On the number of palindromically rich words

Amy Glen

School of Engineering & IT
Murdoch University, Perth, Australia

amy.glen@gmail.com
http://amyglen.wordpress.com

59th AustMS Annual Meeting @ Flinders University

Special Session: Combinatorics

Sept 28 – Oct 1, 2015

What are rich words?

- Vague Answer: finite and infinite words that are "rich" in palindromes in the utmost sense.
- A *palindrome* is a finite word that reads the same backwards as forwards.

  Examples: `eye, civic, radar, glenelg`

The following result is well-known in the field of combinatorics on words.

Theorem (Droubay, Justin, Pirillo 2001)

Any finite word w of length $|w|$ contains at most $|w| + 1$ distinct palindromes (including the empty word ε).

- Inspired by this result, we initiated a unified study of finite and infinite words that are characterised by containing the maximal number of distinct palindromes.
- Such words are called rich words in view of their palindromic richness.

Rich words

Definition (G., Justin, Widmer, Zamboni 2009)

A finite word w is said to be *rich* if w contains exactly $|w|+1$ distinct palindromes (including ε).

Examples

- $abac$ is rich, whereas $abca$ is not rich.
- There exist many rich words in the English language – predominantly a consequence of most letters going unrepeated in a given English word. For example:
  - `rich` is rich.
  - `poor` is rich too!
  - But `plentiful` is not rich.
- On the preceding slide, only the following 10 words are not rich: `known`, `combinatorics`, `including`, `inspired`, `infinite`, `that`, `characterised`, `containing`, `maximal`, `distinct`.

  This was easy to determine without counting palindromic factors . . . but how?

Essentially, a finite (or infinite) word is rich if and only if a new palindrome is introduced at each new position.

Example:

$abaabaaabaaaabaaaaab \cdots \underline{a}baabaaabaaaabaaaaab \cdots a\underline{b}aabaaabaaaabaaaaab \cdots \underline{aba}$

Rich words also have the following characteristic properties, established by Droubay, Justin, Pirillo (2001) and G., Justin, Widmer, Zamboni (2009).

Characteristic Properties of Rich Words

For any finite or infinite word w, the following conditions are equivalent:

i) w is rich;

ii) every prefix of w has a unioccurrent palindromic suffix (and equivalently, when w is finite, every suffix of w has a unioccurrent palindromic prefix);

iii) for each factor u of w, every prefix (resp. suffix) of u has a unioccurrent palindromic suffix (resp. prefix);

iv) for each palindromic factor p of w, every complete return to p in w is a palindrome.

Basic properties

- If a finite word w is rich, then its *reversal* $\widetilde{w}$ is also rich.

  Example: $w = aabac$ and $\widetilde{w} = cabaa$ are both rich.

- If w and w' are rich with the same set of palindromic factors, then they are *abelianly equivalent*, i.e., $|w|_x = |w'|_x$ for all letters x.

- For any rich word w, there exist letters $x, z \in \mathrm{Alph}(w)$ such that wx and zw are rich.

- Palindromic closure preserves richness.

  The *palindromic closure* of a word v, denoted by v^+, is the unique shortest palindrome beginning with v.

  Examples:
  $(race)^+ = race\,car$
  $(tops)^+ = top\,spot$
  $(party)^+ = party\,trap$
  $(tie)^+ = tie\,it$
  $(abac)^+ = abacaba$
  $(glen)^+ = glenelg$ ...looking forward to dinner there tonight!

More about palindromic closure

- Palindromic closure is one way of extending a rich word into a longer one.
- If we iteratively apply palindromic closure, we can obtain infinite rich words.
- The *iterative palindromic closure operator* Pal is defined as follows:

$$Pal(\varepsilon) = \varepsilon \text{ (empty word)} \quad \text{and} \quad Pal(wx) = (Pal(w)x)^+$$

for any word w and letter x.

Example: $Pal(aba) = \underline{a}\,\underline{b}\,a\,\underline{a}\,b\,a$

- Now suppose $\Delta = x_1 x_2 x_3 x_4 \cdots$ is an infinite word over a 2-letter alphabet $\{a, b\}$.

Then

$$Pal(\Delta) := \lim_{n\to\infty} Pal(x_1 x_2 \cdots x_n)$$

is a rich infinite word over $\{a, b\}$ since palindromic closure (and hence Pal) preserves richness.

All such words are called characteristic (or standard) Sturmian words.

Sturmian words

A well-known example of a characteristic Sturmian word is the infinite Fibonacci word $\mathbf{f}$:

$$\Delta = (ab)^\infty = (\underline{a}b)(ab)(ab)\cdots \longrightarrow \Delta = (ab)^\infty = (a\underline{b})(ab)(ab)\cdots$$
$$\longrightarrow \Delta = (ab)^\infty = (ab)(\underline{a}b)(ab)\cdots \longrightarrow \Delta = (ab)^\infty = (ab)(a\underline{b})(ab)\cdots$$
$$\longrightarrow \Delta = (ab)^\infty = (ab)(ab)(ab)\cdots \longrightarrow \mathbf{f} = \underline{a}\underline{b}a\underline{a}ba\underline{b}aaba\cdots$$

- Note that the Fibonacci word is aperiodic (i.e., not ultimately periodic).
- More generally, a characteristic Sturmian word $\boldsymbol{w}$ is aperiodic if and only if its so-called directive word Δ over $\{a, b\}$ does not ultimately degenerate into an infinitely repeated single letter; otherwise, $\boldsymbol{w}$ is purely periodic.
- Aperiodic Sturmian words are characterised by having factor complexity function $\mathcal{C}(n) = n + 1$ for all $n \geq 0$. [Morse & Hedlund 1940]

  The factor complexity function of a finite or infinite word w, denoted by $\mathcal{C}_w(n)$, counts the number of distinct factors of w of each length $n \geq 0$.
- Morse & Hedlund (1940) also showed that an infinite word $\boldsymbol{w}$ is eventually periodic $\Leftrightarrow \mathcal{C}_{\boldsymbol{w}}(n) < n + 1$ for some $n \in \mathbb{N}^+$.

Episturmian words are rich too

$\{a, b\} \longrightarrow \mathcal{A}$ (finite alphabet) gives *standard episturmian words*

Theorem (Droubay, Justin, Pirillo, 2001)

An infinite word $\boldsymbol{s}$ over $\mathcal{A}$ is a standard episturmian word if and only if there exists an infinite word $\Delta = x_1x_2x_3\cdots$ over $\mathcal{A}$ such that

$$\boldsymbol{s} = Pal(\Delta) = \lim_{n\to\infty} Pal(x_1x_2\cdots x_n).$$

Example

$\Delta = (abc)^\infty = abcabcabc\cdots$ directs the so-called *Tribonacci word*:

$$\boldsymbol{r} = \underline{a}\underline{b}a\underline{c}aba\underline{a}bacaba\underline{b}acabaabacaba\underline{c}abaabaca\cdots$$

All such words are known to have linear complexity.

Question: What other sorts of complexity functions are possible for rich words?

On the complexity of rich words

The *palindromic factor complexity* of a finite or infinite word $\boldsymbol{w}$, denoted by $\mathcal{P}_{\boldsymbol{w}}(n)$, counts the number of distinct palindromic factors of $\boldsymbol{w}$ of each length n.

Bucci, De Luca, G., Zamboni (2008) established the following connection between palindromic richness and complexity.

Theorem

For any infinite word $\boldsymbol{w}$ whose set of factors is closed under reversal, the following conditions are equivalent:

i) *all complete returns to palindromes in $\boldsymbol{w}$ are palindromes;*

ii) $\mathcal{P}_{\boldsymbol{w}}(n) + \mathcal{P}_{\boldsymbol{w}}(n+1) = \mathcal{C}_{\boldsymbol{w}}(n+1) - \mathcal{C}_{\boldsymbol{w}}(n) + 2$ *for all* $n \in \mathbb{N}$.

This result can be viewed as a characterisation of recurrent rich infinite words since any rich infinite word is recurrent if and only if its set of factors is closed under reversal.

- From the preceding theorem, we deduce that any infinite word with (sub)linear factor complexity has bounded palindromic complexity since the first difference $\mathcal{C}(n+1) - \mathcal{C}(n)$ is bounded for any such infinite word.

- Many examples of rich infinite words have sublinear factor complexity, such as (epi)Sturmian words and periodic rich infinite words.

  The latter take the form $v^{\infty} = vvvv\cdots$ where $v = pq$ and all circular shifts of v are rich.

- There also exist recurrent rich infinite words with non-sublinear complexity, but such words are not as easy to find.

  For example, the infinite word generated by iterating the morphism: $a \mapsto abab$, $b \mapsto b$ on the letter a, namely

$$abab^2abab^3abab^2abab^4abab^2abab^3abab^2abab^5\cdots,$$

  is a recurrent rich infinite word whose factor complexity grows quadratically with n.

- Another example that was indicated to us by J. Cassaigne is the infinite word word generated by iterating $a \mapsto aab$, $b \mapsto b$ on the letter a:

$$aabaabbaabaabbbaabaabbaabaabbbbaabaabbaabaabbbaabaabbaabaabbbbb\cdots$$

It is a recurrent rich infinite word and its complexity is equivalent to $n^2/2$.

- In the case of non-recurrent infinite words, the rich word $aba^2ba^3ba^4ba^5b\cdots$ has factor complexity of the order n^2.

  More generally, if $f(n) = n^k$ for some constant k, then the infinite word $a^{f(1)}ba^{f(2)}ba^{f(3)}b\cdots$ is rich and has factor complexity of the order $n^{1+1/k}$.

- Actually, we can obtain all kinds of interesting complexity functions in between linear and quadratic.

  For instance, if we let $f(n)$ be any strictly increasing function taking positive integer values, then the word

$$a^{f(1)}ba^{f(2)}ba^{f(3)}b\cdots$$

  is rich with factor complexity depending on f.

Open Question: *Does there exist a rich infinite word with exponential factor complexity, or even anything more than quadratic?*

Enumeration of rich words

- Can we determine a closed formula for the number of rich words of length n over an arbitrary finite alphabet?
- Let $R_k(n)$ denote the number of rich words of length n over a k-letter alphabet.
- Trivial upper bound: $R_k(n) \leq k^n$ (where k^n is the number of words of length n over a k-letter alphabet).
- It is easy to check that all words of length at most 3 are rich, i.e., $R_k(n) = k^n$ for $n = 1, 2, 3$.
- However, for large n, there are far fewer rich words than k^n.
- The following table enumerates $R_k(n)$ for the first few values of k and small n.

$R_k(n)$: number of rich words of length n over a k-letter alphabet

$k \backslash n$	0	1	2	3	4	5	6	7	8	9	10	11
2	1	2	4	8	16	32	64	128	252	488	932	1756
3	1	3	9	27	75	201	513	1269	3033	7047	15903	
4	1	4	16	64	232	784	2464	7336	20776			

- Do you notice any pattern? Can you determine a formula for $R_k(n)$? No?!
- It's a difficult (open) problem.
- Sloane's Online Encyclopaedia of Integer Sequences (OEIS) gives nothing!

Polynomial growth

What we can show is that the number of rich words of length n grows at least polynomially with the size of the alphabet $\mathcal{A}$.

Theorem

Let $\mathcal{A}$ be a finite alphabet consisting of at least 3 letters. Then the number of rich words of length n over $\mathcal{A}$ grows at least polynomially with the size of $\mathcal{A}$.

That is to say, by going to larger and larger alphabets, we get polynomial growth of arbitrarily high degree.

Basic idea

Let $\mathcal{A}$ be a k-letter alphabet with $k \geq 3$.

For any fixed letter $a \in \mathcal{A}$, we define the insertion morphism

$$\phi_a : x \mapsto xa \quad \text{for all letters } x \in \mathcal{A}.$$

Any such morphism preserves richness, i.e., $\phi_a(w)$ is rich for any finite or infinite rich word w [G., Justin, Widmer, Zamboni, 2009].

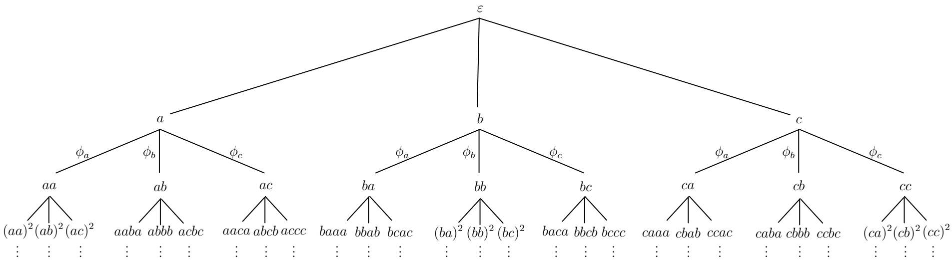
Now we construct an infinite k-ary tree rooted at the empty word ε (level 0) as follows.

- The k nodes at level 1 are the k letters of $\mathcal{A}$.
- For every $n \geq 2$, the k^n nodes at level n are obtained by applying the k insertion morphisms on $\mathcal{A}$ to each of the k^{n-1} nodes at level $n-1$.

For example, when $\mathcal{A} = \{a, b, c\}$, we use the following three insertion morphisms:

$$\phi_a : \begin{cases} a & \mapsto & aa \\ b & \mapsto & ba \\ c & \mapsto & ca \end{cases}, \qquad \phi_b : \begin{cases} a & \mapsto & ab \\ b & \mapsto & bb \\ c & \mapsto & cb \end{cases}, \qquad \phi_c : \begin{cases} a & \mapsto & ac \\ b & \mapsto & bc \\ c & \mapsto & cc \end{cases}$$

to obtain the following infinite ternary tree . . .



- Since insertion morphisms preserve richness, all nodes in the tree are rich words.
- And the rich words at level n have length 2^{n-1} since $|\phi_x(w)| = 2|w|$ for any letter x and word w.
- Moreover, no rich word appears more than once in the tree.

  This follows from the injectivity of insertion morphisms (i.e., $\phi_a(w) = \phi_a(w')$ if and only if $w = w'$) together with the fact that

  $$\phi_a(w) = \phi_b(w) \text{ for some word } w \text{ if and only if } a = b.$$

- Thus, at the n-th level in the tree, there are exactly 3^n distinct rich words of length 2^{n-1}.
- More generally, when $|\mathcal{A}| = k$, the n-th level of the infinite k-ary tree consists of exactly k^n distinct rich words of length 2^{n-1} for each $n \in \mathbb{N}^+$.

Asymptotic exponential growth on two letters

- On a 2-letter alphabet, we can show asymptotic exponential growth of rich words, as follows.
- Let n be a positive integer and let

$$n = n_1 + n_2 + \cdots + n_k$$

be any partition of n into k parts, where we write the parts in non-decreasing order $n_1 \leq n_2 \leq \cdots \leq n_k$.

- Then the word

$$a^{n_1} b a^{n_2} b \cdots a^{n_k} b b^{n-k}$$

is easily verified to be a rich word of length $2n$.

- Furthermore, every partition of n results in a unique rich word w of length $2n$.
- For every $n \geq 1$, the number of rich words of length $2n$ produced in this way is thus equal to the number $p(n)$ of partitions of n.

Asymptotic exponential growth on two letters . . .

A classical result of Hardy and Ramanujan (1918) gives the asymptotic expansion:

$$p(n) \sim \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}.$$

Thus the number of rich words of length $2n$ grows at least as fast (asymptotically) as the above expression for $p(n)$.

Open Question: *Are there exponentially many rich words of each length?*

Thank You!