The Failure of Ogden’s Lemma for Well-Nested Multiple Context-Free Grammars

Makoto Kanazawa

August 29, 2014

Let us say that a language $L$ has the weak Ogden property if there are natural numbers $k$ and $p$ such that for every $z \in L$ and $S \subseteq [1, |z|]$ with $|S| \geq p$, there are strings $u_1, \ldots, u_{k+1}, v_1, \ldots, v_k$ satisfying the following conditions:

1. $z = u_1 v_1 \ldots u_k v_k u_{k+1}$,
2. $S \cap [u_1 v_1 \ldots u_{i-1} v_{i-1} u_i] + 1, |u_1 v_1 \ldots u_i v_i| \neq \emptyset$ for some $i \in [1, k]$, and
3. $u_1 v_1^n \ldots u_k v_k^n u_{k+1} \in L$ for all $n \geq 0$.

The following example is inspired by Lemma 5.4 of Greibach 1978, “The strong independence of substitution and homomorphic replication”, but is much simpler. (One can write a non-branching 8-MCFG for Greibach’s example.) My “weak Ogden property” is also weaker than Greibach’s notion of strong iterativity.

**Theorem 1.** There is an $L \in 3$-MCFL(1) that does not satisfy the weak Ogden property.

**Proof.** Consider the following non-branching 3-MCFG:

\[
A(\varepsilon) \leftarrow \\
A(b x_1 c) \leftarrow A(x_1) \\
B(x_1, \varepsilon) \leftarrow A(x_1) \\
B(a x_1 d, b x_2 c) \leftarrow B(x_1, x_2) \\
C(x_1, x_2, \varepsilon) \leftarrow B(x_1, x_2) \\
C(x_1, a x_2 d, b x_3 c) \leftarrow C(x_1, x_2, x_3) \\
C(x_1, x_2, x_3, \varepsilon) \leftarrow C(x_1, x_2, x_3) \\
D(x_1, a x_2 d) \leftarrow D(x_1, x_2) \\
S(x_1, x_2) \leftarrow D(x_1, x_2)
\]
The language $L$ of this grammar consists of all and only strings of the form
\[ a^{i_1}b^{i_2}c^{i_3}d^{i_4}a^{i_5}b^{i_6}c^{i_7}d^{i_8}\ldots a^{i_n}b^{i_{n-1}}c^{i_n}d^{i_n}, \]
where $n \geq 2$ and $i_0, \ldots, i_n \geq 0$.

Now suppose $L$ has the weak Ogden property, and let $k$ and $p$ be the numbers satisfying the required conditions. Let
\[ z = ad^jbcd^kac^{j+2}d^{j+1}a^{j+2}b^{j+2}c^{j+2}d^{j+2}, \]
and let $S$ consist of the positions in $z$ occupied by $\$$. Note that $|S| = p$. By the weak Ogden property, there must be strings $u_1, \ldots, u_k, v_1, \ldots, v_k$ such that $z = \ldots = u_kv_ku_{k+1}, u_1v_1^\ldots u_kv_k^2u_{k+1} \in L$ for all $n$, and $v_i$ contains at least one occurrence of $\$$. Let us write $z(n)$ for $u_1v_1^\ldots u_kv_k^2u_{k+1}$. We consider two cases, depending on the number of occurrences of $\$$ in $v_i$. Each case leads to a contradiction.

**Case 1.** $v_i$ contains just one occurrence of $\$$. Then $v_i = x\$$y$, where $x$ is a suffix of $a^{j+1}b^{j+1}c^{j+1}d^{j+1}$ and $y$ is a prefix of $a^{j+2}b^{j+2}c^{j+2}d^{j+2}$ for some $j \in [0, p - 1]$. Note that $z(3)$ contains $\$$yx\$$yx\$$ as a factor, so we must have $yx = a^{j+1}b^{j+1}c^{j+1}d^{j+1}$ for some $l \geq 0$.

**Case 1.1.** $l = 0$. In this case, $yx = \varepsilon$, so $v_i = \$$. Since $u_1v_1\ldots u_{i-1}v_{i-1}u_i$ precedes $\$$ in $z$, it must end in $d$. This means that the last non-empty string $w$ in the list $u_1, v_1, \ldots, u_{i-1}, v_{i-1}, u_i$ ends in $d$. Since $w$ is a suffix of $u_1v_1^\ldots u_{i-1}v_{i-1}u_i$, the latter string must also end in $d$. But this contradicts the fact that
\[ z(2) = u_1v_1^2\ldots u_{i-1}v_{i-1}^2u_i\$$u_{i+1}v_{i+1}^2\ldots u_kv_k^2u_{k+1} \]
is of the form [1].

**Case 1.2.** $l \geq 1$. In this case, $yx = a^{j+1}b^{j+1}c^{j+1}d^{j+1}$ contains $bc$ as a factor, so either $x$ contains $cd^{j+1}$ as a suffix or $y$ contains $a^{j+2}b$ as a prefix. In the former case, $l = j + 1$, so $y$ cannot contain $a^{j+2}b$ as a prefix, which means that $x$ must contain $bc^{j+1}$ as a suffix, contradicting $yx = a^{j+1}b^{j+1}c^{j+1}d^{j+1}$. In the latter case, $l = j + 2$, so $x$ cannot contain $cd^{j+1}$ as a suffix, which means that $y$ must contain $a^{j+2}b^{j+1}c$ as a prefix, contradicting $yx = a^{j+2}b^{j+2}c^{j+2}d^{j+2}$.

**Case 2.** $v_i$ contains at least two occurrences of $\$$.
Then we can write $v_i = x\$$a^{l+1}b^{l}c^{l}d^{l+1}\$$a^{m+1}b^{m}c^{m}d^{m+1}\$$y$, where $1 \leq l < m < p - 1$, $x$ is a suffix of $a^{l+1}c^{l}d^{l+1}$, and $y$ is a prefix of $a^{m+2}b^{m+1}c^{m+1}d^{m+2}$. Since
\[ \$$a^{m+1}b^{m}c^{m}d^{m+1}\$$y\$$a^{l+1}b^{l}c^{l}d^{l+1} \]
is a factor of $z(2)$, we must have
\[ yx = a^{l+1}b^{l+1}c^{m+1}d^{l+1}. \]
Since $y$ is a prefix of $a^{m+2}b^{m+1}c^{m+1}d^{m+2}$ and $l < m + 2$, $y$ must be a prefix of $a^l$. It follows that $x$ has $b^{m+1}c^{m+1}d^l$ as a suffix. But then $b^{m+1}c^{m+1}d^l$ must be a suffix of $a' b^{l-1} c^{l-1} d^l$, contradicting the fact that $l - 1 < m + 1$. 

Since every language in Weir’s control language hierarchy satisfies the weak Ogden property (Palis and Shende 1995, “Pumping lemmas for the control language hierarchy”), the language $L$ above is an example of a well-nested 3-MCFL that lies outside of Weir’s control language hierarchy.