Chapter 28

Free word order and MCFLs

Mark-Jan Nederhof
University of St Andrews

It was recently shown that the MIX language, over a three-symbol alphabet, is generated by a multiple context-free grammar. This paper investigates generalizations of the MIX language to alphabets of any size, as well as generalizations of the above-mentioned grammar. Presented are theoretical results that shed new light on the relation between these languages and these grammars. Precise conjectures are formulated that would further narrow down this relation. It is explained that validity of these conjectures would greatly enhance our understanding of the abilities of grammatical formalisms to describe free word order.

1 Introduction

Since the earliest attempts to build formal grammars describing the syntactic structure of natural languages, a central aim has been to identify formalisms with the appropriate generative power. If a formalism is too weak, it cannot describe all languages, or requires inelegant or an unreasonably large number of grammar rules to describe natural language phenomena. If a formalism is too strong, and allows description of phenomena that are unlike those in any natural language, then this causes its own problems. For example, the formalism may offer too little guidance to a linguist building a grammar by hand, and the search space may be too large for an algorithm to effectively learn a grammar from examples. Moreover, parsing and recognition algorithms of very powerful formalisms tend to have time complexities that are too high to be useful for practical purposes.

Against the backdrop of the Chomsky hierarchy (Chomsky 1959), the notion of *mildly context-sensitive grammars* was an attempt to identify properties required of an appropriate formalism for describing syntax, and to motivate tree adjoining grammars as a prime example of such a grammar formalism (Joshi 1985). The specified properties included formal requirements and informal characterizations.

In the years that followed, other formalisms were shown to be equivalent to tree adjoining grammars (Vijay-Shanker & Weir 1994), adding to the evidence that the tree adjoining languages are a natural class. In addition, more powerful formalisms have been found that clearly satisfy the formal requirements of mildly context-sensitive grammars and also appeared to satisfy the informal characterizations. The most notable are the multiple context-free grammars (MCFGs) (Seki et al. 1991), generating the multiple context-free languages (MCFLs). For an overview, see Kallmeyer (2010).

Joshi (1985) singles out one particular artificial language as posing a potential challenge to the theory of mildly context-sensitive grammars. This is the language of strings over $\{b_1, b_2, b_3\}$ such that, for some m, each of the three symbols occurs exactly m times, in any order. It is most commonly known as the MIX language (Gazdar 1988), referred to below as $\mathbf{MIX}_3$ to be able to put it in a broader context later. It is also known as the Bach language, after Bach (1981), although its properties have been studied at least since Aho & Ullman (1972: Exercise 2.6.3c).

The $\mathbf{MIX}_3$ language represents an extreme case of free word order, which appears to be irrelevant to any natural language. Joshi (1985) conjectured that $\mathbf{MIX}_3$ was not a tree adjoining language, consistent with the idea that tree adjoining grammars constitute an appropriate restriction of the power of context-sensitive grammars in order to model natural languages. The conjecture was finally proved by Kanazawa & Salvati (2012).

However, this leaves open the question whether $\mathbf{MIX}_3$ can be generated by other formalisms that are generally considered to be mildly context-sensitive, such as the MCFGs. A recently published result by Salvati (2015) shows the answer to be positive, by a proof via the $\mathbf{O}_2$ language. $\mathbf{MIX}_3$ and $\mathbf{O}_2$ are rationally equivalent, which means that if one is an MCFL then so is the other.

In this paper we will broaden the investigation to the $\mathbf{O}_n$ languages (Fischer & Rosenberg 1968). For fixed $n \geq 1$, there are $2n$ symbols $a_1, \ldots, a_n, \overline{a_1}, \ldots, \overline{a_n}$. A string is in $\mathbf{O}_n$ if, for each i, the number of occurrences of a_i equals the number of occurrences of $\overline{a_i}$. Similarly, a string is in the generalized $\mathbf{MIX}_n$ language over $\{b_1, \ldots, b_n\}$ if, for some m, each of the n symbols occurs exactly m times. For each n, $\mathbf{O}_n$ and $\mathbf{MIX}_{n+1}$ are rationally equivalent. Formally, $T_1(\mathbf{O}_n) = \mathbf{MIX}_{n+1}$ and $T_2(\mathbf{MIX}_{n+1}) = \mathbf{O}_n$, where T_1 and T_2 are two rational transductions. (For language L and rational transduction T we let $T(L) = \{w \mid \exists v \in L\ (v, w) \in T\}$.) We can specify T_1 and T_2 by finite-state transducers M_1 and M_2, each with a single state. For M_1, the transitions are labeled $a_i : b_i$, for $1 \leq i \leq n$, and $\overline{a_1} \cdots \overline{a_n} : b_{n+1}$. For M_2, the transitions are labeled $b_1 \cdots b_i : a_i$ and $b_{i+1} \cdots b_{n+1} : \overline{a_i}$ for $1 \leq i \leq n$.

In this paper, we formulate the family of MCFGs G_n that are conjectured to generate $\mathbf{O}_n$. If our conjectures are true, this would imply the remarkable finding that the MCFLs include all permutation closures of regular languages; by the *permutation closure* of a language L we mean the set of all strings that are permutations of strings in L. This implication was mentioned before by Salvati (2015) on the basis of Latteux (1979).[1] Note that this would also mean that the MCFLs, unlike the context-free languages, are closed under the operation of permutation closure.

[1] The relevant result is Proposition III.12 of Latteux (1979), which states that the permutation closure of a regular language with n symbols is in the closure of $\mathbf{MIX}_{n+1}$ under homomorphism, inverse homomorphism and intersection with regular language. Note that MCFLs are closed under these three operations. The proof appears to require the following correction in the sixth line: "$g(h^{-1}(c(w^*)) \cap R)$ où $R = \{a_1, \ldots, a_k\}^* {w'_1}^* \cdots {w'_p}^*$".

The importance of our investigation is that it sheds more light on the apparent incongruence between the generative power of some formalisms commonly considered to be mildly context-sensitive, and the observation that extreme free word order does not seem to occur in natural languages.

2 The O_n languages

Let n be a positive integer. The alphabet Σ_n consists of the $2n$ pairwise distinct symbols $a_1, \ldots, a_n, \overline{a_1}, \ldots, \overline{a_n}$. The length of a string w is denoted by $|w|$. For a string w over Σ_n and symbol $a \in \Sigma_n$, $|w|_a$ denotes the number of occurrences of a in w.

The *imbalance* of $w \in \Sigma_n^*$, denoted by imb(w), is the n-tuple $(|w|_{a_1} - |w|_{\overline{a_1}}, \ldots, |w|_{a_n} - |w|_{\overline{a_n}})$. In other words, for each i $(1 \leq i \leq n)$, we match the number of occurrences of a_i against the number of occurrences of $\overline{a_i}$, and the difference, which can be positive or negative, is one value in the imbalance. The expression 0^n denotes the tuple of n zeros. If w is such that imb$(w) = 0^n$, then we say that w is *balanced*.

For each n, the language $\mathbf{O}_n$ is defined to be the set of balanced strings, or formally $\mathbf{O}_n = \{w \in \Sigma_n^* \mid \text{imb}(w) = 0^n\}$.

3 MCFGs

We use terminology related to the MCFGs from Seki et al. (1991). This formalism is largely equivalent to the string-based LCFRSs from Vijay-Shanker, Weir & Joshi (1987).

A *multiple context-free grammar* (MCFG) is a tuple $G = (\Sigma, N, S, R)$, where Σ is a finite set of *terminals*, N is a finite set of *nonterminals* $(\Sigma \cap N = \emptyset)$, and $S \in N$ is the *start symbol*. Each nonterminal is associated with a positive integer, called its *fanout*. The start symbol has fanout 1.

Further, R is a finite set of *rules*, each of the form:

$$A_0(s_1, \ldots, s_{k_0}) \quad \rightarrow \quad A_1(x_1, \ldots, x_{m_1})\, A_2(x_{m_1+1}, \ldots, x_{m_2}) \, \cdots \, A_r(x_{m_{r-1}+1}, \ldots, x_{m_r})$$

where each A_i $(0 \leq i \leq r)$ has fanout k_i, $m_i = \sum_{j:1 \leq j \leq i} k_j$ $(1 \leq i \leq r)$, $x_1, \ldots, x_{m_r}$ are pairwise distinct variables, and each s_j $(1 \leq j \leq k_0)$ is a string consisting of variables and terminals. Moreover, each variable x_i $(1 \leq i \leq m_r)$ occurs exactly once in the *left-hand side* $A_0(s_1, \ldots, s_{k_0})$ and the left-hand side contains no other variables. The value r is called the *rank* of the rule.

An *instance* of a rule is obtained by choosing a string over Σ^* for each variable in the rule, and then replacing both occurrences of each variable by the chosen string. Let $\hat{\Sigma}_G$ denote the set of symbols of the form $A(w_1, \ldots, w_k)$, where k is the fanout of $A \in N$ and $w_1, \ldots, w_k \in \Sigma^*$; we refer to such a symbol as an *instance* of a nonterminal. The binary 'derives' relation $\Rightarrow_G$ over $\hat{\Sigma}_G^*$ is defined by $\phi_1 \hat{A}\, \phi_2 \Rightarrow \phi_1 \phi\, \phi_2$ if $\hat{A} \rightarrow \phi$ is a rule instance, with $\hat{A} \in \hat{\Sigma}_G$ and $\phi_1, \phi_2, \phi \in \hat{\Sigma}_G^*$. The reflexive, transitive closure of $\Rightarrow_G$ is $\Rightarrow_G^*$. The language *generated* by G is $L(G) = \{w \in \Sigma^* \mid S(w) \Rightarrow_G^* \varepsilon\}$, where ε is the empty string.

The largest fanout of any nonterminal in a given MCFG is called the fanout of the grammar, and the largest rank of any rule is called the rank of the grammar. We call a MCFG *binary* if the rank is at most 2. Every MCFG can be brought into binary form, at the expense of a higher fanout (Rambow & Satta 1999). We assume that all considered MCFGs are *reduced*, which means that each nonterminal is involved in at least one derivation $S(w) \Rightarrow_G^* \varepsilon$. We say a MCFG is in *normal form* if terminals occur only in rules with rank 0, or in other words if a rule contains variables or terminals, but not both at the same time. A grammar can be brought into normal form without increasing the fanout (Seki et al. 1991: Lemma 2.2).

4 MCFGs and the $\mathbf{O}_n$ languages

In order to relate the $\mathbf{O}_n$ languages to the languages generated by MCFGs, we start with a negative result.

Theorem 1 *For any $n \geq 2$, the language $\mathbf{O}_n$ is not generated by any MCFG with fanout strictly smaller than n.*

For the proof, assume that $\mathbf{O}_n$, for some n, is generated by a MCFG G with fanout $n-1$ or smaller. Without loss of generality, assume G is in normal form. We first show that if there are a nonterminal A and strings $w_1, \ldots, w_k, w'_1, \ldots, w'_k$ such that $A(w_1, \ldots, w_k) \Rightarrow_G^* \varepsilon$ and $A(w'_1, \ldots, w'_k) \Rightarrow_G^* \varepsilon$, then $\mathrm{imb}(w_1 \cdots w_k) = \mathrm{imb}(w'_1 \cdots w'_k)$. A sketch of the proof is as follows. Suppose that $\mathrm{imb}(w_1 \cdots w_k) \neq \mathrm{imb}(w'_1 \cdots w'_k)$. Then let $w \in \Sigma^*$ be such that $S(w) \Rightarrow_G^* \phi_1 A(w_1, \ldots, w_k)\phi_2 \Rightarrow_G^* \varepsilon$; such a derivation must exist as we assumed grammars are always reduced. Moreover $w \in \mathbf{O}_n$ by our initial assumption. Now replace the subderivation of $A(w_1, \ldots, w_k)$ by the subderivation of $A(w'_1, \ldots, w'_k)$; this is possible by the context-freeness of MCFGs. Thereby we obtain $S(w') \Rightarrow_G^* \phi_1 A(w'_1, \ldots, w'_k)\phi_2 \Rightarrow_G^* \varepsilon$, for some w' such that $\mathrm{imb}(w) - \mathrm{imb}(w') = \mathrm{imb}(w_1 \cdots w_k) - \mathrm{imb}(w'_1 \cdots w'_k)$. But since $\mathrm{imb}(w_1 \cdots w_k) - \mathrm{imb}(w'_1 \cdots w'_k) \neq 0^n$ and $\mathrm{imb}(w) = 0^n$, we must have $\mathrm{imb}(w') \neq 0^n$, which violates the assumption that $L(G) = \mathbf{O}_n$.

We conclude that each nonterminal A can be associated with a unique n-tuple τ_A such that $A(w_1, \ldots, w_k) \Rightarrow_G^* \varepsilon$ implies $\mathrm{imb}(w_1 \cdots w_k) = \tau_A$. Let d be $\max_{A, \tau_A = (d_1, \ldots, d_n), i} |d_i|$, or in other words, the largest absolute point-wise imbalance in any of the n pairs $(a_i, \overline{a_i})$ of symbols for any nonterminal A. Let ρ be the rank of the grammar.

Now consider the balanced string:

$$w = \overline{a_1}^{(n-1)m}(a_1 \cdots a_n)^m \overline{a_2}^{(n-1)m}(a_1 \cdots a_n)^m \cdots (a_1 \cdots a_n)^m \overline{a_n}^{(n-1)m}$$

for $m = \rho(2d+n)+d$. In a derivation of w, from the root downwards, consider the first nonterminal instance of the form $A(w_1, \ldots, w_k)$ where $2d + n \leq$

$|w_1 \cdots w_k|_{\overline{a_1}} < \rho(2d+n)$; such an instance always exists, as the number of occurrences of $\overline{a_1}$ in the left-hand side of a rule instance and a nonterminal instance in its right-hand side can differ by at most a factor ρ. This implies $d + n \leq |w_1 \cdots w_k|_{a_1} < \rho(2d+n) + d = m$, by the definition of d. As w contains $n-1$ substrings of the form $(a_1 \cdots a_n)^m$ and a_1 occurs nowhere else, it follows that $w_1, \ldots, w_k$ must contain at least some non-empty parts of these substrings $(a_1 \cdots a_n)^m$, including at least $(d+n)-(n-1) = d+1$ occurrences of each of $a_2, \ldots, a_n$, due to $k \leq n-1$ by our assumptions. No substring $(a_1 \cdots a_n)^m$ can be entirely included in any of w_1, ..., w_k however, as $|w_1 \cdots w_k|_{a_1} < m$. Moreover, $w_1 \cdots w_k$ must include at least $(d+1)-d = 1$ occurrence of each of $\overline{a_2}, \ldots, \overline{a_n}$, so that some non-empty part of each of the n substrings $\overline{a_i}^{(n-1)m}$ of w must be included in $w_1, \ldots, w_k$. This is impossible because $k < n$, which completes the proof. ■

As the above proof of non-existence fails for fanout greater than or equal to n, one may suspect the following.

Conjecture 1 *For any $n \geq 2$, the language $\mathbf{O}_n$ is generated by a binary MCFG of fanout n.*

For $n \geq 2$, let the binary MCFG G_n of fanout n with alphabet Σ_n be defined by the following rules:

$$
\begin{array}{rcl}
S(x_1 \cdots x_n) & \to & A(x_1, \ldots, x_n) \\
A(w_1, \ldots, w_n) & \to & \varepsilon, \text{ for all } w_1, \ldots, w_n \in \Sigma_n \cup \{\varepsilon\} \text{ such that} \\
 & & |w_1 \cdots w_n| \leq 2 \text{ and } \mathrm{imb}(w_1 \cdots w_n) = 0^n \\
A(s_1, \ldots, s_n) & \to & A(x_1, \ldots, x_n)\ A(y_1, \ldots, y_n), \text{ for all non-empty} \\
 & & s_1, \ldots, s_n \text{ such that } |s_1 \cdots s_n| = 2n,\ s_1 \text{ starts with } x_1, \\
 & & x_1 \cdots x_n \text{ and } y_1 \cdots y_n \text{ are subsequences of } s_1 \cdots s_n, \\
 & & \text{and no } s_i \text{ has a substring of the form } x_j x_{j+1} \text{ or } y_j y_{j+1}
\end{array}
$$

We now wish to strengthen Conjecture 1 to:

Conjecture 2 *For any $n \geq 2$, $L(G_n) = \mathbf{O}_n$.*

Even stronger is the following:

Conjecture 3 *For any $n \geq 2$, and $w_1, \ldots, w_n \in \Sigma_n^*$ such that $\mathrm{imb}(w_1 \cdots w_n) = 0^n$, we have $A(w_1, \ldots, w_n) \Rightarrow_{G_n}^* \varepsilon$.*

It is clear that if Conjecture 3 is true, then so is Conjecture 2. The added value of Conjecture 3 is that it would exclude the possibility of making a 'wrong' derivation step, as long as all nonterminal instances contain arguments that together are balanced. For example, how a string $w \in \mathbf{O}_n$ is divided into n parts in the first step using an instantiated rule $S(w_1 \cdots w_n) \to A(w_1, \ldots, w_n)$, with $w = w_1 \cdots w_n$, would not affect whether we can complete the derivation.

A proof of Conjecture 3 would likely be by induction, first on the length of $w_1 \cdots w_n$, and second on the number of arguments among $w_1, \ldots, w_n$ that are ε. The base case is obviously $w_1 = \ldots = w_n = \varepsilon$.

The inductive step is straightforward if at least one of the arguments, say w_i, is ε. Two subcases can be distinguished. In the first, at least one of the remaining arguments, say w_j, has length 2 or greater. Assume without loss of generality that $1 < i < j$. We can then use a rule of the form $A(s_1, \ldots, s_n) \to A(x_1, \ldots, x_n)\ A(y_1, \ldots, y_n)$, where:

$$s_k = \begin{cases} x_k y_k & \text{if } 1 \leq k < i \\ y_k & \text{if } k = i \\ x_{k-1} y_k & \text{if } i < k < j \\ x_{k-1} y_k x_k & \text{if } k = j \\ x_k y_k & \text{if } j < k \leq n \end{cases}$$

In the required rule instance, we would replace each y_k ($1 \leq k \leq n$) by ε, replace each x_k by w_k if $k < i$ or $j < k$, replace each x_{k-1} by w_k if $i < k < j$, and replace x_{j-1} and x_j by non-empty strings w' and w'', respectively, such that $w_j = w'w''$. We also use the rule $A(\varepsilon, \ldots, \varepsilon) \to \varepsilon$, together with the inductive hypothesis for a nonterminal instance in which one argument fewer is ε.

In the second subcase, all the non-empty arguments are of length 1. There must then be two arguments, say w_i and w_j, that are a_ℓ and $\overline{a_\ell}$, respectively, for some ℓ ($1 \leq \ell \leq n$). We can then use a rule of the form $A(s_1, \ldots, s_n) \to A(x_1, \ldots, x_n)\ A(y_1, \ldots, y_n)$, where each s_k ($1 \leq k \leq n$) is $x_k y_k$. In the required rule instance, we would replace each x_k and y_k ($1 \leq k \leq n$) by w_k and ε, respectively, if $k \notin \{i, j\}$, and by ε and w_k, respectively, if $k \in \{i, j\}$. We also use a rule of the form $A(v_1, \ldots, v_n) \to \varepsilon$, with $v_i = a_\ell$ and $v_j = \overline{a_\ell}$, and $v_k = \varepsilon$ for $k \notin \{i, j\}$, together with the inductive hypothesis.

The inductive step is less straightforward if $w_1, \ldots, w_n$ are all non-empty. We then need to show that there is a sequence of $2n$ strings $v_1, \ldots, v_{2n}$ such that:

- there is a sequence of positive integers $k_1, \ldots, k_n$ such that for each i ($1 \leq i \leq n$) we have $w_i = v_{m_{i-1}+1} \cdots v_{m_i}$, where $m_i = \sum_{j:1 \leq j \leq i} k_j$ ($0 \leq i \leq n$), and $m_n = 2n$, and
- there is a permutation $u_1, \ldots, u_n,\ u'_1, \ldots, u'_n$ of $v_1, \ldots, v_{2n}$, such that $\text{imb}(u_1 \cdots u_n) = \text{imb}(u'_1 \cdots u'_n) = 0^n$, $|u_1 \cdots u_n| > 0$ and $|u'_1 \cdots u'_n| > 0$.

In words, a balanced string divided into n non-empty parts can be further divided into $2n$ smaller parts, and in particular the i-th part is divided into k_i smaller parts, and the $2n$ smaller parts can be partitioned to form two other balanced (but non-empty) strings, each again divided into n parts.

Special treatment can be given to cases where $w_1, \ldots, w_n$ are all non-empty, but $\text{imb}(v_1 \cdots v_k) = 0^n$ for a proper non-empty subset $\{v_1, \ldots, v_k\}$ of $\{w_1, \ldots, w_n\}$, by deriving:

$$A(w_1, \ldots, w_n) \Rightarrow A(v_1, \ldots, v_k, \varepsilon, \ldots, \varepsilon)\ A(u_1, \ldots, u_n)$$

for some $u_1, \ldots, u_n$, which allows use of the inductive hypothesis. Hence the interesting case that remains is where $\mathrm{imb}(v_1 \cdots v_k) = 0^n$ does *not* hold for any proper non-empty subset $\{v_1, \ldots, v_k\}$ of $\{w_1, \ldots, w_n\}$.

5 Special cases

5.1 The $\mathbf{O}_1$ language

The case $n = 1$ has been ignored in the above. It is straightforward to show that $\mathbf{O}_1$ is generated by an MCFG of fanout 1, but this grammar G_1 has a slightly different structure from the grammars G_n ($n \geq 2$) that were defined above. The rules of G_1 are $S(xy) \to S(x)\, S(y)$, $S(a_1 x \overline{a_1}) \to S(x)$, $S(\overline{a_1} x a_1) \to S(x)$, and $S(\varepsilon) \to \varepsilon$.

The central observation in the proof by induction concerns strings in $\mathbf{O}_1$ of the form $a_1 w a_1$ or of the form $\overline{a_1} w \overline{a_1}$. In the first case, the imbalance of the prefix a_1 is a positive number and the imbalance of the prefix $a_1 w$ is a negative number. This implies that there must be a proper prefix $a_1 w'$ of $a_1 w$ whose imbalance is 0^1, which means we can use the rule $S(xy) \to S(x)\, S(y)$ and the inductive hypothesis for two shorter strings. The second case is symmetric.

5.2 The $\mathbf{O}_2$ language

Conjecture 3 restricted to $n = 2$ was proved by Salvati (2015), using arguments involving considerable sophistication. The proof is geometric in nature, interpreting the imbalance of a series of prefixes of a string in $\mathbf{O}_2$ of increasing length as a path in 2-dimensional space. The use of the complex exponential function seems to make the proof difficult to generalize to higher dimensions.

An alternative proof is due to Nederhof (2016). It is similarly geometric in nature, but avoids the complex exponential function. Its core argument divides 2-dimensional space into an ‘above’ and a ‘below’. We will refer to this as the ‘partition argument’. Before the argument can be applied, the paths must first be brought into a normal form.

The proof requires all four binary rules of G_2:

$$\begin{array}{lcl} A(x_1 y_1, x_2 y_2) & \to & A(x_1, x_2)\, A(y_1, y_2) \\ A(x_1 y_1, y_2 x_2) & \to & A(x_1, x_2)\, A(y_1, y_2) \\ A(x_1 y_1 x_2, y_2) & \to & A(x_1, x_2)\, A(y_1, y_2) \\ A(x_1, y_1 x_2 y_2) & \to & A(x_1, x_2)\, A(y_1, y_2) \end{array}$$

5.3 The O_3 language

Nederhof (2016) also sketches a potential generalization of the proof to $n = 3$. The partition argument now relies on the (three plus six) rules:

$$A(x_1y_1, x_2y_2, y_3x_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1y_1, y_2x_2, x_3y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(y_1x_1, x_2y_2, x_3y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$

$$A(x_1y_1x_2, x_3y_2, y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1y_1x_2, y_2, x_3y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1y_1, x_2y_2x_3, y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1y_1, y_2, x_2y_3x_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1, y_1x_2y_2, y_3x_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1, y_1x_2, y_2x_3y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$

In addition, we need a 'corkscrew argument', which requires three further rules:

$$A(x_1y_1x_2y_2, x_3, y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1, y_1x_2y_2x_3, y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$
$$A(x_1, y_1, y_2x_2y_3x_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$

It is remarkable that these 12 rules are only a portion of the 22 binary rules of G_3. If however we remove any of the last three rules, then we cannot always use the inductive hypothesis. For example, if we remove:

$$A(x_1y_1x_2y_2, x_3, y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$

then we can no longer handle $A(\overline{a_3}\,\overline{a_2}\,\overline{a_2}\,\overline{a_2}a_3a_1a_1a_2,\, a_2\overline{a_1},\, a_2\overline{a_1})$. The same applies to the group of six rules. For example, if we remove:

$$A(x_1y_1x_2, x_3y_2, y_3) \to A(x_1, x_2, x_3)\ A(y_1, y_2, y_3)$$

then we can no longer handle $A(a_3a_2a_2a_1\overline{a_3}a_1\overline{a_2},\, \overline{a_1}\,\overline{a_3}\,\overline{a_1},\, a_3\overline{a_2})$. This can be verified by mechanically matching these nonterminal instances against the remaining rules. We have not been able to ascertain that more than one of the first three rules is necessary to always be able to apply the inductive hypothesis. Hence we cannot exclude the possibility at this time that only 10 binary rules would suffice.

The main difficulty in obtaining a complete proof of Conjecture 3 restricted to $n = 3$ pertains to the partition argument. This would again depend on a normal form for paths, this time in 3-dimensional space. It seems much more difficult than before to show that the normal form can be obtained while preserving appropriate invariants.

6 Conclusions

It is trivial to show that $\mathbf{O}_1$ is generated by G_1, whereas it took great efforts to find the first proof that $\mathbf{O}_2$ is generated by G_2. The second proof of the same result seems to create realistic prospects that a proof may one day be found that (a subgrammar of) G_3 generates $\mathbf{O}_3$, but considerable challenges lie ahead. Very little is known for $n \geq 4$.

Acknowledgements

This work came out of correspondence with Giorgio Satta. Gratefully acknowledged are also fruitful discussions with Sylvain Salvati, Vinodh Rajan, and Markus Pfeiffer.

References

Aho, A. V. & J. D. Ullman. 1972. *Parsing*. Vol. 1 (The Theory of Parsing, Translation and Compiling). Englewood Cliffs, N.J.: Prentice-Hall.

Bach, E. 1981. Discontinuous constituents in generalized categorial grammars. In *Proceedings of the eleventh Annual Meeting of the North Eastern Linguistic Society*, 1–12.

Chomsky, N. 1959. On certain formal properties of grammars. *Information and Control* 2. 137–167.

Fischer, M. J. & A. L. Rosenberg. 1968. Real-time solutions of the origin-crossing problem. *Mathematical Systems Theory* 2. 257–263.

Gazdar, G. 1988. Applicability of indexed grammars to natural languages. In U. Reyle & C. Rohrer (eds.), *Natural language parsing and linguistic theories*, 69–94. D. Reidel Publishing Company.

Joshi, A. K. 1985. Tree adjoining grammars: how much context-sensitivity is required to provide reasonable structural descriptions? In D. R. Dowty, L. Karttunen & A. M. Zwicky (eds.), *Natural language parsing: psychological, computational, and theoretical perspectives*, 206–250. Cambridge University Press.

Kallmeyer, Laura. 2010. *Parsing beyond context-free grammars*. Springer-Verlag.

Kanazawa, M. & S. Salvati. 2012. MIX is not a tree-adjoining language. In *50th Annual Meeting of the Association for Computational Linguistics, proceedings of the conference*, 666–674. Jeju Island, Korea.

Latteux, M. 1979. Cônes rationnels commutatifs. *Journal of Computer and System Sciences* 18. 307–333.

Nederhof, Mark-Jan. 2016. A short proof that O_2 is an MCFL. In *54th Annual Meeting of the Association for Computational Linguistics, proceedings of the conference*, vol. 1, 1117–1126. Berlin.

Rambow, O. & G. Satta. 1999. Independent parallelism in finite copying parallel rewriting systems. *Theoretical Computer Science* 223. 87–120.

Salvati, S. 2015. MIX is a 2-MCFL and the word problem in $\mathbb{Z}^2$ is captured by the IO and the OI hierarchies. *Journal of Computer and System Sciences* 81. 1252–1277.

Seki, H., T. Matsumura, M. Fujii & T. Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science* 88. 191–229.

Vijay-Shanker, K. & D. J. Weir. 1994. The equivalence of four extensions of context-free grammars. *Mathematical Systems Theory* 27. 511–546.

Vijay-Shanker, K., D. J. Weir & A. K. Joshi. 1987. Characterizing structural descriptions produced by various grammatical formalisms. In *25th Annual Meeting of the Association for Computational Linguistics, proceedings of the conference*, 104–111. Stanford, California, USA.