Exploring Thue’s 1914 paper

On the transformation of strings according to given rules

James Power
Department of Computer Science,
National University of Ireland, Maynooth

HaPoC, October 29, 2013
Thue’s 1914 Paper

Why this paper?
- Source of Thue (and Semi-Thue) systems
- Nearly 100 years old
- Translation?
Exploring Thue’s 1914 paper: Introduction

Models of Computation: 1936 - the annus mirabilis

1900 1910 1920 1930 1940 1950 1960

A. Church
A.M. Turing
E.L. Post
S.C. Kleene
K. Godel
Some questions about Thue’s paper

1. What is Thue’s 1914 paper about?
2. Why was Thue doing this in 1914?
3. How did this paper influence later work?
RECURSIVE UNSOLVABILITY OF A PROBLEM OF THUE

EMIL L. POST

Alonzo Church suggested to the writer that a certain problem of Thue [6] might be proved unsolvable by the methods of [5]. We proceed to prove the problem recursively unsolvable, that is, unsolvable in the sense of Church [1], but by a method meeting the special needs of the problem.
Influence of Thue’s paper

- Hits the big time in 1947
  - First example of an undecidable problem outside the class of ’36
Influence of Thue’s paper

- Hits the big time in 1947
  - First example of an undecidable problem outside the class of ’36

- But:
  - Post’s paper only depends on a small section (1-2 pages) of Thue’s paper
Influence of Thue’s paper

- Hits the big time in 1947
  - First example of an undecidable problem outside the class of ’36

- But:
  - Post’s paper only depends on a small section (1-2 pages) of Thue’s paper
  - Not mentioned in Church’s bibliography (1936)
  - Not mentioned by Markov (1947)
  - Not mentioned by Chomsky (1950s)
Influence of Thue’s paper

- Hits the big time in 1947
  - First example of an undecidable problem outside the class of ’36

- But:
  - Post’s paper only depends on a small section (1-2 pages) of Thue’s paper
  - Not mentioned in Church’s bibliography (1936)
  - Not mentioned by Markov (1947)
  - Not mentioned by Chomsky (1950s)
  - Not mentioned by ...

Axel Thue: Biography

- 1863: Born Tønsberg, Norway
- 1889: Degree at Univ. of Oslo
- 1891-2: visited Leipzig and Berlin
- 1894: teacher of mechanics, Trondheim technical college
- 1903 Prof. of Applied Mechanics, Univ. of Oslo

From “Axel Thue” by Viggo Brun, Selected Mathematical Papers, 1977
Axel Thue: Works on symbol-sequences

Thue published *four* papers on “strings”:

- Über unendliche Zeichenreihen, 20 pp., 1906.
- Die Lösung eines Spezialfalles eines generellen logischen Problems, 38 pp., 1910.
- Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, 65 pp., 1912.
- Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, 32 pp., 1914.

All published in *Christiana Videnskabs-Selskabs Skrifter*
Axel Thue: Works on symbol-sequences

Thue published *four* papers on “strings”:

- About infinite sequences of symbols, 20 pp., 1906.
- Die Lösung eines Spezialfalles eines generellen logischen Problems, 38 pp., 1910.
- On the relative position of equal parts in certain sequences of symbols, 65 pp., 1912.
- Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, 32 pp., 1914.

All published in *Christiana Videnskabs-Selskabs Skrifter*

*Axel Thue’s papers on repetitions in words: a translation*
Axel Thue: Works on symbol-sequences

Thue published *four* papers on “strings”:

- About infinite sequences of symbols, 20 pp., 1906.
- The solution of a special case of a general logical problem, 38 pp., 1910.
- On the relative position of equal parts in certain sequences of symbols, 65 pp., 1912.
- Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, 32 pp., 1914.

All published in *Christiana Videnskabs-Selskabs Skrifter*

*Trees and term rewriting in 1910: On a paper by Axel Thue*
Thue published *four* papers on “strings”:

- About infinite sequences of symbols, 20 pp., 1906.
- The solution of a special case of a general logical problem, 38 pp., 1910.
- On the relative position of equal parts in certain sequences of symbols, 65 pp., 1912.
- Problems on the transformation of sequences of symbols according to given rules, 32 pp., 1914.

All published in *Christiana Videnskabs-Selskabs Skrifter*

*Thue’s 1914 paper: a translation*
J. Power, arXiv:1308.5858
Thue Systems

A Thue system: Given some fixed alphabet, a sequence of pairs of the form:

\[ A_1 \leftrightarrow B_1 \]
\[ A_2 \leftrightarrow B_2 \]
\[ \ldots \]
\[ A_n \leftrightarrow B_n \]

Equivalence: For any strings \( P \) and \( Q \), we write \( P = Q \) when \( P \) can be transformed into \( Q \) by a series of operations, each involving a substitution of some \( A_i \) for \( B_i \) (or vice versa).
Thue and the general context

From §1:

- In this paper I will deal with a problem concerning the transformation of symbol sequences using rules.
- This problem, [...] is a special case of one of the most fundamental problems that can be posed.
- Since this task seems to be extensive and of the utmost difficulty, I must be satisfied with only treating the question in a piecewise and fragmentary manner.
Thue and decision problems

Word problem for semi-groups: pg. 494
For any arbitrary given sequences $A$ and $B$, to find a method, where one can always decide in a predictable number of operations, whether or not two arbitrary given symbol sequences are equivalent in respect of sequences $A$ and $B$.

Word problem for monoids: pg. 498
Given an arbitrary sequence $R$, to find a method where one can always decide in a finite number of investigations whether or not two arbitrary given sequences are equivalent with respect to $R$. 
Word problem for semi-groups: pg. 494
For any arbitrary given sequences $A$ and $B$, to find a method, where one can always decide in a predictable number of operations, whether or not two arbitrary given symbol sequences are equivalent in respect of sequences $A$ and $B$.

Word problem for monoids: pg. 498
Given an arbitrary sequence $R$, to find a method where one can always decide in a finite number of investigations whether or not two arbitrary given sequences are equivalent with respect to $R$. 
Das Identitätsproblem: 
Suppose some element of a group is given in terms of the composition of its generators, we want to give a method to decide in a finite number of steps, whether it is the identity element or not.

Das Transformationsproblem: 
Given any two elements of a group $S$ and $T$, we want a method to answer the question whether $S$ and $T$ can be transformed into one another...

Max Dehn, *Über unendliche diskontinuierliche Gruppen* 
Mathematische Annalen 71(3), 1911.

Given a diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: to give a process according to which it can be decided in a finite number of operations whether the equation is solvable in rational integers.

David Hilbert, *Mathematische Probleme* 
Göttinger Nachrichten, 1900, pp. 253-297
Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?

A: Yes, $aabbcab = abbcaba$.

(This is equality modulo $R$)
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?

$aabbcab$
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$.
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?

$\text{aabbcab}$
Alternative formulation:

- Instead of rules we’re given some *identity string*, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?

\[ abbcaba \]
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?

$abbcaba$
From semi-groups to monoids

Alternative formulation:

- Instead of rules we’re given some identity string, $R$
- Then we define $P = Q$ when we can get from $P$ to $Q$ by inserting/deleting occurrences of $R$.

Example: let $R \equiv abbcab$.

Q: is $aabbcab = abbcaba$?

A: Yes, $aabbcab = abbcaba$.

(This is equality modulo $R$)
Overlaps

Problem: What do we do if there are overlaps?

- Example: let $R \equiv abbcab$, as before.
- Consider: $acabbcabbcabca$
- Option #1: $acabbcabbcabca$
- Option #2: $acabbcabbcabca$

Note: this is only possible because $R$ itself has an overlap: $R \equiv abbcababbcababbcabca$

c.f Thue 1906/1912: producing overlap-free strings
Problem: What do we do if there are overlaps?

- Example: let $R \equiv abbcab$, as before.
- Consider: $acabbcabbcabca$
- Option #1: $acabbcabbcabca$
- Option #2: $acabbcabbcabca$
- Note: this is only possible because $R$ itself has an overlap:
  
  $R \equiv abbcab$
  
  $abbcab \equiv R$

  c.f Thue 1906/1912: producing overlap-free strings
Overlaps

Problem: What do we do if there are overlaps?

- Example: let $R \equiv abbcab$, as before.
- Consider: $acabbcabbcabca$
- Option #1: $acabbcabbcabca$
- Option #2: $acabbcabbcabca$

Note: this is only possible because $R$ itself has an overlap:

$$R \equiv abbcab$$

$$abbcab \equiv R$$

C.f Thue 1906/1912: producing overlap-free strings
Overlaps: general statement

Thue asks: what kind of strings overlap themselves?

\[ R \equiv abbcab \]

\[ abbcab \equiv R \]
Thue asks: what kind of strings overlap themselves?

\[ R \equiv abbcab \]
\[ abbcab \equiv R \]

They must have the general form

\[ R \equiv CU \]
\[ UD \equiv R \]
Thue asks: what kind of strings overlap themselves?

\[ R \equiv abbcab \]

\[ abbcab \equiv R \]

They must have the general form

\[ R \equiv CU \]

\[ UD \equiv R \]

But then

\[ CR \equiv CUD \]

\[ \equiv RD \]
Thue asks: what kind of strings overlap themselves?

\[ R \equiv abbcab \]

\[ abbcab \equiv R \]

They must have the general form

\[ R \equiv CU \]

\[ UD \equiv R \]

But then

\[ CR \equiv CUD \]

\[ \equiv RD \]

That is, \( C \equiv D \)

(i.e \( C \) and \( D \) are equal modulo \( R \))
Overlaps: the big picture
Overlaps: the big picture
Overlaps: the big picture
Overlaps: the big picture

\[ C \quad R \quad D \]

\[ C \quad C \quad \alpha \quad \beta \quad C \quad C \quad C \quad \alpha \]

James Power, NUI Maynooth  
HaPeC 2013
Overlaps: the big picture

So we have $C \equiv \alpha \beta$ and $D \equiv \beta \alpha$
Overlaps: the big picture

Which means

\[ R \equiv C^n \alpha \equiv (\alpha \beta)^n \alpha \equiv \alpha \beta \alpha \beta \alpha \cdots \alpha \beta \alpha \equiv \alpha (\beta \alpha)^n \equiv \alpha D^n \]
And thus the overlap

\[ U \equiv (\alpha \beta)^{n-1} \alpha \equiv \alpha (\beta \alpha)^{n-1} \]
Overlaps: Euclid?

And thus the overlap

\[ U \equiv (\alpha \beta)^{n-1} \alpha \equiv \alpha (\beta \alpha)^{n-1} \]

But now \( U \) has the form \( U \equiv C_1 U_1 \equiv U_1 D_1 \)
A “Euclidean algorithm” for strings

Given some string $R$ we can thus define strings $U_i$, $C_i$ and $D_i$ such that:

\[
R \equiv U_0 \equiv C_1 U_1 \equiv U_1 D_1 \\
U_1 \equiv C_2 U_2 \equiv U_2 D_2 \\
\vdots \\
U_{r-1} \equiv C_r U_r \equiv U_r D_r
\]

with each $U_i$ maximal where two subsequences $U_r$ in the sequence can never have a common part.
An algorithm to generate equations from an identity string:

- Given an identity string $R \equiv abbcab$, as before.

  First overlap:
  
  $R \equiv abbcab$
  
  $abbcab \equiv R$
  
  So $abbc = bcab$.

  Second overlap:
  
  $R \equiv abbcab$
  
  $bcabab \equiv R$
  
  So $abbca = cabab$.
Thue’s “completion algorithm”: example

An algorithm to generate equations from an identity string:

- Given an identity string $R \equiv abbcab$, as before.
- First overlap:

  $R \equiv abbcab$
  
  $abbcab \equiv R$
Thue’s “completion algorithm”: example

An algorithm to generate equations from an identity string:

- Given an identity string $R \equiv abbcab$, as before.
- First overlap:

  $R \equiv abbcab$

  $abbcab \equiv R$

  So $abbc = bcab$
An algorithm to generate equations from an identity string:

- **Given an identity string** \( R \equiv abbcab \), as before.

- **First overlap:**
  \[
  R \equiv abbcab \\
  abbcab \equiv R
  \]
  So \( abbc = bcab \)

- **Second overlap:**
  \[
  R \equiv abbcab \\
  bcabab \equiv R
  \]
Thue’s “completion algorithm”: example

An algorithm to generate equations from an identity string:

- Given an identity string \( R \equiv abbcab \), as before.

  \[
  \begin{align*}
  R &\equiv abbcab \\
  abbcab &\equiv R \\
  \end{align*}
  \]

  So \( abbc = bcab \).

- Second overlap:

  \[
  \begin{align*}
  R &\equiv abbcab \\
  bcabab &\equiv R \\
  \end{align*}
  \]

  So \( abbca = cabab \).
Capture a time when algebra moves to language theory
Important topics, clearly addressed
Algorithms: Euclidean, completion, ‘regular’-like, ...

But: de-coupled from e.g. Hilbert, Post, ...
“Passed by reference” into computing history