the Voynich manuscript

Kevin Knight
Information Sciences Institute
University of Southern California

sources for this talk:

Mary D'Imperio, The Voynich Manuscript, An Elegant Enigma (1978)
Prescott Currier, Some Important New Statistical Findings (1976)
Rene Zandbergen, Currier A and B: Two Different Languages? (1997)
http://www.voynich.ms/forum/
experiments at USC/ISI

MIT / September 2009
Some People Involved with the Voynich Manuscript

Wilfrid Michael Voynich - book dealer
Ethel Boole, daughter of George Boole
Rudolf II - Holy Roman Emperor
Hans P. Kraus - book dealer
Roger Bacon, “first scientist”
Athanasius Kircher, German Jesuit super-scholar
William Newbold, Polymath, PhD UPenn
William Friedman, WWII cryptanalyst
Voynich Manuscript – VMS, for short
  – What is it?
  – Where did it come from?
  – What does it mean?
What is it?

- Medieval illustrated manuscript
- Approx. 235 pages on vellum material
- Color drawings of plants, nymphs, stars, etc.
- Approx. 38,000 words written in an unknown script
- Undeciphered!!! Meaning is unknown
- Currently owned by Yale University
Apparent sections of VMS

<table>
<thead>
<tr>
<th>Section “Name”</th>
<th># of word tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herbal</td>
<td>11,938</td>
</tr>
<tr>
<td>Astrological</td>
<td>2,594</td>
</tr>
<tr>
<td>Biological</td>
<td>6,915</td>
</tr>
<tr>
<td>Cosmological</td>
<td>679</td>
</tr>
<tr>
<td>Pharmacological</td>
<td>5,111</td>
</tr>
<tr>
<td>Pure Text (&quot;Stars&quot;)</td>
<td>10,682</td>
</tr>
</tbody>
</table>
The Pictures: Herbal

Many pictures look like grafting.

Sunflower? Would date VMS as post-1492.
The Pictures: Astrological
The Pictures: Astrological

What is this?

Datable clothing?
The Pictures: Biological

Small nudes in baths

Interconnecting tubes of liquids
The Pictures:
pharmacological medicine jar?
History of Voynich Manuscript

1864  Ethel Boole born in England
1865  WV born in Lithuania
1885  WV imprisoned, Polish nationalist
1890  WV & EB meet, marry in 1902
1898  WV publishes first book list
1912  WV acquires VMS in “ancient castle”
1914  WV moves to USA, opens bookshop
1919  WV sends photostatic copies of VMS
1919  Copying reveals de Tepencz signature
1919  WV writes to Bohemian State Archvs
1921  WV presents VMS + Marci letter mentioning Bacon, $160k price
1921  Newbold & WV announce decipherment
1930  WV dies. VMS placed in vault, $100k
1931  VMS appraised at $19,400
1960  Ethel dies, VMS to secretary Ann Nill
1961  NY dealer Hans Kraus buys for $24,500
1969  Kraus donates VMS to Yale
1972  Brumbaugh finds WV letters in BSA
200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive

Wilfrid Michael Voynich
book dealer

William Newbold,
Polymath, PhD UPenn

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price
1921  Newbold & WV announce decipherment
Reverend and Distinguished Sir; Father in Christ:

This book bequeathed to me by an intimate friend, I destined for you, my very dear Athanasius [Kircher], as soon as it came into my possession, for I was convinced that it could be read by no one except yourself. The former owner of this book once asked your opinion by letter ... Accept now this token ...

Dr Raphael, tutor in the Bohemian language to Ferdinand III, then King of Bohemia, told me the said book had belonged to the Emperor Rudolf and that he presented the bearer who brought him the book 600 ducats. He believed the author was Roger Bacon, the Englishman. On this point I suspend judgment ...

At the command of your reverence, Joannes Marcus Marci of Cronland

Prague, 19 August, 1665(6?)

Emperor Rudolf, paid 600 ducats for VMS

Roger Bacon (1214-94) “first scientist”

Kircher, super-scholar, recipient of this letter

??? owned VMS before Marci

“I’m Not Francis Bacon”
History of Voynich Manuscript

1576-1612  Rudolf II purchases VMS

16xx  Marci inherits VMS from ??

1665  Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price

1921  Newbold & WV announce decipherment

1961  NY dealer Hans Kraus buys for $24,500

1972  Brumbaugh finds WV letters in BSA

200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive

1608-1622  J. de Tepenecz signs VMS in Bohemian court

1630s  George Baresch owns VMS

1650  GB sends letter to Kircher

1665  Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies
History of Voynich Manuscript

1576-1612  Rudolf II purchases VMS

1608-1622  J. de Tepenecz signs VMS in Bohemian court

1912  WV acquires VMS in “ancient castle”

1914  WV moves to USA, opens bookshop

1919  WV sends photostatic copies of VMS

1919  Copying reveals Tepenecz signature

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price

1921  Newbold & WV announce decipherment

16xx  Marci inherits VMS from ??

1665  Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies
History of Voynich Manuscript

1864 Ethel Boole born in England
1865 WV born in Lithuania
1885 WV imprisoned, Polish nationalist
1890 WV & EB meet, marry in 1902
1898 WV publishes first book list
1912 WV acquires VMS in “ancient castle”
1914 WV moves to USA, opens bookshop
1919 WV sends photostatic copies of VMS
1919 Copying reveals de Tepenecz signature
1921 WV presents VMS + Marci letter mentioning Bacon, $160k price
1921 Newbold & WV announce decipherment
1930 WV dies. VMS placed in vault, $100k
1931 VMS appraised at $19,400
1960 Ethel dies, VMS to secretary Ann Nill
1961 NY dealer Hans Kraus buys for $24,500
1972 Brumbaugh finds WV letters in BSA
200x Zandbergen finds 1639 Baresch letter in newly online Kircher archive
History of Voynich Manuscript

1864 Ethel Boole born in England
1865 WV born in Lithuania
1885 WV imprisoned, Polish nationalist
1890 WV & EB meet, marry in 1902
1898 WV publishes first book list
1898 WV publishes first book list
1912 WV acquires VMS in “ancient castle”
1914 WV moves to USA, opens bookshop
1919 WV sends photostatic copies of VMS
1919 Copying reveals de Tepenecz signature
1921 WV presents VMS + Marci letter mentioning Bacon, $160k price
1921 Newbold & WV announce decipherment
1930 WV dies. VMS placed in vault, $100k
1931 VMS appraised at $19,400
1960 Ethel dies, VMS to secretary Ann Nill
“Castle” revealed as Villa Mondragone
1961 NY dealer Hans Kraus buys for $24,500
1969 Kraus donates VMS to Yale
History of Voynich Manuscript

1576-1612 Rudolf II purchases VMS

1608-1622 J. de Tepenecz signs VMS in Bohemian court

“Barschius” owns VMS between J. de Tepenecz and Marci

16xx Marci inherits VMS from ??

1665 Marci sends VMS to Kircher with letter

1665-80 Kircher owns VMS

1680 Kircher dies

1864 Ethel Boole born in England

1865 WV born in Lithuania

1885 WV imprisoned, Polish nationalist

1890 WV & EB meet, marry in 1902

1898 WV publishes first book list

1898 WV publishes first book list

1912 WV acquires VMS in “ancient castle”

1914 WV moves to USA, opens bookshop

1919 WV sends photostatic copies of VMS

1919 Copying reveals de Tepenecz signature

1919 WV writes to Bohemian State Archvs

1921 WV presents VMS + Marci letter mentioning Bacon, $160k price

1921 Newbold & WV announce decipherment

1930 WV dies. VMS placed in vault, $100k

1931 VMS appraised at $19,400

1960 Ethel dies, VMS to secretary Ann Nill

“Castle” revealed as Villa Mondragone

1961 NY dealer Hans Kraus buys for $24,500

1969 Kraus donates VMS to Yale

1972 Brumbaugh finds WV letters in BSA

1576-1612 Rudolf II purchases VMS

1608-1622 J. de Tepenecz signs VMS in Bohemian court

“Barschius” owns VMS between J. de Tepenecz and Marci
History of Voynich Manuscript

1576-1612  Rudolf II purchases VMS

1608-1622  J. de Tepenecz signs VMS in Bohemian court

1630s   George Baresch owns VMS sends letter to Kircher

1639   GB writes Kircher again

16xx   Marci inherits VMS from GB

1665   Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies

1864  Ethel Boole born in England

1865  WV born in Lithuania

1885  WV imprisoned, Polish nationalist

1890  WV & EB meet, marry in 1902

1898  WV publishes first book list

1912  WV acquires VMS in “ancient castle”

1914  WV moves to USA, opens bookshop

1919  WV sends photostatic copies of VMS

1919  Copying reveals de Tepenecz signature

1919  WV writes to Bohemian State Archvs

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price

1921  Newbold & WV announce decipherment

1930  WV dies. VMS placed in vault, $100k

1931  VMS appraised at $19,400

1960  Ethel dies, VMS to secretary Ann Nill “Castle” revealed as Villa Mondragone

1961  NY dealer Hans Kraus buys for $24,500

1969  Kraus donates VMS to Yale

1972  Brumbaugh finds WV letters in BSA

200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive

16xx  Marci inherits VMS from GB

1665  Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies

200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive

1576-1612  Rudolf II purchases VMS

1608-1622  J. de Tepenecz signs VMS in Bohemian court

1630s   George Baresch owns VMS sends letter to Kircher

1639   GB writes Kircher again

16xx   Marci inherits VMS from GB

1665   Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies

1864  Ethel Boole born in England

1865  WV born in Lithuania

1885  WV imprisoned, Polish nationalist

1890  WV & EB meet, marry in 1902

1898  WV publishes first book list

1912  WV acquires VMS in “ancient castle”

1914  WV moves to USA, opens bookshop

1919  WV sends photostatic copies of VMS

1919  Copying reveals de Tepenecz signature

1919  WV writes to Bohemian State Archvs

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price

1921  Newbold & WV announce decipherment

1930  WV dies. VMS placed in vault, $100k

1931  VMS appraised at $19,400

1960  Ethel dies, VMS to secretary Ann Nill “Castle” revealed as Villa Mondragone

1961  NY dealer Hans Kraus buys for $24,500

1969  Kraus donates VMS to Yale

1972  Brumbaugh finds WV letters in BSA

200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive

1576-1612  Rudolf II purchases VMS

1608-1622  J. de Tepenecz signs VMS in Bohemian court

1630s   George Baresch owns VMS sends letter to Kircher

1639   GB writes Kircher again

16xx   Marci inherits VMS from GB

1665   Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies

1864  Ethel Boole born in England

1865  WV born in Lithuania

1885  WV imprisoned, Polish nationalist

1890  WV & EB meet, marry in 1902

1898  WV publishes first book list

1912  WV acquires VMS in “ancient castle”

1914  WV moves to USA, opens bookshop

1919  WV sends photostatic copies of VMS

1919  Copying reveals de Tepenecz signature

1919  WV writes to Bohemian State Archvs

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price

1921  Newbold & WV announce decipherment

1930  WV dies. VMS placed in vault, $100k

1931  VMS appraised at $19,400

1960  Ethel dies, VMS to secretary Ann Nill “Castle” revealed as Villa Mondragone

1961  NY dealer Hans Kraus buys for $24,500

1969  Kraus donates VMS to Yale

1972  Brumbaugh finds WV letters in BSA

200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive

1576-1612  Rudolf II purchases VMS

1608-1622  J. de Tepenecz signs VMS in Bohemian court

1630s   George Baresch owns VMS sends letter to Kircher

1639   GB writes Kircher again

16xx   Marci inherits VMS from GB

1665   Marci sends VMS to Kircher with letter

1665-80  Kircher owns VMS

1680  Kircher dies

1864  Ethel Boole born in England

1865  WV born in Lithuania

1885  WV imprisoned, Polish nationalist

1890  WV & EB meet, marry in 1902

1898  WV publishes first book list

1912  WV acquires VMS in “ancient castle”

1914  WV moves to USA, opens bookshop

1919  WV sends photostatic copies of VMS

1919  Copying reveals de Tepenecz signature

1919  WV writes to Bohemian State Archvs

1921  WV presents VMS + Marci letter mentioning Bacon, $160k price

1921  Newbold & WV announce decipherment

1930  WV dies. VMS placed in vault, $100k

1931  VMS appraised at $19,400

1960  Ethel dies, VMS to secretary Ann Nill “Castle” revealed as Villa Mondragone

1961  NY dealer Hans Kraus buys for $24,500

1969  Kraus donates VMS to Yale

1972  Brumbaugh finds WV letters in BSA

200x  Zandbergen finds 1639 Baresch letter in newly online Kircher archive
Newbold Decipherment

- Marci letter → Bacon → Cabala → “letter doubling” cipher
- Create $22^2 = 484$ Latin letter pairs AA...XX
  - these letter pairs are the cipher alphabet
- Assign each plaintext Latin letter to a set of cipher-alphabet letter pairs (B → AQ, RT, ...)
- This gives the encipherer some freedom, while the recipient can still decipher by using the table
- Cleverly encipher plaintext in such a way as to construct a “cover” message that looks like Latin, to fool readers
Newbold System

• Example:
  a n n … → DO MI NU … → DOMINU …

• Too hard to assemble good “cover” text!

• So, make cipher letter-pairs overlap:
  a n n … → AD DB BR … → ADBR …

• Also difficult, possibly too easy to decipher

• So, employ anagramming:
  a n n … → OM DO MI … → DO OM MI … → DOMI …

• Now can construct a plausible looking “cover” text in Latin for our secret message (also in Latin) – an ingenious system, to be sure!!
Newbold Decipherment

Hmm, by the method, both plaintext and ciphertext should be in Latin letters…

But the VMS doesn’t have Latin letters…
William Newbold, Polymath, PhD UPenn

… 409cc89 …

“artist’s rendition”
William Newbold, Polymath, PhD UPenn

apparent ciphertext

real ciphertext: DOMI...

“artist’s rendition”
Let's Decipher with Newbold!

apparent ciphertext

real ciphertext

doubling

non-deterministic anagramming

lookup in $22^2$ table

non-deterministic mapping from 11 Latin letters to full 22
Let's Decipher with Newbold!

Of course the $22^2$ table isn't given, so we have to build it up through cryptanalysis. Wow, this is a lot of work!
1300 real ciphertext “letters” in first 3 lines

Decipherment of those first lines:
“Í, Roger Bacon, have written this…”
(in Latin)

Anagramming sets of 55 letters is sometimes required.

Slow but steady progress… Andromeda galaxy, ovaries & ova … so Bacon must have had a microscope & telescope, hundreds of years before they were discovered!
The Text

• Approx. 38,000 words, unknown script
• Writing style similar to 15\textsuperscript{th} century Florentine “humanist” hand
• Between 23 and 40 distinct characters
• No corrections, likely to have been copied
• Writing was done after illustrations
Another medieval manuscript, just for calibration…
Introduction to Astrology and Its Use in Weather Prediction, Medicine, and Agriculture, in English. Manuscript on Paper. 1490.
Alphabet: Currier/D’Imperio

Transcription

Maybe this is really

There are several transcription schemes to choose from.
Alphabet: Currier/D’Imperio

Transcription

Variations of \( \text{.characters} \), or separate characters?
Are these ligatures?  
Is キ just a fancy way of writing キ？

If you didn’t know English, how would you know if キ was the same as キ？

Suppose キ never occurred. Would that be evidence?  
Suppose キ did occur, with the same contexts as キ (e.g., *shing)?  
Suppose キ did occur, but never in the same context as キ？

Another common motif:
# Letter Frequencies

<table>
<thead>
<tr>
<th>count</th>
<th>letter</th>
<th>count</th>
<th>letter</th>
<th>count</th>
<th>letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>25468</td>
<td>O</td>
<td>2886</td>
<td>2</td>
<td>148</td>
<td>U</td>
</tr>
<tr>
<td>20227</td>
<td>C</td>
<td>1752</td>
<td>N</td>
<td>96</td>
<td>6</td>
</tr>
<tr>
<td>17655</td>
<td>9</td>
<td>1413</td>
<td>B</td>
<td>74</td>
<td>Y</td>
</tr>
<tr>
<td>14281</td>
<td>A</td>
<td>1046</td>
<td>J</td>
<td>52</td>
<td>K</td>
</tr>
<tr>
<td>12973</td>
<td>8</td>
<td>950</td>
<td>Q</td>
<td>31</td>
<td>G</td>
</tr>
<tr>
<td>11008</td>
<td>S</td>
<td>908</td>
<td>X</td>
<td>17</td>
<td>L</td>
</tr>
<tr>
<td>10471</td>
<td>E</td>
<td>591</td>
<td>T</td>
<td>14</td>
<td>H</td>
</tr>
<tr>
<td>10026</td>
<td>F</td>
<td>524</td>
<td>*</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6716</td>
<td>R</td>
<td>431</td>
<td>V</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>5994</td>
<td>P</td>
<td>316</td>
<td>I</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5423</td>
<td>4</td>
<td>217</td>
<td>W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4501</td>
<td>Z</td>
<td>157</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4076</td>
<td>M</td>
<td>156</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 63k character tokens
# Most Frequent Words

<table>
<thead>
<tr>
<th>count</th>
<th>word</th>
<th>count</th>
<th>word</th>
<th>count</th>
<th>word</th>
</tr>
</thead>
<tbody>
<tr>
<td>863</td>
<td>8AM</td>
<td>836</td>
<td>8AM</td>
<td>212</td>
<td>OFAM</td>
</tr>
<tr>
<td>537</td>
<td>OE</td>
<td>535</td>
<td>4OFCC9</td>
<td>211</td>
<td>8AN</td>
</tr>
<tr>
<td>501</td>
<td>SC89</td>
<td>469</td>
<td>AM</td>
<td>191</td>
<td>4OFAE</td>
</tr>
<tr>
<td>426</td>
<td>ZC89</td>
<td>396</td>
<td>SOE</td>
<td>186</td>
<td>ZOE</td>
</tr>
<tr>
<td>363</td>
<td>OR</td>
<td>350</td>
<td>AR</td>
<td>177</td>
<td>OFCC9</td>
</tr>
<tr>
<td>344</td>
<td>SC9</td>
<td>318</td>
<td>8AR</td>
<td>174</td>
<td>SCC9</td>
</tr>
<tr>
<td>308</td>
<td>4OFCC9</td>
<td>305</td>
<td>4OFCC89</td>
<td>172</td>
<td>SCOE</td>
</tr>
<tr>
<td>283</td>
<td>ZC9</td>
<td>279</td>
<td>4OFAN</td>
<td>155</td>
<td>S9</td>
</tr>
<tr>
<td>272</td>
<td>4OFC89</td>
<td>270</td>
<td>89</td>
<td>155</td>
<td>OPC89</td>
</tr>
<tr>
<td>262</td>
<td>4OFAM</td>
<td>260</td>
<td>AE</td>
<td>154</td>
<td>OPAM</td>
</tr>
<tr>
<td>253</td>
<td>8AE</td>
<td>243</td>
<td>2</td>
<td>152</td>
<td>4OFAR</td>
</tr>
<tr>
<td>219</td>
<td>SOR</td>
<td></td>
<td></td>
<td>151</td>
<td>9</td>
</tr>
</tbody>
</table>

etc

**Totals:**

- 8116 word types
- 38k word tokens
## Word Length Distributions

<table>
<thead>
<tr>
<th>Length</th>
<th>Distribution</th>
<th>Length</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.02</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>2</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td>3</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>0.23</td>
<td>4</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>0.21</td>
<td>5</td>
<td>0.11</td>
</tr>
<tr>
<td>6</td>
<td>0.12</td>
<td>6</td>
<td>0.09</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>7</td>
<td>0.11</td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>8</td>
<td>0.08</td>
</tr>
<tr>
<td>9</td>
<td>0.003</td>
<td>9</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>10</td>
<td>0.03</td>
</tr>
<tr>
<td>11</td>
<td>0.0001</td>
<td>11</td>
<td>0.01</td>
</tr>
<tr>
<td>12</td>
<td>0.00007</td>
<td>12</td>
<td>0.006</td>
</tr>
<tr>
<td>13</td>
<td>0.00002</td>
<td>13</td>
<td>0.002</td>
</tr>
<tr>
<td>35</td>
<td>0.00002</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Counts on word types
Features of the Text

• 115 (out of 8116) word types appear doubled at least once

  ... 40Ifcc89 40Ifcc89 ...

• 8 words appear tripled

  ... 40Ifcc89 40Ifcc89 40Ifcc89 ...

  ... 40Ifcc89 40Ifcc89 40Ifcc89 ...

  ... 40Ifcc89 40Ifcc89 40Ifcc89 ...

However, very few repeated word bigrams and word trigrams!
No word trigram appears more than 5 times.
Some Theories About the Text

- Cryptogram
- Phonetic writing system
- Philosophical language
- Outsider art
- Glossolalia
- Hoax
Cryptogram

- Newbold (1921)
- Manly (1931) critique of Newbold
- Feely (1945), abbreviated Latin
- Strong (1945), polyalphabetic cipher, no details
  - might fall into hands of enemies of USA!
- Brumbaugh (1972), numerological box
- Several attempts in the 1990s
William Friedman

- Most famous American cryptographer of World War II
  - broke key ciphers, including Japanese “Purple” code, led proto-NSA
- VMS Study Group (1944-46)
  - developed transcription alphabet
  - group disbanded after the war
- 2nd VMS Study Group (1962)
  - at RCA
- Included his VMS theory in paper on another topic
  - paper shortened due to space constraints
  - VMS theory included in a footnote, as an anagram, to establish “invention date”

Theory

VMS written in a synthetic “philosophical” language
"Writing in Tongues"

suggested in Kennedy & Churchill, 2005

- Glossolalia (Speaking in tongues)
  - Christian New Testament, Pentecost
  - People spoke tongues foreign to themselves

- Writing in Tongues?
  - Medium Helene Smith, investigated by Theodore Flournoy (1896)
  - Under a trance, Smith was able to converse with Martians
  - She learned their language and could speak and write it
  - Looked like a genuine language
  - Grammar closer to French than you might expect

Smith’s Martian
Hoax

• Previous hoaxes:
  – Hitler diaries
  – Vinland map

• Voynich Manuscript:
  – How?
  – Why?
  – Who?
How?

• Gordon Rugg *(Scientific American, 2004)*
  – Proposed Cardan grille
  – Elizabethan espionage tool
  – If applied with randomness injected, claimed to generate VMS-like text
KPMG Forensic’s 2006 Survey of Fraud in Australia and New Zealand

Most Popular Motives for Fraud:

- greed/lifestyle (54%)
- gambling (22%)
- personal financial pressure (5%)
- other (5%)
- not specified (3.5%)
- opportunity (0.4%)
- substance abuse (0.4%)
Who?

BUT: what if Voynich knew that?

BUT: same signature in other docs

de Tepenecz signature suspiciously found during overexposure

Marci letter very convenient

faked to add a Roger Bacon connection?

BUT: Baresh letter later found in Kircher archive also mention Bacon

member of Society of Friends of Russian Freedom

said to have faked passports

Needed $ → who doesn’t?

tricky → said to have traded newer, “better” books for monks’ old dirty ones

suggested in Kennedy & Churchill, 2005

spoke 18 languages
Experiments

• Can computers help us make sense of VMS?
• Is VMS a kind of letter substitution cipher?
  – Originally in Latin?
  – English?
  – Ukrainian?
  – Ukrainian written without vowels?
• Are there patterns of any sort?
Substitution Cipher

ingcmpnqsnwf cv fpn owoktvcv
hu ihgznwfv rqcffnw cw owgcnwf
kowazoanv ...
Substitution Cipher

ingcmpnqsnwf cv fpn owoktvcv

hu ihgzsnwfv rqcffnw cw owgcnwf

kowazoanv ...
Substitution Cipher

e e e e e the
ingcmpnqswnwf cv fpn owoktv cv
  e e e e
hu ihgzsnwfv rqcffn w cw owgc n w f
  
e
kowazo anv ...
Substitution Cipher

e he e the
ingcmpnqswnf cv fpn owoktvcv
 e e e t

hu ihgzswnwfv rqcffnw cw owgcnwf
e
kowazoanv ...
Substitution Cipher

e     he  e     of the
ingcmpnqsnwf cv fpn owoktvcv
     e     e     e  t
hu ihgzsnwfv rqcffnw cw owgcnwf
     e
kowazoanv ...
Substitution Cipher

e he e of the fof
ingcmnpqsnwf cv fpn owoktvcv
e f o e o oe t
hu ihgzsnwfv rqcfafnw cw owgcnwf
ef
kowazoanv ...
Substitution Cipher

e he e of the
ingcmpnqsnowf cv fpn owoktvcv
e e et
t hu ihgzsnwfv rqcffnw cw owgcnwf
e
kowazoanv ...
Substitution Cipher

e he e e is the sis
ingcmpnqsnwf cv fpn owoktvcv
es i e i iet
hu ihgznwfv rqcffnw cw owgcnwf
es
kowazoanv ...
Substitution Cipher

e he e is the sis
ingcmpnqsnwf cv fpn owoktvcv
es i e i ie t
hu ihgznwfv rqcffnw cw owgcnwf
es
kowazoanv ...
Decipherment is the analysis of documents written in ancient languages ...

Substitution Cipher

Cryptodict

<table>
<thead>
<tr>
<th>abacdefb</th>
<th>ACADEMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>abacdefb</td>
<td>DEDICATE</td>
</tr>
<tr>
<td>abacdefb</td>
<td>MEMBRANE</td>
</tr>
<tr>
<td>abacdefc</td>
<td>ELECTRIC</td>
</tr>
<tr>
<td>abacdefc</td>
<td>TUTELAGE</td>
</tr>
<tr>
<td>abacdefd</td>
<td>ANARCHIC</td>
</tr>
<tr>
<td>abacdefd</td>
<td>EVERYDAY</td>
</tr>
<tr>
<td>abacdefe</td>
<td>ANALYSES</td>
</tr>
<tr>
<td>abacdefe</td>
<td>ANALYSIS</td>
</tr>
<tr>
<td>abacdeff</td>
<td>EYEGLASS</td>
</tr>
</tbody>
</table>

Decipherment is the analysis of documents written in ancient languages ...
Generative models

Spanish letter trigram model

\[\text{quovadebrerte} \ldots\]

a \rightarrow \{all Voynich letters\}
b \rightarrow \{all Voynich letters\}
c \rightarrow \{all Voynich letters\}
\ldots
z \rightarrow \{all Voynich letters\}
_ \rightarrow _

Probabilistic model that substitutes VMS letters for Latin letters. Initially uniform.

Train on Spanish web text. Parameters fixed.

EM Algorithm.

\[
\text{argmax } P(\text{VMS}) = \text{argmax } \sum_{\theta} P(\text{latin}) P(\text{VMS} \mid \text{Latin}) \\
\theta \quad \theta \quad \text{latin}
\]

EM method demonstrated on many decipherment tasks in [Knight et al 2006].

Easy experiments in Carmel finite-state package:

```
carmel --train-cascade corpus latin.wfsa subst.wfst
```

Returns trained devices & Viterbi decipherment.

\[\text{VAS929FAEARAPAM} \ldots\]
## Substitution Cipher

<table>
<thead>
<tr>
<th>Input</th>
<th>Best decipherment assuming plaintext is Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>cevzren cnegr qry vatravbfb uvqnytb qba dhvwbgr qr yn znapun ...</td>
<td>primera parte del ingenioso hidalgo don quijote de la mancha ...</td>
</tr>
<tr>
<td>VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR ZOE89 ...</td>
<td>decos acho es imen des dena denal y des denta ...</td>
</tr>
</tbody>
</table>

If plaintext is assumed to be Latin:
quiss squm is onum pom quss hates s qum hatis ...
Hypothesize Other Source Languages

- Pre-collect language models for 80 languages
- Decipher against each
- See which decoding run yields highest probability
No one shall be arbitrarily deprived of his property
Niemand se eiendom sal arbitrêr afgeneem word nie
Asnjeri nuk duhet të privohet arbitrarisht nga pasuria e tij
لا يجوز تجريد أحد من ملكه تعسفا
Janiw khitisa utaps oraqpês inaki aparaskapati
Arrazoirik gabe ez zaio inori bere jabegoa kenduko
Den ebet ne vo tennet e berc'hentiezg digantañ diouzh c'hoant
Никой не трябва да бъде произволно лишен от своята собственост
Ningú no serà privat arbitràriament de la seva propietat
任何人的财产不得任意剥夺。
Di a so prupiità ùn ni pò essa privu nimu di modu tirannicu
Nitko ne smije samovoljno biti lišen svoje imovine
Nikdo nesmi být svévolně zbaven svého majetku
Ingen må vilkårligt berøves sin ejendom
Niemand mag willekeurig van zijn eigendom worden beroofd
Nul ne peut être arbitrairement privé de sa propriété
Nimmen mei samar fan syn eigendom berôve wurde
Ninguín será privado arbitrariamente da súa propiedad
Niemand darf willkürlich seines Eigentums beraubt werden
Канеңіз дән мәпорәи на стерәшә ивәйәрәта ән идәкәйәә яу
Avavégui ndojep'e'а va'eräi oimehăicha reintе imбäе teëva
Ba wanda za a kwace wa dukiyarsa ba tare da cikakken dalili ba
Senkit sem lehet tulajdonától önkényesen megfosztani
Engan má eftir geðþötta svipta eign sinni
Tak seorang pun boleh dirampas hartanya dengan semena-mena
Necuno essera private arbitrariamente de su proprietate
Ní féidir a mhaoin a bhaint go forlámhach de duine ar bith
Al neniu estu arbitre forprentitа lia proprio
Kelleltki ei tohi tema vara meelevaldseл ära võtta
Eingin skal hissini vera fyri ongартко
Me kua ni dua e kovei vua na nona iyau
Keltään älköön mielivaltaisesti riistettäkö hänen omaisuuttaan
### Unknown Source Language

<table>
<thead>
<tr>
<th>Input</th>
<th>Best guess of plaintext language</th>
<th>Best decipherment</th>
</tr>
</thead>
<tbody>
<tr>
<td>cevzren cnegr qry vatravbf b uvqnytb qba dhvwbgr qr yn znapun ...</td>
<td>Spanish</td>
<td>primera parte del ingenioso hidalgo don quijote de la mancha ...</td>
</tr>
<tr>
<td>VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR ZOE89 ...</td>
<td>Romanian</td>
<td>nonsense</td>
</tr>
</tbody>
</table>
### Consonantal Writing

<table>
<thead>
<tr>
<th>Input</th>
<th>Best guess of plaintext language</th>
<th>Best decipherment</th>
</tr>
</thead>
</table>
| ceze ceg qy ataf uqyt qa dwg q y zapu | Spanish                          | prmr prt dl ngns hdlg dn qvt d l mnch ...
| VAS92 9FAE AR APAM ZOE ZOR9 QOR92 9 FOR ZOE89 ... | more nonsense                  |                             |
Generative models

• Okay, that didn’t work…

• Let’s devise looser generative models, to mine for patterns.
Generative models

Trigram model over \{a, b, _\}

- \(a \rightarrow \{\text{all Voynich letters}\}\)
- \(b \rightarrow \{\text{all Voynich letters}\}\)
- \(_ \rightarrow _\)

Initially uniform

What parameter settings result in highest P(corpus) ?

\(\rightarrow\) EM algorithm.

VAS92_9FAE_AR_APAM_...
Generative models

Trigram model over \{a, b, _\}

\[\text{a} \rightarrow \{\text{all English letters}\}\]

\[\text{b} \rightarrow \{\text{all English letters}\}\]

\[\_ \rightarrow \_\]

\text{in the town where i was \ldots}\]
Generative models

Trigram model over \{a, b, _\}

Initially uniform

What parameter settings result in highest P(corpus)? → EM algorithm.

Sample tagging with learned model:

\[ \text{a b _ b b a _ b a b b _ } \]
\[ \text{in _ the _ town _ } \]
\[ \text{b b a b a _ a _ ... } \]
\[ \text{where _ i _ ... } \]

\[ \text{in _ the _ town _ where _ i _ was ... } \]
Generative models

Trigram model over \{a, b, _\}

\[
a a _ b a b _ a b a a _ ... \\
\]

What parameter settings result in highest P(corpus)?

\rightarrow EM algorithm.

ssample tagging with learned model:

V A S 9 2 _ 9 F A E _ A R _

V A S 9 2 _ 9 F A E _ A R _

A P A M _ Z O E _ Z O R 9 _ ...

Initially uniform

V A S 9 2 _ 9 F A E _ A R _ A P A M _ ...
Generative models

Trigram model over \{a, b, _\}

\[
a a _ b a b _ a b a a _ \ldots
\]

What parameter settings result in highest P(corpus) ?
→ EM algorithm.

Initially uniform

Sample tagging with learned model:

\[
\begin{array}{c}
 b b b b a _ a b b a _ b a _ \\
 V A S 9 2 _ 9 F A E _ A R _ \end{array}
\]

\[
\begin{array}{c}
 b b b a _ b b a _ b b b a _ \ldots \\
 A P A M _ Z O E _ Z O R 9 _ \ldots \end{array}
\]
Generative models

P(letter | tag)

English

a → A E O U
b → B C D H I K L M N P R S T V W X

Voynich

a → 9
b → 8 4 2 1 0 F E C O P S Z

P(tag | letter)

P(a)

0 1 4 SWYXQC AFCBIO8V*2 HEGTRKU6JD93NML
Generative models

Bigram model over \{a, b\}

\[
a \rightarrow \{ \text{all Voynich words!} \}
\]

\[
b \rightarrow \{ \text{all Voynich words!} \}
\]

What parameter settings result in highest P(corpus)?
→ EM algorithm.

VAS92 9FAE AR APAM ZOE ZOR9 QRC2 9 ...
Generative models

Bigram model over \{a, b\}

\[ a \rightarrow b \rightarrow a \rightarrow a \rightarrow \ldots \]

Do words with similar contexts have similar spellings?!

That would be very interesting.

\[ a \rightarrow 8AM \rightarrow SOE \rightarrow \ldots \]

\[ b \rightarrow 8AM \rightarrow SC89 \rightarrow \ldots \]

VAS92 9FAE AR APAM ZOE ZOR9 QRC2 9 \ldots
Generative models

Bigram model over \{a, b\}

\[
\begin{align*}
&\text{a} \rightarrow \text{b} \\
&\text{b} \rightarrow \text{a} \\
&\text{a} \text{ a} \text{ b} \text{ a} \text{ b} \text{ a} \\
&\text{a} \text{ a} \text{ a} \text{ a} \text{ a} \\
\end{align*}
\]

Do words with similar contexts have similar spellings?!

That would be very interesting.

Sample tagging with learned model:

\[
\begin{align*}
&\text{a} \text{ a} \text{ a} \text{ a} \text{ a} \text{ a} \text{ a} \text{ a} \\
&\text{VAS92 9FAE AR APAM ZOE ZOR9} \\
&\text{a} \text{ a} \text{ a} \text{ a} \text{ a} \text{ a} \text{ a} ... \\
&\text{QRC2 9 FOR ZOE89 2OR9} ... \\
\end{align*}
\]

WAIT, WHAT?
Generative models

Voynich words tagged as “a”

Voynich words tagged as “b”
Generative models

Voynich words tagged as “a”

Voynich words tagged as “b”

← pages →
Generative models

Voynich words tagged as “a”

Voynich words tagged as “b”

Herbal  Astro  Bio  Pharma  Stars
Known since Capt. Currier’s analysis (1976):
Two “languages” (in the formal sense).
Several handwriting styles, supposedly similar breakdown.
captain currier's
"Two Languages"

Pages w/ Herbal
drawings

Help! I'm tired.
For every pair of pages, how similar are they to each other?

Rene Zandbergen (1997)
Focus Further Experiments on Voynich-B (Bio & Stars)

- Consistent vocabulary
- Still plenty of words

- Let’s try models that divide **words** into classes
- 10 classes
10 classes of words: English
10-class tagging of Voynich-B
Class-Tag Sequences

- Tagging of first VMS page:
  - fgdhfgidbjccccbeeeahfgeeabeeahfgdbjjccbeahfgjjjjc chfgbjjjcccbjjjcbjjccbeahfgbjcbjccbjcbididcbbcjjccc ccccbccbeaidbjjjcccbjjccbeahfgbjjjjjccchfgdbjjjjjccchfgbjjjc b e a b i d i d h f g d i d i d i d h f g d b j j j c b j c c c b j c c c b e a h f h f h f g b j c b e e e a h f g b j e a i d i d b j c b j c b j c h f g b j j c c c c c c b j j c b j c b e a h f g d i d i d b j c b j j j j j c b j c b j j c c c b j c b j c c c c c i d b j c c c c b j c c c c b j c c c c c c b j c c c c c c b j c c c c c i d b j c c c c c c b j c c c c c c b j c c c c c c b j c c c c c c b j j c b j c b j c b e a b j c c c c c b j c c c c c c c c c c i d i d i d b j j c b j c c c c i d i d i d h f g b j c c c c c c c c c c c c c c c c c c c c c c c c e a h f g h f g e a i

- 14-grams found in 10-class tagging:
  - 25 c c c c c c c c c c c c c c
  - 9 i d i d i d i d b e a h f g
  - 7 i d i d i d i d i d i d
  - 7 i d i d h f g e e a h f g e
  - 7 e a h f g e a h f g e a i d
  - 6 j c c c c c c c c c c c c
10 classes of words: Voynich-B

Tags per page.
10 classes of words: Voynich-B

Tags per page.

"Bio" words vs. "stars" words
• Voynich Manuscript
  – What it is → pretty clear
  – Where it came from → less clear
  – What it means → totally unclear

• Lots of room for empirical, unsupervised computer techniques
  – Character analysis (e.g., ligatures)
  – Determining relations between words and pictures
  – Identification of “topics”
  – More cipher types
thank you