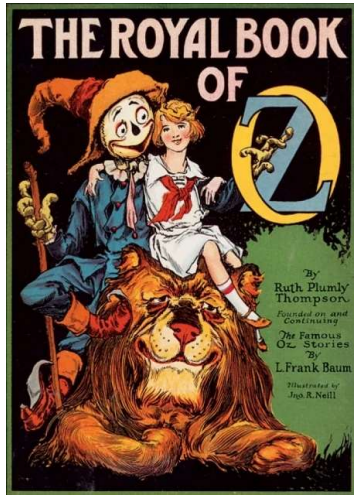
Who wrote Royal Book of Oz?



Erica Klarreich “Statistical tests are unraveling knotty literary mysteries”, *Science News* Dec 2003.

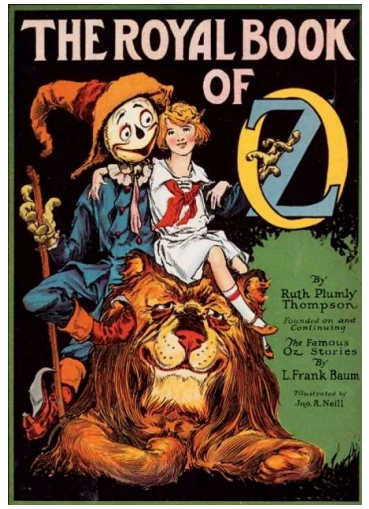
Books of Oz

L. Frank Baum (1900-20)

26 Oz books after 1920, most written by Ruth Plumly Thompson

??

(published 1921 under Baum's name, but did he truly write it?)



Poll

How many Books of Oz have you read?

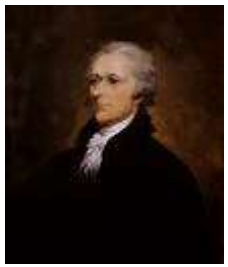
(a) 0 (b) 1-5 (c) 6-20 (d) All of them

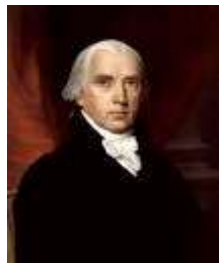
Stylometry

Seeks to attribute authorship (e.g. in coauthored texts like *Federalist Papers*, *King James Bible*) via quantitative analyses of author styles.

Surprising finding: Style is easier to spot from usage of *function words ("to", "with", "then", "however", etc.)* than from less common words.

Example: Alexander Hamilton used "upon" 10 times more frequently than James Madison.





Method

the (6.7%)	with (0.7%)	up (0.3%)	into (0.2%)	just (0.2%)
and (3.7%)	but (0.7%)	no (0.3%)	now (0.2%)	very (0.2%)
to (2.6%)	for (0.7%)	out (0.3%)	down (0.2%)	where (0.2%)
a/an (2.3%)	at (0.6%)	what (0.3%)	over (0.2%)	before (0.2%)
of (2.1%)	this/these (0.5%)	then (0.3%)	back (0.2%)	upon (0.1%)
in (1.3%)	so (0.5%)	if (0.3%)	or (0.2%)	about (0.1%)
that/those (1.0%)	all (0.5%)	there (0.3%)	well (0.2%)	after (0.1%)
it (1.0%)	on (0.5%)	by (0.3%)	which (0.2%)	more (0.1%)
not (0.9%)	from (0.4%)	who (0.3%)	how (0.2%)	why (0.1%)
as (0.7%)	one/ones (0.3%)	when (0.2%)	here (0.2%)	some (0.1%)

Most frequent 50 function words in Baum/Thompson Oz books

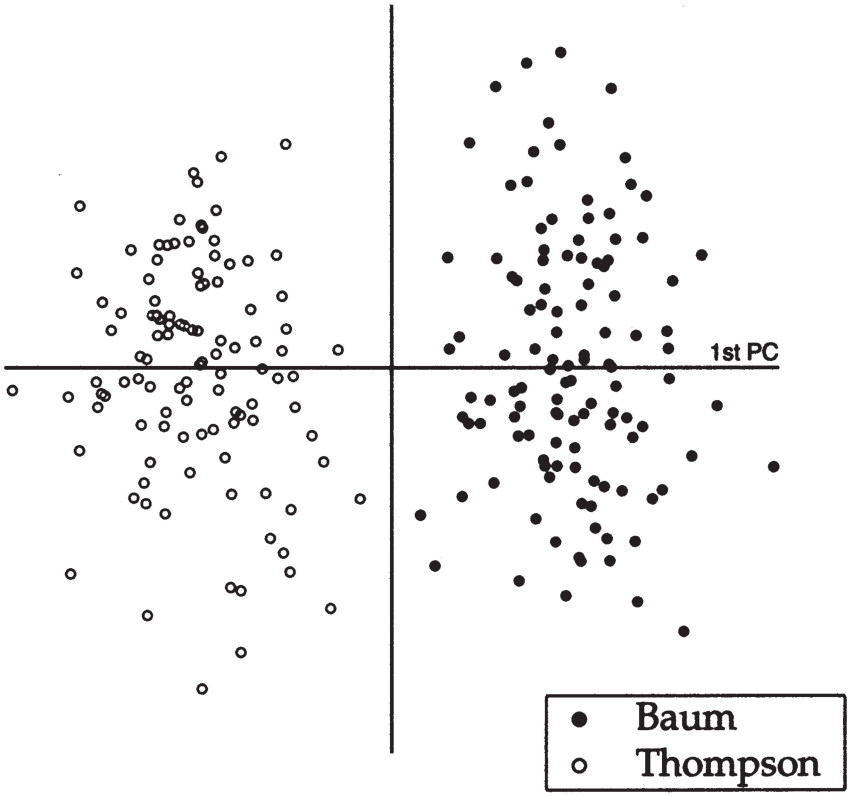
1) Divide all Oz books (except “royal book of oz”) into 223 text blocks of 5000 words ea.

2) For each block, measure #occurrences for each function word.
Obtain 223 vectors in $\mathfrak{R}^{50}$

3) Compute their 2D approximation and visualize in 2D

J.N.G. Binongo, Chance Magazine 2003
http://dh.obdurodon.org/Binongo-Chance.pdf

Method



1) Divide all Oz books (except “royal book of oz”) into 223 text blocks of 5000 words ea.

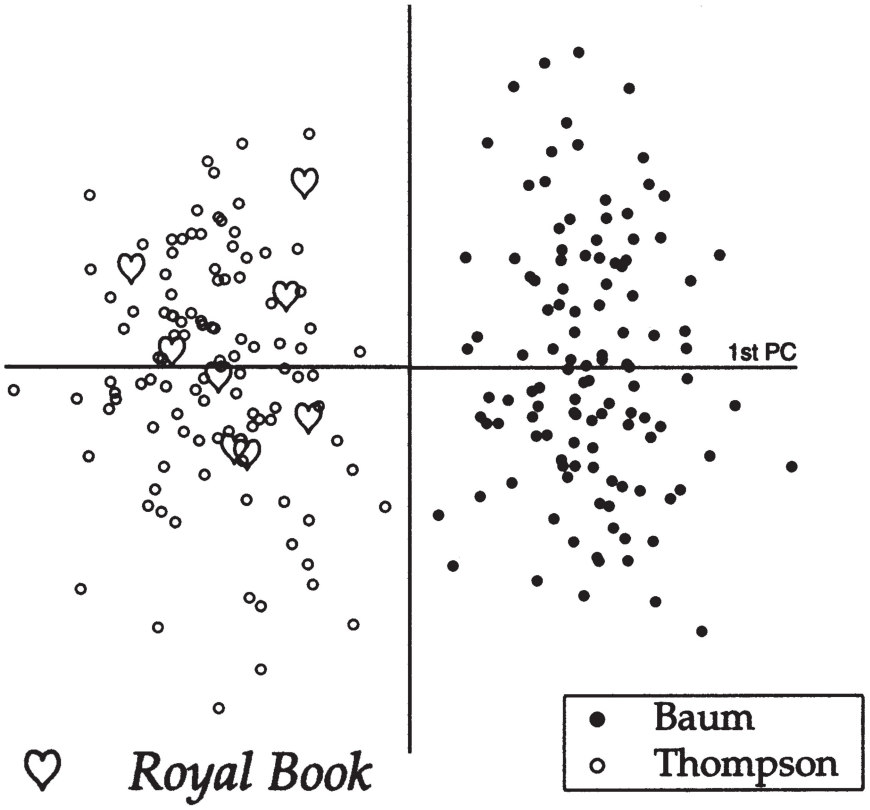
2) For each block, measure #occurrences for each function word.
Obtain 223 vectors in $\mathfrak{R}^{50}$

3) Compute their 2D approximation and visualize in 2D

Heldout data: Non-Oz writings by Baum and Thompson!
Falls neatly onto the correct side; see Binongo’s paper.

J.N.G. Binongo, Chance Magazine 2003
http://dh.obdurodon.org/Binongo-Chance.pdf

Method



1) Divide all Oz books (except “royal book of oz”) into 223 text blocks of 5000 words ea.

2) For each block, measure #occurrences for each function word.
Obtain 223 vectors in $\mathfrak{R}^{50}$

3) Compute their 2D approximation and visualize in 2D

Ruth Plumly Thompson!

J.N.G. Binongo, Chance Magazine 2003
http://dh.obdurodon.org/Binongo-Chance.pdf