# Sublinear Ambiguity

**Klaus Wich**
E-mail: wich@informatik.uni-stuttgart.de

Institut für Informatik, Universität Stuttgart,
Breitwiesenstr. 20-22, 70565 Stuttgart.

**Abstract.** A context-free grammar $G$ is ambiguous if there is a word that can be generated by $G$ with at least two different derivation trees. Ambiguous grammars are often distinguished by their degree of ambiguity, which is the maximal number of derivation trees for the words generated by them. If there is no such upper bound $G$ is said to be ambiguous of infinite degree. By considering how many derivation trees a word of at most length $n$ may have, we can distinguish context-free grammars with infinite degree of ambiguity by the growth-rate of their ambiguity with respect to the length of the words. It is known that each cycle-free context-free grammar $G$ is either exponentially ambiguous or its ambiguity is bounded by a polynomial. Until now there have only been examples of context-free languages with inherent ambiguity $2^{\Theta(n)}$ and $\Theta(n^d)$ for each $d \in \mathbb{N}_0$. In this paper first examples of (linear) context-free languages with nonconstant sublinear ambiguity are presented.

## 1 Introduction

A context-free grammar $G$ is ambiguous if there is a word that can be generated by $G$ with at least two different derivation trees. Ambiguous grammars are often distinguished by their degree of ambiguity, which is the maximal number of derivation trees for the words generated by them. If there is no such upper bound $G$ is said to be ambiguous of infinite degree.

In [5] and [6] the ambiguity function has been introduced as a new tool for examining the ambiguity of cycle-free context-free grammars. The ambiguity function maps the natural number $n$ to the maximal number of derivation trees which a word of at most length $n$ may have. It has been shown there that for cycle-free context-free grammars the ambiguity function is either an element of $2^{\Theta(n)}$ or of $\mathcal{O}(n^d)$ for a $d \in \mathbb{N}_0$ which can be effectively constructed from $G$

$L$ has inherent ambiguity $\Theta(f)$ if there is a grammar with an ambiguity function in $\mathcal{O}(f)$ and each grammar that generates $L$ has an ambiguity function in $\Omega(f)$. Languages with inherent ambiguity $2^{\Theta(n)}$ and with inherent ambiguity $\Theta(n^d)$ for each $d \in \mathbb{N}_0$ have been presented in [4].

It is easy to prove that the above mentioned infinite ambiguities are exactly the ones that can occur, of course not inherently, in *right linear* grammars over a *single letter alphabet*. In that sense sublinear ambiguity requires a more complicated structure.

In this article first examples of context-free grammars having sublinear ambiguity are presented (they are even linear). These grammars have logarithmic and square-root ambiguity, respectively. Moreover it is shown that these ambiguities are inherent for the corresponding languages.

## 2 Preliminaries

Let $\Sigma$ be a finite alphabet. For a word $w \in \Sigma^*$, a symbol $a \in \Sigma$, and $n \in \mathbb{N}$ the length of $w$ is denoted by $|w|$, the number of $a$'s in $w$ is denoted by $|w|_a$. The empty word is denoted by $\varepsilon$. The set $\Sigma^{\leq n}$ denotes all words over $\Sigma$ with length up to $n$. The cardinality of a set $S$ is denoted by $|S|$.

A *context-free grammar* is a quadruple $G = (N, \Sigma, P, S)$, where $N$ and $\Sigma$ are finite disjoint alphabets of nonterminals and terminals, respectively, $S \in N$ is the start symbol, and $P \subseteq N \times (N \cup \Sigma)^*$ is a finite set of productions. We usually write $A \to \alpha$ or $(A \to \alpha)$ for the pair $(A, \alpha)$. We write $A \to \alpha \mid \beta$ as an abbreviation for the two productions $A \to \alpha$, $A \to \beta$.

For a context-free grammar $G = (N, \Sigma, P, S)$ and $\alpha, \beta \in (N \cup \Sigma)^*$, we say that $\alpha$ derives $\beta$ in one step, denoted by $\alpha \Rightarrow_G \beta$, if there are $\alpha_1, \alpha_2, \gamma \in (N \cup \Sigma)^*$ and $A \in N$ such that $\alpha = \alpha_1 A \alpha_2$, $\beta = \alpha_1 \gamma \alpha_2$ and $(A \to \gamma) \in P$. We say that $\alpha$ derives $\beta$ leftmost in one step if in the definition above $\alpha_1 \in \Sigma^*$.

Let $\Rightarrow_G^+$ denote the transitive closure of $\Rightarrow_G$, and $\Rightarrow_G^*$ denote the reflexive closure of $\Rightarrow_G^+$. For $\alpha, \beta \in (N \cup \Sigma)^*$ and $\pi \in P^*$ we write $\alpha \Rightarrow_G^\pi \beta$ if $\alpha$ derives $\beta$ by the sequence of leftmost steps indicated by $\pi$. We call $\pi$ a *parse* from $\alpha$ to $\beta$ in this case. The language generated by $G$ is defined by $L(G) = \{\, w \in \Sigma^* \mid S \Rightarrow_G^* w \,\}$. If the grammar is clear from the context, the subscript $G$ is omitted. A language $L$ is said to be *context-free* if there is a context-free grammar $G$ with $L = L(G)$. Let $G = (N, \Sigma, P, S)$ be a context-free grammar, and $\alpha \in (N \cup \Sigma)^*$. We say that $\alpha$ is a *sentential form* if $S \Rightarrow^* \alpha$. The grammar $G$ is said to be *cycle-free* if there is no $A \in N$ such that $A \Rightarrow^+ A$.

**Definition 1.** *Let $G = (N, \Sigma, P, S)$ be a context-free grammar, $w \in \Sigma^*$ and $n \in \mathbb{N}$. We define the ambiguity of $w$, and the ambiguity function $am_G : \mathbb{N}_0 \to \mathbb{N}_0$ as follows:*

$$am_G(w) := |\{\pi \in P^* \mid S \Rightarrow_G^\pi w\}|$$
$$am_G(n) := \max\left\{am_G(w) \mid w \in \Sigma^{\leq n}\right\}$$

Note that for a grammar $G$ which contains cycles the set $parse_G(A, \beta)$ may be infinite. But for all cycle-free grammars $G$ the ambiguity function $am_G$ is a total mapping $am_G : \mathbb{N}_0 \to \mathbb{N}_0$. Note that $am_G(w) = 0$ for all $w \notin L(G)$.

**Definition 2.** *Let $f : \mathbb{N}_0 \to \{r \in \mathbb{R} \mid r > 0\}$ be a total function and $L$ a context-free language. We call $L$ inherently $f$-ambiguous if*

*(i) for all context-free grammars $G$ such that $L = L(G)$ we have $am_G = \Omega(f)$, and*

*(ii) there is a grammar $G_0$ such that $L = L(G_0)$ and $am_{G_0} = \mathcal{O}(f)$.*

Note that we have defined inherent complexity *classes* for languages here. Let $L$ be an $f$-ambiguous context-free language such that $f(n) > 1$ for some $n \in \mathbb{N}$. This does not imply that each grammar $G$ with $L(G) = L$ has a word of at most length $n$ with at least $f(n)$ derivation trees. In fact there are grammars for $L$ which generate all words up to length $n$ unambiguously.

## 3 Sublinear Languages

Let $\Sigma = \{0, 1\}$ be an alphabet. We denote $0^{i-1}1$ by $[i]$ for all $i \in \mathbb{N}$. We define the following two languages:

$$\begin{aligned} L_{\log} := \{[i_1] \ldots [i_{2m-1}] \mid\ & m \in \mathbb{N};\ i_1, \ldots, i_{2m-1} \in \mathbb{N};\ \exists 1 \leq k \leq m : \\ & ((\forall \ell < k : 2i_\ell = i_{2m-\ell}) \land (\forall m \geq \ell > k : i_\ell = 2i_{2m+1-\ell}))\} \end{aligned}$$

and analogously

$$\begin{aligned} L_{\sqrt{}} := \{[i_1] \ldots [i_{2m-1}] \mid\ & m \in \mathbb{N};\ i_1, \ldots, i_{2m-1} \in \mathbb{N};\ \exists 1 \leq k \leq m : \\ & ((\forall \ell < k : i_\ell + 1 = i_{2m-\ell}) \land (\forall m \geq \ell > k : i_\ell = i_{2m+1-\ell} + 1))\} \end{aligned}$$

Let $w = [i_1] \ldots [i_{2m-1}] \in L_{\log}$ for some $i_1, \ldots, i_{2m-1} \in \mathbb{N}$ and some $m \in \mathbb{N}$. For $1 \leq k \leq 2m - 1$ we call $[i_k]$ the $k$-th block of $w$. The blocks are pairwise correlated from the borders to the middle. One block $k$ $(1 \leq k \leq m)$ is not forced to have a correlation. When passing this free blockt, the direction of the correlation is reversed. For $L_{\log}$ the quotient of the correlated numbers is 2, for $L_{\sqrt{}}$ their difference is 1. The situation is illustrated in the following diagram. Arrows indicate correlations forced by the definition of the language:

$$[i_1] \ldots [i_{k-1}]\, \underbrace{[i_k]}_{free}\, [i_{k+1}] \ldots [i_m][i_{m+1}] \ldots [i_{2m-k}][i_{2m-k+1}] \ldots [i_{2m-1}]$$

The languages $L_{\log}$ and $L_{\sqrt{}}$ are generated by the subsequently defined grammars $G_{\log}$ and $G_{\sqrt{}}$, respectively. For $sub \in \{\log, \sqrt{}\}$ we define

$$G_{sub} := (\{A, B, C, D, S\}, \{0, 1\}, P_{sub}, S) \text{ as follows:}$$

$$
P_{\log} := \{ \begin{array}{llll} S \to 1S01 & \mid 0A01 & \mid B & \\ A \to 0A00 & \mid 1S00 & & \\ B \to 0B & \mid 1C & \mid 1 & \\ C \to 01C1 & \mid 00D1 & \mid 011 & \\ D \to 00D0 & \mid 01C0 & \mid 010 & \} \end{array}
\qquad \Bigg| \qquad
P_{\surd} := \{ \begin{array}{llll} S \to 1S01 & \mid 0A1 & \mid B & \\ A \to 0A0 & \mid 1S00 & & \\ B \to 0B & \mid 1C & \mid 1 & \\ C \to 01C1 & \mid 0D1 & \mid 011 & \\ D \to 0D0 & \mid 01C0 & \mid 010 & \} \end{array}
$$

It is easily verified that these grammars generate the languages defined above. The derivation starts with a (possibly empty) finite number of cycles in the nonterminals $S$ and $A$ which produces the blocks to the left of the free block and the corresponding blocks at the right end of the word. Eventually the production $S \to B$ is applied. The nonterminal $B$ generates the free block. Finally either the derivation is terminated with the production $B \to 1$, or with $B \to 1C$ we begin to produce blocks with the opposite correlation to the right of the free block, by using the nonterminals $C$ and $D$.

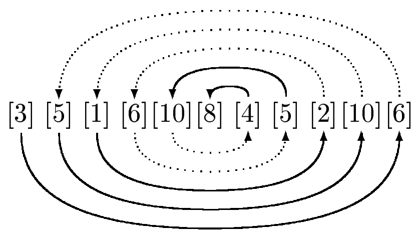
### 3.1 Sublinear Ambiguity of the Presented Grammars

In this section we prove that $am_{G_{\log}} \in \mathcal{O}(\log\ n)$. Analogously $am_{G_{\surd}} \in \mathcal{O}(\sqrt{n})$ can be shown, which however will not be done here.

**Definition 3.** *Let $i_1, \ldots, i_{2m-1} \in \mathbb{N}$ for some $m \in \mathbb{N}$, $w = [i_1] \ldots [i_{2m-1}]$, and $1 \leq \ \ell \leq \ m$.*

- *The word $w$ has a* forward correlation *at block $\ell$ if and only if $\ell < m$ and $2i_\ell = i_{2m-\ell}$. It has a* forward crack *if and only if $\ell < m$ and $2i_\ell \neq i_{2m-\ell}$.*
- *The word $w$ has a* backward correlation *at block $\ell$ if and only if $1 < \ell \leq \ m$ and $i_\ell = 2i_{2m+1-\ell}$. It has a* backward crack *if and only if $1 < \ell \leq \ m$ and $i_\ell \neq 2i_{2m+1-\ell}$.*
- *Block $\ell$ is* isolated *in $w$ if and only if block $\ell$ has neither a forward nor a backward correlation.*

*Example 1.* We illustrate these definitions by the following diagram. The relevant relations between blocks are on a spiral from the leftmost block to the block in the middle, indicated by solid and dotted arrows.




The blocks 1, 2, and 3 have a forward correlation. Block 4 and 5 have a forward crack. Forward correlations and cracks are indicated by solid and dotted arrows from left to right, respectively. Blocks 2, 3, and 4 have a backward crack, blocks 5 and 6 have a backward correlation, again indicated by dotted and solid arrows, this time from right to left. Block 4 is isolated, since it has neither a forward nor a backward correlation.

**Definition 4.** *Let* $m, r \in \mathbb{N}$, $i_1, \ldots, i_{2m-1} \in \mathbb{N}$, *and* $w = [i_1] \ldots [i_{2m-1}]$.

- $(r * w) := [ri_1] \ldots [ri_{2m-1}]$
- $z_m = \begin{cases} [1] & \textit{for } m = 1 \\ [1](4 * z_{m-1})[2] & \textit{for } m > 1 \end{cases}$

  *For example* $z_4 = [1][4][16][64][32][8][2]$.
- $L_{even,r} := \{(r * z_m) \mid m \in \mathbb{N}\}$
- $L_{even} := \bigcup_{r \in \mathbb{N}} L_{even,r}$
- $L_{min} := L_{even,1}$.

Note that $L_{min} = \{z_m \mid m \in \mathbb{N}\}$, and that a word is in $L_{even}$ if and only if it has no cracks.

For a word with an isolated block we know that this block has to be derived by the nonterminal $B$ and therefore the derivation of the whole word is completely determined. In general cracks provide information about the position of the free block. But the language definition does not require the existence of cracks. Hence $L_{even} \subseteq L_{\log}$. For a word $w \in L_{even}$ any block up to the one in the middle can be produced by nonterminal $B$. For example in the word $z_3 = [1][4][16][8][2]$ either [1], [4], or [16] is the free block. This gives ambiguity 3. Hence, for each $m, r \in \mathbb{N}$, the word $(r * z_m)$ has $m$ derivations. Moreover we will prove that $z_m$ is the shortest word in $L_{\log}$ with $m$ derivations, which inspired the name $L_{min}$. Due to the free block the forward and backward correlations are interlocked. Therefore in a word without cracks the length of the blocks is strictly increasing along the spiral, while the ambiguity is proportional to the number of blocks. Thus the ambiguity is sublinear.

**Lemma 1.** *Let* $w \in L_{\log}$ *and* $w \notin L_{min}$. *Then there is a word* $w' \in \Sigma^*$ *with* $|w'| < |w|$ *and* $am_{G_{\log}}(w') = am_{G_{\log}}(w)$.

*Proof.* We distinguish three cases.

**Case 1:** $w \in L_{even}$.
For some $m, r \in \mathbb{N}$ we have $w = (r * z_m)$. Since $w \notin L_{min}$ we have $r > 1$. Thus we obtain $|z_m| < r|z_m| = |(r * z_m)| = |w|$. Moreover $am_{G_{\log}}(z_m) = m = am_{G_{\log}}((r * z_m)) = am_{G_{\log}}(w)$.

**Case 2:** $w$ has a block $\ell$ with a forward crack.
For some $m \in \mathbb{N}$ we have $|w|_1 = 2m-1$, which is the number of blocks in $w$. Since block $\ell$ has a forward crack, by definition $\ell < m$. Moreover block $\ell$ cannot be generated by the nonterminals $S$ and $A$. Therefore block $\ell$ is either produced by nonterminal $B$ or by the nonterminals $C$ and $D$. In both cases blocks $\ell + 1$ up to block $m$ are generated by $C$ and $D$. Since $\ell < m$ there is at least one such block. But then the derivation after generating block $\ell$ is completely determined by the blocks $\ell + 1$ up to block $m$. That is, by erasing these and their correlated blocks

we obtain a word $w'$ which consists of $2\ell - 1$ blocks from $w$, and which has the same ambiguity as $w$. Hence we obtain $|w'| < |w|$ and $am_{G_{\log}}(w') = am_{G_{\log}}(w)$.

**Case 3:** $w$ has a block $\ell$ with a backward crack.
For some $m \in \mathbb{N}$ we have $|w|_1 = 2m - 1$. Since block $\ell$ has a backward crack, by definition $\ell > 1$. Moreover block $\ell$ cannot be generated by the nonterminals $C$ and $D$. Therefore block $\ell$ is either produced by nonterminal $B$ or by the nonterminals $S$ and $A$. In both cases blocks 1 up to block $\ell - 1$ are generated by $S$ and $A$. Since $\ell > 1$, there is at least one such block. But then the derivation until generating block $\ell$ is completely determined by the blocks 1 up to block $\ell - 1$. That is, by erasing these and their correlated blocks we obtain a word $w'$ which consists of $2(m - \ell) + 1$ blocks from $|w|$ and which has the same ambiguity as $w$. Hence we obtain $|w'| < |w|$ and $am_{G_{\log}}(w') = am_{G_{\log}}(w)$.

**Theorem 1.** $\forall j \in \mathbb{N}\ \forall w \in L_{\log}: \quad |w| < |z_j|$ *implies* $am_{G_{\log}}(w) < j$

*Proof.* Let $w$ be a shortest word such that $am_{G_{\log}}(w) >= j$. Since $am_{G_{\log}}(z_j) = j$ we observe that $|w| \le |z_j|$. Moreover Lemma 1 implies that $w$ is in $L_{min}$ and hence $w = z_i$ for some $i \in \mathbb{N}$. Since $am_{G_{\log}}(z_i) = i$ we get $i \ge j$. Now $|z_i| = |w| \le |z_j|$ implies $i \le j$. Thus we obtain $i = j$, that is $w = z_j$ which proves Theorem 1.

By Theorem 1 we obtain the following table

| ambiguity | shortest word | length |
|---|---|---|
| 1 | $z_1 = [1]$ | 1 |
| 2 | $z_2 = [1][4][2]$ | 7 |
| 3 | $z_3 = [1][4][16][8][2]$ | 31 |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $i$ | $\ldots$ | $\frac{1}{2}4^i - 1$ |

If we proceed analogously for $L_{\sqrt{}}$ we obtain

**Theorem 2.**

$$am_{G_{\log}}(n) = \lfloor \log_4(2n+2) \rfloor = \mathcal{O}(\log\ n)$$

$$am_{G_{\sqrt{}}}(n) = \left\lfloor \frac{1}{4} + \sqrt{\frac{1}{2}n + \frac{1}{16}} \right\rfloor = \mathcal{O}(\sqrt{n})$$

### 3.2 Inherence of the Sublinearity

In this section we will prove that the language $L_{\log}$ has inherent logarithmic ambiguity. We already proved that logarithmic ambiguity is sufficient to generate the language. Thus we have to prove that less than logarithmic ambiguity does not suffice. First we prove a technical lemma.

**Lemma 2.** *Let $w = [i_1] \dots [i_{2m-1}]$ for some $m \in \mathbb{N}$ and $i_1, \dots, i_{2m-1} \in \mathbb{N}$, and let $1 \leq n \leq \frac{1}{3}(m-1)$. Then*

$$i_{m-3n} = i_{m+3n} \text{ and } i_m = i_{m+2n} \text{ and } i_{m+1} = i_{m+1-2n} \text{ implies } w \notin L_{\log}.$$

*Proof.* By definition $w$ has a forward crack at block $m - 3n$. Now assume $w \in L_{\log}$. Then all blocks numbered $m - 3n + 1$ up to $m$ must have a backward correlation. In particular $i_{m+1-2n} = 2i_{m+2n}$ and $i_m = 2i_{m+1}$. But then $i_m = 2i_{m+1} = 2i_{m+1-2n} = 4i_{m+2n} = 4i_m$ is a contradiction.

The lemma above is important because it tells us that in a word of $L_{\log}$ a sequence consisting of $2n$ blocks cannot be repeated too often in the vicinity of the middle block.

**Theorem 3.** *$L_{\log}$ has inherent logarithmic ambiguity.*

*Proof.* Let $G = (N, \Sigma, P, S)$ be an arbitrary context-free grammar such that $L(G) = L_{\log}$. We will apply Ogden's iteration lemma for context-free grammars (see [1, Lemma 2.5]). Let $p$ be the constant of Ogden's iteration lemma for $G$. We define $s := p + 1$ and $r := s! + s$. For each $m \in \mathbb{N}$, and $1 \leq n \leq 2m - 1$, we define $i_{m,n}$ such that $[i_{m,n}]$ is the $n$-th block of $z_m$. Let

$$S_m := \{[ri_{m,1}] \dots [ri_{m,\ell-1}][si_{m,\ell}][ri_{m,\ell+1}] \dots [ri_{m,2m-1}] \mid 1 \leq \ell \leq m\} \subseteq L_{\log}.$$

Now for some $m \in \mathbb{N}$ and $1 \leq \ell \leq m$ we consider the word

$$z := [ri_{m,1}] \dots [ri_{m,\ell-1}][si_{m,\ell}][ri_{m,\ell+1}] \dots [ri_{m,2m-1}] \in S_m.$$

Corresponding to Ogden's Lemma we mark all the 0's in the $\ell$-th block. Then we can write $z = uvwxy$ such that for a nonterminal $A$ we have $S \Rightarrow_G^* uAv$, $A \Rightarrow_G^* vAx$ and $A \Rightarrow_G^* w$. By the iteration theorem $v$ or $x$ lie completely inside the 0's of block $\ell$. Assume $v$ lies completely in the 0's of block $\ell$ and $|x|_1 > 0$. Now $|x|_1$ is even, because otherwise by pumping only once we would obtain a word with an even number of blocks, which is impossible by the definition of the language. But then after pumping up $m + 3$ times we obtain a word which has enough repeated occurrences of a sequence of $2n$ blocks for some $n \in \mathbb{N}$, such that the condition of Lemma 2 is satisfied. Thus $x$ cannot contain 1's in this case. The case that $x$ lies completely in block $\ell$ and $|v|_1 > 0$ is treated analogously. Hence $x$ and $v$ cannot contain 1's. Thus both $x$ and $v$ lie completely in the 0's of one block, respectively. Assume $x$ and $v$ do not lie in the same block and $x \neq \varepsilon$ and $v \neq \varepsilon$. That is, block $\ell$ can be pumped together with a block $\ell'$. Assume $\ell' \leq m$ then after one pumping step we obtain a word with two isolated blocks, which is a contradiction. Assume $\ell' > m$ then after one pumping step we obtain a word with a forward crack in block $2m - \ell'$ and a backward crack in block $2m - \ell' + 1$ again a contradiction. Note that in both blocks the correlation is either destroyed if it held before, or its partner is block $\ell$ and then due to the choice of $s$ and $r$ the crack is not repaired by one pumping step. Hence $x$ and $v$ either both lie inside block $\ell$ or the one which doesn't is the empty word.

Thus only block $\ell$ is pumped up. And by repeated pumping we can repair the cracks in block $\ell$ and obtain $(r * z_m)$. That is, all the words in $S_m$ can be pumped up to yield $(r * z_m)$. Now assume that among the derivation trees obtained by this method there are two which are equal. Then we can pump two different blocks $1 \le \ell_1, \ell_2 \le m$ independently. Thus by pumping once in both blocks we obtain a word with two isolated blocks, which is a contradiction.

Finally we have proved that $(r * z_m)$ has at least $m$ derivation trees. Now the length of $(r * z_m)$ increases exponentially with respect to $m$. Hence the ambiguity is logarithmic with respect to the length of the word.

The proof that $L_{\sqrt{}}$ is inherently square-root ambiguous is analogous.

## 4 Conclusion

Here we have presented first examples of linear context-free languages with nonconstant sublinear ambiguity. By concatenation we can get some other sublinear ambiguities. Is it possible to find nonconstant sublogarithmic ambiguity? Can we characterize the possible complexity classes? These questions are deeply connected with the structure of the intersection of context-free languages. To see this we consider the languages $L_1 := \{1^i 0^{2i} \mid i \in \mathbb{N}\}$ and $L_2 := \{0^i 1^{2i} \mid i \in \mathbb{N}\}$. Now we define the unambiguous languages $L_1' := 0L_1^*$ and $L_2' := L_2^* 0^*$. The language $L_1' \cap L_2'$ contains only $\mathcal{O}(\log\ n)$ words with a length up to $n$. Of course $L_1' \cup L_2'$ has the degree of ambiguity 2, but ambiguity is “needed” only logarithmic many times. The languages $L_1'$ and $L_2'$ are slightly modified versions of languages found in [3]. The main question was how sublinear “density” of the intersection can be transformed into an inherent degree of ambiguity. The idea was to concatenate $L_1^*$ and $L_2^*$ buffered with a free block to interconnect the correlations and hide the factorization. This led to the (non-linear) language $L_1^*\, 1^+\, L_2^*$ which is a context-free language with inherent logarithmic ambiguity.

Recall that intersections of context-free languages can have a very complicated structure. If we denote the set of computations of a Turing machine $M$ by sequences of configurations, where every second configuration is written in reverse, then we obtain the set of valid computations. In [2, Lemma 8.6] it is shown that this set is the intersection of two linear languages.

Thus if our method of transforming the “density” of an intersection into an inherent degree of ambiguity can be generalized, we can hope for a variety of sublinear ambiguities.

*Acknowledgements* Thanks to Prof. Dr. Friedrich Otto, Dr. Dieter Hofbauer, and Gundula Niemann for proofreading, valuable discussions and LaTeX tips.

## References


1. J. Berstel. *Transductions and Context-Free Languages.* Teubner, 1979.
2. J.E. Hopcroft, J.D. Ullman. *Introduction to Automata Theory, Formal Languages, and Computation.* Addison-Wesley, 1979.
3. R. Kemp. A Note on the Density of Inherently Ambiguous Context-free Languages. *Acta Informatica* 14, pp. 295–298, 1980.
4. M. Naji. *Grad der Mehrdeutigkeit kontextfreier Grammatiken und Sprachen.* Diplomarbeit, FB Informatik, Johann–Wolfgang–Goethe–Universität, Frankfurt am Main, 1998.
5. K. Wich. *Kriterien für die Mehrdeutigkeit kontextfreier Grammatiken.* Diplomarbeit, FB Informatik, Johann–Wolfgang–Goethe–Universität, Frankfurt am Main, 1997.
6. K. Wich. Exponential Ambiguity of Context-free Grammars. *Proc. 4th Int. Conf. on Developments in Language Theory '99*, World Scientific, Singapore, to appear.