Exploring Thue's 1914 paper on the transformation of strings according to given rules

James F. Power
Dept. of Computer Science, NUI Maynooth

To be presented at HaPoC 2013

Rarely has any paper in the history of computing been given such a prestigious introduction as that given to Axel Thue's paper by Emil Post in 1947 [Pos47]:

> "Alonzo Church suggested to the writer that a certain problem of Thue [Thu14] might be proved unsolvable ..."

However, only the first two pages of Thue's paper are directly relevant to Post's proof, and, in this abstract, I hope to shed some light on the remaining part, and to advocate its relevance for the history of computing.

Thue Systems Thue's 1914 paper is the last of four he published that directly relate to the theory of words and languages [Ber95, ST00]. In this 1914 paper, Thue introduces a system consisting of pairs of corresponding strings over a fixed alphabet:

$$\begin{array}{ccccc} A_1, & A_2, & A_3, & \ldots, & A_n \\ B_1, & B_2, & B_3, & \ldots, & B_n, \end{array}$$

and poses the problem: given two arbitrary strings P and Q, can we get one from the other by replacing some substring A_i or B_i by its corresponding string? Post called these systems of "Thue type" and proved this problem to be recursively unsolvable.

Reception of Thue's Work Thue's earlier work was not widely cited but often rediscovered independently [Hed67], and something similar seems to have happened with the 1914 paper.

For example, Thue is not among the 547 authors in Church's 1936 *Bibliography of Symbolic Logic*, nor is Thue cited in Post's major work on tag systems, correspondence systems, or normal systems before 1947. His work appears to have had no direct influence on the development of formal grammars by Chomsky in the 1950s. Most subsequent references to Thue's paper (where they exist) note it only for providing a definition of Thue systems.

Thue's awareness Thue explicitly understood the general meta-mathematical context (that we now associate with *Hilbert's programme*), describing the problem as being of relevance to one of the "most fundamental problems that can be posed".

Further, he phrases the problem in terms that have become quite familiar in the post-1936 world:

> "... to find a method, where one can always calculate in a predictable number of operations, ..."

This language parallels that used in Hilbert's 10th problem in 1900 and places Thue's work firmly in what we would now regard as computing, rather than pure algebra.

Foundations of Language Theory Having posed the general problem in §II of his paper, Thue then presents an early example of a proof of (what we would now call) *termination* and *local confluence* for a system where the rules are non-overlapping and non-increasing in size.

When reducing some string P, we must find some occurrence of A_i and replace it with B_i. A difficulty arises if there is an overlap: some substring CUD in P, such that A_i matches both CU and UD, and thus choosing one option will eliminate our ability to later choose the other.

In §IV, Thue presents the string U as a *common divisor* of CU and UD and then shows how we can apply **Euclid's algorithm** to derive a Thue system from this. Euclid's algorithm had been considerably generalised throughout the 19th century, but here the string U "measures" the strings CU and UD just as Euclid's lines measure each other (*Elements*, Book 10, proposition 3).

Thue derives another *algorithm* in §V which, given two strings P and Q will derive those strings equivalent to them, and gradually reduce them to a core set of irreducible strings, providing a solution to the word problem in a restricted case. He investigates variants of these presentations *based on their syntactic properties* in §VI and gives some examples in §VII.

We remark that from the identity $CU \equiv UD$ we can derive rules of the form $CU \to UD$, and that this template is precisely what Post termed *normal form* for his rewriting systems.

Thue's "completion" algorithm In §VIII of his paper Thue develops an algorithm to derive a system of equations from any given sequence R. This is interesting not just for its structure (the algorithm iterates until it reaches a fixed point) but also for its use of overlapping sequences as a generation mechanism.

Starting from some given identity sequence R we can identify all pairs where $R \equiv CU \equiv UD$, and then add the rules $C \leftrightarrow D$ to the Thue system. We can then apply these rules using R as a starting symbol to derive a further set of identity sequences $R_1, R_2, \ldots$. These, in turn, can be factored based on overlaps to provide a further set of rules $C_i \leftrightarrow D_i$ and so on. Since all R_i have the same length, as do all C_i and D_i, this process is guaranteed to terminate.

This is similar to, but *not* the Knuth-Bendix algorithm: there is no explicit concept of well-ordering, for example. However, it certainly contains many of the "basic features" of the algorithm as described by Buchberger [Buc87], and could be considered, under restrictive conditions, as an embryonic version of it.

References

[Ber95] Jean Berstel. Axel Thue's papers on repetitions in words: a translation. Publications du LaCIM, Université du Québec à Montréal, 1995.

[Buc87] Bruno Buchberger. History and basic features of the critical-pair/completion procedure. *Journal of Symbolic Computation*, 3(12):3–38, 1987.

[Hed67] G.A. Hedlund. Remarks on the work of Axel Thue on sequences. *Nordisk Matematisk Tidskrift*, 15:148–150, 1967.

[NSST77] Trygve Nagell, Atle Selberg, Sigmund Selberg, and Knut Thalberg, editors. *Selected Mathematical Papers of Axel Thue*. Universitetsforlaget, Oslo, 1977.

[Pos47] Emil L. Post. Recursive unsolvability of a problem of Thue. *Journal of Symbolic Logic*, 12(1):1–11, March 1947.

[ST00] M. Steinby and W. Thomas. Trees and term rewriting in 1910: On a paper by Axel Thue. *EATCS Bull.*, 72:256–269, 2000.

[Thu14] Axel Thue. Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln. *Christiana Videnskabs-Selskabs Skrifter, I. Math.-naturv. Klasse*, 10, 1914. Reprinted in [NSST77, pp. 493-524].