

# Vergleich von Schriftsystemen: Jaccard-Index und Concordance Index

Schriftsysteme können auf verschiedene Art und Weise verglichen werden. Dazu gehört der Vergleich des Zeicheninventars bezüglich Anzahl und graphischer Ähnlichkeiten von Schriftzeichen, der Leserichtung, der Häufigkeitsverteilung von Schriftzeichen oder der phonetischen Lautwerte von Schriftzeichen.

Im Folgenden werden verschiedene sowohl bekannte als auch noch nicht entzifferte Schriftsysteme auf die graphischen Ähnlichkeit von einzelnen Zeichen untereinander verglichen. Um die Methodik zu testen werden zuerst einmal bekannte Alphabetschriften verglichen, da ihre entwicklungsgeschichtliche Verwandtschaft zum größten Teil bekannt ist und auch die Zeichenlisten vollständig bekannt sind. Dazu gehören Alphabete, die aus dem griechischen Alphabet bzw. aus dem lateinischen Alphabet abstammen. Erst nach der Verifizierung der Methode werden weniger bekannte oder auch bisher unbekannte Schriftsysteme verglichen. Dazu gehören die alpinen Runenschriften, die Proto-Byblosschrift, und die Indusschrift.

## Methodik

Schriftsysteme lassen sich immer paarweise vergleichen, sobald man zwei oder mehr Zeichenlisten zur Verfügung hat. Bei mehr als zwei Zeichenlisten können aber auch alle Schriftsysteme als ein großer Datensatz gemeinsam untersucht werden. Für beide Fälle gibt es verschiedene statistische Maße den Grad der Ähnlichkeit zu bestimmen.

Ein Vergleich von jeweils zwei Zeichenlisten (Schriften S1 und S2) erfolgt über das Auszählen von graphisch identischen Zeichen. Die Anzahl graphisch identischer Zeichen T ist ein Wert, der ins Verhältnis zur Gesamtanzahl von Zeichen in beiden Systemen (A1 bzw. A2) gesetzt wird. Dies erlaubt den Jaccard Index (J) wie folgt zu bestimmen.

$$J = T / (A1 + A2 - T)$$

Der Jaccard Index liegt immer zwischen Null und Eins. Je größer der Index ist, desto ähnlicher sind sich die beiden Schriftsysteme bezüglich ihrer Zeichenlisten. Möchte man mehr als zwei Schriften vergleichen, dann müssen deren Zeichenlisten in allen Kombinationen verglichen werden, um die jeweiligen Jaccard Indizes zu bestimmen. Bei n Schriften ergeben sich  $n \times (n-1)/2$  Kombinationen.

Eine etwas andere Herangehensweise wäre der Vergleich einer jeden Zeichenliste mit der Gesamtliste aller Zeichen aus allen untersuchten Schriftsystemen, die sich graphisch unterscheiden. Jedes Graphem tritt in der Gesamtliste nur einmal auf, unabhängig davon, welche Lautwerte es in den jeweiligen Schriften repräsentiert. Beim Vergleich von jeweils zwei Schriftsystemen wird wieder die Anzahl der übereinstimmenden, graphisch identischen Zeichen (T) ermittelt, so wie auch beim Paarweisen Vergleich. Zusätzlich gibt es aber auch noch eine Anzahl von Zeichen, die in keinem der beiden Schriften auftreten (U). Die beiden Werte (T, U) können nun ins Verhältnis zur Anzahl aller Grapheme in der Gesamtliste G gesetzt werden. Die

Berechnung erfolgt nach der Methode des einfachen Vergleichskoeffizienten (SMC: Simple-Matching-Coefficient).

$$\text{SMC} = (T + U) / G$$

Der Unterschied zwischen dem Jaccard Index und dem SMC liegt in der Berücksichtigung von Zeichen, die in keinem der beiden verglichenen Schriften auftreten.

Die Verwendung des SMC hat den Nachteil, dass das Ergebnis von der Auswahl der untersuchten Schriftsysteme abhängt. Die Auswahl ist prinzipiell selektiv und damit willkürlich. Wird eine weitere Schrift mit einer etwas anderen Zeichenliste hinzugenommen oder eine Schrift aus herausgenommen und nicht verwendet, dann müssen alle Berechnungen erneut durchgeführt werden. Das bedeutet aber auch, dass das Ergebnis sich ändern kann, je nach dem, wie viele Zeichenlisten verglichen werden. Diese Unsicherheit tritt beim Jaccard Index nicht auf. Deswegen ist bei der folgenden Untersuchungen der Jaccard Index verwendet worden, um Schriftsysteme bezüglich ihrer Zeichenlisten zu vergleichen.

## Lateinische Alphabete

Um die folgende Untersuchung übersichtlich zu gestalten, sind vier moderne Alphabete (Englisch, Deutsch, Dänisch, Tschechisch) beispielhaft mit dem Lateinischen Alphabet der Spätantike und untereinander verglichen worden. Die Auswahl ist willkürlich und erfordert deswegen die Verwendung des Jaccard Indizes, so wie oben beschrieben. Die folgende Tabelle listet die Anzahl an Zeichen für jedes Alphabet auf und die Grapheme der Großbuchstaben. Einzige Ausnahme ist das deutsche Eszett, welches als gesondertes Zeichen beim deutschen Alphabet mit berücksichtigt wurde.

**Tabelle 1:**

Latein (Spätantike) 23 Zeichen	Englisch 26 Zeichen	Deutsch 30 Zeichen	Dänisch 29 Zeichen	Tschechisch 40 Zeichen
A	A	A	A	A
B	B	B	B	B
C	C	C	C	C
D	D	D	D	D
E	E	E	E	E
F	F	F	F	F
G	G	G	G	G
H	H	H	H	H
I	I	I	I	I
	J	J	J	J
K	K	K	K	K
L	L	L	L	L
M	M	M	M	M
N	N	N	N	N
O	O	O	O	O
P	P	P	P	P
Q	Q	Q	Q	Q
R	R	R	R	R
S	S	S	S	S
T	T	T	T	T

	U	U	U	U
V	V	V	V	V
	W	W	W	W
X	X	X	X	X
Y	Y	Y	Y	Y
Z	Z	Z	Z	Z
		Ä	Æ	Á
		Ö	Ø	Č
		Ü	Å	Ď
		ß		É
				Ě
				Í
				Ň
				Ó
				Ř
				Š
				Ú
				Ů
				Ý
				Ž

Zwischen jeweils zwei Alphabeten gibt es eine unterschiedliche Anzahl von identischen Zeichen (s. Spalte "Gemeinsam"), wie aus Tabelle 2 zu entnehmen ist. Die Ergebnisse sind nach dem Jaccard Index absteigend sortiert.

**Tabelle 2: Die Unsicherheitswerte dT und dA sind mit einer Standardabweichung vergleichbar, die entsteht, wenn die Anzahl T bzw. A um ein Zeichen unsicher sind.**

Schrift1	Schrift2	Anzahl1	Anzahl2	Gemeinsam	Jaccard Index	dT	dA
Englisch	Dänisch	26	29	26	0,897	0,063	0,030
Latein	Englisch	23	26	23	0,885	0,070	0,033
Englisch	Deutsch	26	30	26	0,867	0,060	0,028
Latein	Dänisch	23	29	23	0,793	0,060	0,026
Deutsch	Dänisch	30	29	26	0,788	0,053	0,023
Latein	Deutsch	23	30	23	0,767	0,057	0,025
Englisch	Tschechisch	26	40	26	0,650	0,040	0,016
Dänisch	Tschechisch	29	40	26	0,605	0,036	0,014
Deutsch	Tschechisch	30	40	26	0,591	0,035	0,013
Latein	Tschechisch	23	40	23	0,575	0,038	0,014

Demnach weisen das Englische und das Dänische Alphabet die größte Ähnlichkeit untereinander auf, gefolgt von Latein-Englisch und Englisch-Deutsch. Alle vier Alphabete können als eine Gruppe betrachtet werden, da die Unterschiede im Jaccard Index (Werte liegen zwischen 0,867 und 0,897) kleiner sind als die Unsicherheit von einem Zeichen bei den gemeinsamen Zeichen (dT ca. 0,6) oder bei der Zeichenanzahl (dA ca. 0,3). Nur das Tschechische Alphabet weist zu den anderen Schriften einen viel kleineren Jaccard Index auf. Dies liegt daran, dass im Tschechische Alphabet besonders viele Sonderzeichen verwendet werden, um die Phonetik einer Slawischen Sprache wiederzugeben, während Englisch, Dänisch und Deutsch zu den Germanischen Sprachen gehören.

## Norditalische und runische Schriften

Die Indizes sind wie folgt (für weitere Überlegungen siehe weiter unten):

Schrift1	Schrift2	Anzahl1	Anzahl2	Gemeinsam	Jaccard Index
Lep	Raet	225	183	85	0,263
Lep	Etr	225	177	71	0,215
Raet	Run	183	125	52	0,203
Raet	Etr	183	177	60	0,200
Lep	Run	225	125	53	0,178
Etr	Run	177	125	37	0,140

## Indusschrift

Die Indusschrift gehört zu den noch nicht entzifferten Schriftsystemen, welches zwischen 2600 und 1700 v.Chr. im heutigen Indien und Pakistan verwendet wurde. Die Liste der Schriftzeichen enthält ca. 702 Zeichen, von denen aber möglicherweise einige Allographen von anderen Zeichen sind. Da viele Zeichen nur selten, teilweise nur einmal, im Corpus der Inschriften auftreten, können einige sich ähnelnde Grapheme nicht sicher als Allographen bestimmt bzw. als verschiedene Schriftzeichen unterschieden werden. Dies macht die folgende Untersuchung etwas unsicher, weswegen die Ergebnisse als vorläufig zu betrachten sind.

Vergleicht man das Zeicheninventar der Orte Mohenjo-daro, Harappa, Lothal, Dholavira, Kalibangan und Chanhujodaro miteinander, dann ergeben sich zwei Ortsgruppen, die jeweils den höchsten Jaccard Index aufweisen. Die erste Gruppe setzt sich aus Mohenjo-daro und Harappa mit 307 gemeinsamen Zeichen zusammen und die zweite Gruppe besteht aus Lothal und Dholavira mit 97 gemeinsamen Zeichen.

**Tabelle 3:**

Ort1	Ort2	Anzahl1	Anzahl2	Gemeinsam	Jaccard Index	dT	dA
Mohenjo-daro	Harappa	538	430	307	0,464	0,002	0,001
Dholavira	Lothal	159	149	97	0,460	0,007	0,002
Dholavira	Kalibangan	159	125	75	0,359	0,006	0,002
Lothal	Chanhujodaro	149	113	69	0,358	0,007	0,002
Lothal	Kalibangan	149	125	72	0,356	0,007	0,002
Dholavira	Chanhujodaro	159	113	68	0,333	0,007	0,002
Kalibangan	Chanhujodaro	125	113	58	0,322	0,007	0,002
Harappa	Dholavira	430	159	139	0,309	0,003	0,001
Harappa	Lothal	430	149	129	0,287	0,003	0,001
Mohenjo-daro	Dholavira	538	159	150	0,274	0,002	0,001
Mohenjo-daro	Lothal	538	149	135	0,245	0,002	0,000
Harappa	Kalibangan	430	125	101	0,222	0,003	0,000
Harappa	Chanhujodaro	430	113	94	0,209	0,003	0,000
Mohenjo-daro	Kalibangan	538	125	110	0,199	0,002	0,000
Mohenjo-daro	Chanhujodaro	538	113	100	0,181	0,002	0,000

Da jedes Zeicheninventar eines Ortes eine Teilmenge der gesamten Zeichenliste der Indusschrift mit 702 Zeichen ist und das Hinzufügen oder Entfernen eines Ortes die Grundgesamtheit (G) aller Zeichen nicht verändert, ist auch die Verwendung des SMC möglich.

Tabelle 4:

Schrift1	Schrift2	G	T	A1	U1	A2	U2	d	SMC
Dholavira	Lothal	702	97	159	62	149	52	491	0,838
Kalibangan	Chanhujodaro	702	58	125	67	114	56	521	0,825
Lothal	Chanhujodaro	702	69	149	80	113	44	509	0,823
Lothal	Kalibangan	702	72	149	77	125	53	500	0,815
Dholavira	Kalibangan	702	75	159	84	125	50	493	0,809
Dholavira	Chanhujodaro	702	68	159	91	113	45	498	0,806
Harappa	Dholavira	702	139	430	291	159	20	252	0,557
Harappa	Lothal	702	129	430	301	149	20	252	0,543
Harappa	Kalibangan	702	101	430	329	125	24	248	0,497
Mohenjodaro	Harappa	702	307	538	231	430	123	41	0,496
Harappa	Chanhujodaro	702	94	430	336	113	19	253	0,494
Mohenjodaro	Dholavira	702	150	538	388	159	9	155	0,434
Mohenjodaro	Lothal	702	135	538	403	149	14	150	0,406
Mohenjodaro	Kalibangan	702	110	538	428	125	15	149	0,369
Mohenjodaro	Chanhujodaro	702	100	538	438	113	13	151	0,358

Hierbei tritt die zweite Gruppe noch deutlicher hervor, die nun aus den Orten Lothal, Dholavira, Kalibangan und Chanhujodaro besteht. Ihre SMC liegen zwischen 0,806 und 0,838. Es gibt 47 Zeichen, die in jeden der vier Orte mindestens einmal auftreten (von insgesamt 278 Zeichen, die mindestens einmal an einen der vier Orte verwendet werden). Die Gruppe setzt sich deutlich im SMC Wert von Mohenjodaro und Harappa ab, der unter 0,56 liegt, wenn Mohenjodaro oder Harappa mit einem anderen Ort verglichen wird.

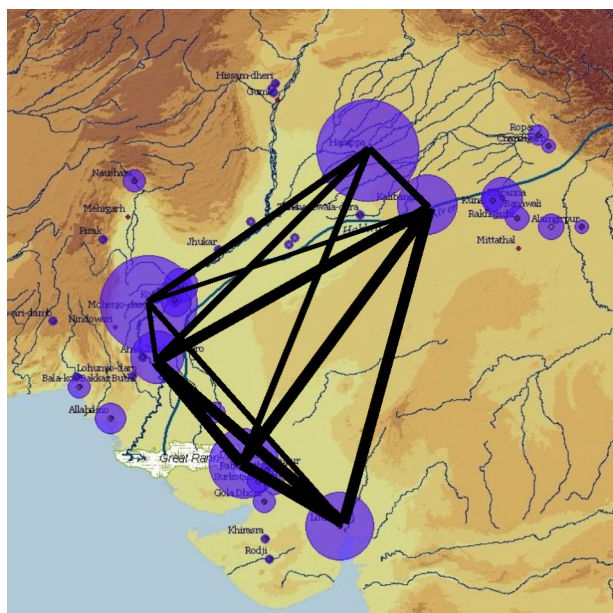


Abbildung: SMC zwischen Zeichenlisten von Orten der Induskultur mit mehr als 60 Inschriften. Die Stärke der Linien entspricht dem SMC. Lothal, Dholavira, Kalibangan und Chanhujodaro bilden eine Gruppe mit ungefähr doppelt so hohen SMC Werten wie zwischen anderen Orten.

# Concordance Index I<sub>C</sub> (Übereinstimmungs- oder Integrationsrate)

Ausgangslage: Rätisch, Lepontisch und Etruskisch sind bereits anerkannterweise sog. norditalische Schriftsysteme, d.h. ihre Verwandtschaft untereinander im Sinne eines gemeinsamen Ursprungsalphabets ist unbestritten. Hingegen ist die Zugehörigkeit der Runenschrift zur norditalischen Alphabetfamilie ein Streitpunkt, der seit gut 150 Jahren diskutiert wird.

Es geht darum die "Nähe" zwischen den vier Schriftsystemen - Rätisch, Lepontisch, Etruskisch und Runen - zu operationalisieren, also auf diese Frage eine Antwort in Form einer Zahl, also einer „Integrationszu finden.

## Ausschnitt aus den Schriftsystemen

Raet		Lep		Etr		Run	
A		B		C		D	
1	⌘ Raet A <sub>1</sub> ←	39	⌘ Lep A <sub>22</sub> ←	6	⌘ Etr A <sub>3</sub> ← (Buffa 6)		
2	⌘ Raet A <sub>1</sub> →	40	⌘ Lep A <sub>22</sub> →	5	⌘ Etr A <sub>3</sub> → (Buffa 6)		
3	∨ Raet A <sub>2</sub> ← (0)	47	∨ Lep A <sub>26</sub> →				
4	∨ Raet A <sub>2</sub> →	46	∨ Lep A <sub>26</sub> ←				
5	⌘ Raet A <sub>3</sub> ←	43	⌘ Lep A <sub>24</sub> ←	3	⌘ Etr A <sub>2</sub> ← (Buffa 5)		
6	⌘ Raet A <sub>3</sub> →	44	⌘ Lep A <sub>24</sub> →	4	⌘ Etr A <sub>2</sub> → (Buffa 5)		
7	∨ Raet A <sub>4</sub> ←						
8	∨ Raet A <sub>4</sub> →						
9	⌘ Raet A <sub>5</sub> ←						
10	⌘ Raet A <sub>5</sub> →						
11	∨ Raet A <sub>6</sub> ←						
12	∨ Raet A <sub>6</sub> → (0)						
13	⌘ Raet A <sub>7</sub> ←						
14	⌘ Raet A <sub>7</sub> →						
15	∨ Raet A <sub>8</sub> ← (0)						
16	∨ Raet A <sub>8</sub> →						
17	⌘ Raet A <sub>9</sub> ← (0)	13	⌘ Lep A <sub>7</sub> ←				
18	⌘ Raet A <sub>9</sub> →	14	⌘ Lep A <sub>7</sub> →				
19	∨ Raet A <sub>10</sub> ← (0)						
20	∨ Raet A <sub>10</sub> → (0)						
21	⌘ Raet A <sub>11</sub> ←	17	⌘ Lep A <sub>9</sub> ←				
22	⌘ Raet A <sub>11</sub> → (0)	18	⌘ Lep A <sub>9</sub> →				
23	∨ Raet A <sub>12</sub> ←						
24	∨ Raet A <sub>12</sub> → (0)						
25	⌘ Raet A <sub>13</sub> ←						
26	⌘ Raet A <sub>13</sub> →						
27	⌘ Raet A <sub>14</sub> ←	25	⌘ Lep A <sub>14</sub> ←	1	⌘ Etr A <sub>1</sub> ←		
28	⌘ Raet A <sub>14</sub> →	26	⌘ Lep A <sub>14</sub> →	2	⌘ Etr A <sub>1</sub> →		
29	⌘ Raet A <sub>15</sub>						
30	⌘ Raet A <sub>16</sub> ← (Cam) (Ziehaus & Rix 1998:299)						
31	⌘ Raet B <sub>1</sub> (Cam)						
32	⌘ Raet B <sub>2</sub> (Cam, 0, PC 10)						
33	⌘ Raet C <sub>1</sub> (Cam)			20	⌘ Etr C <sub>3</sub> ←	38	⌘ Run K <sub>6</sub> ←
34	⌘ Raet C <sub>2</sub> (Cam, 0, PC 10; F	207	⌘ Lep U <sub>11</sub> →	19	⌘ Etr C <sub>2</sub> → = ⌘ Etr J <sub>3</sub> → (Ven)	32	⌘ Run K <sub>1</sub> →
35	⌘ Raet E <sub>1</sub> ←	48	⌘ Lep E <sub>1</sub> ←	22	⌘ Etr E <sub>1</sub> ←		
36	⌘ Raet E <sub>1</sub> →	49	⌘ Lep E <sub>1</sub> →	23	⌘ Etr E <sub>1</sub> → (0)		
37	⌘ Raet E <sub>2</sub> ←	58	⌘ Lep E <sub>6</sub> ←				
38	⌘ Raet E <sub>2</sub> →	59	⌘ Lep E <sub>6</sub> →				
39	⌘ Raet E <sub>3</sub> ←	50	⌘ Lep E <sub>2</sub> ←				
40	⌘ Raet E <sub>3</sub> →	51	⌘ Lep E <sub>2</sub> →				
41	⌘ Raet E <sub>4</sub> ←						
42	⌘ Raet E <sub>4</sub> → (0)						
43	⌘ Raet E <sub>5</sub> ← (0)						
44	⌘ Raet E <sub>5</sub> →						
45	⌘ Raet V <sub>1</sub> ←	1	⌘ Lep A <sub>1</sub> ← = ⌘ Lep V <sub>1</sub> ←	26	⌘ Etr V <sub>1</sub> ←	15	⌘ Run A <sub>1</sub> ←
46	⌘ Raet V <sub>1</sub> →	2	⌘ Lep A <sub>1</sub> → = ⌘ Lep V <sub>1</sub> →	27	⌘ Etr V <sub>1</sub> → (0)	16	⌘ Run A <sub>1</sub> →

### Statistica di sintesi

L'alfabeto retico-camuno contiene **183** varianti grafiche, di cui

**85** hanno una corrispondenza grafica perfetta con l'alfabeto leponzio,

**60** hanno una corrispondenza grafica perfetta con l'alfabeto etrusco-venetico,

**52** hanno una corrispondenza grafica perfetta con l'alfabeto runico (Futhark antico).

l'alfabeto leponzio contiene **225** varianti grafiche di cui

**85** hanno una corrispondenza grafica perfetta con l'alfabeto retico-camuno,

**71** hanno una corrispondenza grafica perfetta con l'alfabeto etrusco-venetico,

**53** hanno una corrispondenza grafica perfetta con l'alfabeto runico (Futhark antico).

l'alfabeto etrusco-venetico contiene **177** varianti grafiche di cui

**60** hanno una corrispondenza grafica perfetta con l'alfabeto retico-camuno,

**71** hanno una corrispondenza grafica perfetta con l'alfabeto leponzio,

**37** hanno una corrispondenza grafica perfetta con l'alfabeto runico (Futhark antico).

l'alfabeto runico (Futhark antico) contiene **125** varianti grafiche di cui

**52** hanno una corrispondenza grafica perfetta con l'alfabeto retico-camuno,

**53** hanno una corrispondenza grafica perfetta con l'alfabeto leponzio,

**37** hanno una corrispondenza grafica perfetta con l'alfabeto runico etrusco-venetico.

Im konkreten Fall der norditalienischen Sprachen lässt sich ein Vergleichsindex  $I_C$  definieren, ein Wert, zwischen 0 und 1, der angibt, wie stark die Schriftsysteme als Gesamtes in Zusammenhang oder Übereinstimmung stehen. Bei mehreren Schriftsätzen, die miteinander verglichen werden bedeutet der Wert des Indexes 1, wenn die Schriftsysteme identisch sind, resp. 0, wenn sie völlig voneinander unabhängig sind.

Der Index kann auch für beliebig viele andere Vergleichsmengen verwendet werden. Hier werden die Fälle den Vergleich für 2, 3 oder 4 Schriftsysteme behandelt, es kann aber beliebig auf Vergleiche mit mehr Schriftsystemen erweitert werden.

Menge der Schriftzeichen (analog die norditalienischen Schriftsystem)

A Latein (Spätantike)      C Deutsch

B Englisch      D Dänisch

A Latein (Spätantike) 23 Zeichen	B Englisch 26 Zeichen	C Deutsch 30 Zeichen	D Dänisch 29 Zeichen
A	A	A	A
B	B	B	B
C	C	C	C
D	D	D	D
E	E	E	E
D	D	D	D
E	E	E	E
F	F	F	F
G	G	G	G
H	H	H	H
I	I	I	I
	J	J	J
K	K	K	K
L	L	L	L
M	M	M	M
N	N	N	N
O	O	O	O
P	P	P	P
Q	Q	Q	Q
R	R	R	R
S	S	S	S
T	T	T	T
	U	U	U
V	V	V	V
	W	W	W
X	X	X	X
Y	Y	Y	Y
Z	Z	Z	Z
		Ä	Æ
		Ö	Ø
		Ü	Å
		ß	



## Definitionen

$A \cap B$  alle gemeinsame Schriftzeichen von Latein und Englisch

$A \cup B$  alle verschiedenen Schriftzeichen von Latein oder Englisch

$G$  Grundmengen  $A \cup B \cup C \cup D$  = der gesamte Schriftsatz der vier Sprachsystemen

$A^c$  alle Schriftzeichen, die nicht in A (Latein) vorkommen (Komplementmenge) usw.

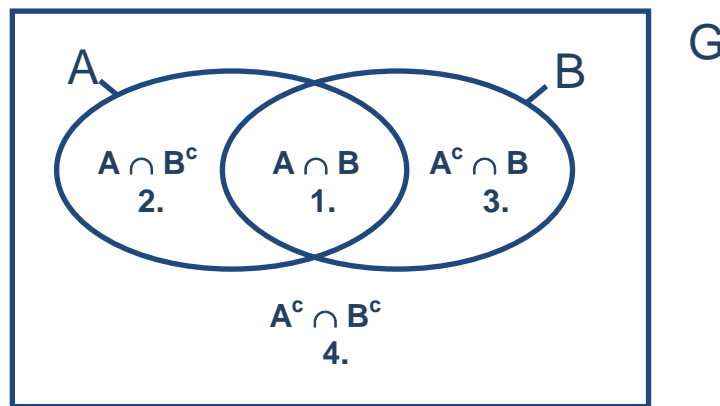
$|A|$  = die Anzahl der verschiedenen (unterscheidbaren) lateinischen Schriftzeichen usw.

## Beispiele

Berechnung des Concordance Index  $I_c$  für Latein (A) und Englisch (B)

Betrachten wir die Schriftsysteme A, B und C in einem allgemeinen Zusammenhang

Grundmenge G mit zwei Teilmengen A und B: es existieren 4 unabhängige (disjunkte) Teilmengen.



Teilmenge		Anzahl $ \dots $	Gewichtung (*)	Teilsummen
1.	$A \cap B$	$ A \cap B $	1	$1 \cdot  A \cap B $
2.	$A \cap B^c$	$ A \cap B^c $	0	$0 \cdot  A \cap B^c $
3.	$A^c \cap B$	$ A^c \cap B $	0	$0 \cdot  A^c \cap B $
4.	$A^c \cap B^c$	$ A^c \cap B^c $	0	$0 \cdot  A^c \cap B^c $
			Total	$\Sigma$
Total	$G = A \cup B$	$ G $	1	$1 \cdot  G $

(\*) Die Gewichtung gibt an, wie oft ein Schriftzeichen zusätzlich in einem andern Schriftsystem vorkommt.

Definition Concordance Index  $I_c = \frac{\Sigma}{1 \cdot |G|}$

konkret für Schriftsysteme A und B:

Teilmenge		Anzahl  ...	Gewichtung	Teilsummen
1.	$A \cap B$	23	1	$1 \cdot 23$
2.	$A \cap B^c$	0	0	$1 \cdot 0$
3.	$A^c \cap B$	3	0	$1 \cdot 0$
4.	$A^c \cap B^c$	0	0	$0 \cdot 0$
			Total	23
Total	$G = A \cup B$	26	1	$1 \cdot 26 = 26$

$$A \cap B = \{ A, B, C, D, \dots I, K, \dots T, V, X, Y, Z \}$$

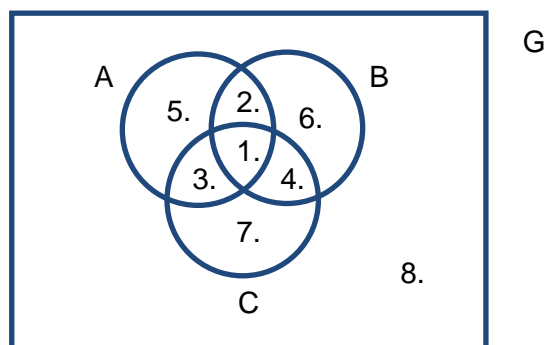
$$A \cap B^c = \{ \}$$

$$A^c \cap B = \{ J, U, W \}$$

$$A^c \cap B^c = \{ \}$$

$$\text{Concordance Index } I_c = \frac{23}{1 \cdot 26} = 0.885 = 88.5\%$$

Betrachten wir die Schriftsysteme A, B und C in einem allgemeinen Zusammenhang



Teilmenge		Anzahl  ...	Gewichtung	Teilsummen
1.	$A \cap B \cap C$	$ A \cap B \cap C $	2	$2 \cdot  A \cap B \cap C $
2.	$A \cap B \cap C^c$	$ A \cap B \cap C^c $	1	$1 \cdot  A \cap B \cap C^c $
3.	$A \cap B^c \cap C$	$ A \cap B^c \cap C $	1	$1 \cdot  A \cap B^c \cap C $
4.	$A^c \cap B \cap C$	$ A^c \cap B \cap C $	1	$1 \cdot  A^c \cap B \cap C $
5.	$A \cap B^c \cap C^c$	$ A \cap B^c \cap C^c $	0	0
6.	$A^c \cap B \cap C^c$	$ A^c \cap B \cap C^c $	0	0
7.	$A^c \cap B^c \cap C$	$ A^c \cap B^c \cap C $	0	0
8.	$A^c \cap B^c \cap C^c$	$ A^c \cap B^c \cap C^c $	0	0
			Total	$\Sigma$
Total	$G = A \cup B \cup C$	$ G $	2	$2 \cdot  G $

$$\text{Concordance Index } I_c = \frac{\Sigma}{2 \cdot |G|}$$

konkret für Schriftsysteme A, B und C:

Teilmenge		Anzahl  ...	Gewichtung	Teilsummen
1.	$A \cap B \cap C$	23	2	$2 \cdot 23$
2.	$A \cap B \cap C^c$	4	1	$1 \cdot 4$
3.	$A \cap B^c \cap C$	0	1	$1 \cdot 0$
4.	$A^c \cap B \cap C$	3	1	$1 \cdot 3$
5.	$A \cap B^c \cap C^c$	0	0	$0 \cdot 0$
6.	$A^c \cap B \cap C^c$	0	0	$0 \cdot 0$
7.	$A^c \cap B^c \cap C$	4	0	$0 \cdot 4$
8.	$A^c \cap B^c \cap C^c$	0	0	$0 \cdot 0$
			Total	53
Total	$G = A \cup B \cup C$	30	3	$2 \cdot 30 = 60$

$$A \cap B \cap C = \{A, B, C, D, \dots I, K, \dots T, V, X, Y, Z\} \quad A \cap B^c \cap C^c = \{ \}$$

$$A \cap B \cap C^c = \{ \} \quad A^c \cap B \cap C^c = \{ \}$$

$$A \cap B^c \cap C = \{ \} \quad A^c \cap B \cap C = \{J, U, W\}$$

$$A^c \cap B^c \cap C = \{\ddot{A}, \ddot{O}, \ddot{U}, \beta\} \quad \text{Concordance Index } I_c = \frac{53}{2 \cdot 30} = 0.883 = 88.3\%$$

Betrachten wir alle Schriftsysteme A, B, C und D in einem allgemeinen Zusammenhang

Teilmengen		Anzahl  ...	Gewichtung	Teilsummen
1.	$A \cap B \cap C \cap D$	...	3	...
2.	$A \cap B \cap C \cap D^c$	...	2	...
3.	$A \cap B \cap C^c \cap D$	...	2	...
4.	$A \cap B^c \cap C \cap D$	...	2	...
5.	$A^c \cap B \cap C \cap D$	...	2	...
6.	$A \cap B^c \cap C^c \cap D$	...	1	...
7.	$A^c \cap B \cap C^c \cap D$	...	1	...
8.	$A^c \cap B^c \cap C \cap D$	...	1	...
9.	$A \cap B \cap C^c \cap D^c$	...	1	...
10.	$A \cap B^c \cap C \cap D^c$	...	1	...
11.	$A^c \cap B \cap C \cap D^c$	...	1	...
12.	$A^c \cap B^c \cap C^c \cap D$	...	0	0
13.	$A^c \cap B^c \cap C \cap D^c$	...	0	0
14.	$A^c \cap B \cap C^c \cap D^c$	...	0	0
15.	$A \cap B^c \cap C^c \cap D^c$	...	0	0
16.	$A^c \cap B^c \cap C^c \cap D^c$	...	0	0
			Total	$\Sigma$
Total	$G = A \cup B \cup C \cup D$	G	4	$3 \cdot  G $

$$\text{Concordance Index } I_c = \frac{\Sigma}{3 \cdot |G|}$$

konkret für Schriftsysteme A, B, C und D

Teilmenge		Anzahl  ...	Gewichtung	Teilsummen
1.	$A \cap B \cap C \cap D$	23	3	$3 \cdot 23$
2.	$A \cap B \cap C \cap D^c$	0	2	$2 \cdot 0$
3.	$A \cap B \cap C^c \cap D$	0	2	$2 \cdot 0$
4.	$A \cap B^c \cap C \cap D$	0	2	$2 \cdot 0$
5.	$A^c \cap B \cap C \cap D$	3	2	$2 \cdot 3$
6.	$A \cap B^c \cap C^c \cap D$	0	1	$1 \cdot 0$
7.	$A^c \cap B \cap C^c \cap D$	0	1	$1 \cdot 0$
8.	$A^c \cap B^c \cap C \cap D$	0	1	$1 \cdot 0$
9.	$A \cap B \cap C^c \cap D^c$	0	1	$1 \cdot 0$
10.	$A \cap B^c \cap C \cap D^c$	0	1	$1 \cdot 0$
11.	$A^c \cap B \cap C \cap D^c$	0	1	$1 \cdot 0$
12.	$A^c \cap B^c \cap C^c \cap D$	3	0	$0 \cdot 3$
13.	$A^c \cap B^c \cap C \cap D^c$	4	0	$0 \cdot 4$
14.	$A^c \cap B \cap C^c \cap D^c$	0	0	$0 \cdot 0$
15.	$A \cap B^c \cap C^c \cap D^c$	0	0	$0 \cdot 0$
16.	$A^c \cap B^c \cap C^c \cap D^c$	0	0	$0 \cdot 0$
			Total	75
Total	$G = A \cup B \cup C \cup D$	33	3	99

Concordance Index  $I_c = \frac{75}{3 \cdot 33} = 0.758 = 75.8\%$

- |   |   |
|---|---|
| 1. $A \cap B \cap C \cap D = \{A, B, C, D, \dots I, K, \dots T, V, X, Y, Z\}$ | 9. $A \cap B \cap C^c \cap D^c = \{\dots\}$                                     |
| 2. $A \cap B \cap C \cap D^c = \{\dots\}$                                     | 10. $A \cap B^c \cap C \cap D^c = \{\dots\}$                                    |
| 3. $A \cap B \cap C^c \cap D = \{\dots\}$                                     | 11. $A^c \cap B \cap C \cap D^c = \{\dots\}$                                    |
| 4. $A \cap B^c \cap C \cap D = \{\dots\}$                                     | 12. $A^c \cap B^c \cap C^c \cap D = \{\text{Æ}, \emptyset, \text{Å}\}$          |
| 5. $A^c \cap B \cap C \cap D = \{J, U, W\}$                                   | 13. $A^c \cap B^c \cap C \cap D^c = \{\text{Ä}, \text{Ö}, \text{Ü}, \text{ß}\}$ |
| 6. $A \cap B^c \cap C^c \cap D = \{\dots\}$                                   | 14. $A^c \cap B \cap C^c \cap D^c = \{\dots\}$                                  |
| 7. $A^c \cap B \cap C^c \cap D = \{\dots\}$                                   | 15. $A \cap B^c \cap C^c \cap D^c = \{\dots\}$                                  |
| 8. $A^c \cap B^c \cap C \cap D = \{\dots\}$                                   | 16. $A^c \cap B^c \cap C^c \cap D^c = \{\dots\}$                                |

## Grenzbetrachtungen

Fall 1 Alle 4 Schriftsysteme wären gleich, z.B. Latein:

$A \cap B \cap C \cap D = \{ A, B, C, D, \dots I, K, \dots T, V, X, Y, Z \}$  mit  $|A \cap B \cap C \cap D| = 23$  und Gewichtung 3

alle anderen Kombinationen sind  $\{\dots\}$  mit  $|\{\dots\}| = 0$

Die Grundmenge  $G = A \cup B \cup C \cup D = \{ A, B, C, D, \dots I, K, \dots T, V, X, Y, Z \}$  mit  $|G| = 23$

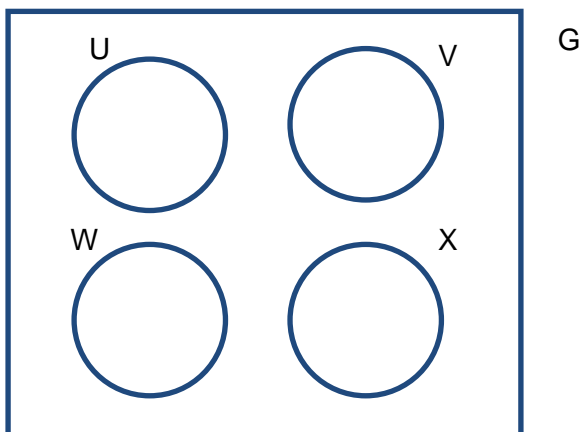
Concordance Index  $I_c = \frac{3 \cdot 23}{3 \cdot 23} = 1.0 = 100\%$  Übereinstimmung 100%

Fall 2 Alle 4 Schriftsysteme wären komplett verschieden:

1.  $A \cap B \cap C \cap D = \{ \}$  mit  $|A \cap B \cap C \cap D| = 0$  und Gewichtung 3

2. alle Fälle mit Teil  $\dots U \cap V \dots = \{ \}$  (zwei nicht komplemente Mengen kommen vor)

3. alle Fälle mit Teil z.B.  $U^c \cap V^c \cap W^c \cap X = X$



(drei komplemente Mengen ergeben die vierte Menge, die keine Elemente enthält, die zusätzlich in einer anderen Menge vorkommen (da ohne gemeinsame Schriftzeichen, da alle Mengen disjunkt/Elemente-fremd sind), also Gewichtung 0. Diese bilden alle keinen Beitrag zur totalen Summe

4. Fall  $U^c \cap V^c \cap W^c \cap X^c = (A \cup B \cup C \cup D)^c = G^c = \{ \}$  mit Gewichtung 0

Die Grundmenge  $G = A \cup B \cup C \cup D$

mit  $|G| = |A| + |B| + |C| + |D|$  Summe aller verschiedenen Zeichen, die grösser als 0 ist.

Concordance Index  $I_c = \frac{0}{3 \cdot |G|} = 0.0 = 0\%$  Übereinstimmung 0%

Zur Ausgangslage zurück:

Die Concordance Index  $I_c$  (Integrationsrate) der vier Norditalienischen Sprachen

**Concordance Index  $I_c$**  für A = Raetisch, B = Lepontisch, C = Etruskisch und D = Runen

Teilmengen		Anzahl  ...	Gewichtung	Teilsummen
1	$A \cap B \cap C \cap D$	34	3	102
2	$A \cap B \cap C \cap D^c$	28	2	56
3	$A \cap B \cap C^c \cap D$	11	2	22
4	$A \cap B^c \cap C \cap D$	9	2	18
5	$A^c \cap B \cap C \cap D$	13	2	26
6	$A \cap B^c \cap C^c \cap D$	18	1	18
7	$A^c \cap B \cap C^c \cap D$	11	1	11
8	$A^c \cap B^c \cap C \cap D$	7	1	7
9	$A \cap B \cap C^c \cap D^c$	32	1	32
10	$A \cap B^c \cap C \cap D^c$	10	1	10
11	$A^c \cap B \cap C \cap D^c$	18	1	18
12	$A^c \cap B^c \cap C^c \cap D$	52	0	0
13	$A^c \cap B^c \cap C \cap D^c$	86	0	0
14	$A^c \cap B \cap C^c \cap D^c$	102	0	0
15	$A \cap B^c \cap C^c \cap D^c$	69	0	0
16	$A^c \cap B^c \cap C^c \cap D^c$	0	0	0
			Total	320
Total	$G = A \cup B \cup C \cup D$	500		1500

**Concordance Index  $I_c$**  **0.21333** **21.3%**

**Concordance Index  $I_c$**  für A = Raetisch, B = Lepontisch und C = Etruskisch

Teilmengen		Anzahl  ...	Gewichtung	Teilsummen
1	$A \cap B \cap C$	62	2	124
2	$A \cap B \cap C^c$	43	1	43
3	$A \cap B^c \cap C$	19	1	19
4	$A^c \cap B \cap C$	31	1	31
5	$A^c \cap B^c \cap C$	93	0	0
6	$A^c \cap B \cap C^c$	113	0	0
7	$A \cap B^c \cap C^c$	87	0	0
8	$A^c \cap B^c \cap C^c$	0	0	0
			Total	217
Total	$G = A \cup B \cup C \cup D$	448		896

**Concordance Index  $I_c$**  **0.24219** **24.2%**

Die Abweichung mit vier Schriftsätzen (mit Runenzeichen) von drei Schriftsätzen (ohne Runenzeichen)

ist nur  $r = \frac{0.242 - 0.213}{0.242} = 0.120 = 12.0\%$ , Runisch ist mit 88.0% integriert.