ISSN 0105-8517

Circularity Testing of Attribute Grammars Requires Exponential Time: A Simpler Proof

by
Neil D. Jones

DAIMI PB-107
January 1980

Computer Science Department
AARHUS UNIVERSITY
Ny Munkegade – DK 8000 Aarhus C – DENMARK
Telephone: 06 – 12 83 55

Circularity Testing of Attribute Grammars Requires Exponential Time: A Simpler Proof

Neil D. Jones
Computer Science Department
University of Aarhus
8000 Aarhus C, Denmark

Keywords: attribute grammars, circularity testing, time complexity.

In [JOR75] it was shown that the high time complexity of Knuth's algorithm for testing attribute grammars for circularity ([Knu71]) is no accident. It was proved that there is a constant $c > 0$ such that any deterministic Turing Machine which correctly tests for circularity must run for more than $2^{cn/\log n}$ steps on infinitely many attribute grammars (AGs) (the size of an AG is the number of symbols required to write it down). The proof was rather complex; the purpose of this note is to provide a simpler one.

<u>Construction</u> Let Z be an arbitrary one tape deterministic Turing machine such that

i) Z halts in at most 2^n steps on any input of length n.

Greek alpha, omega

ii) Z has start state α, accepting state ω, and accepts by entering a one-state loop at .

iii) In the first move Z writes # (blank) at position 1 of its tape, and thereafter never changes this symbol.

Given any input $x = a_1 \ldots a_n$ of length n, an accepting computation by Z on x can be described by a matrix M as in Figure 1. Row i of M is Z's tape contents at time t.

[Figure 1]

Now one of three cases applies to each matrix entry M_{it}:

a) $i = 1$, $t > 1$ and $M_{it} = \#$ or $\overset{\omega}{\#}$

b) $t = 1$ and M_{it} is the i-th symbol of $\overset{\alpha}{a}_1 a_2 \ldots a_n \# \ldots \#$

c) $i > 1$, $t > 1$ and $M_{it} = f(M_{i-1,\, t-1}, M_{i,\, t-1}, M_{i+1,\, t-1})$
where f is a function depending only on Z (since Z is deterministic).

We first construct an AG G_x such that

i) $L(G_x) \neq \emptyset$ iff Z accepts x

ii) G_x has only inherited attributes:

$i, t \in \{1, \ldots, 2^n\}$

$s \in \{y \mid y \text{ a suffix of } \overset{\alpha}{a}_1 a_2 \ldots a_n\}$

iii) the number of productions of G_x is independent of x

iv) certain constraints on the attribute values must be satisfied in order to apply some productions.

The notation is borrowed from [Wat77]; for example

$A \downarrow x \uparrow y{*}y \rightarrow A \downarrow x \uparrow z \;\; B \downarrow z{+}1 \uparrow y$ indicates that A has one inherited attribute x and one synthesized attribute z, that B has one inherited attribute w and one synthesized attribute y, and that corresponding to production $A_1 \rightarrow A_2 B$ we have attribute equations:

$$z(A_1) = y(B) * y(B),\quad z(A_2) = z(A_1) \text{ and } w(B) = z(A_2) + 1.$$

G_x has nonterminals S and M^A, N^A where A is a tape symbol of Z. Its productions are, for each A, B, C, D:

Greek epsilon

I	$S \to N^{\overset{\omega}{\#}} \downarrow 1 \downarrow 2^n \downarrow \overset{\alpha}{a}_1 a_2 \ldots a_n$	
II	$N^{\#} \downarrow i \downarrow t \downarrow s \to \epsilon$	if $i = 1$ and $t > 1$
III	$N^A \downarrow i \downarrow t \downarrow s \to M^A \downarrow i \downarrow t \downarrow s$	if $t = 1$
	$M^A \downarrow i \downarrow t \downarrow s \to \epsilon$	if $i = 1$ and $s = Au$ for some u
	$M^A \downarrow i \downarrow t \downarrow s \to M^A \downarrow i-1 \downarrow t \downarrow u$	if $i > 1$ and $s = au$ for some a
	$M^{\#} \downarrow i \downarrow t \downarrow s \to \epsilon$	if $i > 1$ and $s = \epsilon$
IV	$N^{f(B,C,D)} \downarrow i \downarrow t \downarrow s \to N^B \downarrow i-1 \downarrow t-1 \downarrow s \; N^C \downarrow i \downarrow t-1 \downarrow s \; N^D \downarrow i+1 \downarrow t-1 \downarrow s$ if $i > 1$, $t > 1$	

The effect of group III is that $N^A \downarrow i \downarrow 1 \downarrow s \Rightarrow^* \epsilon$ iff A is the i-th symbol of $\overset{\alpha}{a}_1 a_2 \ldots a_n \# \ldots \#$ (note that $s = \overset{\alpha}{a}_1 a_2 \ldots a_n$).

<u>Lemma 1</u> $L(G_x) \neq \emptyset$ iff Z accepts x.

<u>Proof</u> If Z accepts x, let M be the matrix corresponding to its computation on x. A derivation of the empty string is easily constructed.

If $L(G_x) \neq \emptyset$, let T be any attributed derivation tree which satisfies the constraints. Associate with node labeled $N^A \downarrow i \downarrow t \downarrow s$ the assertion "$M_{i,t} = A$". Note that a subtree of T has root $N^A \downarrow i \downarrow 1 \downarrow s$ iff A is the i'th symbol of $\overset{\alpha}{a}_1 a_2 \ldots a_n \# \ldots \#$. Thus the assertion holds for the lowest N^A nodes. Clearly if it holds for the N^B, N^C and N^D nodes of production IV, then it also holds for its left side. Thus it holds for $N^{\overset{\omega}{\#}}$ in production I. Thus Z enters state ω and so accepts x.

□ Lemma

Greek sigma

Greek pi

<u>Lemma 2</u> There is an AG G'_x such that

i) $L(G'_x) \neq \emptyset$ iff Z accepts x

ii) G'_x has only one inherited attribute y ranging over $\Sigma^{h|x|}$ for some $h > 0$ and some alphabet Σ

iii) the number of productions of G'_x is independent of x

iv) each production of G'_x has one of these forms:

I	$A \downarrow y \rightarrow \epsilon$	
II	$A \downarrow y \rightarrow B \downarrow y \; C \downarrow y$	
III	$A \downarrow za \rightarrow B \downarrow bz$	where $a, b \in \Sigma$. Note that this is a constraint on the A attribute.
IV	$S \rightarrow A \downarrow y_0$	where $y_0 \in \Sigma^{h|x|}$ is a constant string

<u>Proof</u> First, represent attributed nonterminal $A \downarrow i \downarrow t \downarrow s$ of G_x by the string $<\bar{i}, \bar{t}, s>$ where $\bar{i}, \bar{t}$ are the binary representations of i and t, and let $\Sigma = \{<, >, ",", 0, 1\} \cup \{A \mid A \text{ is a tape symbol of } Z\}$. By padding with zeroes we can ensure that $|<\bar{i}, \bar{t}, s>| = h|x|$ for some h and all x. Note that the right side attribute representation of each production π of G_x may be computed from $<\bar{i}, \bar{t}, s>$ by a finite state sequential machine M_π, which reads $<\bar{i}, \bar{t}, s>$ from right to left. By adding a few nonterminals it is thus eady to construct an AG equivalent to G_x which has only productions of the forms:

S	$\rightarrow A \downarrow <\bar{1}, \overline{2^n}, \overset{\alpha}{a}_1 a_2 \ldots a_n>$	
$A \downarrow y$	$\rightarrow B \downarrow y \; C \downarrow y$	
$A \downarrow y$	$\rightarrow \epsilon$	
$A \downarrow y$	$\rightarrow B \downarrow M(y)$	where M is a finite state sequential machine which accepts y, and M(y) is its output.

Let M be such a sequential machine with state set Q, start and accepting states $\alpha_m, \omega_m \in Q$ and transition function $\delta: Q \times \Sigma \to Q \times \Sigma$. Without loss of generality M enters its accepting state only after reading "<".

Now G'_x will have a nonterminal A^q for each $q \in Q$. The production $A\downarrow y \to B\downarrow M(y)$ may now be replaced by:

$A\downarrow y \to A^{\alpha_m}\downarrow y$

$A^p\downarrow za \to A^q\downarrow bz$ if $y = za$ and $\delta(p,a) = (q,b)$

$A^{\omega_m}\downarrow y \to B\downarrow y$

□ Lemma

Lemma 3 There is an AG G''_x such that

i) G''_x is circular iff Z accepts x

ii) $\text{size}(G''_x) = 0(|x| \log |x|)$

Proof will be given after the following:

Theorem There is a constant $c > 0$ such that any deterministic Turing Machine which decides whether an AG is circular must run for more than $2^{cn/\log n}$ steps on infinitely many AGs.

Proof of Theorem

Let Z be a Turing Machine as in the first construction. Suppose the theorem is false, so circularity can be decided in time $2^{cn/\log n}$ for all $c > 0$. Then the test "is x accepted by Z" could be done indirectly as follows:

1. Construct G''_x

2. Answer "yes" iff G''_x is circular.

Step 1 can clearly be done in polynomial time $p(|x|)$. By Lemma 3 $\text{size}(G''_x) \leq d|x| \log|x|$ for some $d > 0$ independent of x. Letting $n = |x|$, we see that steps 1 and 2 can be done in time $p(n) + 2^{h(n)}$, where

$$h(n) = \frac{c\, d\, n \log n}{\log(dn \log n)} = \frac{c\, d\, n \log n}{\log n + \log d + \log \log n} \leq c\, d\, n$$

Since $c > 0$ is arbitrary, $L(Z)$ can be accepted in time $p(n) + 2^{n/2}$. This is impossible for all Z, since by Theorem 12.9 of [Hop 79] there is a language accepted in time 2^n but not in time $p(n) + 2^{n/2}$.

<u>Proof</u> of Lemma 3

Assume for simplicity of notation that $\Sigma = \{0, 1\}$ (extension to general Σ is straightforward). G''_x will have the same nonterminals as G'_x. Let $m = h|x|$. Attribute y of nonterminal A in G'_x will be replaced by the $4\,m$ attributes $d(i,a)$ and $u(i,a)$ for $a \in \Sigma$, $1 \leq i \leq m$:

$$A{\downarrow}\, d(1,0)\; d(1,1) \ldots d(m,0)\; d(m,1) \uparrow u(m,1)\; u(m,0) \ldots u(1,1)\; u(1,0)$$

The only attribute-defining expressions will be variable names or the constant 17, so attributes may only be copied or set to a constant value. G''_x has the following productions:

I If G'_x contains: $A{\downarrow}\, y \rightarrow \varepsilon$, then G''_x contains

$$A{\downarrow}\, d(1,0)\; d(1,1) \ldots d(m,0)\; d(m,1) \uparrow d(m,1)\; d(m,0) \ldots d(1,1) d(1,0) \rightarrow \varepsilon$$

[Figure 2]

II If G'_x contains: $A\downarrow y \rightarrow B\downarrow y\ C\downarrow y$, then G''_x contains

$$A\downarrow d(1,0)\ldots d(m,1)\uparrow u(m,1)\ldots u(1,0) \rightarrow$$
$$B\downarrow d(1,0)\ldots d(m,1)\uparrow e(m,1)\ldots e(u,1)$$
$$C\downarrow e(1,0)\ldots e(m,1)\uparrow u(m,1)\ldots u(1,0)$$

[Figure 3]

III If G'_x contains: $A\downarrow za \rightarrow B\downarrow bz$, then G''_x contains

$$A\downarrow d(1,0)\ d(1,1)\ldots d(m,0)\ d(m,1)\uparrow u(m,1)\ u(m,0)\ldots u(1,1)\ u(1,0) \rightarrow$$
$$B\downarrow e(1,0)\ e(1,1)\ d(1,0)\ldots d(m-1,0)\ d(m-1,1)$$
$$\uparrow u(m-1,1)\ u(m-1,0)\ldots u(1,1)\ u(1,0)\ f(1,1)\ f(1,0)$$

where

$$e(1,0)\ e(1,1) = \begin{cases} d(m,a)\ 17 & \text{if } b = 0 \\ 17\ d(m,a) & \text{if } b = 1 \end{cases}$$

$$u(m,1)\ u(m,0) = \begin{cases} 17\ f(1,b) & \text{if } a = 0 \\ f(1,b)\ 17 & \text{if } a = 1 \end{cases}$$

[Figure 4]

IV If G'_x contains: $S \rightarrow A\downarrow a_1 a_2 \ldots a_m$ then G''_x contains:

$$S \rightarrow A\downarrow e(1,0)\ e(1,1)\ldots e(m,0)\ e(m,1)\uparrow u(m,1)\ u(m,0)\ldots u(1,1)\ u(1,0)$$

where

$$e(1,0)\ e(1,1) = \begin{cases} u(m,a_m)\ 17 & \text{if } a_1 = 0 \\ 17\ u(m,a_m) & \text{if } a_1 = 1 \end{cases}$$

and for $i = 2,\ldots,m$

$$e(i,0)\ e(i,1) = \begin{cases} u(i-1,a_{i-1})\ 17 & \text{if } a_i = 0 \\ 17\ u(i-1,a_{i-1}) & \text{if } a_i = 1 \end{cases}$$

[Figure 5]

Note that this last diagram has a cycle containing the dotted lines and the dependencies of the S production, and that no cycle exists if any dotted line is removed. With this in mind we consider the following:

<u>Property</u> Let T' be an attributed parse tree of G'_x with root $A\downarrow a_1 \ldots a_m$, and let T'' be the corresponding parse tree of G''_x. Then T' satisfies the constraints imposed by its type III productions iff for each $i = 1, \ldots, m$ $u(i, a_i)$ is dependent on $d(i, a_i)$.

This is easily verified by induction on the height of T'. It is trivially true for production type I, and follows immediately from the inductive assumption for type II. For type III, consider for example a tree node $A\downarrow a_1 \ldots a_m$ corresponding to the production $A\downarrow z0 \rightarrow B\downarrow 1z$. If $a_m = 1$, then $u(m, a_m) = u(m, 1) = 17$ which is independent of $d(m, a_m)$. If $a_m = 0$ then $u(m, 0) = f(1, b)$, $f(1, b)$ depends on $e(1, b) = e(1, 1)$ by the inductive assumption, and $e(1, 1) = d(m, 0)$. Thus $u(m, 0)$ is dependent on $d(m, 0)$ iff y ends in 0. The other attribute dependencies are trivial.

Consequently G''_x is circular iff G'_x has a parse tree which satisfies all the type IV constraints.

Finally, note that the number of productions of G''_x is independent of x. Each production contains $4m = O(|x|)$ attributes, so the production can be written down in $O(|x| \log |x|)$ symbols (the $\log |x|$ factor comes from the need to name each of the attributes). Thus size $(G''_x) = O(|x| \log |x|)$.

□ Lemma 3

References

[Hop 79] Hopcroft, J. and J. Ullman, "Introduction to Automata Theory, Languages and Computation", Addison-Wesley, Reading, Mass. 1979.

[JOR 75] Jazayeri, M., W.F. Ogden and W.C. Rounds, The Intrinsically exponential complexity of the circularity problem for attribute grammars, Comm. ACM 18, 697-706 (1975).

[Knu 68] Knuth, D.E., Semantics of context-free languages, Mathematical Systems Theory 2, 127-145 (1968).

[Knu 71] Knuth, D.E., Semantics of context-free languages: Correction, Math. Syst. Thy. 5, 95-96 (1971).

[Wat 77] Watt, D.A. The parsing problem for affix grammars, Acta Informatica 8, 1-20 (1977).

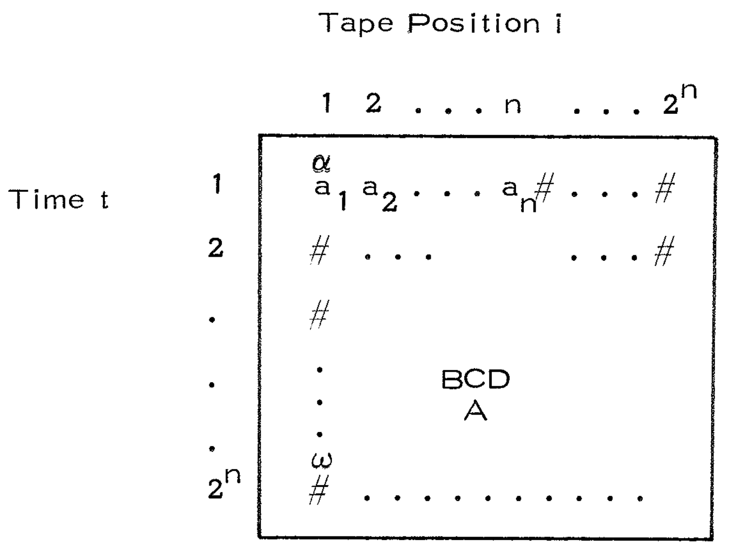

Figure 1. An Accepting Computation

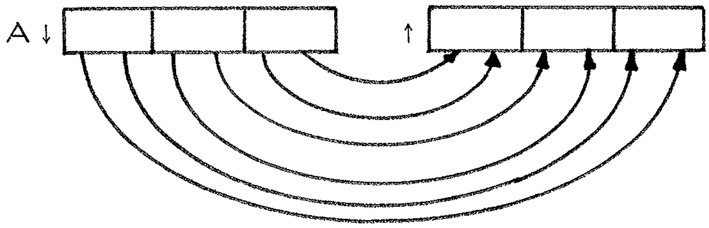

Figure 2. Dependencies for $A\downarrow y \rightarrow \epsilon$, assuming $m = 3$

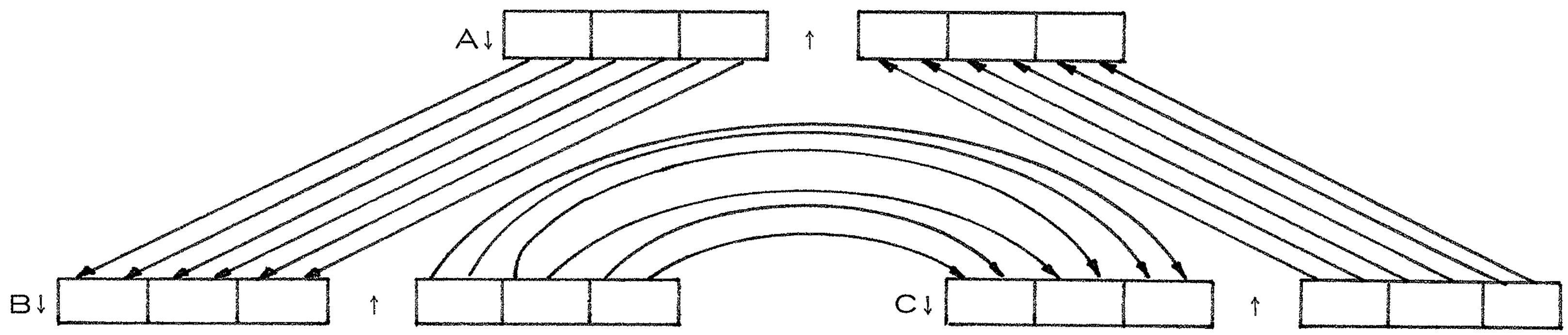

Figure 3. Dependencies for A↓ y → B↓ y C↓ y, assuming m = 3

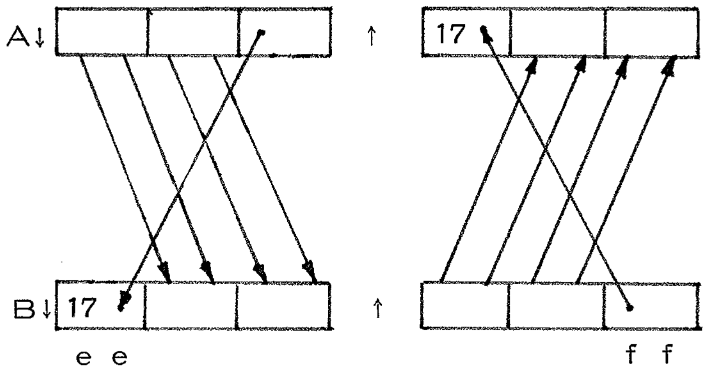

Figure 4. Dependencies for A↓ z1 → B↓ 1z, assuming m = 3

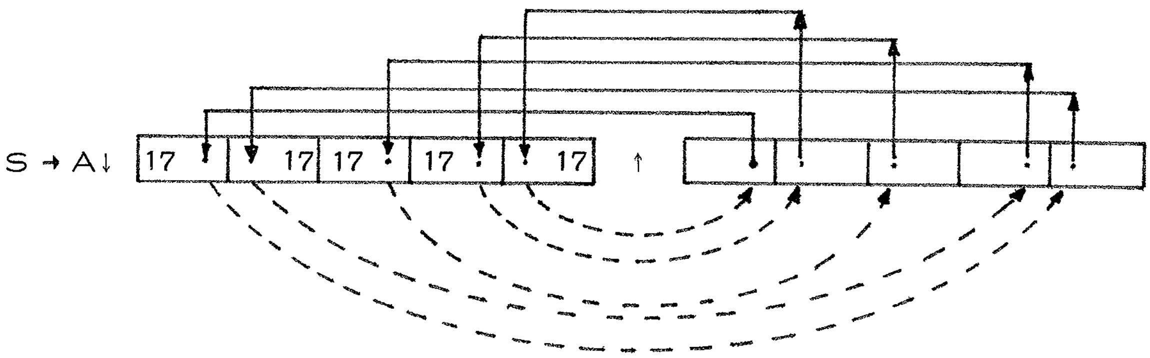

Figure 5. Dependencies for S → A↓ 10110