

# Tarzan and the Learning of Language from Text

Yoav Seginer

## 1 Introduction

While all readers of this volume are probably familiar with Dick de Jongh's work in logic, perhaps not all are aware of Dick's interest in language learning and language acquisition. It was at the end of 2001 that Dick first introduced me to this problem and convinced me it was worthwhile studying. Since then, he has been guiding my work in this field. Now, it is my turn to try to bring some of you just a little closer to the mysteries of language learning. I will try to do so not by conjuring a set of definitions, but by awakening childhood memories.

To the untrained ear, *language learning* and *language acquisition* may sound synonymous, but they are not (or at least, possibly not): while *language learning* may refer to any manner of learning, as long as it results in some linguistic skill, *language acquisition* is restricted to what humans actually learn and to how they actually do it. Until the invention of the computer, language learning beyond language acquisition remained of purely theoretical interest (and of little interest at that). But today, the hope that computers could be taught to perform some linguistic tasks to our satisfaction seems to hinge on the possibility of learning language in a way not quite identical to the way humans do it. The reason is simple - the computer could never be so fully steeped in life's experiences as a child is. This brings to mind Tarzan of the Apes, because Tarzan, though fully a part of jungle life, learned his English from books in the total seclusion of his parents' cabin - like a computer closed inside its box. And there we have another similarity between Tarzan and the modern computer (which shares Tarzan's speed and agility, but not his cleverness): they both learn language from text - the printed words of books rather than the spoken words of people.

Computer scientists would certainly be greatly encouraged in their pursuit of language learning algorithms if they only knew that it could be done. That children can do it is already something, but to know that Tarzan did it, with little to guide him but a large library, would certainly be much better. Of course, even Edgar Rice Burroughs, the author of the Tarzan books, did not claim that the story was necessarily true; but he was convinced that it could be true. He felt that for him that was enough. So should it be for us.

Many over the age of twelve find Tarzan's story too incredible to warrant any serious discussion. Whether it is at all humanly possible to swing from one tree branch to another while holding Jane Porter in one arm, I cannot say, though this may be settled by a mere calculation. The linguistic adventures of Tarzan, however, I hope to show to be far more subtle than we tend to remember them and not as easily refutable as many think. Perhaps it never happened; perhaps it could never have happened; but it may very well be believed, at least for a short while.

## 2 Tarzan's Progress

Burroughs's account of Tarzan's linguistic development, while brief, portrays the process from beginning to end. At the age of ten, Tarzan first discovered the books in his parents' cabin and deciphered the meaning of the first word. The word was BOY and the deciphering method was simple: "Beneath the picture [of a boy] were three little bugs - B,O,Y. And now he had discovered in the text upon the page that these three were repeated many times in the same sequence." He quickly realized that these were "the three little bugs which always accompanied the little ape [i.e. boy]." This seems to make perfect sense to us as it indeed must have seemed to have made perfect sense to Burroughs, for he dedicates more than his customary few lines to the description of the process. It is not surprising, as this is essentially no different from other theories, going back at least as far as St. Augustine <sup>1</sup>, which attempt to explain the acquisition of word meaning.

It took Tarzan almost five years to acquire a basic vocabulary, but "by the time he was fifteen he knew the various combinations of letters which stood for every pictured figure in the little primer and in one or two of the picture books". A basic vocabulary he had, but "of the meaning and use of the articles and conjunctions, verbs and adverbs and pronouns he had but the faintest conception." Another two years passed and "by the time he was seventeen he had learned to read the simple, child's primer and had fully realized the true and wonderful purpose of the little bugs" (including the alphabetic order). Only a year later, at eighteen, "he read fluently and understood nearly all he read in the many and varied volumes on the shelves." In no more than three years, Tarzan, with the help of Burroughs (or Burroughs with the help of Tarzan?) managed to leap from the understanding of single words associated with simple pictures to a complete command of the language. How complete? Enough to read fluently. And why did he understand only "nearly all" he read? One can come up with different explanations, but isn't it simply that in the library of Lord Greystoke there were certain to be more than several books which would have proved a challenge even to those equipped with the very best education?

Many wondered whether Tarzan could have done it, but we wish to know the answer to a far more interesting question: How did Tarzan do it? Burroughs does not tell us and it may very well be that he did not know. He could have left it at that, adopting the popular learning theory that one should take care of the words and the rest will take care of itself. But Burroughs did feel obliged to give some explanation and therefore tells us that in the last stages of learning, Tarzan's progress was rapid "with the help of the great dictionary and the active intelligence of a healthy mind endowed by inheritance with more than ordinary reasoning powers...". To be certain, when Burroughs writes of inheritance he does not refer to the superiority of man over ape but to that of an (English) nobleman over all other living things, and in this deviates from many current theories of language acquisition.

What Tarzan has learned in those three years, except for enriching his knowledge of word meanings, must have included the syntax of the language. Whether he indeed understood all he read, we have nothing but Tarzan's word for, and whether he understood it *correctly*, we will never know. But that his command

---

<sup>1</sup>As quoted on the first page of Wittgenstein's Philosophical Observations.

of syntax was sufficient to read fluently we can learn from the notes and letters he produced when first coming into contact with white people.<sup>2</sup>

### 3 Jungle Thoughts in English Syntax

We tend to remember Tarzan as the muscular jungle-man who upon first meeting the beautiful Jane Porter comments simply but forcibly: *Me Tarzan, you Jane*. But Tarzan never actually said that. In fact, when he and Jane Porter first met, Tarzan did not say a word - he could not speak a word of English. He could write however, and his first love letter is a masterpiece combining the best of jungle directness and English syntax. Yet Tarzan not only could never have said *me Tarzan, you Jane*, he could also never have written it, for where in all the vast jungle could he have found an example to construct such a sentence by, and how could he have produced it when all the English he had ever known was that found in the library of his father, an English nobleman? So rather than invent something completely new, *me Tarzan, you Jane* the like of which he never saw (or heard), Tarzan simply wrote "I am Tarzan of the Apes... You are Jane Porter ..." following carefully the numerous examples he saw in his books. Tarzan may have uttered stupidities, but not ingrammaticalities. To do otherwise would have required inventiveness to a degree which even Tarzan never possessed.

Burroughs seems to have understood this well. His Tarzan (unlike those that followed) blunders at least once in a basic misunderstanding of a speech act, but does so with the perfect grammar of a well-written book. When Tarzan reads the letter he had stolen from Jane Porter's desk in the cabin, he discovers that Jane thinks that *Tarzan of the Apes* who had left a note on the cabin door and the savage white man who had saved them from great dangers are two different people. Jane's confusion is understandable - the white man cannot speak English, while Tarzan of the Apes can write it. These two clearly cannot be one and the same. But Tarzan, of course, knows better. To put things right, Tarzan adds at the bottom of Jane's letter: "I am Tarzan of the Apes" - not forgetting to properly capitalize the "A" of "Apes". No *me Tarzan* is anywhere to be found. Had Tarzan handed back the letter to Jane pointing with his finger to his little addition at the bottom, he and Jane could have laughed together at her foolish mistake and ten chapters would have disappeared from the end of the book. But Tarzan does not seem to understand at all how necessary his personal presence is for the sentence "I am Tarzan" to work its effect. Instead, he stealthily returns the letter to Jane's desk and vanishes into the jungle. When Jane discovers the letter, all she learns is that Tarzan had been there again. Of the white savage she learns nothing, and Tarzan and he are left to roam together through the jungle but separately in Jane's mind.

What the story tells us (if it is true) or what Burroughs wishes to tell us (if he invented it) is simple: language cannot be perfectly learned from text alone; but syntax can, perhaps.

---

<sup>2</sup>Tarzan's earlier writings were produced on "pieces of bark and flat leaves and even smooth stretches of bare earth" and were therefore lost forever.

## 4 Conclusion

We can only conclude: the book is far better than the film. What we remember from the film is a muscular giant who could hardly speak English; but Burroughs, in his book, admires Tarzan for his extraordinary intellect, not for his strength (which he saw simply as necessary adaptation to the jungle life-style), and of all of Tarzan's intellectual achievements, the greatest was no doubt his learning of the English language through the diligent perusal of books. Burroughs never made the mistake of blessing Tarzan with speech, which he could never have learned, not even with an accent. Instead, he made Tarzan work hard to achieve a grammatically perfect command of the written language, with the necessary multitude of gaps in his understanding of its meaning and the manner in which it is to be used. It is easy to doubt the story, but even ninety years later, it is still difficult to prove it false. So did Tarzan really exist? Many a computer scientist surely hopes that even if Tarzan never was, it is still possible that he may have been.