Words, Words, I Am So Sick of Words!

Joseph Malkevitch
Mathematics (York College)

What is mathematics about?

The study of numbers.

The study of shapes.

But it should be viewed as:

The study of patterns.

When we see:

12334421

we usually view it as a number (base 10 or base 5).

When we think about numbers we think about whether they are real, integers, complex, or rational. If an integer, we ask if the number is prime, a perfect square, etc.

However we could view 12334421 as a *word* or a *string* from the alphabet:

a. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

or

b. 1, 2, 3, 4

What patterns do you notice for the string

12334421

Some answers:

Each letter in the alphabet appears twice.

There are two squares: 33; 44

12 appears as a prefix but not a suffix.

21 appears as a suffix.

"Multiplying" or "combining" two strings means that they are concatenated or juxtaposed.

Examples:

x multiplied by xz is xxz or x^2z

xz multiplied by x is xzx

Multiplication does NOT commute.

Vocabulary:

Factor or subword: xz is a factor of zx^2zyx

Square: xyxy is a square as is uu

Cube: abcabcabc is a cube as is u^3

Palindrome: racecar or 121 (read the same forwards and backwards)

Prefix: xz is a prefix of xz^3y

Suffix: zy is a suffix of xz^3y

What area of mathematics do such problems belong to?

Generally:

Discrete mathematics

Combinatorics

More specifically: ***Combinatorics on words***

Norwegian Mathematician:

(Axel Thue (1863-1922))

Also the American mathematician Marston Morse (1892-1977).

Originally developed as theoretical mathematics combinatorics on words has increasingly many applications in biology:

DNA can be viewed as strings from the 4 letter alphabet:

A, C, G, T

with special rules about how pairing occurs!

What pattern do you think generated these numbers?

1, 1, 2, 3, 5, 8, 13, 21, 34, ... ,

These are the famous Fibonacci numbers, where the next term is obtained by summing the previous two. They arise as a way to mode the population of rabbit pairs that don't ever die.

We can express this using a recursion equation:

$$F_n = F_{n-1} + F_{n-2}$$

using the initial conditions: 1 and 1

Note: Since integers commute we also have:

$$F_n = F_{n-2} + F_{n-1}$$

Using analogy we can construct a sequence of Fibonacci words, using two letters, and the "multiplication" given by the recursion, where we concatenate strings.

A next word is gotten by concatenating the previous two:

a, ab, aba, abaab, abaababa, ... ,

The recursion being:

$$f_n = f_{n-1} f_{n-2}$$

What about using:

$$f_n = f_{n-2} f_{n-1}$$

a, ab, aab, abaab, aababaab, ... ,

This is not the same as the prior sequence, because here multiplication does not commute. Does it have interesting properties like the Fibonacci words?

Some properties of the Fibonacci sequence of words, and the "infinite" word formed so that each word of the sequence is a prefix of the infinite word:

$$f = abaababaabaab\ldots$$

* The words of the Fibonacci sequence end alternately in ab and ba

* Suppressing the last two letters of a Fibonacci word, or prefixing the complement of the last two letters, creates a palindrome.

Example: abaababa
abaaba is a palindrome
ababaababa is a palindrome

* The subwords 11 and 000 never occur

* As the size of the initial section of the infinite word grows, the ratio of the number letters to the number of a's and the ratio of the number of a's to b's approaches the Golden Ratio, phi, $(1 + \sqrt{5})/2$

* The infinite Fibonacci word is recurrent; every subword occurs infinitely often.

* If w is a subword of the infinite Fibonacci word, then so is w's reversal

* Two factors of the same length anywhere in the infinite Fibonacci word have the property that the difference between the number of b's they contain never exceeds 1

Question:

Are there infinite length words in an alphabet A that avoid squares?

Easy theorem:

Words over any alphabet with two symbols contain squares!

Amazing theorem:

Axel Thue:

There is a square free word sequence on an alphabet of three letters.

A word is *cube free* if it contains no factor of the form uuu, where u is a finite non-empty word.

A word is *strongly cube free* if it contains no subsequence of the form uua, where u is a finite non-empty word and a is the first symbol in u.

Theorem (Axel Thue)

The Thue-Morse sequence is strongly cube free.

There are many approaches to the Thue-Morse sequence. This one is rather appealing because it involves simple ideas from discrete mathematics and counting in binary.

Let TM_n to denote the nth term of the Thue-Morse sequence where n takes on the values 0, 1, 2, 3, ...

To find the value of TM_n write n in binary.

If this number has an even number of ones set TM_n to 0, otherwise to 1!

n in decimal	*n* in binary	Number of 1's mod 2
0	0	0
1	1	1
2	10	1
3	11	0
4	100	1
5	101	0
6	110	0
7	111	1
8	1000	1
9	1001	0
10	1010	0
11	1011	1
12	1100	0
13	1101	1
14	1110	1
15	1111	0
16	10000	1

So the Thue-Morse sequence is without commas between terms to show an infinite word:

01101001100101101001011001101001....

Note: lots of squares but no cubes.

Here is another lovely approach using the idea of replacing a seed word using morphisms, functions which replace symbols in the alphabet with other symbols.

Morphism Rule:

Replace 0 in the string by 01 and 1 in the string by 10. Use 0 as your starting string. ($0 \longrightarrow 01$ and $1 \longrightarrow 0$)

0
01
0110
01101001
0110100110010110

Is there a morphism approach to the Fibonacci word?

Yes!

$$a \rightarrow ab$$
$$b \rightarrow a$$

$$a \rightarrow ab$$
$$b \rightarrow a$$

a

ab

aba

abaab

abaababa

abaababaabaab

Nifty!

Bibliography:

Crochemore, M. and W. Rytter, Text Algorithms, Oxford University Press, 1994.

Crochemore, M. and C. Hancart, T. Lecroq, Algorithms on Strings, Cambridge U. Press, 2007.

Gusfield, D., Algorithms on Strings, Trees, and Sequences, Cambridge U. Press, New York, 1997.

Lothaire, M., Combinatorics on Words, Addison-Wesley, 1983.

Lothaire, M., Algebraic Combinatorics on Words, Cambridge University Press, New York, 2002.

Lothaire, M. Applied Combinatorics on Words, Cambridge U. Press, New York, 2005.