



# VMS-Text als Hoax

Lisa-Marie Jost, Anika Johnen, Calina Dahl, Elisabeth Reuhl

24.06.2019



## Gordon Rugg (2004)

- Motivation: Voynich Manuskript eigentlich zu komplex um ein Hoax zu sein
- Für Rugg beruht dies eher auf einer Meinung als auf konkreten Beweisen
- Es gibt kein ausgeprägtes Fachwissen über die Imitation eines mittelalterlicher Geheimtextes
  - Es gibt kaum Beispiele für solche Geheimtexte, geschweige denn Hoaxes
- Versuch ein eigenes Hoax-Dokument zu erstellen

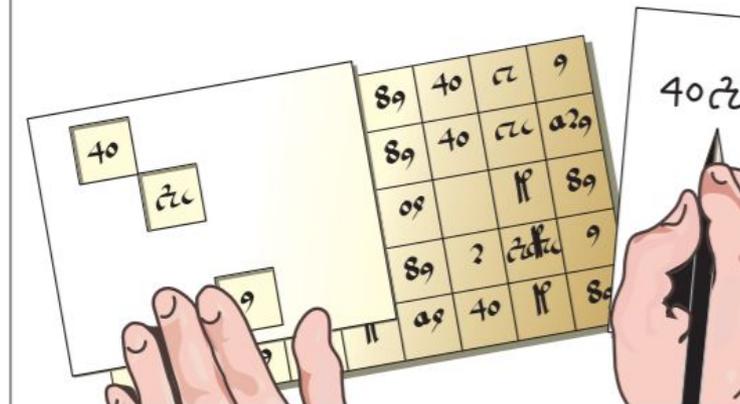


# Tabelle und Gitter Methode/ Cardan Grille

- Tabelle mit bedeutungslosen Silben, die in drei Spalten aufgeteilt sind → Präfix, Stamm, Suffix
- Jede Tabelle hat 40 Reihen und 12 Spalten mit Präfix, Stamm und Suffix
- Um Wörter zu erstellen, werden Gitter mit drei Löchern auf die Tabelle gelegt
- Anordnung der Löcher an jedem Gitter ist unterschiedlich

→ Erstellen von Wörtern mit ähnlicher Silbenstruktur wie im Voynich Manuskript

Prefix	Midfix	Suffix	Prefix	Midfix	Suffix	Prefix	Midfix	Suffix
40	ℳ	09		ℳℳ	89	40	ℳ	9
	ℳℳ	89	0	ℳ	89	40	ℳℳ	029
0	ℳ	09		ℳ	09		ℳ	89
40	ℳ	9	40	ℳ	89	?	ℳℳℳ	9
	ℳℳ	89		ℳ	09	40	ℳ	89







# Tabelle und Gitter Methode/ Cardan Gitter

## Vorgehensweise

- Methode passt zeitlich in das ungefähre Spektrum des Herstellungsdatums (1470-1608)
- Verfassen von drei Tabellen per Hand (ca. 2-3 Stunden)
- Ausschneiden der einzelnen Grilles (ca. 10 Stück in 2-3 Minuten)
- Produktion von ca. 3000 Wörtern mit dieser Methode



# Fazit

- Cardan Grille als mögliche Methode für die Produktion des Voynich Manuskriptes
- Kein Beweis für eine versteckte Nachricht/ Code im Text
- Historische Fakten: Edward Kelley soll das Manuskript angefertigt haben, um den römischen Herrscher Rudolph II zu hintergehen



# Kritik

- Immer noch unklar, ob der produzierte Text auch so genauso aussieht wie der Text im Manuskript.  
→ Rugg müsste genauere Instruktionen und ein Set von Gittern und Tabellen liefern, die einem Paragraphen ähneln.
- Problem der verschiedenen Transkriptionen, die sich in vielen Aspekten unterscheiden.  
  
→ erschweren eine statistische Analyse



## Gordon Rugg/Gavin Taylor (2017)

- Entwicklung einer Software, um die Theorie zu überprüfen, dass das Voynich Manuskript ein Hoax ist
- Software produziert Texte, die dem Voynich Manuskript ähneln
- Angelehnt an Silben und deren Häufigkeit im Voynich Manuskript



# Worthäufigkeiten

- Voynich Manuskript zeigt ähnliche Verteilung der Worthäufigkeit wie natürliche Sprachen
- Zipf's Law: Beschreibt die relative Häufigkeit von Wörtern in einem Text

→ natürliche Sprachen zeigen eine nichtlineare Kurve mit steilem Anfang, gefolgt von langem Auslauf an Wörtern, die nur ein paar mal vorkommen

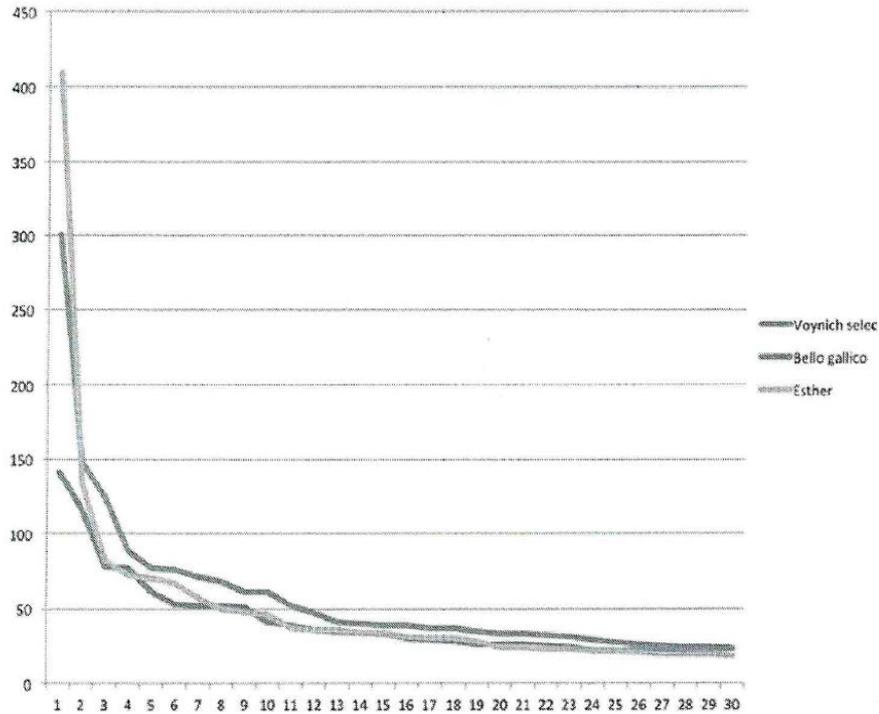


Figure 3. Frequency distributions of the 30 most common words in the Voynich Manuscript, Latin Vulgate translation of the Book of Esther, and Caesar's *De Bello Gallico*.

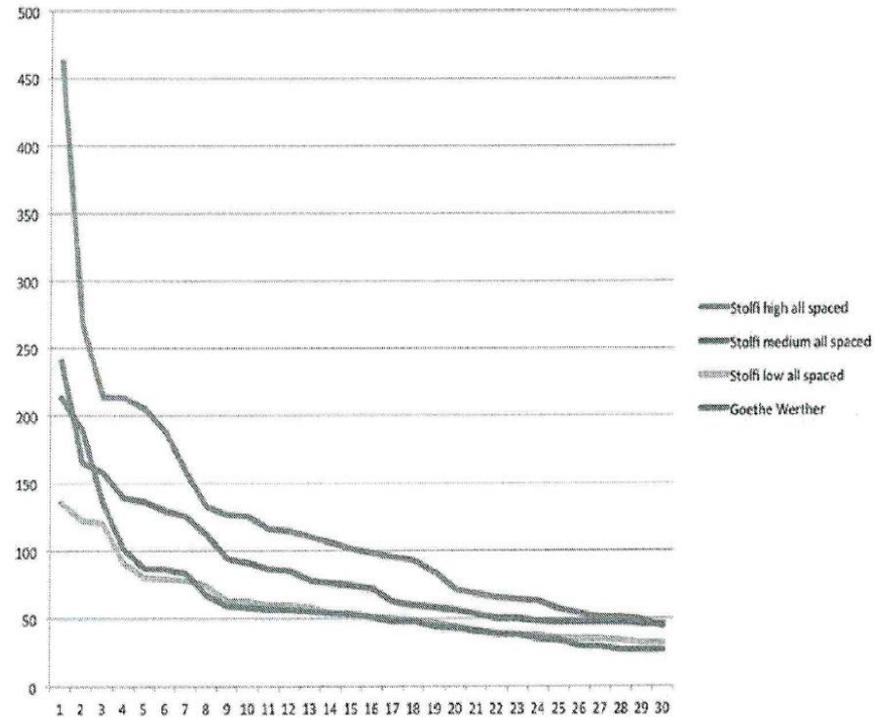


Figure 5. Three versions of meaningless gibberish compared with Goethe's *Werther* in German.



# Worthäufigkeiten

- Aber: die Verteilung sollte nicht überschätzt werden
  - Obwohl natürliche Sprachen dieser Regel folgen, heißt das nicht, dass alles, was der Regel folgt auch eine natürliche Sprache sein muss
  - Mit der Software konnte gezeigt werden, dass eine ähnliche Verteilung auch bei bedeutungslosem Text möglich ist
- Die Verteilung im Voynich Manuskript sollte nicht überbewertet werden
- Kein Beweis dafür, dass eine natürliche Sprache vorliegt



# Wortlängen

- Die Wortlängen im Voynich Manuskript formen eine symmetrische Verteilung (im Gegensatz zu den meisten natürlichen Sprachen)
- Verteilung ist ein Nebenprodukt durch das Generieren mit Tabellen

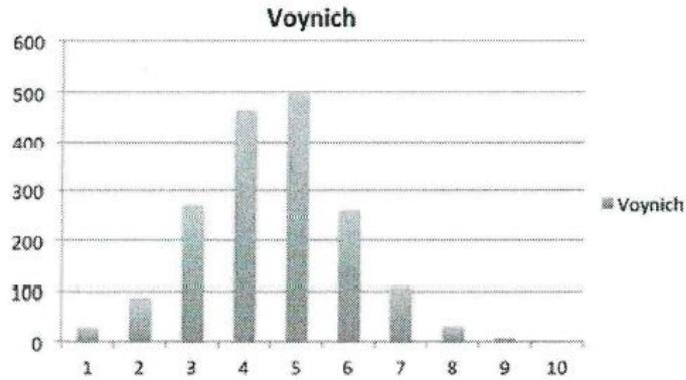


Figure 10. Word length frequencies from the Voynich Manuscript.

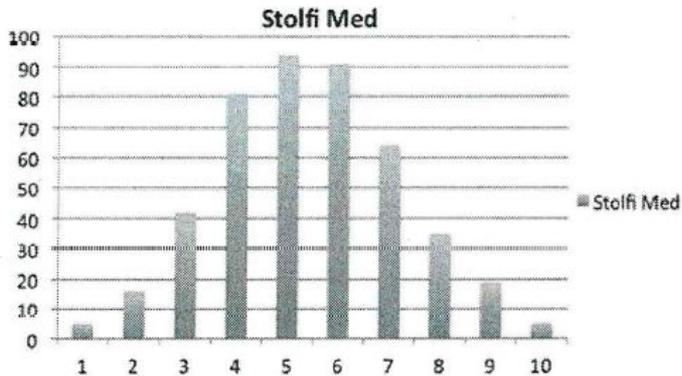
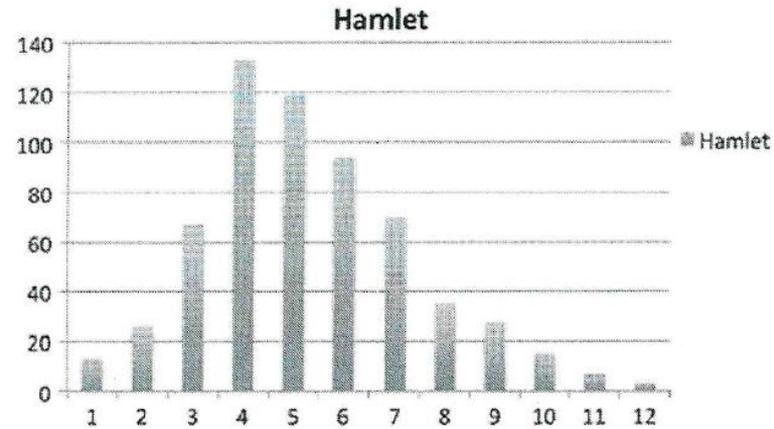


Figure 11. Word length frequencies from medium-structure gibberish text.

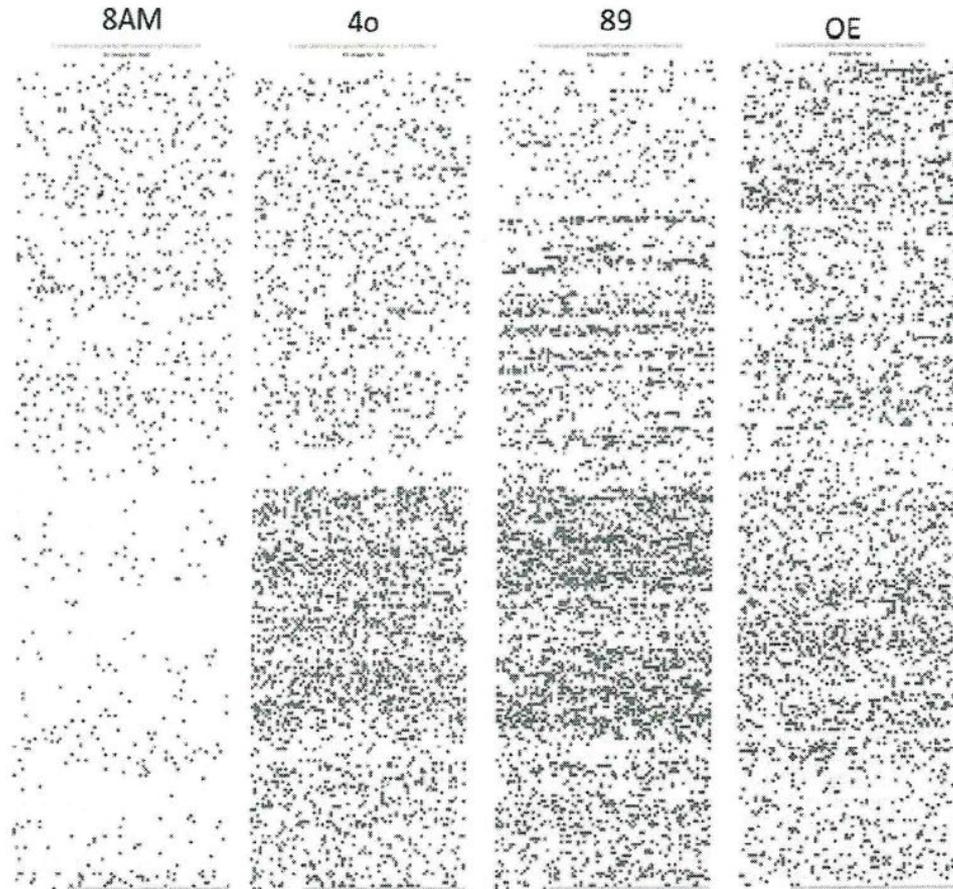


Word length frequency distribution in *Hamlet*.



# Silbenverteilung

- Text wurde nicht durch einen beliebigen Prozess erstellt
- Aber auch der durch die Tabellen und Gitter Methode produzierte Text ist nicht beliebig
- Verteilung der vier häufigsten Silben im Manuskript unterscheidet sich deutlich von der im Deutschen und Englischen
- Abrupte und ausgeprägte Änderungen der Silben Häufigkeit im Voynich Manuskript



**Figure 14.** Distribution of four common syllables in the Voynich Manuscript, across the entire text. *Source:* <http://hydeandrugg.wordpress.com/2013/04/14/visualizing-textual-structures-in-the-voynich-manuscript/>.



# Silbenverteilung

- Nebenprodukt durch das Generieren neuer Tabellen mit bedeutungslosen Silben
  - Dabei können Fehler entstehen
- Unterschiede zwischen Voynich A und B sind schwierig zu erklären, wenn man von einer natürlichen Sprache ausgeht
- Aber leicht zu erklären als Nebenprodukt der manuellen Produktion neuer Tabellen



# Silbenverteilung

- Tabelle muss mit Silben ausgestattet sein, deren Frequenz variiert
- z.B. 'qo' → sehr populär im Manuskript, muss also in vielen Spalten auftreten
- Wenn die Silbe nicht gleichmäßig horizontal und vertikal in der Tabelle verteilt wird, ergibt sich das Vorkommen immer an bestimmten Punkten

→ Diese Art der Verteilung findet sich im Voynich Manuskript



# Wortverteilung

- Cluster von Wörtern in normalen Texten stimmt mit thematischer Struktur überein
- Nutzung der Tabellen und Gitter Methode zeigt, dass auch in bedeutungslosen Texten Worthäufungen auftreten können



# Fazit

- Vorgestellte Verteilungen können durch die Nutzung der Tabellen und Gitter Methode erklärt werden
- Inspiration für die Methode ist das Cardan - Gitter
- Erklärung für nicht passende zeitliche Datierung:
  - Produzenten des Manuskripts haben altes Pergament genutzt, um das Dokument älter erscheinen zu lassen, als es ist
- Oder: Inspiriert durch Schablonen in Italien, die für die Produktion von Kartenspielen genutzt wurden (Layout der Schablonen ist sehr ähnlich )



# Was spricht unserer Meinung nach für/gegen Ruggs Theorie

## PRO

-Methode wurde überprüft indem selbst ein ähnlicher Text erzeugt wurde

## CONTRA

-Methode wurde mit selbst erstelltem Text überprüft, es wurde aber kein Vergleich mit zeitgenössischen Texten die nach dem Cardan Gitter erstellt wurden, gemacht



# Kritik an Ruggs Theorie

-Theorie ist historisch nicht richtig eingeordnet (Cardan Grid wird ins 16. Jhdt. geordnet während alle anderen Merkmale des Manuscripts auf das 15. Jhdt. deuten)

→ Ruggs Argument, dass altes Pergament verwendet wurde ist fragwürdig

-Es werden nur Statistiken für Voynichese belegt, aber es gibt keine Gegenversuche

-Technik ist zu kompliziert für die Zeit

-> Historische Fakten werden missachtet



## Theorie nach Torsten Timm (2015)

*“The main question is whether the text within the manuscript contains a message in an unknown or constructed language, an encoded message using an unknown cipher system or whether it is a pseudo text containing no message at all (Torsten Timm).”*



# Die Haupteigenschaften des Textes

- Wiederholung geänderter Wortfolgen
- Tabellen Theorie / generierte Sprache
- Labels u. Zusammenhänge
- Kontextabhängigkeit

# Nachweise

Der Text enthält:

- Wortähnlichkeiten, Wiederholungen, Zusammenhänge

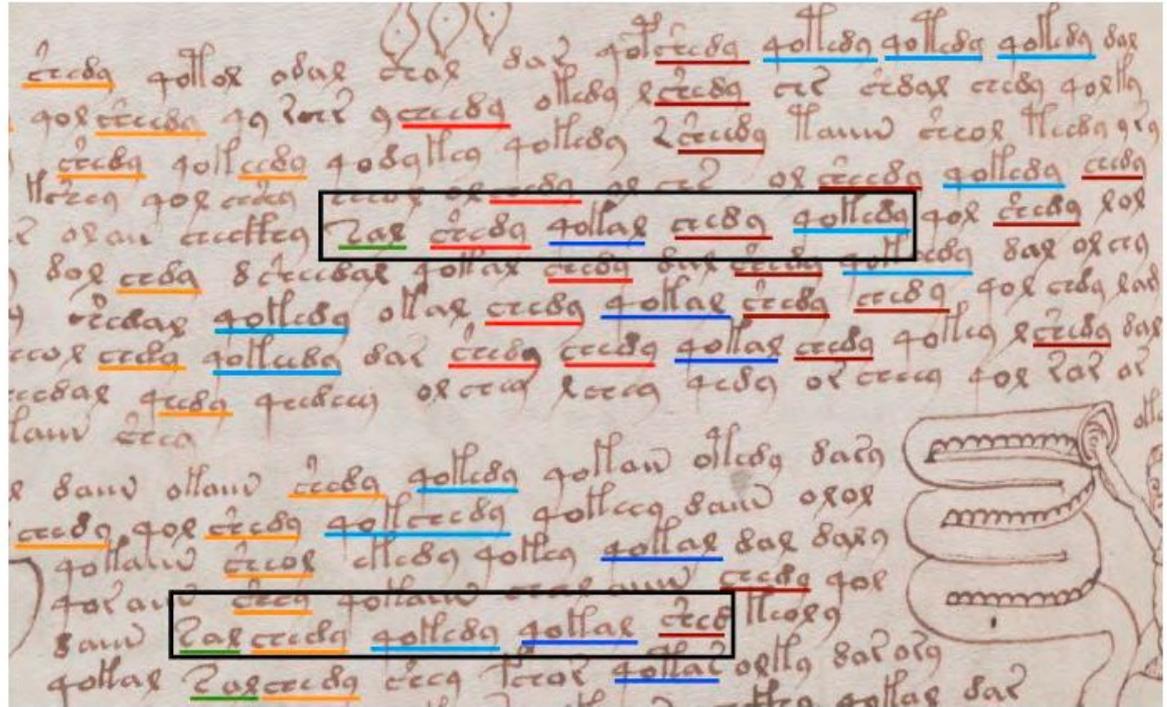


Figure 7: part of page <f77v>



# Methode der Textgenerierung

- der Text entsteht aus sich selbst
- zur Erstellung der Glyphen gibt es bestimmte Regeln
- Die Verbindung der Glyphengruppen ist somit ein Effekt des ständigen Kopierens

“The Text is a copy of itself (Torsten Timm)”





# Konkretisierung

- es handelt sich um einen generierten Text -> Text generiert sich aus sich selbst
  - die Glyphen entstehen aus Erweiterungen u. Veränderungen aus sich selbst
  - zwischen den Glyphen entsteht insofern eine Verbindung, dass sie sich gegenseitig bestimmen
- 
- Currier A und B lassen sich erklären, da sich die Wörter mit dem Text verändern
  - der Text enthält sehr wenig Fehler
  - der Text passt genau in die vorhandene Fläche

“The text layout was made during writing (Torsten Timm)”



## Es handelt sich um einen Text ohne Bedeutung - Ein verschlüsselter Code macht keinen Sinn

1. Es wäre ein simples (lösbares) Verfahren genutzt worden
2. Für einzelne Glyphen mit Bedeutung benötigt man keine so komplexe Sprache
3. Die Glyphengruppen stehen nicht für Buchstaben, da es zu viele Gruppen gibt
4. Die Wortanordnung und die Wiederholungen sprechen gegen einen Text mit Bedeutung





# Theorie nach Torsten Timm/Andreas Schinner (2019)

Vorgehen:

- 1) statistische Netzwerkanalyse von Wortähnlichkeiten des VMS
- 2) Beschreibung eines möglichen Algorithmus zur Generierung von VMS-typischen Texten
- 3) Vergleich der statistischen Eigenschaften der generierten Texte mit denen des VMS

Zusatzmaterialien auf GitHub <https://github.com/TorstenTimm/SelfCitationTextgenerator>  
bzw. DOI: 10.5281/zenodo.2531632



# 1.) Statistische Netzwerkanalyse: Wortähnlichkeiten

- kein grammatisches System, aber Wörter haben innere Struktur und oft ähnliche Glyphenfolgen
- Korrelation von Worthäufigkeit, Wortähnlichkeit und Wortposition
- auf >50 % der Seiten zwei von drei der häufigsten Wörter einander sehr ähnlich
- 85% der Wörter über Ähnlichkeit (pro Schritt Levenshtein-Distanz = 1) miteinander verbunden





# Statistische Netzwerkanalyse: Wortähnlichkeiten

- kein grammatisches System, aber Wörter haben innere Struktur und oft ähnliche Glyphenfolgen
- Korrelation von Worthäufigkeit, Wortähnlichkeit und Wortposition
- auf >50 % der Seiten zwei von drei der häufigsten Wörter einander sehr ähnlich
- 85% der Wörter über Ähnlichkeit (pro Schritt Levenshtein-Distanz = 1) miteinander verbunden
  
- im Vergleich erwartbar bei einer natürlichen Sprache:
  - Cluster statt gleichmäßiges Netzwerk
  - Funktionswörter gleichmäßig verteilt
- im Ähnlichkeits-Netzwerk isolierte Wörter = Konkatenationen
- → z.B.: **<polcheolkain>** = **<pol>** + **<cheol>** + **<kain>**

# Statistische Netzwerkanalyse: Wortähnlichkeiten

- Currier A und B? → nein, graduelle Entwicklung von Zustand A zu B
- für Currier B typische Wörter auch in A, aber nicht andersrum
- z.B. wegen Entwicklung von <ed> diese vermutete Abfassungsreihenfolge:

**Table 2.** Frequency of words in different sections of the VMS. The sections are ordered in such a way that two sections with comparable frequency counts follow each other.

	daiin	qokaiin	chol	cheody	chedy	shedy	qokeedy	Word count
Herbal (A)	403	1	228	8	1	0	0	8,087
Pharma (A)	99	2	45	18	1	1	0	2,529
Astro	23	0	8	8	4	0	0	2,136
Cosmo	36	18	19	7	24	17	4	2,691
Herbal (B)	72	20	13	7	62	35	9	3,233
Recipes (B)	122	114	62	33	190	113	137	10,673
Biological (B)	84	88	14	0	210	247	153	6,911

$$\sum_{M=0}^{\infty} p(M, \ell) = 1 \quad (1)$$

is the normalization condition, and

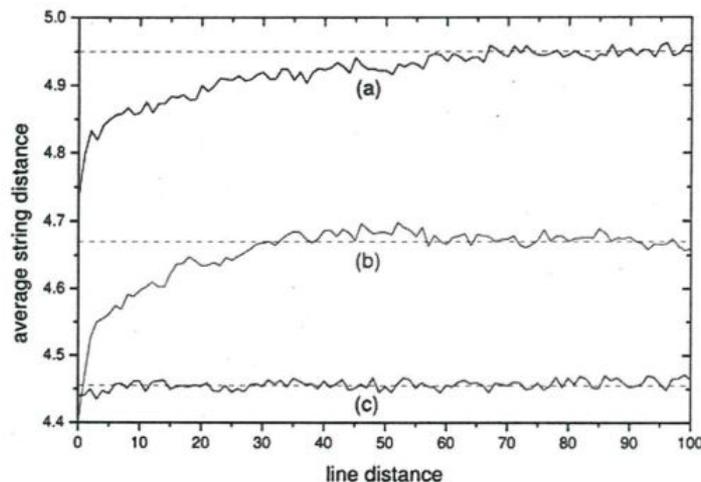
$$\bar{M}(\ell) = \sum_{M=0}^{\infty} Mp(M, \ell) \quad (2)$$

defines the average edit distance. The upper limits of both sums may, of course, be replaced by the maximal word length present in the text.

## Durchschnittliche Editier-Distanz

→ je ähnlicher zwei Wörter sind, desto wahrscheinlicher treten sie nah beieinander im Text auf

Vergleich mit Englisch:



**Figure 4.** Average string metric distance  $\bar{M}$  versus line distance of two tokens in a text. Black line (a): VMS “Recipes” section (f103r–f116v, Currier B); red line (b): VMS “Herbal” section (f1r–f66v, Currier A); blue line (c): the first 10,000 tokens of “Alice in Wonderland”; dashed lines: the respective asymptotic values of  $\bar{M}$ .



## 2) Textgenerierung per Selbstzitation/Autokopieren

- statistische Eigenschaften (Mischung natürlichsprachlicher und dem widersprüchlicher Merkmale) sind Beiprodukte des Textgenerierungs-Algorithmus
- Vorgehen: Quellwort auswählen und abwandeln
- (vermutete/mögliche) Regeln für Änderungen:
  - einen/mehrere Glyphen durch ähnliche Glyphen ersetzen
  - Präfix hinzufügen/entfernen
  - zwei Quellwörter konkatenieren
- plus Entscheidungen nach ästhetischem Empfinden, Willkür, Spontaneität des\*der Schreibers\*in
- Auswahl von selber Seite weil praktisch, vor allem aus ähnlicher Position in Zeilen oberhalb

**Algorithm 1.** Generate line of VMS text

**Require:** boolean value *isFirstLineInParagraph*

**Require:** the length of the line to create as *maxlineLength*

**Ensure:** the generated *line* of text

**while** length of *line* is smaller than *maxlineLength* **do**

**if** *isFirstLineInParagraph* **then**

*sourceWord*  $\leftarrow$  chooseWordFromParagraphInitialLines()

**else**

*sourceWord*  $\leftarrow$  chooseWordFromActualPage()

**end if**

*chance*  $\leftarrow$  generateRandomNumber(100)

*isFirstWord*  $\leftarrow$  isEmpty(*line*)

**if** *isFirstWord* **or** *chance* is between 1 and 20 **then**

*modifiedWord*  $\leftarrow$  tryToAddPrefix(*sourceWord*)

**if** *modifiedWord* = *sourceWord* **then**

*modifiedWord*  $\leftarrow$  tryToRemovePrefix(*sourceWord*)

**end if**

**else if** *chance* is between 21 and 50 **then**

*sourceWord2*  $\leftarrow$  getNextWord(*sourceWord*)

*modifiedWord*  $\leftarrow$  tryToCombine(*sourceWord*, *sourceWord2*)

**if** *modifiedWord* = *sourceWord* **then**

*modifiedWord*  $\leftarrow$  tryToSplit(*sourceWord*)

**end if**

**else if** *chance* is between 51 and 100 **then**

*modifiedWord*  $\leftarrow$  tryToReplaceGlyphsWithSimilarOnes(*sourceWord*)

**end if**

*chance*  $\leftarrow$  generateRandomNumber(100)

**if** *isFirstLineInParagraph* **and** *isFirstWord* **then**

**if** *chance*  $\leq$  94 **then**

*modifiedWord*  $\leftarrow$  addGallowPrefix(*modifiedWord*)

**end if**

**else**

**if** *chance*  $\leq$  94 **and** *modifiedWord* starts with gallow **then**

*modifiedWord*  $\leftarrow$  removeGallowPrefix(*modifiedWord*)

**end if**

**end if**

*freeSpace*  $\leftarrow$  calculateFreeSpace(*line*, *maxlineLength*)

**if** length of *modifiedWord*  $>$  *freeSpace* **then**

*modifiedWord*  $\leftarrow$  cropWord(*modifiedWord*, *freeSpace*)

**end if**

**if** *modifiedWord*  $\neq$  *sourceWord* **then**

*line*  $\leftarrow$  addWordToLine(*line*, *modifiedWord*)

**end if**

**end while**

**return** *line*



## 3) Analyse des generierten Textes

Übereinstimmung von Text-Samples und Rezepte-Sektion des VMS

- Zipf'sche Gesetze: ja
  - **“A fact that should be seen as one of the most significant results of our present work”**
- Analyse Zufallspfad (Random Walk) im Ähnlichkeitsgraphen: ja
- weiträumige Korrelationen: ja
- kleinere Strukturen: ja
- ... : ja, alle statistischen Werte stimmen!
  
- absichtlich kein Finetuning des Algorithmus, z.B. um Verteilung von Präfixen genau zu entsprechen; wäre unwissenschaftliches Vorgehen
- auch kreative Entscheidungen des\*der Schreibers\*in nicht nachgeahmt
- dennoch: 10-20% der Samples stimmen fast exakt mit VMS überein



# Fazit Torsten Timm/Andreas Schinner (2019)

- Autoren sind von Hoax-Hypothese überzeugt
  - durch Algorithmus und seine Text-Samples nachgewiesen, dass Autokopie-Methode möglich ist und dieselben statistischen Ergebnisse bringt
  - plausible Erklärung, warum teils ja, teils nicht linguistische Merkmale
  - plausibel für historischen Kontext
  - relativ leicht ausführbar
  - → Occams Rasiermesser
  - aber Sinnlosig(/-haftig-)keit ist nie beweisbar
- Motivation des\*der Schreibers\*in?
  - vielleicht Geld oder Reputation
  - ohne neue historische Dokumente nicht lösbar



# Was spricht unserer Meinung nach für/gegen Timms (2014) bzw. Timm/Schinnerers (2019) Theorie

- + insgesamt die schlüssigste Theorie
  - + glaubhafte Erklärung der widersprüchlichen (nicht) natürlichsprachlichen statistischen Merkmale
  - es wird kaum auf die Bilder eingegangen
  - 2014: formuliert, als träfe die Hypothese eindeutig zu
  - + 2019: guter, wissenschaftlicher Stil
  - + 2019: Nachprüfbarkeit durch Veröffentlichung des Algorithmus; keine Geheimniskrämerei
  - + 2014: ausführlicher Anhang
  - (Timm und Schinner sind formal keine Geisteswissenschaftler)
- 
- + Erklärung, warum es keine Korrekturen im Text gibt, warum Text so gut in Platz passt
  - ~ Bilder als Ablenkung/Aufwertung
  - schwer vorstellbar, dass jemand soviel Aufwand für einen Text ohne Inhalt getrieben hat
  - ~ unbefriedigend, dass offene Fragen zur Motivation bleiben



# Kritik anderer an Timms (2014) bzw. Timm/Schinners (2019) Theorie

- + **Hermes:** sehr überzeugend, braucht keine Hilfsmittel
- **Hermes:** ist es den Aufwand wert?
- + **Schmeh:** gute/vielversprechende Theorie, erklärt warum (trotz im Mittelalter nicht so komplexer Verschlüsselungssysteme) bisher niemand das VMS entschlüsseln konnte
- + **Schmeh:** zu 2014: “müsste man auf diese Weise generierte Texte mit den gleichen statistischen Methoden wie den Voynich-Text untersuchen und die Ergebnisse vergleichen” [2019 genau das getan]
- + **Schwerdtfeger:** “sehr gute Erklärung aller mir bekannter Eigenschaften des Manuskriptes”, z.B. wegen Erklärung für flüssiges Schriftbild
- **Smith:** “rules do not work without supplemental information on how Voynich words should (and do) look”; Regeln sind z.T. zu allgemein gefasst, sind eher Beobachtungen [2019 Regelkomplexität reduziert]
- **Smith:** Entstehung der Wortstruktur nicht hinreichend erläutert



# Literatur/Blogposts zu Rugg/Taylor

Gordon Rugg (2004): "The mystery of the Voynich Manuscript: New analysis of a famously cryptic medieval document suggests that it contains nothing but gibberish." Scientific American 291.1: 104109.

Gordon Rugg, Gavin Taylor (2017): Hoaxing statistical features of the Voynich Manuscript. Cryptologia 41:3, 247-268.

Nick Pelling: Gordon Rugg, The Search Visualizer, And The Voynich..., in Cipher Mysteries, 2013/04/17, <http://ciphermysteries.com/2013/04/17/gordon-rugg-the-search-visualizer-and-the-voynich>.

Nick Pelling: Gordon Rugg, "The Man Who Cracked The Mystery Of The Voynich Manuscript", Cracks It Once Again (Not)..., in Cipher Mysteries, 2016/09/14, <http://ciphermysteries.com/2016/09/14/gordon-rugg-man-cracked-mystery-voynich-manuscript-cracks-not>.

Klaus Schmech: "Has Gordon Rugg solved the mystery of the Voynich manuscript text?", in Klausis Krypto Kolumne (ScienceBlogs), 2016/10/26, <http://scienceblogs.de/klausis-krypto-kolumne/2016/10/26/has-gordon-rugg-solved-the-mystery-of-the-voynich-manuscript-text-part-1/>.

Wikipedia contributors, "Gordon Rugg," Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/w/index.php?title=Gordon\\_Rugg&oldid=888787589](https://en.wikipedia.org/w/index.php?title=Gordon_Rugg&oldid=888787589) (accessed May 27, 2019).



# Literatur/Blogposts zu Timm/Schinner

Torsten Timm (2015): How the Voynich Manuscript was created. <https://arxiv.org/abs/1407.6639>

Torsten Timm & Andreas Schinner (2019) A possible generating algorithm of the Voynich manuscript, Cryptologia, DOI: 10.1080/01611194.2019.1596999.

<https://github.com/TorstenTimm/SelfCitationTextgenerator>

<https://github.com/TorstenTimm/VoynichTextGenerator/>

Jürgen Hermes: "Voynich Manuskript das Werk eines Autokopisten?", in TEXperimenTales, 29.07.2014, <https://experimentales.hypotheses.org/1076>.

Klaus Schmech: "Neue Theorie: So könnte der Text des Voynich-Manuskripts entstanden sein", in Klausis Krypto Kolumne (ScienceBlogs), 01.08.2014, <http://scienceblogs.de/klausis-krypto-kolumne/2014/08/01/neue-theorie-so-koennte-der-text-des-voynich-manuskripts-entstanden-sein/>

Elias Schwerdtfeger: Ein Autokopist. In Das Voynich-Blog, 30. Juli 2014, <http://voynich.tamagothi.de/2014/07/30/ein-autokopist/>.

Emma Smith: An Objection to Timm (2014), in Agnostic Voynich, 06.02.2016, <https://agnosticvoynich.wordpress.com/2016/02/06/an-objection-to-timm-2014/>.