It was, I think, Huxley who said that six monkeys, set to strum unintelligently on typewriters for millions of millions of years, would be bound in time to write all the books in the British Museum.

J. H. Jeans, 1930, *The Mysterious Universe*
Language as a Complex System

complexio (Latin): combination
plecto (Latin): folded
πλεκτω (Greek): to fold
plek (proto Indo-European root): folded once

“Humans possess a culturally transmitted system of unparalleled complexity in the nature world” [KH 01]
Direction

Word Frequency Analysis
- Zipf
- Mandelbrot

Word Length Analysis
- Mandelbrot
- Miller

Comments on Evolution Simulation
Zipfian Analysis 1935

\[ f_i = \frac{K}{i^A} \]

where \( K \) is a constant and \( f_i \) is the sample frequency of the \( i^{th} \) type in a ranking according to decreasing frequency

**EMPIRICAL**
Modelling – Mandelbrot 1953, 1962

Complexity in the Structure and Evolution of Language

John Bartholomew

\[ f_i = \left( \frac{N + \rho}{i + \rho} \right)^{1+\varepsilon} \]
Ln(Frequency) vs Ln(Rank) for Great Expectations

$y = -1.2047x + 11.041$

Experimental Data
Mandelbrot's Law
Linear (Experimental Data)

Complexity in the Structure and Evolution of Language  John Bartholomew
Experiment

Frequency vs Rank for Epistola Ad Pisones, De Arte Poetica

[Graph showing frequency vs rank with data points and two lines: one representing Zipf's Law and the other Experimental Data.]
Experiment

Ln(Frequency) vs Ln(Rank) for Epistola Ad Pisones, De Arte Poetica

\[ y = -0.8903x + 7.7 \]
Complexity in the Structure and Evolution of Language

John Bartholomew

Experiment

Frequency vs Rank for ABC Radio Transcript

Zipf's Law
Experimental Data
Mandelbrot's Extension
Complexity in the Structure and Evolution of Language

John Bartholomew

Experiment

Ln(Frequency) vs Ln(Rank) for ABC Radio Transcript

\[ y = -0.964x + 7.05 \]

Ln(Frequency)

Ln(Rank)

Experimental Data
Mandelbrot's Extension
Linear (Experimental Data)
A monkey, for example types away randomly at a typewriter with 26 letters and a spacebar.

Each key has an equal probability of being hit with the exception that the monkey is not allowed to type 2 spaces in a row.
Experiment

Frequency vs Length for 10 words
Experiment

Frequency vs Length for 100 words

- Y-axis: Frequency
- X-axis: Length
Experiment

Frequency vs Length for 250 words
Experiment

Frequency vs Length for 500 words

- Exponential Distribution
- Experimental Data
Experiment

Frequency vs Length for 1000 words

Exponential Distribution
Experimental Data
Experiment

Frequency vs Length for 10000 words
Experiment

Frequency vs Length for 100000 words

- Exponential Distribution
- Experimental Data
The probability that a word of length $i$ is typed:

$$P_i = p(space)p(letter)^i$$

The number of possible words of length $i$:

$$A^i$$

where $A$ is the number of letters in the alphabet, 26 in our case

Probability of a particular word of length $i$:

$$P(w, i) = \frac{P_i}{A^i} = \frac{p(space)p(letter)^i}{A^i}$$
Rank these words with increasing length:

Length = 1 ............ Rank [1 \rightarrow A]

Length = 2 ............ Rank \left[ A + 1 \rightarrow \frac{A(1-A^2)}{(1-A)} \right]

Length = 3 ............ Rank \left[ \frac{A(1-A^2)}{(1-A)} + 1 \rightarrow \frac{A(1-A^3)}{(1-A)} \right]

Using:

\[ \sum_{i=1}^{k} A^i = \frac{A(1-A^k)}{(1-A)} \]
Monkey Theory

\[ p(w, i) = \frac{p(\ast)}{p(L)} \left[ \frac{2(A - 1)}{A + 1} \left[ r(w, i) + \frac{A + 1}{2(A - 1)} \right] \right]^{\left[1 - \ln p(L) / \ln A\right]} \]

\[ p(w, i) = B \left( \frac{1}{r(w, i) + \rho} \right)^d \]

\[ B = 0.02 \quad \rho = 0.54 \quad d = 1.01 \]
Monkey Theory

Probability vs Length Rank

- Miller's Prediction
- Experimental results 1000 words
- Experimental results 10000 words

Probability vs Length Rank graph showing the comparison between Miller's Prediction and experimental results for 1000 and 10000 words.
Iterated Learning Model [KH 01]

- Parent teaches child/language meaning pairs
  Meaning Space \{jack, bob, alice, admires, knows\}
  Language Space \{a, b, c, d, e\}
  Meaning/Language pair might be:
  \textit{Jack knows Bob} \quad abcccced
- Child learns according to particular rules
- The child becomes the parent of the next generation
Despite the enormous diversity of language, there exist common structural elements that permeate throughout all communication

- Zipf’s Law & Mandelbrot’s Extension

Are we that much better than monkeys?

- Random fact:
A website entitled The Monkey Shakespeare Simulator, launched on July 1, 2003, contained a Java applet that simulates a large population of monkeys typing randomly, with the stated intention of seeing how long it takes the virtual monkeys to produce a complete Shakespearean play from beginning to end. For example, it produced this partial line from *Henry IV, Part 2*, reporting that it took "2,737,850 million billion billion billion billion monkey-years" to reach 24 matching characters:

*RUMOUR. Open your ears; 9r"5j5&?OWTY Z0d...*

Evolution models – understanding the past, looking to the future
G. A. Miller, *Some effects of intermittent silence*,
The American Journal of Psychology, *70*, pp311-314


