

# Voynich Manuskript

Verschlüsselung

Präsentation von:

Anne, Thomas, Natascha, Judith & Kathrin

# Gliederung

- *Textprozessierung* (Hermes, 2012)
- *Decoding Anagrammed Texts Written in an Unknown Language and Script* (Hauer & Kondrack, 2016)
- William Newbold - steganografische Lösung
- William Friedman - Künstliche Sprache
  - > John Wilkins - erste Idee einer künstlichen Sprache
  - > Edward Foster - künstliche Sprache: *Ro*
- Arcady Balandin und Sergey Averyanov - Plansprache ähnlich wie Lojban
- Grzegorz Jáskiewicz - Buchstabenhäufigkeit- und verteilungsanalyse
- Ata Team Alberta (ATA) - Alttürkisch

# *Textprozessierung* (Hermes, 2012)



# Übersicht

## Entschlüsselungsversuch nach Hermes:

- PIII Methode
- Anwendung auf VM
  - Part I
  - Part II
  - Ausblick

# Trithematische Polygraphia – PIII

- Kryptographische Methode des Benediktinermönchs Johannes Trithemius
- 1506-1508 verfasst (1518 gedruckt)
- In der Polygraphia (Bücher 1-6) werden verschiedene Methoden der Verschlüsselung beschrieben
- PIII besteht aus Listen mit denen Texte generiert werden können

# PIII Listen

- Ein Wortstamm pro Spalte und dazugehörige Endung pro Zeile
- Bezug zu Voynich: Erzeugung des VM durch eine Methode ähnlich der P3

**Wort** im VM → **Buchstabe** im entschlüsselten Text

Spalte	PIII-5	PIII-15
a	pafa	mafra
b	pafe	mafte
c	pafi	maftri
d	pafu	maftru
e	pafu	maftru
f	pafan	maftran
g	pafen	mafren
h	pafin	maftrin
i	pafon	maftron
f	pafun	maftrun
l	pafas	maftral
m	pafes	maftrai
n	pafis	maftril
o	pafos	maftrai
p	pafus	maftrul
q	pafal	maftras
r	pafel	maftrai
s	pafil	maftris
t	pafol	maftrai
u	paful	maftrus
z	pafar	maftraff
v	pafar	maftraff
3	pafar	maftraff
w	pafur	maftraff

# Besonderheiten VM

Nr.	Besonderheit des VM	Erklärung auf Grundlage der P.III-Hypothese
1	Die Zeichen entsprechen keinem bekannten Alphabet.	Trithemius experimentiert mit abweichenden Alphabeten in der P.VI
2-3	Die Zeichen sind zu Wörtern und Paragraphen gruppiert, keine Interpunktion	Die P.III besteht aus Wörtern, die für Buchstaben stehen. Paragraphen könnten für Wörter oder Sätze stehen; Interpunktion ist nicht vorgesehen
4-8	Die Häufigkeitsverteilung der Zeichen entspricht der NaSp – Die Verbundentropie ist extrem niedrig – Die Wortlängen sind eher kurz und binomialverteilt – Die Zipf-Verteilung der Wörter entspricht der von NaSp – Die Information ist innerhalb der Wörter annähernd gleichverteilt	Eine P.III-ähnliche Methode könnte vergleichbare Kennwerte aufweisen, zur Verifikation dieser Hypothese sind computationale Analysen notwendig
9	Die Anzahl verschiedener Wörter ist sehr klein	Könnte durch die beschränkte Anzahl verwendeter Substitutionstabellen erklärt werden
10	Bestimmte Zeichen kommen nur in bestimmten Regionen innerhalb von Wörtern vor	Auch in der P.III werden nur bestimmte Buchstaben als Schlusskonsonanten genutzt, vorletzte oder letzte Position ist immer von Vokal besetzt
11-12	Die Wörter haben morphologisch einen sehr regelmäßigen Aufbau – Wörter unterscheiden sich oft nur durch einzelne Zeichen voneinander	Das sind auch zwei Haupteigenschaften der P.III-Wörter
13	Gleiche oder ähnliche Wörter treten in direkter Nachbarschaft zueinander auf	Ist in der P.III der Fall, wenn eine Spalte der Substitutionstabelle mehrfach hintereinander genutzt wird
14	Verschiedene Seiten scheinen in unterschiedlichen Dialekten verfasst	Könnte durch die Verwendung unterschiedlicher Substitutionstabellen erklärt werden
15-16	Zeilen bilden eine funktionale Einheit – Weitreichende Korrelationen deuten auf einen zufälligen Auswahlprozess hin	Könnte durch einen Auswahlprozess über die Substitutionstabellen erklärt werden

# Vorgehensweise

## Part I:

1. Suche nach dem Verschlüsselungsmechanismus – Welche Methode wäre in der Lage, einen **ähnlichen Text** wie den, der im Voynich Manuskript zu finden ist, zu generieren?

## Part II:

2. Suche nach dem Schlüssel – Für den Fall, dass ein solches Verfahren gefunden wird – wie ließen sich **mögliche Schlüssel** für dieses Verfahren rekonstruieren?

## Ausblick:

3. Suche nach dem Klartext – Wenn eine Option existiert, mögliche Schlüssel zu ermitteln, welchen Inhalt hat der **Klartext**, den sie generieren?



# Methode verifizieren

**Hypothese:** VM wurde mit Methode ähnlich zu PIII chiffriert

**Vorgehensweise:** Vergleich von sprachlichen Eigenschaften 3er Texte:

PIII-chiffrierter Text, VM-Text  
(Transkribierte Version (Currier A/B)),  
Natürlichsprachlicher Text

## Eigenschaften:

- Type-Token-Relation
- Wortlänge
- Wortentropie
- Verbundentropie
- Koinzidenzwert
- Anzahl der Minimalpaare
- Wortzwillinge/drillinge
- Weitreichende Korrelation

# Ergebnisse

- **Type-Token Relation:** VM & P.III Werte nicht signifikant anders als NaSp
- **Wortlänge:** Ähnlichkeiten der Kenndaten für Wortlängen bei VM & P.III
- **Wortentropie:** keine Unterschiede von VM, P.III & NaSp
- **Verbundentropie:** P.III Wert signifikant niedriger als NaSp, aber höher als bei VM
- **Koinzidenzwerte:** VM & P.III Werte ähnlich, Ähnlichkeiten zu Latein/Italiensch
- **Anzahl Minimalpaare:** Bei VM & P.III signifikant häufiger als NaSp
- **Wortzwillinge:** Bei VM & P.III signifikant höher als NaSp
- **Weitreichende Korrelationen:** VM-Wert höher als bei NaSp, nicht bei P.III

# Fazit

- Vergleich statistischer Werte von VM, PIII-Chiffre und NaSp-Vergleichstext  
→ PIII-ähnliches Verschlüsselungsverfahren für VM möglich
- Teils ähnliche Struktur, teils nicht im Vergleich zu NaSp-Textanalyse
- Nächster Schritt: Codebuch erstellen

# Part II – Schlüssel rekonstruieren

PIII Erinnerung: Phantasiewörter  
bestehen aus Stamm + Endung

-Graphemische Methode zur Findung  
der Endungen (Minimalpaare)

-Morphologische Methode zur Findung  
der Stämme

Ziel → Rekonstruktion der  
Substitutionstabelle nach PIII

# Part II - Methode

## Graphemische Methode:

- Minimalpaaranalyse wird angewandt
- Die ermittelten bedeutungsunterscheidenden Grapheme werden in Clustern repräsentiert

### Clustering:

PIII				VM						NaSp
	Cluster 1	Cluster 2	Cluster 3	Carrier A	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Ergebnisse eher schlecht -> zeigt weniger Regelmäßigkeit auf
Positionen	a e i o u	l n r s t x y \$	b c d f g h m p v z	Positionen	⌈⌈⌈⌈ c o d	g??g	f??f	vtca	4	
Partner	a e i o u	l n r s t x y \$	b c d f g h m p v z	Partner	⌈⌈⌈⌈ c	g??g o d f??f c a	τ	ι	4	
Kontexte	a e i o u	l n r s t x y \$	b c d f g h m p v z	Kontexte	⌈⌈⌈⌈ c o d ?	g??g	f??f	vtca	4	
Alle	a e i o u	l n r s t x y \$	b c d f g h m p v z	Alle	⌈⌈⌈⌈ c τ	g??g d f??f	oac	ι	4	
regelmäßiger Aufbau der Wörter der PIII Methode				es finden sich weniger Regelmäßigkeiten als PIII – Tendenz trotzdem zu vielen Minimalpaaren						

# Part II - Methode

## Morphologische Methode

- Zerlegung der Wörter, um Stammmorpheme zu finden
- Methode: Keyword-Trees erstellen, um Wörter im Text als solche zu klassifizieren

1. Zerlegung	2. Häufigkeiten	3. Aggregation																								
nach einem Score-Wert in unterschiedliche Variationen  <u>Pasan</u> = <u>pas-an</u> , <u>pas-a-n</u> , <u>pasa-n</u>	der Zerlegungsvariationen im Text finden  <table border="1"><thead><tr><th>Chiffre</th><th>Stamm + Suffix</th><th>Stamm + 2 Suffixe</th><th>Falscher Stamm</th></tr></thead><tbody><tr><td>pasan</td><td>pas-an [25-82]</td><td>pas-a-n [25-96-56]</td><td>pasa-n [1-56]</td></tr><tr><td>pasa</td><td>pas-a [25-96]</td><td></td><td>pa-sa [0-5]</td></tr><tr><td>husal</td><td>hus-al [24-83]</td><td>hus-a-l [24-96-56]</td><td>husa-l [0-56]</td></tr><tr><td>nas</td><td>n-as [24-89]</td><td>n-a-s [24-96-56]</td><td>na-s [0-56]</td></tr><tr><td>cabalas</td><td>cabal-as [4-89]</td><td>cabal-a-s [4-96-56]</td><td>cab-al-as [44-35-89]</td></tr></tbody></table>	Chiffre	Stamm + Suffix	Stamm + 2 Suffixe	Falscher Stamm	pasan	pas-an [25-82]	pas-a-n [25-96-56]	pasa-n [1-56]	pasa	pas-a [25-96]		pa-sa [0-5]	husal	hus-al [24-83]	hus-a-l [24-96-56]	husa-l [0-56]	nas	n-as [24-89]	n-a-s [24-96-56]	na-s [0-56]	cabalas	cabal-as [4-89]	cabal-a-s [4-96-56]	cab-al-as [44-35-89]	der Morpheme aller möglichen Zerlegungen  -> Anzahl der Wörter in denen ein Morphem potentiell vorkommt. Größte Gruppe mit gleicher Häufigkeit: PIII: 70 Morpheme in 24 Wörtern -> 70 Stammmorpheme
Chiffre	Stamm + Suffix	Stamm + 2 Suffixe	Falscher Stamm																							
pasan	pas-an [25-82]	pas-a-n [25-96-56]	pasa-n [1-56]																							
pasa	pas-a [25-96]		pa-sa [0-5]																							
husal	hus-al [24-83]	hus-a-l [24-96-56]	husa-l [0-56]																							
nas	n-as [24-89]	n-a-s [24-96-56]	na-s [0-56]																							
cabalas	cabal-as [4-89]	cabal-a-s [4-96-56]	cab-al-as [44-35-89]																							

# Part II - Methode

## **VM**

- Gleiche Vorgehensweise
- Ergebnisse weniger eindeutig

## **NaSp**

- Ergebnisse schwieriger zu analysieren
- Morphemanalyse nicht unbedingt für natürlichsprachliche Texte anwendbar

# Part II – Fazit

## Graphemische + Morphologische Analyse

- Regelmäßiger Aufbau deutet auf Ähnlichkeit von PIII und VM hin
- VM KÖNNTE durch ähnliche Methode wie PIII generiert worden sein
- P3 zwangsläufig einfachere Auswertung, da Codebuch bereits vorliegt
- Anwendung struktureller Methoden führt zu eindeutigeren Ergebnissen bei strukturell generierten Texten
- Anwendung von graphemischer und morphologischer Methodik bisher eher auf NaSp angewandt

→ Codebucherstellung für VM schwierig, weil komplex und Klartext nicht vorhanden



# Part III - Ausblick

- Nächster Schritt: Klartext des VMs generieren, aber ohne Codebuch mit PIII nicht machbar
- Erster Schritt Richtung Klartexterstellung → Detektieren von Wiederholungsmustern in VM
- Wortkombinationen und deren Eigenschaften werden untersucht mithilfe des n-Gramm Modells

# Part III - Ausblick

Fazit:

- Geringe Anzahl an Kollokationsmustern
- Besonders im Vergleich mit NaSp
- Verfahren für kryptoanalytischen Angriff ungeeignet

# Hauer & Kondrak (2016)



# Hauer & Kondrak (2016)

- Stellen Algorithmen für die Entschlüsselung vor
  - Klartext monoalphabetisch verschlüsselt
  - Klartext kann zusätzlich durch Anagramme verschlüsselt sein
  - Text könnte in einer Konsonantenschrift geschrieben sein
- Hauptaugenmerk auf Finden der Ursprungssprache
  - Evaluieren 3 verschiedene Algorithmen

# Identifikation der Ursprungssprache

- 3 Algorithmen werden vorgestellt
1. Character Frequency (Buchstabenhäufigkeit)
  2. Decomposition Pattern Frequency (Zersetzungsmusterhäufigkeit)
  3. Trial Decipherment (Versuchsweise Entschlüsselung)

# Character Frequency

- Symbole des Geheimtexts werden ausgezählt
- Die relative Häufigkeit jedes Symbols wird berechnet
- Sprache mit der ähnlichsten Buchstabenhäufigkeit wird als Ursprungssprache ausgesucht

# Decomposition Pattern Frequency

- Types werden in Wörtern ausgezählt
- Types werden der Häufigkeit nach innerhalb des Wortes sortiert
- Sprache mit der ähnlichsten Zersetzungsmusterhäufigkeit wird als Ursprungssprache ausgesucht

- Beispiel:

seems = (2, 2, 1), beams = (1, 1, 1, 1, 1)

neben = (2, 2, 1), Apfel = (1, 1, 1, 1, 1)

# Trial Decipherment

- Entschlüsselt den Geheimtext in jede zu überprüfende Sprache
- Analysiert alle Bigramme im Text
- Versucht einen Schlüssel zu finden, bei dem das Bigramm-Modell möglichst nah an dem der überprüften Sprache ist



# Evaluation

- Algorithmen an Texten aus 380 Sprachen getestet
- Testtexte wurden verschlüsselt
- Anschließend werden die 3 Algorithmen an den Geheimtexten getestet
  
- Character Frequency - 67,9%
- Decomposition Pattern Frequency - 85,5%
- Trial Decipherment - 97,1%

# Anagramm Entschlüsselung

- Nutzen Decomposition Pattern Frequency, um Sprache zu finden
  - Da so trotz Anagrammen die Ursprungssprache gefunden werden kann
- Versucht einen Schlüssel durch Musteräquivalenz zu finden
  - Schlüssel wird immer wieder mutiert, um andere Muster zu finden
- Mutationen ergeben Braumstruktur, die durchsucht wird, um den besten Schlüssel zu finden
- Nutzen Hidden Markov Model, um Anagramme zu entziffern
- Entfernen Vokale aus Texten, um auch Konsonantschriften entschlüsseln zu können

# Evaluation

- Verwenden 5 Sprachen: Englisch, Bulgarisch, Deutsch, Griechisch, Spanisch
  - Korpora haben jeweils ca 50 Millionen Wörter (nur bulgarisch 9 Millionen)
- 
- Mit Anagrammentschlüsselung - 93,8%
  - Mit Konsonantschrift - 73,1%

# Voynich Manuskript

- 43 Seiten (in Currier Alphabet)
- Testen erst die 3 Algorithmen zur Spracherkennung (mit 380 Sprachen)

# Voynich Manuskript – Character Frequency

- Sprache, die am wahrscheinlichsten ist, ist Mazatec
- Mozarabisch (3)
- Italienisch (8)
- Judäo-Spanisch (10)
  
- Algorithmus jedoch nicht sehr reliabel

# Voynich Manuskript – Decomposition Pattern

- Sprache, die am wahrscheinlichsten ist, ist Hebräisch
- Malaiisch (2)
- Arabisch (3)
- Amharisch (4)
  
- Hebräisch mit Abstand auf Platz 1

# Voynich Manuskript – Trial Decipherment

- Rangliste ändert sich je nach Parametereinstellung
- Auf den oberen Rängen immer Hebräisch und Esperanto

# Voynich Manuskript – Alphagramm–Distanz

- Die Reihenfolge der Buchstaben in Wörtern wird betrachtet
- Alphagramm-Distanz von 0, wenn Buchstaben eines Wortes in alphabetischer Reihenfolge
- Es werden alle Buchstabenpaare analysiert
- Jedes Buchstabenpaar erhöht die Distanz um 1, wenn nicht in alphabetischer Reihenfolge
  
- Beispiel: “Der” (0), “Wer” (2)

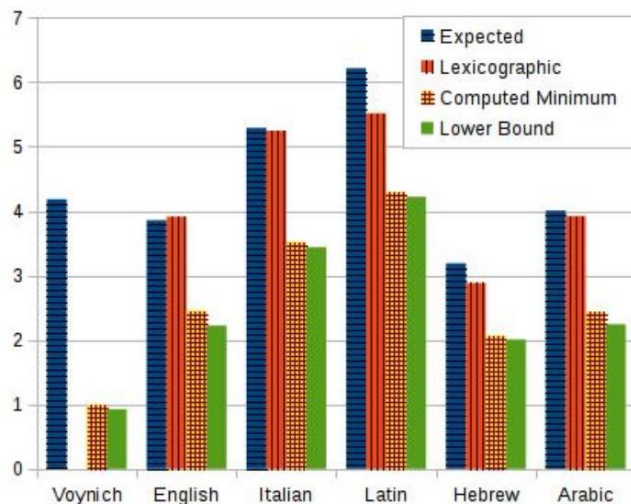


# Voynich Manuskript – Alphagramm-Distanz

- Versuch ein Alphabet zu finden, bei dem die Alphagramm-Distanz am geringsten ist
  - Alle Buchstaben werden erst in der Reihenfolge aufgeschrieben, wie sie im text erscheinen
  - Buchstaben im Alphabet werden getauscht und überprüft, ob Alphagramm-Distanz niedriger wird
- 
- Alphagramm-Distanz von VMS: 0,996

# Voynich Manuskript – Alphagramm–Distanz

- 5 Sprachen zum Vergleich ausgewählt: Englisch, Italienisch, Latein, Hebräisch, Arabisch



- VMS Alphagramm-Distanz viel geringer als bei anderen Sprachen
- Rückschluss: VMS wurde in Anagramme umgeschrieben, die in alphabetischer Reihenfolge stehen

# Voynich Manuskript – Anagramme

- Testen Anagrammentschlüsselung am VMS
- Nutzen Hebräisch als Ursprungssprache
- Erster Satz der analysierten VMS-Passage ist kein zusammenhängender Satz
- Hauer und Kondrak schaffen es mit Google Translate den Satz in “passable english” zu übersetzen

“She made recommendations to the priest, man of the house and me and people.”

# Voynich Manuskript – Anagramme

- Übersetzen einige Wörter aus der Pflanzensektion
- Einige Wörter könnten zu mittelalterlicher Pflanzenkunde passen
- Darunter “narrow”, “farmer”, “light”, “air”, “fire”

# Voynich Manuskript – Fazit

- VMS könnte ein monoalphabetische Verschlüsselung sein
- Ursprungssprache: Hebräisch (Konsonantenschrift)
- Die Worte könnten in Anagramme geändert sein
  
- Ganze Textpassagen lassen sich nicht entschlüsseln
- Einzelne Wörter könnten entschlüsselt werden

# William Newbold – steganografische Lösung

- Dozent der griechischen Philosophie
- beschäftigte sich 1919 mit dem Voynich-Manuskript
- „Michiton oladabas multos te tccr cere portas“ -> „michi dabas Multas portas“
- übersetzt ins Lateinische: „Du zeigtest mir viele Tore“
- er zählte die Buchstaben des Originalsatzes und erhielt die Zahl 22
- er entfernte den Buchstaben „k“ und „x“ und ersetzte diese durch „v“

- so erhielt er ein Geheimentextalphabet, das auch Roger Bacon verwendete

m i c h i d a b a s m u l t a s p o r t a s  
a b c d e f g h i l m n o p q r s t u v y z

- Original in Latein
- Verschlüsselung in einem bilateralen Code
- konnte nicht auf andere Textstellen angewandt werden



- Untersuchung des Voynich-Textes mit dem Mikroskop
- Theorie einer Mikroschrift
- Inhalt versteckt sich in mikroskopisch kleinen Unregelmäßigkeiten der Zeichen
- altgriechische Kurzschriftzeichen
- Im Manuskript befindet sich eine Beschreibung des Andromedanebels und der menschlichen Eizelle

## Kritik:

- Chemiker fanden in den Mikrozeichen nur feine Haarrisse
- Durch die Entschlüsselungsmethode würden zu viele Mehrdeutigkeiten entstehen
- Newbolds Anagrammsystem: er habe die Buchstaben eines Wortes zu anderen Wörtern mit neuem Sinn umgestellt
- Text kann nicht eindeutig zurücktransformiert werden
- zu der Entstehungszeit des Voynich-Manuskripts gab es kein Fernrohr und kein Mikroskop

# William Friedman – Künstliche Sprache

- er war der erste ausgewiesene Experte für Kryptologie, der sich mit dem Voynich-Manuskript befasste
- 1944 Gründung der Arbeitsgruppe FSG (First Study Group)
- Ziel: Entwicklung einer maschinenlesbaren Transkription des Voynich-Manuskriptes mit Hilfe von Lochkarten
- Darstellung der Zeichen durch Großbuchstaben und Zahlen

- Kein Zeichen für das Zeilenende -> Erhebliche Einschränkung der statistischen Auswertung
- Friedman stellte in den 50er Jahren eine neue Hypothese auf
- Bei dem Voynich-Manuskript handelt es sich um eine philosophisch orientierte künstliche oder universelle Sprache

# John Wilkins – erste Idee einer künstlichen Sprache

- Bischof von Chester
- 1668 veröffentlichte er in London eine Abhandlung über eine künstliche, philosophisch orientierte Sprache

„An Essay towards a Real Character and a Philosophical Language“

# Edward Foster – *Ro*

- 1906 entwickelte Edward Foster die künstliche Sprache “*Ro*”
- Idee einer Welthilfssprache -> diese konnte sich jedoch nicht durchsetzen
- alphabetischer Bibliothekscodex
- alle Tiernamen beginnen mit dem Buchstaben **m**
  - alle Wirbeltiere mit **ma**
  - alle Säugetiere mit **mam**
  - alle Huftiere mit **mamb**

- Viele Wörter unterscheiden sich im Voynich-Manuskript nur durch eine Verdoppelung von Buchstaben
- Parallelen zwischen dem Voynich-Manuskript und der künstlichen Sprache *Ro*
- Kritik: Das Voynich-Manuskript entstand vermutlich schon sehr weit vor der ersten Idee einer künstliche Sprache

# Arcady Balandin und Sergey Averyanov

- The Voynich manuscript: New approaches to deciphering via a constructed logical language. 2014.
- These: Plansprache mit Parallelen zu Lojban
- Lojban:
  - Plansprache
  - 1987 von der Logical Language Group entwickelt
  - basiert auf der Prädikatenlogik



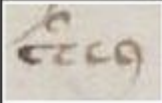
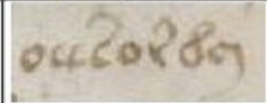
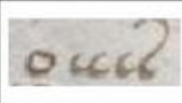



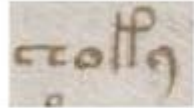
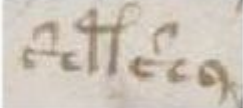


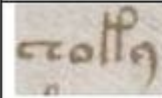
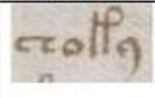
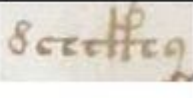


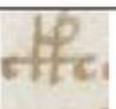
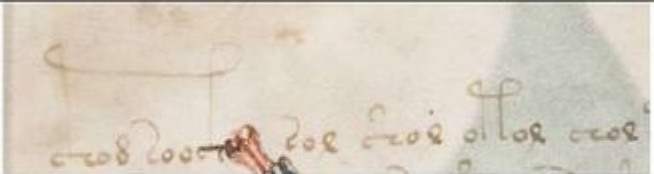
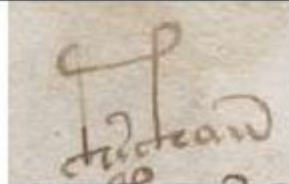
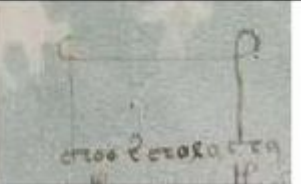
Table 1. Polymorphism of the letter "c".		
		
		
		
		
Ligature of "P" and "c".		
		
	+	
	=	

Table 2. Examples of trans-lexeme ligatures.	
	
	

Baladin; Averyanov. 2014. *The Voynich Manuscript*. P.15f.

- Lojban als mögliche Erklärung für das VM
  - “The dog that jumped to the wooden house belonged to Tom”  
→ Übersetzung in Lojban:  
{lo gerku poi pu plipe lo dinju poi pu se zbasu fi lo mudri **ku ku ku** cu se ralte la tom}
  - Parallele zu Seite 76 im VM auf der ein Wort dreimal hintereinander wiederholt wird:

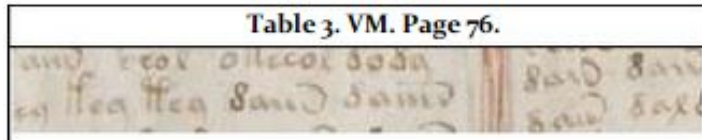






Table 4	
	
The female person. The text says "darseBar" and "olkseS".	The male person. The text says: "soral" and "dotedy".

Table 5	
	
Penis. The title says "sororal"	Same, rotated 90 degrees counterclockwise

→ Abbildung eines Mannes

Baladin; Averyanov. 2014. *The Voynich Manuscript*. P.16ff.



- “olkseš”- intestines - earth
- “otedy” - mouth - water
- “otol”- nose - air
- “dsedy” - penis - fire

*Baladin; Averyanov. 2014. The Voynich Manuscript. P.16ff.*

- Männliche Begriffe mit gleichem Wortstamm “sor”
- Weibliche Begriffe mit gleichem Wortstamm “dar”
- Wörter für die vier Elemente
- Schlussfolgerung und Überlegungen zum Autor
  - Plansprache aus dem 15.Jhd.
  - Autor mit Weitblick in die Zukunft
  - Überlegung: Leonardo Da Vinci (geb. 15.04.1452 in Florenz)

## Kritik:

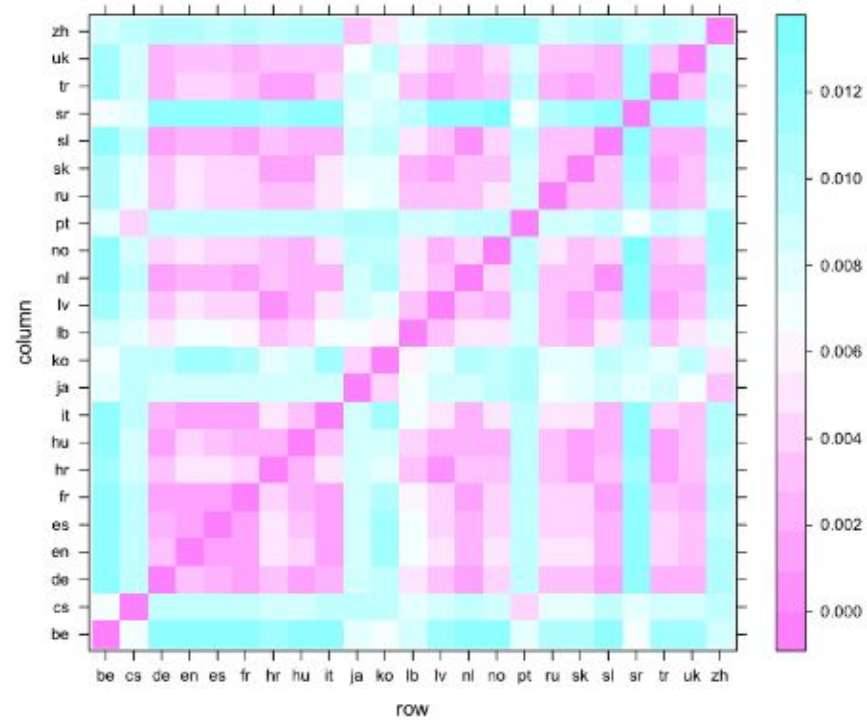
- Eigene Transkription
- Nur ein sehr willkürlich gewähltes Beispiel für Ähnlichkeit zu Lojban
- Sehr subjektive Interpretationen der Bilder
- Nicht nachvollziehbarer Schluss auf Leonardo Da Vinci als potenzieller Autor des VM

# Grzegorz Jaskiewicz

- Analysis of Letter Frequency Distribution in the Voynich Manuscript. 2011.
- Ziel: Finden einer Sprache, die der im VM verwendeten Sprache ähnlich ist.
- Technik: Untersuchung der Buchstabenhäufigkeit und -verteilung im VM und Vergleiche mit natürlichen Sprachen
- Anwendungen der Buchstabenhäufigkeiten und -verteilungsanalyse in
  - der Kryptoanalyse
  - der Datenkompression
  - Usability-Design
  - der Computerlinguistik

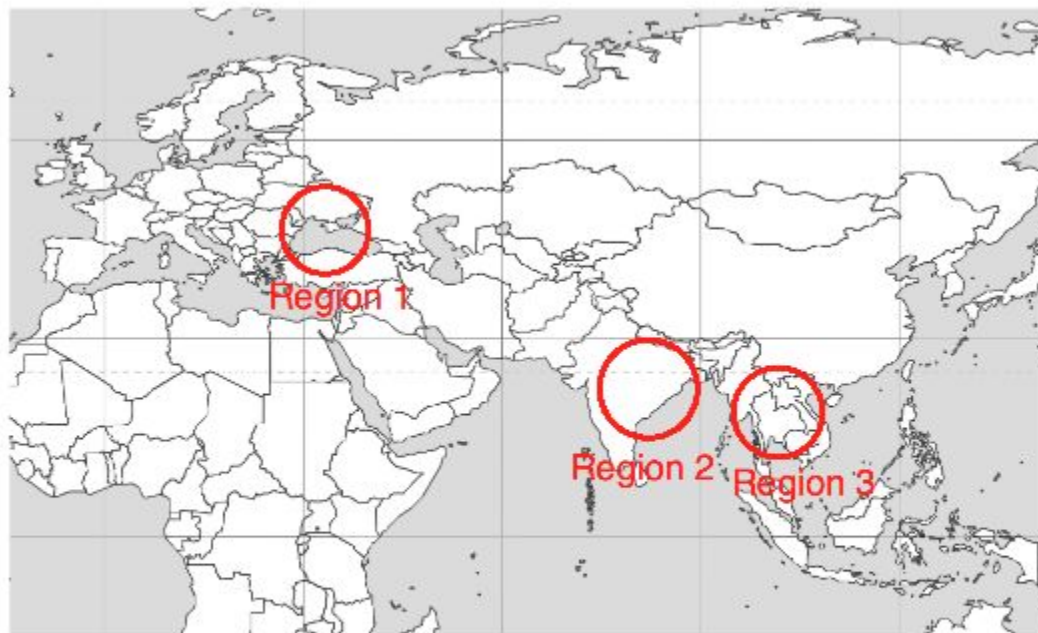
- Vergleichstexte in verschiedenen Sprachen von Wikipedia
  - Sammlung mit Hilfe eines Screenscrapers
  - Sprache in einzelnen Artikeln nicht immer konsistent
- Nutzung eines eigens entwickelten Algorithmus
  - schreibt der Gemeinsamkeit verschiedener Sprachen einen bestimmten numerischen Wert zu





**Fig. 5.** Similarity of languages based on letter frequency

- Ergebnisse:
  - VM-Sprache enthält verhältnismäßig wenig Vokale
  - Besondere Ähnlichkeit der im VM verwendeten Sprache mit:
    - Moldawisch
    - Karakalpak
    - Karbadisch-Tscherkessisch
    - Kannada
    - Thai



**Fig. 6.** Regions indicated by character distribution similarity

- Schlussfolgerung:
  - Wahrscheinlicher Ursprung der VM-Sprache in Asien
- Theorie:
  - europäischer Reisender in China
  - schrieb Informationen in erfundenem Alphabet nieder
  - Einflüsse seiner Muttersprache im Bezug auf Buchstabenhäufigkeit und -verteilung

## Kritik:

- Sehr subjektive Interpretation der Ergebnisse zum Entstehen des VM
- Markierte Punkte und genannte ähnliche Sprachen zum VM passen nicht mit der China-Theorie zusammen

# Ata Team Alberta (ATA) – Alttürkisch

- behaupten 30% des Voynich-Manuskriptes entschlüsselt und übersetzt zu haben
- die Schrift ist eine Art Alttürkisch, geschrieben in einem poetischen Stil
- der Autor habe die Worte so buchstabiert, wie er sie gehört hat
- Muster entspricht der linguistischen Struktur des Türkischen
- rhythmische Struktur -> daher die vielen Theorien, es handle sich um eine alte Sprache
- einige Zeichen des Manuskriptes werden auch in mehreren protoeuropäischen und frühen semitischen Sprachen verwendet

# Quellen

- <https://www.heise.de/tp/features/Das-Voynich-Manuskript-das-Buch-das-niemand-lesen-kann-3420065.html> (Stand 02.06.2019 15:00 Uhr)
- <http://raetsel-der-menschheit.info/71.html> (Stand 02.06.2019 17:30 Uhr)
- <http://www.openculture.com/2019/02/has-the-voynich-manuscript-finally-been-decoded.html> (Stand 12.06.2019 13:00 Uhr)
- Roitzsch, Erich H. P. (2008). Das Voynich-Manuskript : ein ungelöstes Rätsel der Vergangenheit, Verl. - Haus Monsenstein und Vannerdat, Münster.
- Arcady Balandin and Sergey Averyanov. 2014. The Voynich manuscript: New approaches to deciphering via a constructed logical language. URL: [https://www.academia.edu/6683518/The\\_Voynich\\_manuscript\\_written\\_in\\_Lojban\\_A\\_new\\_study](https://www.academia.edu/6683518/The_Voynich_manuscript_written_in_Lojban_A_new_study) (Stand 14.06.2019)
- Grzegorz Jaskiewicz. 2011. Analysis of letter frequency distribution in the Voynich manuscript. In International Workshop on Concurrency, Specification and Programming (CS&P'11), pages 250–261. URL: [http://www.deepsky.com/~merovech/voynich/voynich\\_manchu\\_reference\\_materials/PDFs/CSP2011250.pdf](http://www.deepsky.com/~merovech/voynich/voynich_manchu_reference_materials/PDFs/CSP2011250.pdf) (Stand 14.06.2019)

# Quellen

- Hauer, B. & Kondrak, G. (2016). Decoding Anagrammed Texts Written in an Unknown Language and Script in *Transactions of the Association for Computational Linguistics*, vol. 4, S. 75–86