Gesetz zur Gruppenbildung - Beispielsammlung

SINNVOLLE SÄTZE TROTZ CHAOTISCHER ABFOLGE IHRER WORTBAUSTEINE

WORUM ES GEHT

Im Beitrag: Link → SPRACHVERGLEICHE wurde die Frage behandelt, ob die vier grundverschiedenen Sprachen: Deutsch, kroatisch, finnisch und italienisch in der Abfolge ihrer «ersten Wortbausteine» - gebildet jeweils aus «Vokal + Vokal», bzw. «Vokal + Konsonant», bzw. «Konsonant + Vokal», bzw. «Konsonant + Konsonant» - gewisse Ähnlichkeiten oder gar Gemeinsamkeiten aufweisen. Betreffend die Ergebnisse wird auf den obigen LINK verwiesen.

Bei der Behandlung dieser Frage wurde nebenbei auch die Abfolge der genannten Wortbausteine erfasst. Damit stellt sich HIER eine andere **FRAGE:**

Gehorcht die Abfolge der «ersten Wortbausteine» eines Textes dem «Gesetz zur Gruppenbildung»?

Sollte dies zutreffen, hätte es zur Konsequenz, dass das aus der realen Abfolge der «ersten Wortbausteine» gebildete «Stichprobenprofil», mit dem aus deren Häufigkeitsverteilung entwickelten «logischen Referenzprofil» quasi deckungsgleich sein müsste (hoher Plausibilitätsgrad!). Findet sich keine (angenäherte) Übereinstimmung, interessiert die Frage, wie dies zu erklären ist.

VORGEHEN

Die Ermittlung der «Paketbildung» für das «logische Referenzprofil» und für das «Stichprobenprofil» erfolgte in der früher detailliert beschriebenen Weise (siehe QR – Code am Schluss dieses Berichts). Das gilt auch für die interessierenden Plausibilitätsgrade je nach Wortbaustein im Einzelnen sowie für den Gesamt – Plausibilitätsgrad einer Sprache. Als «Sprache» wurden dazu die drei dort untersuchten Kurztexte pro Sprache (Texte à je ca. 100 Worte) nahtlos zu einem Gesamttext (à je ca. 250 – 300 Worte) zusammengefasst.

ERGEBNISSE IM ÜBERBLICK

TABELLE 1 zeigt den Gesamt – Plausibilitätsgrad ($P_{n=...}$) basierend auf der Anzahl (n) Wortbausteine der zusammengefügten Texte je Sprache. Zudem den jeweils auf 1000 Wortbausteine hochgerechneten effektiven Plausibilitätsgrad ($P_{ZF=...}$) als Folge des Zuschlagsfaktors (ZF). Siehe dazu: Link → ZUR GLEICHWERTIGKEIT VON PLAUSIBILITAETSGRADEN.

SPRACHE	ANZAHL WORTE (n)	PLAUSIBILITÄTSGRAD (%) P (n = ...)	ZUSCHLAGSFAKTOR (ZF)	PLAUSIBILITÄTSGRAD (%) P (ZF =...; n = 1000)
DEUTSCH	287	46.3	1.145	53.1
KROATISCH	306	46.9	1.138	53.4
FINNISCH	233	37.4	1.168	43.7
ITALIENISCH	309	41.2	1.137	42.3

TABELLE 1: PLAUSIBILITÄTGRADE (%)

Ausgangspunkt für die Ergebnisse laut Tabelle 1 waren die prozentualen Häufigkeiten der vier Wortbausteine (p.m.: V+V, V+K, K+V, K+K) je nach Sprache. Diese Häufigkeiten sind in **TABELLE 2** dargestellt.

SPRACHE	Vokal + Vokal (%)	Vokal + Konsonant (%)	Konsonant + Vokal (%)	Konsonant + Konsonant (%)
DEUTSCH	6	23.8	60.4	9.6
KROATISCH	0	15.5	56.1	28.4
FINNISCH	7	13,6	68.9	10.5
ITALIENISCH	0.3	22.1	68.5	9.1

TABELLE 2: HÄUFIGKEITEN DER "ERSTEN WORTBAUSTEINE" (%)

Für die «Sprache deutsch» kann die detaillierte Auswertung in den **TABELLEN 3 bis 6** im **ANHANG** (Querformat) verfolgt werden.

INTERPRETATION

Gemessen am «Gesetz zur Gruppenbildung» ist die Abfolge der Wortbausteine als «chaotisch» zu bezeichnen. Die hier generell fehlende Übereinstimmung von «logischem Referenzprofil» und «Stichprobenprofil» kann bekanntlich drei (bzw. vier, siehe c 2)) mögliche Gründe haben:

a) Die Abfolge für das Referenzprofil wird ungenau erfasst (Messfehler, Ablesefehler – z.B. bei der Materialprüfung) → ist HIER kein Argument.

b) Die Gewichtung (Häufigkeiten) der Daten/Merkmale/etc. verschiebt sich während der Dauer der Datenerfassung (Stimmungsumschwung im Laufe einer längeren Befragungsperiode) → ist HIER kein Argument.

c 1) Die unter sich gleichgewichtig oder ungleichgewichtig verteilten Daten/Merkmale/etc. sind in der Urne (aus welcher die Werte nach dem FAIREN ZUFALLSPRINZIP gezogen werden), NICHT HOMOGEN verteilt (Clusterbildung statt homogener Verteilung der Mischanteile → Betonkies vor Mischvorgang, Salatsauce, etc.) → ist HIER kein Argument.

c 2) In der Umkehrung von Fall c 1) kann man sich auch vorstellen, dass die «Mischung» (also die Verteilung der Daten/Merkmale/etc. in der Urne) zwar eine homogene ist, dass aber daraus KEINE FAIR - ZUFÄLLIGE Ziehung erfolgt – sondern dass diese SELEKTIV vorgenommen wird. Dies läge beispielsweise auch dann vor, wenn – evtl. mit böser Absicht – ein dem Gesetz von Benford entsprechender Zahlenverlauf «selektiv» verändert würde.

Für die HIER untersuchte Buchstabenfolge kann der Fall c 2) als Erklärung für die tiefen Plausibilitätsgrade dienen:

Damit ein sinnvoller Text entsteht, genügt es nicht, dass seine «Erstbuchstaben je Wort» in der richtigen Häufigkeitsverteilung vorliegen, sondern sie müssen auch SELEKTIV – RICHTIG - und nicht «fair – zufällig» gezogen werden.

Um damit auf den Titel dieses Berichts zurückzukommen:

Sinnvolle Sätze entstehen nicht trotz chaotischer, sondern DANK chaotischer, hier selektiver Abfolge ihrer Wortbausteine!

Ob diese Einsicht von irgendwelcher praktischen Bedeutung ist, sei dahingestellt.

Weiteres zum «Gesetz zur Gruppenbildung» siehe:

Februar 2026 / Ba.

ANHANG

PROMILLE - ANTEILE ZUR GRUPPENBILDUNG NACH DEFINIERTEM ZAHLEN - MIX IN DER URNE

Gruppenbildung "m"	Promille Zahlenband 1	Promille Zahlenband 2	Promille Zahlenband 3	Promille Zahlenband 4	Promille Zahlenband 5	Promille Zahlenband 6	Promille Zahlenband 7	Promille Zahlenband 8	Promille Zahlenband 9	Promille Zahlenband 10	Promille ALLE Zahlenbänder 1 bis 10
m = 1	53	136	93	79	0	0	0	0	0	0	360
m = 2	3.6	31.3	57.4	7.9	0	0	0	0	0	0	100.4
m = 3	0.3	8.2	35.4	0.6	0	0	0	0	0	0	42.7
m = 4	0.1	2	20.8	0.1	0	0	0	0	0	0	23.7
m = 5	0	0.6	12.7	0	0	0	0	0	0	0	13.9
m = 6	0	0.1	7.9	0	0	0	0	0	0	0	7.5
m = 7	0	0	4.8	0	0	0	0	0	0	0	5.1
m = 8	0	0	2.8	0	0	0	0	0	0	0	1.8
m = 9	0	0	1.8	0	0	0	0	0	0	0	2.1
m= 10	0	0	0.9	0	0	0	0	0	0	0	1.1
NICHT ERFASSTE ANZAHL DATEN (alle für m > 10); gemittelt	Bei extrem unterschiedlichen Gewichtungen der Urnenzahlen kommt es vor, dass bis zur maximal erfassten Gruppenbildung (bis m = 10) nicht alle 1000 Daten erfasst werden. Diese (gemittelte!) ANZAHL - bestehend aus Gruppen mit m > 10 - ist rot eingetragen.										19
ERFASSTE ANZAHL DATEN (alle bis m = 10; gemittelt)	61.5	234.8	589.3	97	0	0	0	0	0	0	982.6 (GEMITTELTE ANZAHL AUS 20 x 1000)

Hinweis:
Da das Total (blau + rot) aus je 20 (mal 1000) getrennt gemittelten Ergebnissen besteht, muss dieses nicht zwingend = 1000 betragen!

TABELLE 3: Logisches Referenzprofil «DEUTSCH»

FEHLER - BANDBREITEN ZUR GRUPPENBILDUNG NACH DEFINIERTEM ZAHLEN - MIX IN DER URNE (ergänzend zu TABELLE 5)

Gruppenbildung "m"	Zahlenband 1	Zahlenband 2	Zahlenband 3	Zahlenband 4	Zahlenband 5	Zahlenband 6	Zahlenband 7	Zahlenband 8	Zahlenband 9	Zahlenband 10	Promille ALLE Zahlenbänder 1 bis 10
m = 1	5.68	10.88	11.1	8.3	0	0	0	0	0	0	21.48
m = 2	1.49	5.86	6.08	2.67	0	0	0	0	0	0	9.06
m = 3	0.43	2.13	5.01	0.64	0	0	0	0	0	0	5.44
m = 4	0.21	2.13	4.26	0.21	0	0	0	0	0	0	5.54
m = 5	0	0.85	3.3	0	0	0	0	0	0	0	3.41
m = 6	0	0.21	2.35	0	0	0	0	0	0	0	2.45
m = 7	0	0	1.49	0	0	0	0	0	0	0	2.24
m = 8	0	0	1.39	0	0	0	0	0	0	0	1.07
m = 9	0	0	1.17	0	0	0	0	0	0	0	1.81
m= 10	0	0	1.17	0	0	0	0	0	0	0	1.17
m = 10+	Bei extrem unterschiedlichen Gewichtungen der Urnenzahlen kommt es vor, dass bis zur maximal erfassten Gruppenbildung (bis m = 10) nicht alle 1000 Daten erfasst werden. Diese (gemittelte!) ANZAHL - bestehend aus Gruppen mit m > 10 - ist rot eingetragen.										0

TABELLE 4: Fehlerbandbreite zu Tabelle 3

PROMILLE - ANTEILE ZUR GRUPPENBILDUNG AUS GELIEFERTER DATENABFOLGE

Gruppenbildung "m"	PROMILLEANTEILE Zahl1	PROMILLEANTEILE Zahl 2	PROMILLEANTEILE Zahl 3	PROMILLEANTEILE Zahl 4	PROMILLEANTEILE Zahl 5	PROMILLEANTEILE Zahl 6	PROMILLEANTEILE Zahl 7	PROMILLEANTEILE Zahl 8	PROMILLEANTEILE Zahl 9	PROMILLEANTEILE Zahl 10	PROMILLEANTEILE aller Zahlen 1 bis 10
m = 1	17	51	30	18	0	0	0	0	0	0	116
m = 2	0	5	20	4	0	0	0	0	0	0	29
m = 3	0	0	12	0	0	0	0	0	0	0	12
m = 4	0	2	3	0	0	0	0	0	0	0	5
m = 5	0	0	2	0	0	0	0	0	0	0	2
m = 6	0	0	4	0	0	0	0	0	0	0	4
m = 7	0	0	1	0	0	0	0	0	0	0	1
m = 8	0	0	2	0	0	0	0	0	0	0	2
m = 9	0	0	0	0	0	0	0	0	0	0	0
m= 10	0	0	0	0	0	0	0	0	0	0	0
NICHT ERFASSTE ANZAHL DATEN (alle für m > 10)	Bei extrem unterschiedlichen Gewichtungen der Urnenzahlen kommt es vor, dass die maximal erfasste Gruppenbildung (bis m = 10) nicht alle GELIEFERTEN Daten umfasst. Diese ANZAHL - bestehend aus Gruppen mit m > 10 - ist in "Zeile rechts unten" rot angegeben.										713
ERFASSTE ANZAHL DATEN (alle bis m = 1	17	69	175	26	0	0	0	0	0	0	287
GEWICHTUNG DER GELIEFERTEN DATEN (Pr	59.23	240.42	609.76	90.59	0	0	0	0	0	0	1000

TABELLE 5: Stichprobenprofil «DEUTSCH»

PLAUSIBILITÄTSGRAD [%] FÜR DIE GELIEFERTE ZAHLENFOLGE

(Hilfsspalten überdeckt oder Zahlen weiss)	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	
Gruppenbildung "m"	Zahl1	Zahl 2	Zahl 3	Zahl 4	Zahl 5	Zahl 6	Zahl 7	Zahl 8	Zahl 9	Zahl 10	
m = 1	35.9	40.8	36.6	25.5	0	0	0	0	0	0	
m = 2	0	19.7	39	76.5	0	0	0	0	0	0	
m = 3	100	0	39.5	100	0	0	0	0	0	0	
m = 4	100	100	18.1	100	0	0	0	0	0	0	
m = 5	0	100	21.3	0	0	0	0	0	0	0	
m = 6	0	100	72.1	0	0	0	0	0	0	0	
m = 7	0	0	100	0	0	0	0	0	0	0	
m = 8	0	0	100	0	0	0	0	0	0	0	
m = 9	0	0	0	0	0	0	0	0	0	0	
m= 10	0	0	100	0	0	0	0	0	0	0	
ADDIERTE PROZENTE PRO ZAHL	610.3	3077.8	8540.6	1071	0	0	0	0	0	0	PLAUSIBILITÄT GESAMT [%]
ERFASSTE ANZAHL DATEN (alle bis m = 10)	17	69	175	26	0	0	0	0	0	0	287
GEMITTELTE PLAUSIBILITÄT [%] PRO ZAHL	35.9	44.6	48.8	41.2							46.3
HILFSZEILE	35.9	44.6	48.8	41.2	0	0	0	0	0	0	

TABELLE 6: Plausibilitätsgrade «DEUTSCH»

Februar 2026