}_{C\cup B} \leq \mathsf{H}_{\text{Ł}} \leq \mathsf{H}_{\text{K}} \leq \min\{\mathsf{H}_{\text{Ł}^*}, \mathsf{H}_{\text{K}^*}\} \leq \max\{\mathsf{H}_{\text{Ł}^*}, \mathsf{H}_{\text{K}^*}\} \leq \mathsf{H}_{(C\cup B)^*}$

The identity $\mathsf{H}_{\text{Ł}} = \mathsf{H}_{\text{K}}$ is a consequence of equation (4) and Property 4.2. For a proof of the inequality $\max\{\mathsf{H}_{\text{Ł}^*}, \mathsf{H}_{\text{K}^*}\} \geq \mathsf{H}_{(C\cup B)^*}$ observe that Theorem 2.3.2 shows $\text{Ł}^* \cdot \text{K}^* \supseteq (C \cup B)^*$, whence Property 4.2 yields the assertion. □

As a byproduct of the subsequent estimates of $\mathsf{H}_{\text{Ł}^*}$ and H_{K^*} we get the identity $\mathsf{H}_{\text{Ł}} = \mathsf{H}_{\text{K}} = \min\{\mathsf{H}_{\text{Ł}^*}, \mathsf{H}_{\text{K}^*}\}$ (see Cor. 4.11 below), whereas we shall show in Example 4.12 that the other inequalities in equation (23) are independent of each other.

Since $C \cup B$ is a code, Corollary 4.4 implies a necessary an sufficient condition for the entropies in equation (23) to coincide.

Proposition 4.10. *The equality* $\mathsf{H}_{C\cup B} = \mathsf{H}_{(C\cup B)^*}$ *holds if and only if* $\mathfrak{s}_{C\cup B}(t) < 1$ *for all* $t \in [0, \mathsf{rad}\,C \cup B)$.

Next we consider the case when $\mathfrak{s}_{C\cup B}(t_1) = 1$ for some $t_1 \in [0, \mathsf{rad}\,C \cup B]$. We know from the considerations in Section 4.3 and from Property 4.3 that this value is closely connected to the entropy of the star language Ł^*. In particular, $t_1 = \mathsf{rad}\,(C \cup B)^*$ if t_1 exists.

The following table shows the dependencies of the values related to the entropies $\mathsf{H}_{Ł} = \mathsf{H}_{K}, \mathsf{H}_{Ł^*}$ and H_{K^*} from the value which takes on the function $\mathfrak{s}_B$ at our critical point $t_1 = \mathsf{rad}\,(C \cup B)^*$.

	$\mathfrak{s}_B(t_1) < \frac{1}{n}$	$\mathfrak{s}_B(t_1) = \frac{1}{n}$	$\mathfrak{s}_B(t_1) > \frac{1}{n}$
$\mathfrak{s}_{Ł}(t_1)$	$= 1$	$= 1$	< 1
$\mathfrak{s}_{K}(t_1)$	< 1	$= 1$	$= 1$
$\mathfrak{s}_{Ł}(t)$ for $t \in (t_1, \mathsf{rad}\,Ł]$	> 1	—	< 1
$\mathfrak{s}_{K}(t)$ for $t \in (t_1, \mathsf{rad}\,Ł]$	< 1	—	> 1
$\mathsf{rad}\,Ł = \mathsf{rad}\,K$	$\geq \mathsf{rad}\,(C \cup B)^*$	$= \mathsf{rad}\,(C \cup B)^*$	$\geq \mathsf{rad}\,(C \cup B)^*$
$\mathsf{rad}\,Ł^*$	$= \mathsf{rad}\,(C \cup B)^*$	$= \mathsf{rad}\,(C \cup B)^*$	$= \mathsf{rad}\,Ł$
$\mathsf{rad}\,K^*$	$= \mathsf{rad}\,K$	$= \mathsf{rad}\,(C \cup B)^*$	$= \mathsf{rad}\,(C \cup B)^*$

We give some explanations.

Proof. The results of Rows 1 and 2 follow from equations (12), (14), (15), (20) and the identity $\mathfrak{s}_K(t) = \sum_{i=0}^{n-1} \mathfrak{s}_B(t) \cdot (\mathfrak{s}_{Ł}(t))^i$.

Since $\mathfrak{s}_{C\cup B}(t) > 1$ for $t \in (t_1, \mathsf{rad}\,Ł]$ Rows 3 and 4 follow from equations (13) and (16). Observe that $\mathfrak{s}_B(t_1) = \frac{1}{n}$ and $\mathfrak{s}_{C\cup B}(t_1) = 1$ imply $t_1 = \mathsf{rad}\,Ł = \mathsf{rad}\,K$.

Finally, Properties 4.1 and 4.3 in connection with the preceding rows yield the results for $\mathsf{rad}\,Ł^*$ and $\mathsf{rad}\,K^*$. □

We rephrase our results in terms of entropies of the languages Ł, K, $Ł^*$ and K^*.

Corollary 4.11. *Let $\mathfrak{s}_{C\cup B}(t_1) = 1$. Then the following holds.*

(1) *If $\mathfrak{s}_B(t_1) < \frac{1}{n}$ then $\mathsf{H}_{Ł} = \mathsf{H}_{K} = \mathsf{H}_{K^*} \leq \mathsf{H}_{Ł^*}$.*
(2) *If $\mathfrak{s}_B(t_1) = \frac{1}{n}$ then $\mathsf{H}_{Ł} = \mathsf{H}_{K} = \mathsf{H}_{K^*} = \mathsf{H}_{Ł^*}$.*
(3) *If $\mathfrak{s}_B(t_1) > \frac{1}{n}$ then $\mathsf{H}_{Ł} = \mathsf{H}_{K} = \mathsf{H}_{Ł^*} \leq \mathsf{H}_{K^*}$.*

In particular, we have always $\mathsf{H}_{Ł} = \mathsf{H}_{K} = \min\{\mathsf{H}_{Ł^*}, \mathsf{H}_{K^*}\}$.

We conclude this section by computing the entropies $\mathsf{H}_{Ł_i}, \mathsf{H}_{K_i}, \mathsf{H}_{Ł_i^*}$ and $\mathsf{H}_{K_i^*}$ ($i = 1, \ldots, 8$) for the Łukasiewicz languages given in Examples 4.8 and 4.9 and their counterparts $K_i := b \cdot \bigcup_{j=0}^{i} Ł_i^j$ ($i = 1, 2, 3$) and $K_i := B_i \cup B_i \cdot Ł_i$ ($i = 4, \ldots, 8$).

These examples show that all possible cases in equation (23) really occur.

Example 4.12. We present our results in the table below. The value of t_1 is always $\frac{1}{3}$ except for $i = 8$ when $\mathfrak{s}_{C_8 \cup B_8}(\mathsf{rad}\, C_8 \cup B_8) < 1$.

	n	$\mathfrak{s}_B(t_1) \lesseqgtr \frac{1}{n}$	$\mathsf{H}_{C\cup B}$	$\mathsf{H}_{\text{Ł}} = \mathsf{H}_{\text{K}}$	$\mathsf{H}_{\text{Ł}^*}$	H_{K^*}	$\mathsf{H}_{(C\cup B)^*}$
$i = 1$	2	$\frac{1}{3} < 1/2$	0	$h_3(\frac{1}{2})$	1	$h_3(\frac{1}{2})$	1
$i = 2$	3	$\frac{1}{3} = 1/3$	0	1	1	1	1
$i = 3$	4	$\frac{1}{3} > 1/4$	0	$h_3(\frac{3}{4})$	$h_3(\frac{3}{4})$	1	1
$i = 4$	2	$\frac{23}{27} > 1/2$	$h_3(\frac{1}{2})$	$h_3(\frac{1}{2})$	$h_3(\frac{1}{2})$	1	1
$i = 5$	2	$\frac{4}{27} < 1/2$	$h_3(\frac{1}{2})$	$h_3(\frac{1}{2})$	1	$h_3(\frac{1}{2})$	1
$i = 6$	2	$\frac{1}{3} < 1/2$	1	1	1	1	1
$i = 7$	2	$\frac{2}{3} > 1/2$	1	1	1	1	1
$i = 8$	2		$h_3(\frac{3}{4})$	$h_3(\frac{3}{4})$	$h_3(\frac{3}{4})$	$h_3(\frac{3}{4})$	$h_3(\frac{3}{4})$

Observe that $0 < h_3(\frac{1}{2}) < 1$ and $0 < h_3(\frac{3}{4}) < 1$.

In conclusion, one should remark that in the case of entropy of Łukasiewicz languages a similar situation as in the case of their Bernoulli measures appears. In order to achieve maximum possible entropy for both Łukasiewicz languages Ł and K it is necessary and sufficient to choose basic codes C and B whose power series $\mathfrak{s}_C(t)$ and $\mathfrak{s}_B(t)$ behave in agreement with the composition parameter n of the Łukasiewicz language.

Acknowledgements. My thanks are due to Jeanne Devolder who helped me in simplifying my first rather lengthy proof of Proposition 1.3 and to the unknown referee for many helpful comments.

References

[1] J.-M. Autebert, J. Berstel and L. Boasson, *Context-Free Languages and Pushdown Automata,* in *Handbook of Formal Languages*, edited by G. Rozenberg and A. Salomaa. Springer-Verlag, Berlin **1** (1997) 111–174.
[2] J. Berstel and D. Perrin, *Theory of Codes.* Academic Press, Orlando (1985).
[3] N. Chomsky and G.A. Miller, Finite-state languages. *Inform. Control* **1** (1958) 91–112.
[4] J. Devolder, M. Latteux, I. Litovski and L. Staiger, Codes and Infinite Words. *Acta Cybernetica* **11** (1994) 241–256.
[5] S. Eilenberg, *Automata, Languages and Machines, Vol. A.* Academic Press, New York (1974).
[6] H. Fernau, Valuations of Languages, with Applications to Fractal Geometry. *Theoret. Comput. Sci.* **137** (1995) 177–217.
[7] G. Hansel, D. Perrin and I. Simon, Entropy and compression, in *STACS'92*, edited by A. Finkel and M. Jantzen. *Lect. Notes Comput. Sci.* **577** (1992) 515–530.
[8] R. Johannesson, *Informations theorie.* Addison-Wesley (1992).
[9] J. Justesen and K. Larsen, On Probabilistic Context-Free Grammars that Achieve Capacity. *Inform. Control* **29** (1975) 268–285.

[10] F.P. Kaminger, The noncomputability of the channel capacity of context-sensitive languages. *Inform. Control* **17** (1970) 175–182.
[11] W. Kuich, On the entropy of context-free languages. *Inform. Control* **16** (1970) 173–200.
[12] M. Li and P.M.B. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, New York (1993).
[13] L. Staiger, On infinitary finite length codes. *RAIRO-Inf. Theor. Appl.* **20** (1986) 483–494.
[14] L. Staiger, Ein Satz über die Entropie von Untermonoiden. *Theor. Comput. Sci.* **61** (1988) 279–282.
[15] L. Staiger, Kolmogorov complexity and Hausdorff dimension. *Inform. Comput.* **103** (1993) 159–194.

Communicated by H.J. Hoogeboom.
Received August 7, 2002. Accepted August 5, 2004.