# On Independent Pumpability

Marcus Kracht
*Department of Linguistics, UCLA*
*PO Box 951543*
*405 Hilgard Avenue*
*Los Angeles, CA 90095-1543*
*USA*
`kracht@humnet.ucla.edu`

**Abstract**

In an unpublished note, [Manaster-Ramer *et al.*, 1992], it is shown that context free languages satisfy a pumping property that asserts the existence of a linear number of pumping strings rather than just one, as in the original theorem proved by Ogden. It is shown here that this property is stronger than Ogden's. Additionally, we show that for every $k$ there are continuously many languages that have $k$ pumping pairs but not $k+1$ many pumping pairs.

## 1 Ogden's Characterisation

The well-known pumping lemma asserts that if $L$ is context free language then there is a number $k$ such that every $L$-string $\vec{x}$ of length at least $k$ possesses a decomposition

(1) $\vec{x} = \vec{u}\vec{y}\vec{v}\vec{z}\vec{w}$

such that

(2) $\{\vec{u}\vec{y}^n\vec{v}\vec{z}^n\vec{w} : n \in \mathbb{N}\} \subseteq L$

The pair $\langle \vec{y}, \vec{z} \rangle$ is a called a **pumping pair** of $\vec{x}$. A stronger version of this lemma is due to [Ogden, 1968].

**Lemma 1 (Ogden's Lemma)** *Let $L$ be a context free language. Then there exists a number $n_L$ such that for every string $\vec{x} \in L$: if $P$ is a set of at least $n_L$ occurrences of letters in $\vec{x}$ then there exists a pumping pair containing at least one member of $P$ and at most $n_L$ of them.*

If $L$ is a language, let $L_n$ denote the set of strings that are in $L$ and have length $n$. The following is from [Ogden *et al.*, 1985].

**Lemma 2 (Interchange Lemma)** *Let $L$ be a context free language. Then there exists a real number $c_L$ such that for every natural number $n$ and every set $Q \subseteq L_n$ there is $k \geq \ulcorner |Q|/(c_L n^2) \urcorner$, and strings $\vec{x}_i$, $\vec{y}_i$, $\vec{z}_i$, $i < k$, such that*

1. *for all $i < k$: $\vec{x}_i \vec{y}_i \vec{z}_i \in Q$,*
2. *for all $i < j < n$: $\vec{x}_i \vec{y}_i \vec{z}_i \neq \vec{x}_j \vec{y}_j \vec{z}_j$,*
3. *for all $i < i < k$: $|\vec{x}_i| = |\vec{x}_j|$, $|\vec{y}_i| = |\vec{y}_j|$, and $|\vec{z}_i| = |\vec{z}_j|$.*
4. *for all $i < k$: $n > |\vec{x}_i \vec{z}_i| > 0$, and*
5. *for all $i, j < k$: $\vec{x}_i \vec{y}_j \vec{z}_i \in L_n$.*

Note that if the sequence of numbers $L_n/n^2$ is bounded, then the language satisfies the Interchange Lemma. For assume that for $n_0$ we have $L_{n_0}/n_0^2 \leq c$. Then set $c_L := \max\{c \cdot |L_m| \cdot m^2 : m \leq n_0\}$. Then for every subset $Q$ of $L_n$, $\ulcorner |Q|/(c_L n^2) \urcorner \leq 1$. However, with $k = 1$ the conditions above become empty.

**Theorem 3** *Every language $L$ where $\lim_{n\to\infty} |L_n|/n^2$ is bounded satisfies the Interchange Lemma. In particular, every one–letter language satisfies the Interchange Lemma.*

Also, the following is useful:

**Lemma 4** *Suppose that $L$ and $L'$ satisfy the Interchange Lemma; then so does $L \cup L'$.*

**Proof.** Let $c_L$ and $c_{L'}$ be the constants of the Interchange Lemma for $L$ and $L'$, respectively. Put $d := 2\max\{c_L, c_{L'}\}$. Let $Q$ be a set of strings from $L \cup L'$ of length $n$ for a given $n$. Then either $|Q \cap L| \geq |Q \cap L'|$ or $|Q \cap L| \leq |Q \cap L'|$. Without loss of generality the first is the case. Then by the assumption on $L$ there is a $k \geq \ulcorner |Q \cap L|/(c_L n^2) \urcorner$ and strings $\vec{x}_i$, $\vec{y}_i$, and $\vec{z}_i$, $i < k$, satisfying the assertion of the Interchange Lemma for $L$. It is clear that they also satisfy the assertion for $L \cup L'$. Furthermore, $k \geq \ulcorner |Q \cap L|/(c_L n^2) \urcorner \geq \ulcorner |Q|/(2c_L n^2) \urcorner \geq \ulcorner |Q|/(dn^2) \urcorner$, and so $k$ is of the required size. □

## 2 Independent Pumpability

[Manaster-Ramer *et al.*, 1992] have shown that a stronger theorem holds: if $L$ is context free then there is a number $k_L$ such that if we select in a given $L$-string any number $\geq k_L N$ of positions there are at least $N$ disjoint independent pumping pairs each containing at least but less than $k_L$ of the designated positions. I shall expose the definition of what it is to have two independent pumping pairs. It means that the string $\vec{x}$ can be decomposed either as

(3) $\vec{x} = \vec{u}_1\vec{y}_1\vec{u}_2\vec{y}_2\vec{v}\vec{z}_2\vec{w}_2\vec{z}_1\vec{w}_1$

such that

(4) $\{\vec{u}_1\vec{y}_1^m\vec{u}_2\vec{y}_2^n\vec{v}\vec{z}_2^n\vec{w}_2\vec{z}_1^m\vec{w}_1 : m, n \in \mathbb{N}\} \subseteq L$

or as

(5) $\vec{x} = \vec{u}_1\vec{y}_1\vec{u}_2\vec{z}_1\vec{v}\vec{y}_2\vec{w}_2\vec{z}_2\vec{w}_1$

such that

(6) $\{\vec{u}_1\vec{y}_1^m\vec{u}_2\vec{z}_1^m\vec{v}\vec{y}_2^n\vec{w}_2\vec{z}_2^n\vec{w}_1 : m, n \in \mathbb{N}\} \subseteq L$

Notice that the pumping pairs cannot cross. Thus, we may not use a decomposition of the following form:

(7) $\vec{u}_1\vec{y}_1\vec{u}_2\vec{y}_2\vec{v}\vec{z}_1\vec{w}_2\vec{z}_2\vec{w}_1$

In the latter case we say that the two occurrences of the pumping pairs **cross**. So, we say that a set of $k$ occurrences of pumping pairs is **independent** if

① all pairs are disjoint,

② no two pairs cross each other, and

③ all pairs can be repeated any number of times independently of the others.

**Definition 5** *A language $L$ is said to be* ***independently*** $k$***-pumpable*** *if there is a constant $p_L$ such that for every $\vec{x} \in L$: if we designate $kp_L$ positions in $\vec{x}$ then there are $k$ independently pumpable pairs each involving at least one and less than $p_L$ of the designated positions.*

The question now is whether this fine structure really defines translates into a hierarchy of languages even if we require the Interchange Property. The answer will be affirmative, showing in particular that there are languages that satisfy Ogden's Lemma and the Interchange Property but not the property of [Manaster-Ramer *et al.*, 1992].

**Theorem 6** *For every $k$ there is a language which is independently $k$-pumpable but not independently $k+1$-pumpable.*

**Proof.** Put

(8) $$L_k := \{\mathsf{a}_0^{n_0}\Pi_{0<i<k+1}\mathsf{a}_i^{n_i} : \text{for some } i \leq k\text{: } n_i = 0 \text{ or there is } p \in \mathbb{N}\text{:} n_0 = 2^p + \sum_{i=1}^{k} n_i\}$$

So, the language is that of all strings in the letters $\mathsf{a}_i$, $i < k+1$, in which all occurrences of $\mathsf{a}_i$ precede all occurrences of $\mathsf{a}_j$ whenever $i < j$, where either one of the letters does not occur or otherwise the number of occurrences of $\mathsf{a}_0$ is the sum of the occurrences of the other letters plus some power of 2. Let us notice right away that this language is not context free. Its intersection with $\mathsf{a}_0^*\mathsf{a}_1$ is the language $\{\mathsf{a}_0^{2^p+1}\mathsf{a}_1 : p \in \mathbb{N}\}$, which is not context free.

I first establish that $L_k$ is independently $k$-pumpable $p_{L_k} = 3$. In fact, take a string $\vec{x}$ and designate $2k$ positions. (Case 1.) Some letter does not occur in $\vec{x}$. Then randomly select $k$ occurrences of pairs of the form $\langle\varepsilon, b\rangle$, where $b$ is a letter that has a designated occurrence in that pair. Then these pairs can be independently pumped, as there are no restrictions on the strings of this type. To make them nested rather than crossed, proceed as follows. Pair each letter occurrence with the first occurrence of the empty string. (Case 2.) All letters occur in $\vec{x}$. Notice then that any occurrence of $\mathsf{a}_0$ can be paired with any occurrence of an $\mathsf{a}_i$, $i > 0$, into a pumping pair. These pairs are all independent; it can be arranged that they do not cross. Basically, suppose $\langle\mathsf{a}_0, b\rangle$ and $\langle\mathsf{a}_0, c\rangle$ are pumping pairs, and we have a decomposition

(9) $$\vec{x} = \vec{u}_1\mathsf{a}_0\vec{v}_1\mathsf{a}_0\vec{w}b\vec{v}_1c\vec{v}_2$$

Then pair the first occurrence of $\mathsf{a}_0$ with $c$ rather than $b$. All pairs will then be nested.

As there are at least $k$ occurrences of letters containing a letter different from $\mathsf{a}_0$, we have $k$ such pairs. Notice that any letter occurrences can be made to occur in one such pair.

The language $L_k$ is not independently $k+1$-pumpable, though. For take the strings $\mathsf{a}_0^{2^p+k}\mathsf{a}_1\cdots\mathsf{a}_k$. These strings do not contain $k+1$ independent pumping pairs, no matter what positions are designated. For if there are $k+1$ pumping pairs, then one of them contains only occurrences of $\mathsf{a}_0$. But such a pair cannot be dropped unless it has size $2^p - 2^q$ for some $q < p$. But then it cannot be pumped up even once since $\mathsf{a}^{2\cdot 2^p - 2^q + k}\mathsf{a}_1\cdots\mathsf{a}_k \notin L_k$. □

**Proposition 7** *$L_k$ satisfies the Interchange Lemma.*

**Proof.** The languages $\mathsf{a}_0^*\mathsf{a}_1^*\cdots\mathsf{a}_g^*$ all satisfy the Interchange Lemma (and so do all homomorphic images thereof) so by Lemma 4 it is enough to establish that the following language satisfies the Interchange Lemma:

(10) $$M_k := \{\mathsf{a}_0^{n_0}\Pi_{0<i<k+1}\mathsf{a}_i^{n_i} : \text{there is } p \in \mathbb{N}{:}n_0 = 2^p + \sum_{0<i<k+1} n_i\}$$

Let $Q \subseteq M_k$ a set of strings of length $n$. All of them have the form $\mathsf{a}_0\vec{z}\mathsf{a}_k$, and so with this decomposition the Interchange Lemma is trivially satisfied. □

# 3 Further Extensions

**Proposition 8** *If both L and L′ are independently k-pumpable, then so is L ∪ L′.*

The proof is really straightforward. If the constant for $L$ is $p$ and the constant for $L'$ is $q$, then the constant $L \cup L'$ is $r := \max\{p, q\}$. For given $\vec{x}$ and $rk$ positions, either $\vec{x} \in L$ or $\vec{x} \in L'$, and in either case we find $k$ occurrences of independent pumping pairs involving at least one but less than $r$ of the positions.

**Proposition 9** *Let L and L′ be languages over disjoint alphabets. Suppose that L is not independently k-pumpable. Then also L ∪ L′ is not independently k-pumpable.*

**Proof.** Suppose that $L \cup L'$ is independently $k$-pumpable with constant $p_L$. Pick a string $\vec{x} \in L$. Designate $kp_L$ positions. By assumption there are $k$ occurrences of independently pumpable pairs in $\vec{x}$. The effect of pumping or depumping these pairs does not yield the empty string (their combined length is less than that of $\vec{x}$) and is therefore over the alphabet of $L$, which is disjoint from that of $L'$. Hence all the strings obtained by pumping the pairs of $\vec{x}$ are all in $L$. Hence $L$ is independently $k$-pumpable as well. □

**Theorem 10** *For every k, there are continuously many languages which are independently k-pumpable but not independently $k + 1$-pumpable. Moreover, all these languages satisfy the Interchange Lemma.*

**Proof.** It has been shown in [Kracht, 2004] that for every $\Omega \subseteq \mathbb{N}$ the language $L_\Omega := \{\mathsf{b}_0^m \mathsf{b}_1^n : m \neq n \text{ or } m \in \Omega\}$ is $k$ pumpable for every $k$. Then $L_\Omega \cup L_k$ is $k$ pumpable by Lemma 8 but not $k + 1$ pumpable by Lemma 9. It is clear that $L_k \cup L_\Omega \neq L_k \cup L_\Psi$ iff $\Omega \neq \Psi$. $L_\Omega$ has the Intercange Property by Lemma 3. By Lemma 4, $L_k \cup L_\Omega$ has the Interchange Property. □

# References


[Kracht, 2004] Marcus Kracht. Too Many Languages Satisfy Ogden's Lemma. In *Proceedings of the 27th Pennsylvania Linguistics Colloquium*, number 10.1 in Penn Linguistics Club Working Papers, pages 115 – 121, 2004.

[Manaster-Ramer *et al.*, 1992] Alexis Manaster-Ramer, M. Andrew Moshier, and R. Suzanne Zeitman. An Extension of Ogden's Lemma. Manuscript. Wayne State University, 1992.

[Ogden *et al.*, 1985] R. Ogden, R. J. Ross, and K. Winkelmann. An "Interchange Lemma" for Context Free Languages. *SIAM Journal of Computing*, 14:410 – 415, 1985.

[Ogden, 1968] R. W. Ogden. A helpful result for proving inherent ambiguity. *Mathematical Systems Sciences*, 2:191 – 194, 1968.