

# Ausgliederungsreihenfolge der Indogermania auf Grundlage des LIV-2\*

Hans J. Holm

هانس هولم

ζᾶνσ ζᾶνσ

한스 홀름

# Gliederung



1 Motivation: Traditionelle vs. quantitative Methoden



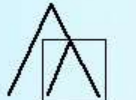
2 Fachprüfung: Sprachwandel quantitativ



3 Anwendung: Unsere Basiserkenntnis vs. andere Versuche



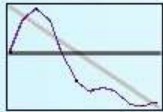
4 Erweiterung: Mehrstufige Aufgliederungen



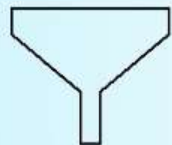
5 Neuerungen ? Am Beispiel Arisch



6 Störfaktor Streuung



7 Störfaktor Verteilung: Ideale vs. Zipf vs. LIV (meine Daten)  
Analyse und Folgerung



8 Ergebnisse: 8.1 Ausgliederungsergebnis  
8.2 Diskussion möglicher Fehler  
8.3 Teste an realer Geografie und  
mit Ringes "recalcitrant features"  
8.4 Schluss - Rückblick  
- Zusammenfassung, Lehren

## 1.1 Traditionelle vgl. Sprachwissenschaft

in 200 Jahren keinen Konsens erzielt, z.B.:

- NW - Block?
- Germanisch?
- Balto-Slavisch?
- Italo-Keltisch?
- Albanisch?
- Armenisch : Griechisch?
- Indo-Hethitisch?
- Nuristani : Indo-Iranisch?

*In den Augen der Öffentlichkeit machen das jetzt die Bioinformatiker*

*jedoch ...*

## 1.2 Quantitative Ansätze ...

... oft nicht besser, wenn sie

- bereits die Erkennung "Griechisch" vs. "Germanisch" als Evidenz hinstellen;

*Billigargument!* Gray&Atkinson, Rexová et al.

## 1.2 Quantitative Ansätze ...

... oft nicht besser, wenn sie

- eine feste Ersetzungsrate voraussetzen (offene oder verdeckte Glottochronologie);

Vertreter: Gray & Atkinson, trotz ggf. Behauptung!  
Alle UPGMA\*-Anwendungen

## 1.2 Quantitative Ansätze ...

... oft nicht besser, wenn sie

- oberflächliche Ähnlichkeit der genealogischen Verwandtschaft gleichsetzen.

Vertreter: Fast alle !

*Dabei ist ...*

## 1.3 Verständnislose Ablehnung

(Was ich nicht weiß - )  
"das können wir nicht wissen"

(Typische Schutzbehauptung altväterlicher Professoren)

*pave ..*

*genauso unwissenschaftlich, wie ...*

# 1.4 Unverstandene Falschanwendung

## Einige 'Milchmädchen-Mathematiker':

- 1954 Porzig
- 1992 I. Dyen "... based on the simple! principle that the lexico-statistical percentage indicates the degree of relatedness."
- 1995 F. Kammerzell (Ägyptologe) in IF 99: "simples! Kalkül"
- 1998 J.H. Holst in HS (ein paar alb.-gri. Isoglossen).

## Einige unverbesserliche 'Glottochronologen':

- D. Wunderlich (Germanist em.) im Internet ...
- Chr. Lehmann, U-EF ("Linguist") lehrt (!) Glottochronologie ...
- P. Forster (Mathematiker), Beispiel später ..
- Gray & Atkinson
- G. Starostin ..

*Wissenschaftliches Vorgehen: Zunächst exakte Analyse  
der quantitativen Zusammenhänge ...*



## 2.1 Situationsanalyse: Was geschieht?

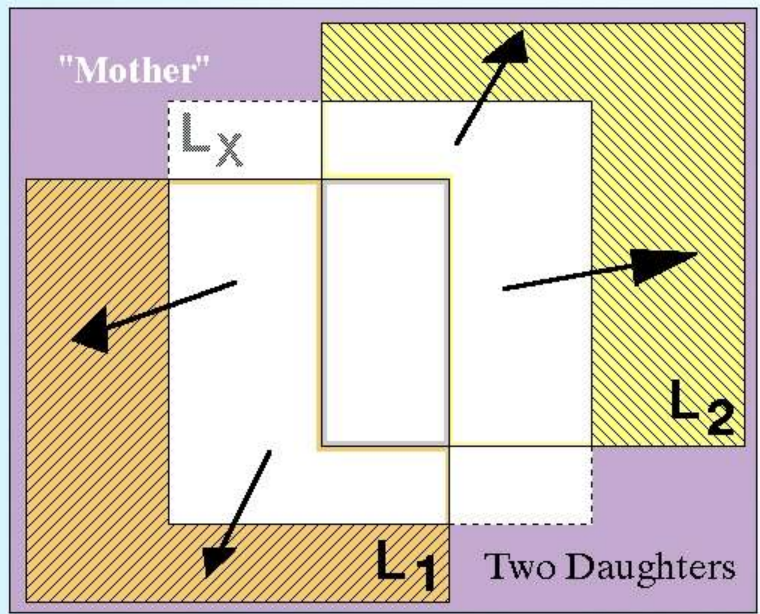
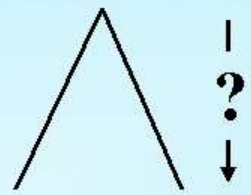


### Definition:

"Zwei Sprachen sind verwandt, wenn sie aus einer gemeinsamen Quellsprache hervorgegangen sind"  
(Metapher "Mutter- : Tochter"-sprache)

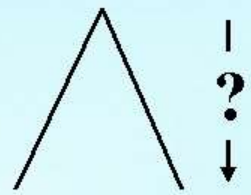
***bildlich ...***

# 2.2 Situationsanalyse: Was geschieht?

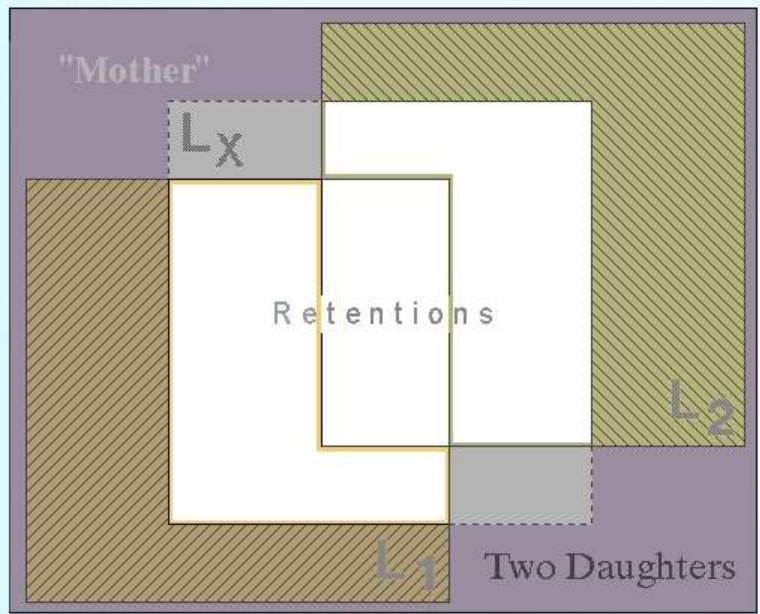


*Was gibt's da zu zählen? ...*

# 2.2 Situationsanalyse: Was geschieht?



Die Muttersprache  $L_x$  sei ausgestorben

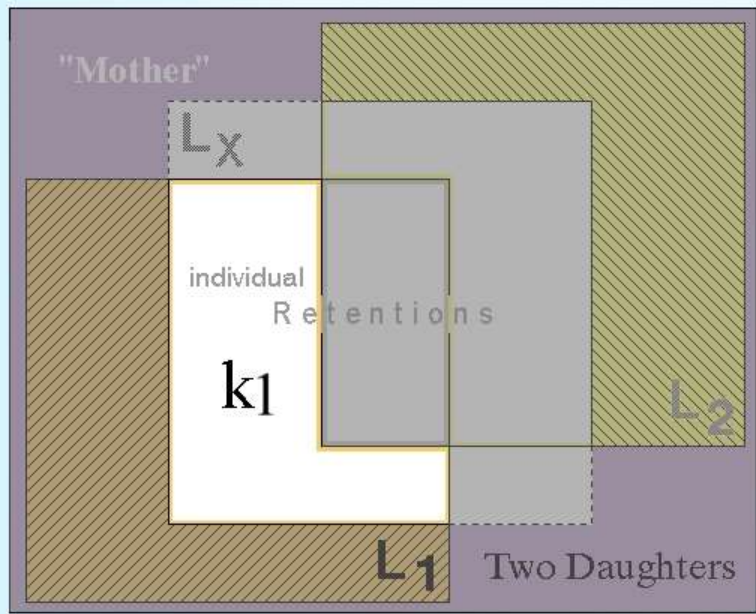
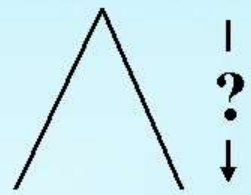


Es bleiben greifbar ..

N Originalmerkmale

*davon ...*

# 2.2 Situationsanalyse: Was geschieht?

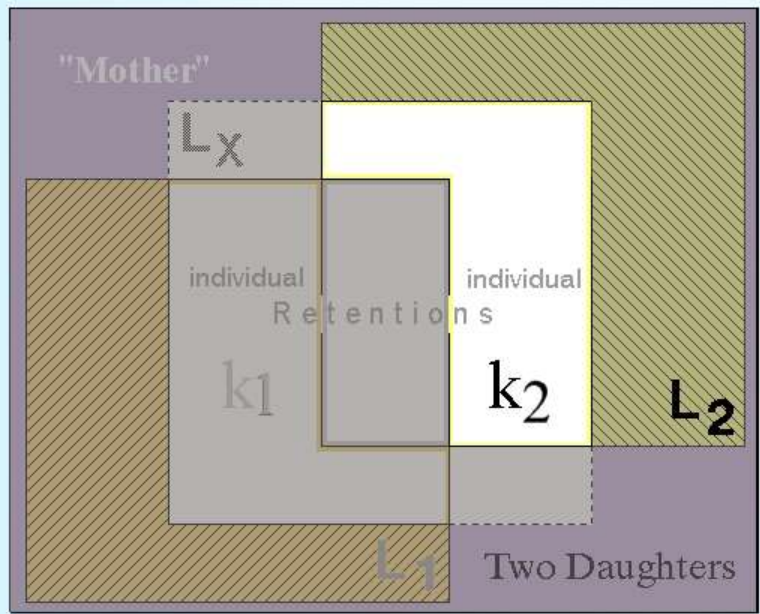
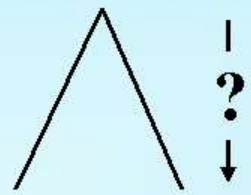


Wir können unterscheiden und zählen: Restliche

- N Originalmerkmale, davon
- k1 in L1 allein erhalten,

*und ...*

# 2.2 Situationsanalyse: Was geschieht?



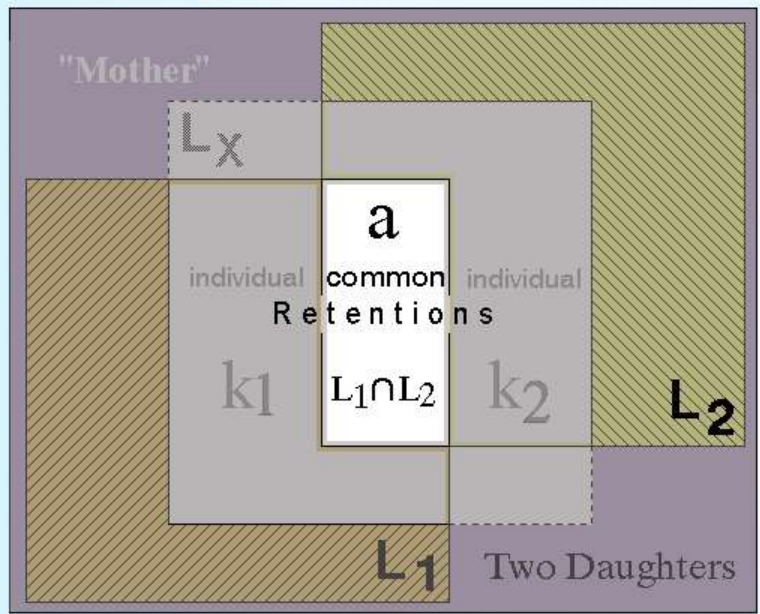
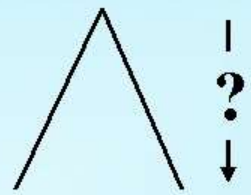
Wir können unterscheiden  
und z ä h l e n : Restliche

N Originalmerkmale, davon

• k<sub>2</sub> in L<sub>2</sub> " "

*sowie ...*

# 2.2 Situationsanalyse: Was geschieht?

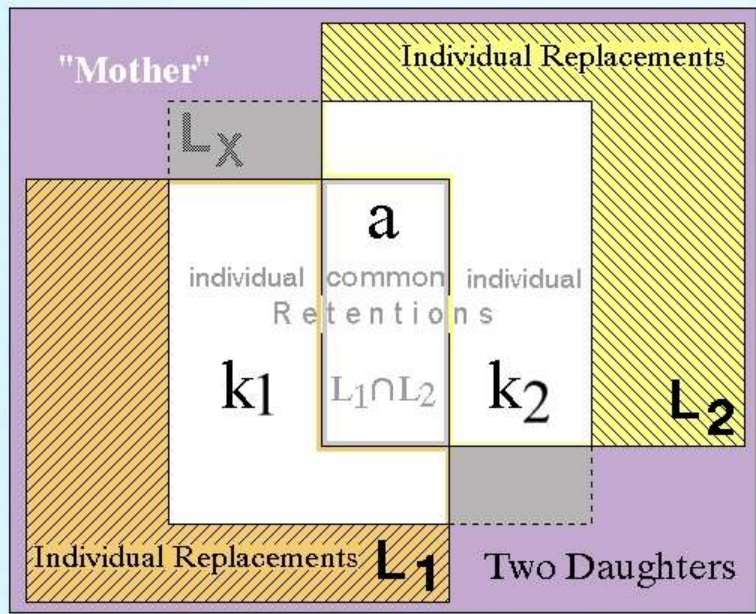
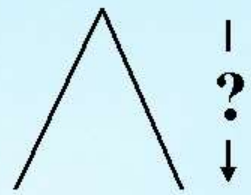


Wir können unterscheiden und zählen:

□ • a in beiden gemeinsam.

*alles Übrige ...*

# 2.2 Situationsanalyse: Was geschieht?

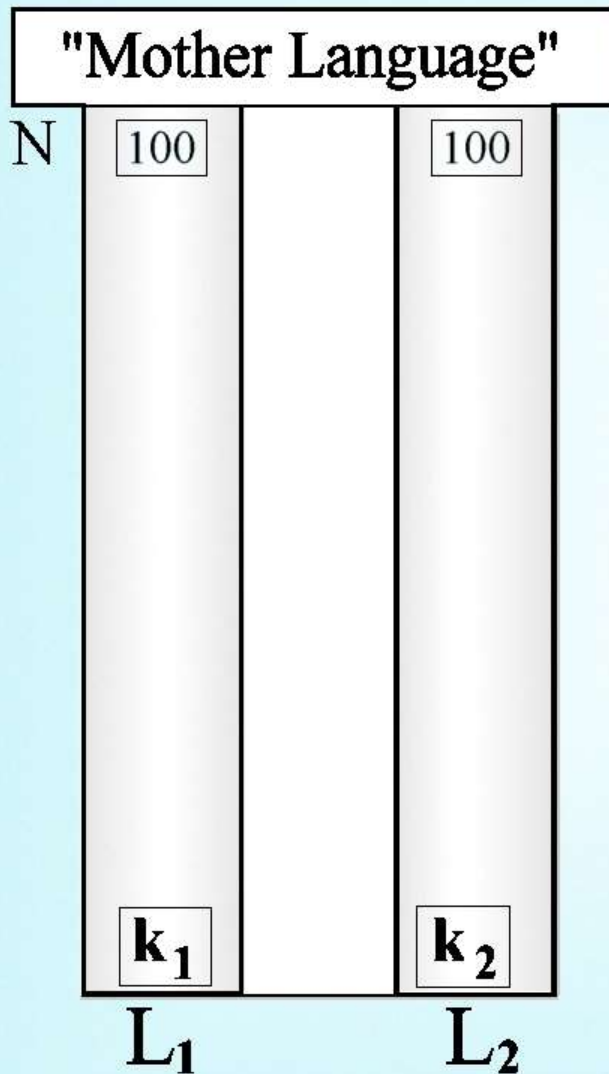
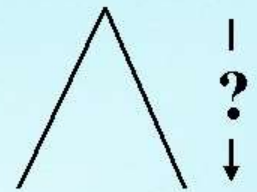


Wir können unterscheiden

  Alles Übrige geneuert.

*Einfacher: die beiden Töchter  
als zwei Säulen nebeneinander ...*

## 2.3 Ausgangssituation



Ausgangsmenge N

Für den Anfang = 100 (%)

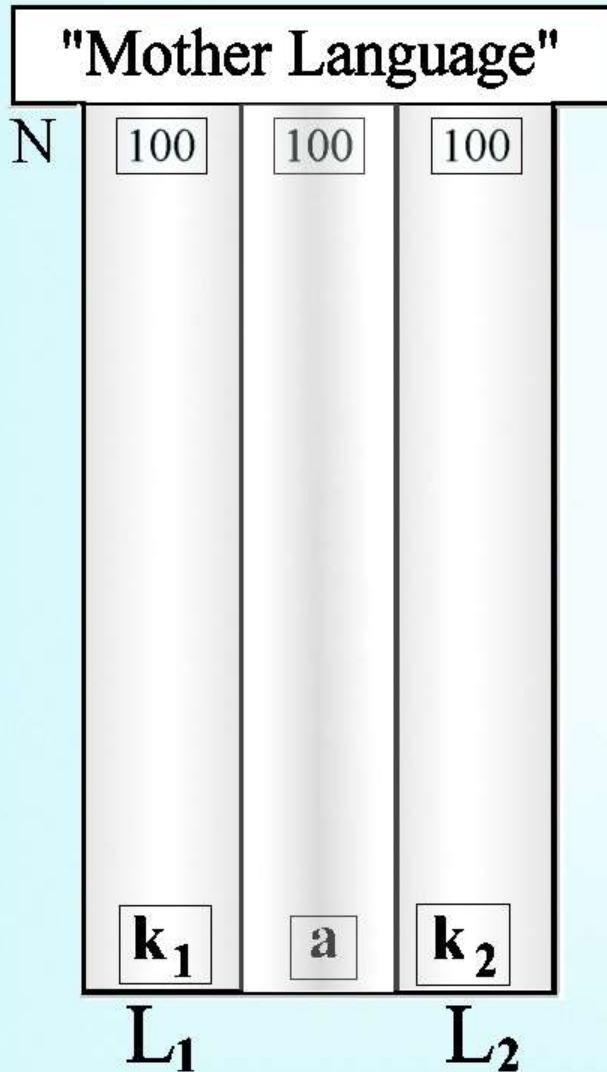
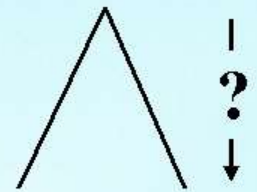
Davon befinden sich bei Ausgliederung  
in beiden Tochttersprachen natürlich  
noch

-  $k = 100\%$  Originalmerkmale

*und damit ...*



## 2.3 Ausgangssituation

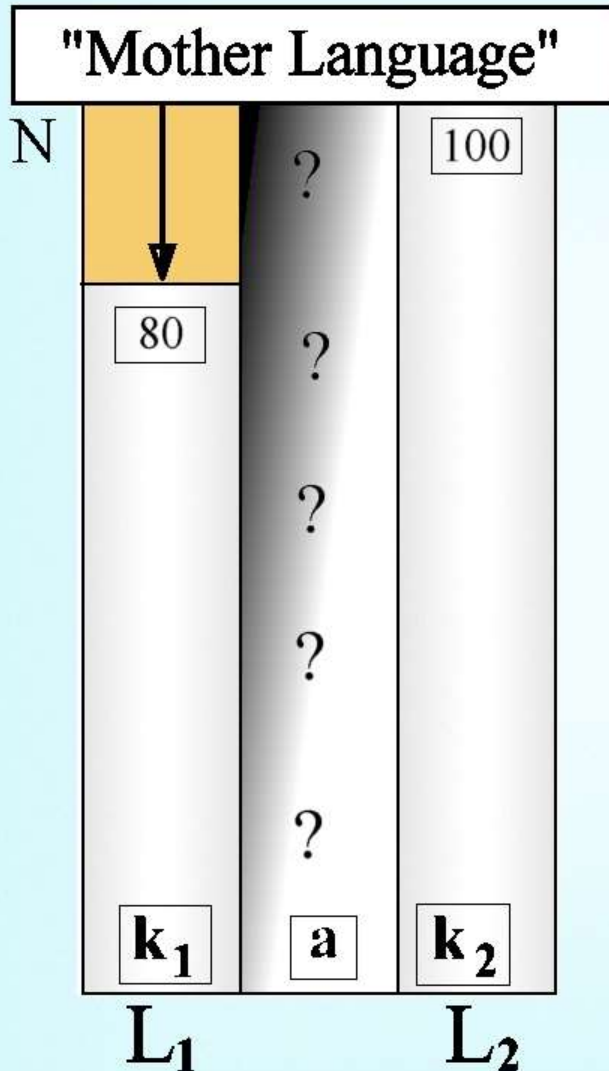
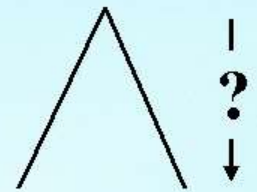


und damit

- a = 100 % Übereinstimmungen.

*Aber es bleibt nicht so ...*

## 2.4 Erste Tochter wird selbständig



Zunächst verliert Tochter  $L_1$   
20 % ihrer ursprünglichen  
Merkmale ..

Unsere Parameter:

$N = 100 \% =$  Ausgangswert

$k_2 = 100 \% =$  Ausgangswert

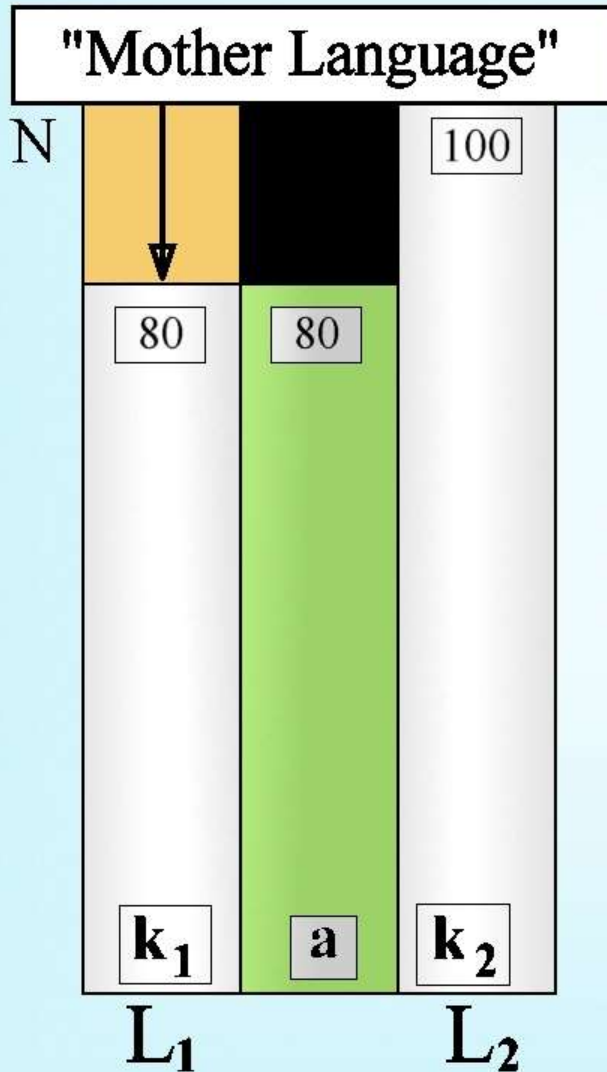
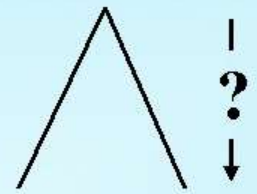
$k_1 = 80 \%$

$a = ?$

***Preisfrage:***

***Mit wieviel Gemeinsamkeiten  
können wir rechnen ?***

## 2.4 Erste Tochter wird selbständig

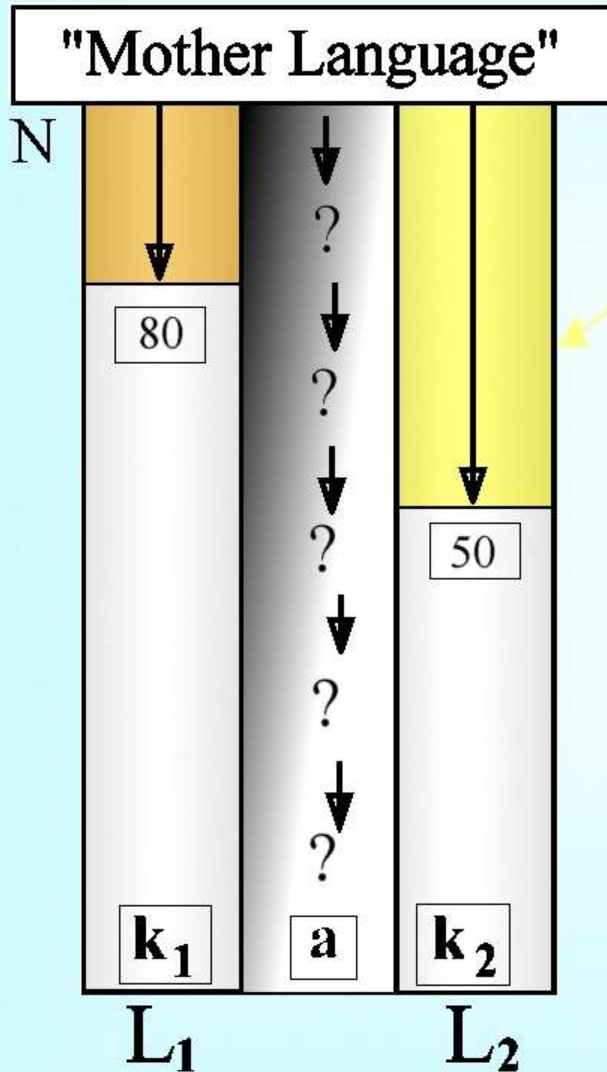


Die richtige Antwort lautet natürlich

$$a = 80\%$$

*Da sich jede Sprache wandelt,  
kommen wir nun zur zweiten Tochter ...*

## 2.4 Zweite Tochter wird selbständig



Auch Tochter L2 ersetzt nun - der Einfachheit halber - 50 % ihrer Merkmale.

Unsere Parameter:

$N = 100 \% =$  Ausgangswert

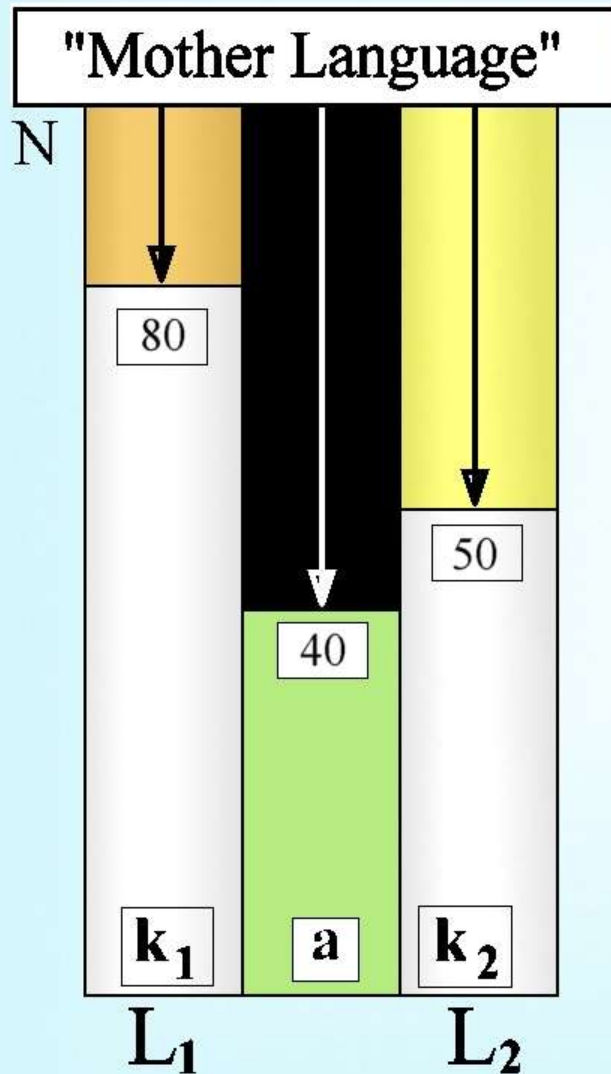
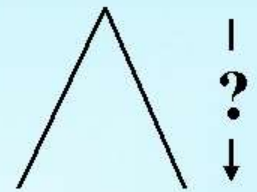
$k_1 = 80 \%$

$k_2 = 50 \%$

$a = ?$

*Mit wievielen Gemeinsamkeiten ist jetzt noch zu rechnen ??*

## 2.4 Zweite Tochter wird selbständig

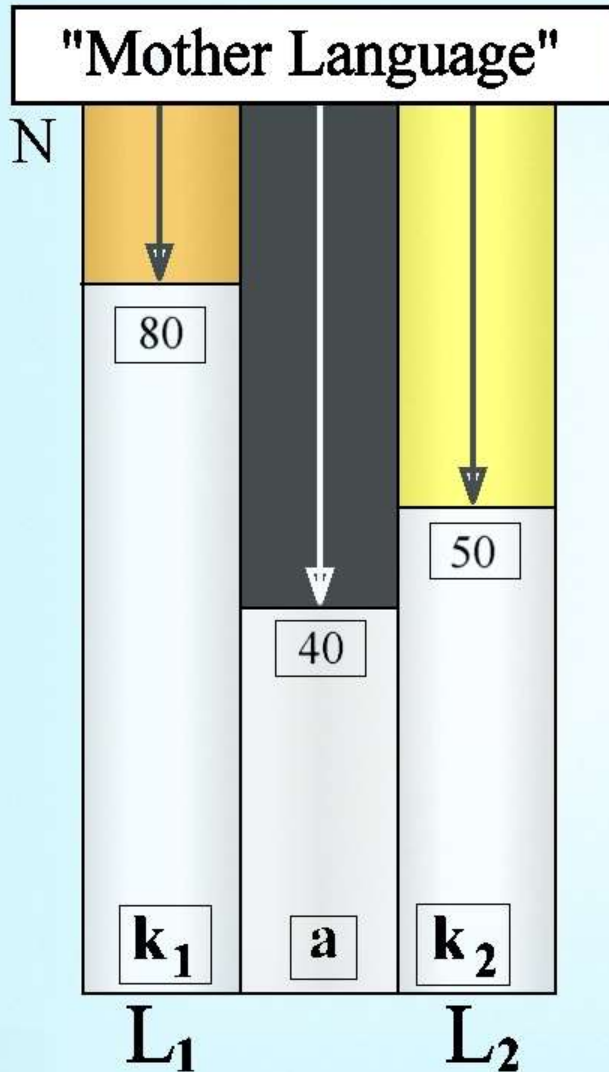
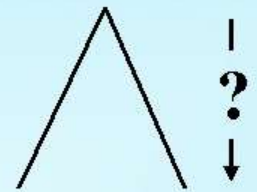


Die richtige Antwort lautet natürlich

a ~ 40 % \*

*Fassen wir zusammen ...*

## 2.5 Zusammenfassung



'a' durch schlichte Prozentrechnung:

als 50% von 80

oder  $\frac{50}{100}$  von 80 = 40

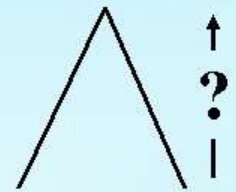
---

das sind nur unsere bereits bekannten Parameter

$$\frac{k_1 \cdot k_2}{N} = a$$

*Was nützt uns das?*

## 3.1 Was nützt uns das?

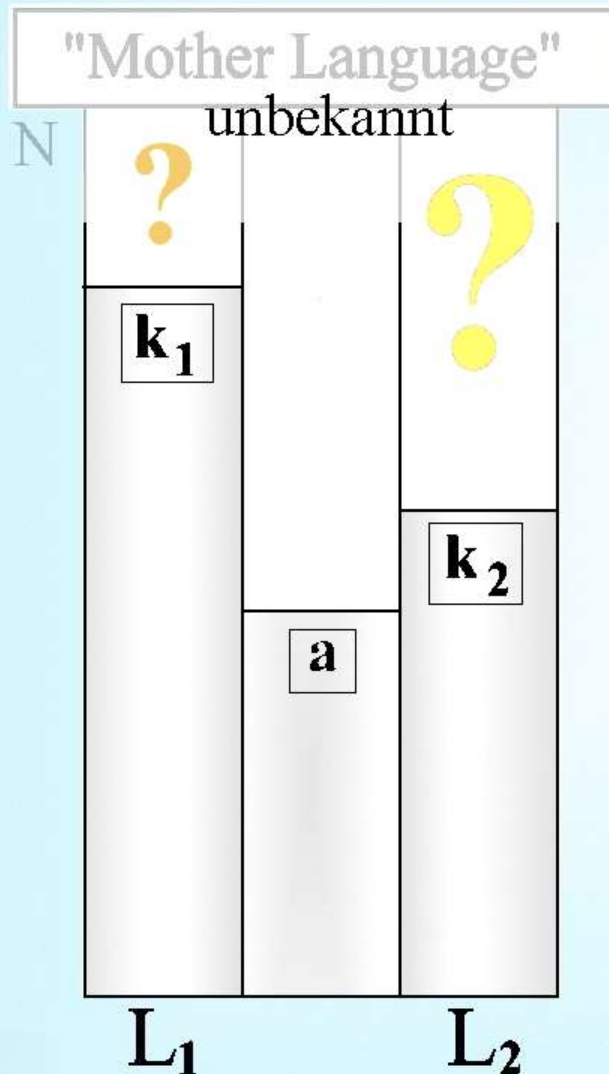


### Realität umgekehrt!

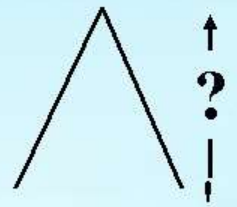
Wir können zwar für jede Sprachenpaarung (z.B. im LIV)

- die belegten Kognaten 'k'
- deren Gemeinsamkeiten 'a' auszählen;

Der zugrundeliegende Ausgangswert 'N' ist uns aber zunächst unbekannt.



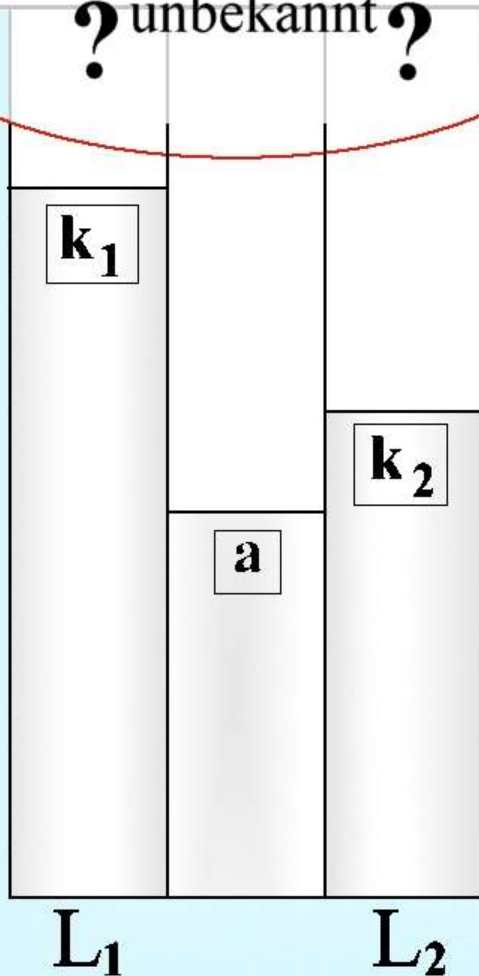
## 3.2 Berechnung der Ausgangsmenge



"Mother Language"

? unbekannt ?

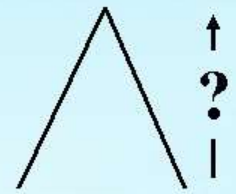
N



Dabei ist diese Ausgangsmenge 'N' viel wichtiger: Da sie nämlich nur abnehmen kann, bedeutet



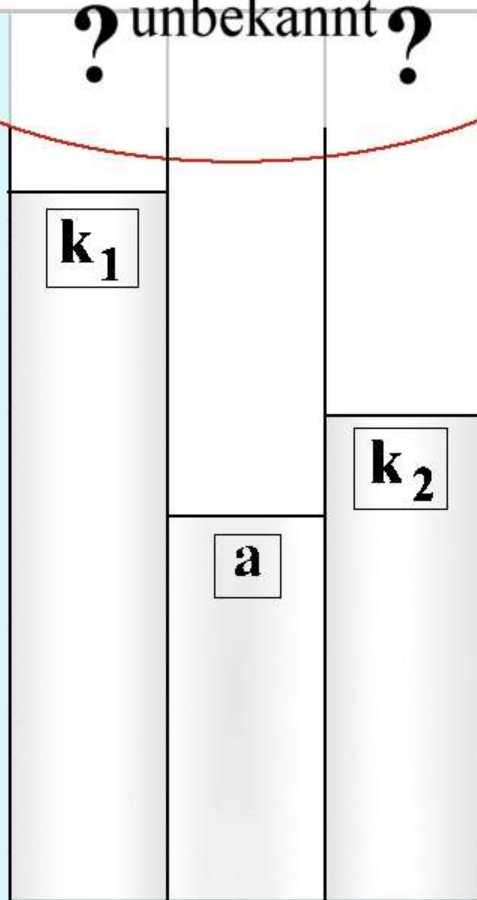
## 3.2 Berechnung der Ausgangsmenge



"Mother Language"

? unbekannt ?

N



$L_1$

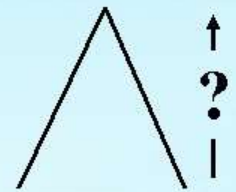
$L_2$

Dabei ist diese Ausgangsmenge 'N' viel wichtiger: Da sie nämlich nur abnehmen kann, bedeutet

hohe Ausgangsmenge = frühe Aufspaltung,  
niedrigere " = spätere Aufspaltung,

*Und wie erhalten wir diese Ausgangsmenge?*

## 3.2 Berechnung der Ausgangsmenge



"Mother Language"  
? unbekannt ?

N erhalten wir durch  
einfaches Umstellen der Parameter:

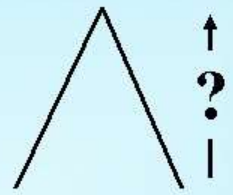
$$\frac{k_1 \cdot k_2}{N} = a$$

zu

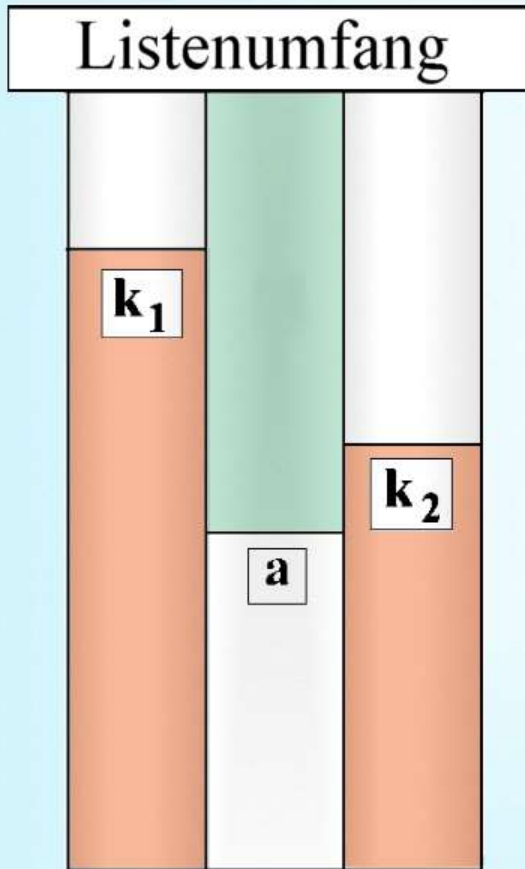
$$\frac{k_1 \cdot k_2}{a} = \hat{N}$$

*Wir stellen fest ...*

## 3.3 Kernfeststellung & Abgrenzung



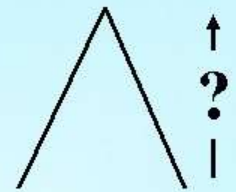
Entscheidend: Der Urzustand  $\hat{N}$  lässt sich nur schätzen bei Einbeziehung aller drei Ausgangswerte !!!



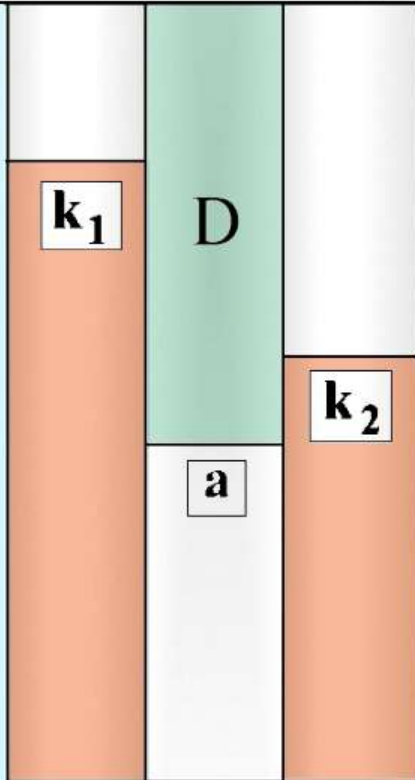
$$\frac{k_1 \cdot k_2}{a} = \hat{N}$$

*Was aber wird uns weltweit angeboten?*

## 3.3 Kernfeststellung & Abgrenzung



Listenumfang

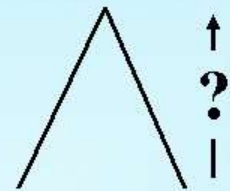


Alle anderen weltweit kursierenden Versuche nutzen maximal 2 Werte:

- nur beide k ("character-based")
- oder gar nur 'D' ("Distance-based")
- Milchmädchen reicht schon 'a'

*Dazu einige Beispiele der letzten Zeit ...*

# 3.4 Beispiele anderer Ansätze\*



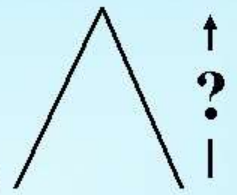
Year	Author(s)	Aim(s)	Data		Method	Assessment
			Chars.	Langs.		
2000	Lohr M PhD (Prof. of Engl.)	Glottochronology	128	18 Kel i.a. (incomplete)	Old Distance M's	- too few langs - not comparable
2003	Forster & Todt (Mathematician)	- proof method - Glottochronology	35(!)	6 Kel i.a. 'homespun'	'Network'	- mere surface structure (Rev. by Eska/Ringe)
2002 passim	McMahon A&R (Prof. of Engl., BioInf)	Interdisciplinary cooperation	<div style="border: 2px solid black; padding: 5px;">                     "I. Dyen lists"                       200 ~ Swadesh                      type                       84 recent (!)                 </div>		Old Distance M's	- unrooted - false relations
2003	Rexová et al. (Biologists)	"Optimality Criterion"			Max. Parsimony	- diff. odd results - self contradicting
2003	Gray & Atkinson (Biologists)	- Glottochronology - Anatolia Urheimat			MrBayes (Posterior propability)	- odd time depths - wrong relations
2006	Ellison/Kirby (Mathematicians)	lexical forms			Gram numbers (!)	- unqualified - wrong relations
2002 passim	Ringe D (IE-ist) varying mathematicians	subgrouping of 1st order IE	>300	24 oldest IE (only publ. 2007!)	"Warnow "Maximum Compatibility"	- poor results (examples later)

\*ausführlich besprochen in: Journal of Quantitative Linguistics 14-2/2007: 167-214

*Zwischenbilanz ...*

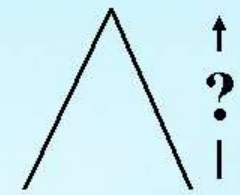
## 3.5 Zwischenbilanz

- keine dieser Methoden stellt korrekt den Urzustand her!
  - Bestenfalls rezente Oberflächen-Struktur.

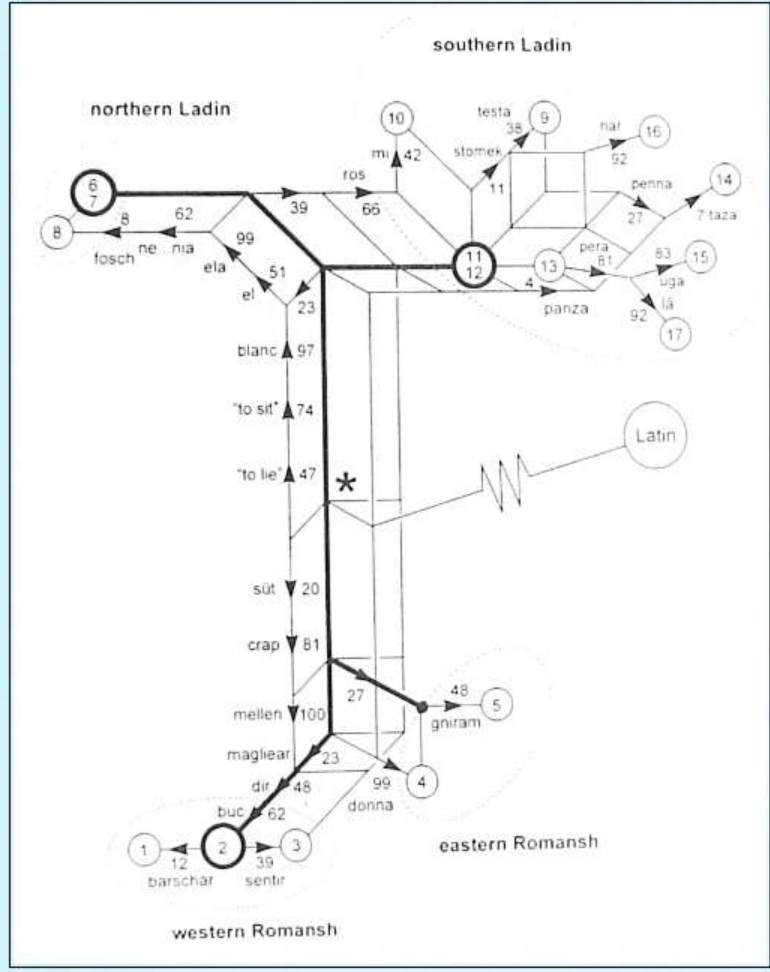


*Beispiel ...*

# 3.5 zum Beispiel



- Bestenfalls rezente Oberflächen-Struktur. Beispiel:

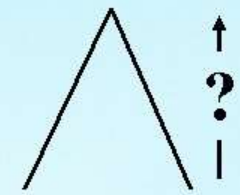


## Rumantsch . aus

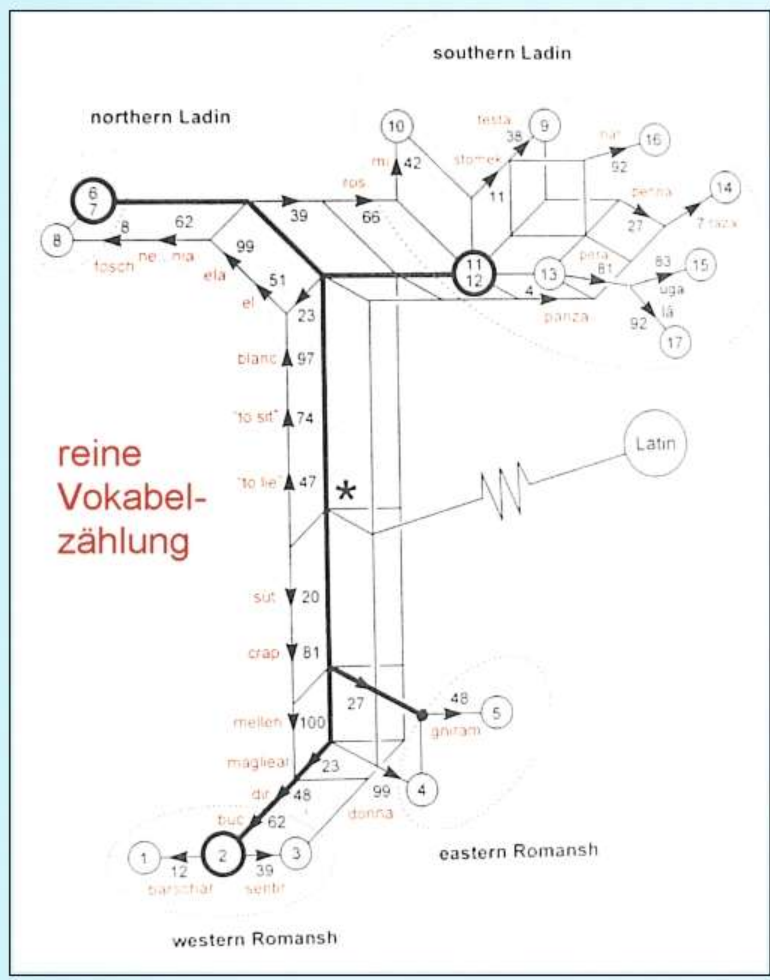
(Forster P, Toth A, Bandelt H-J:  
Evolutionary Network Analysis of Word  
Lists: Visualising the Relationships between  
Alpine Romance Languages. In JQL 5-3/  
1998:174-87).

*im Einzelnen ...*

# 3.5 zum Beispiel



rezente Oberflächen-Struktur.



## Rumantsch . aus

(Forster P, Toth A, Bandelt H-J:  
Evolutionary Network Analysis of Word Lists: Visualising the Relationships between Alpine Romance Languages. In JQL 5-3/ 1998:174-87).

Unsinn

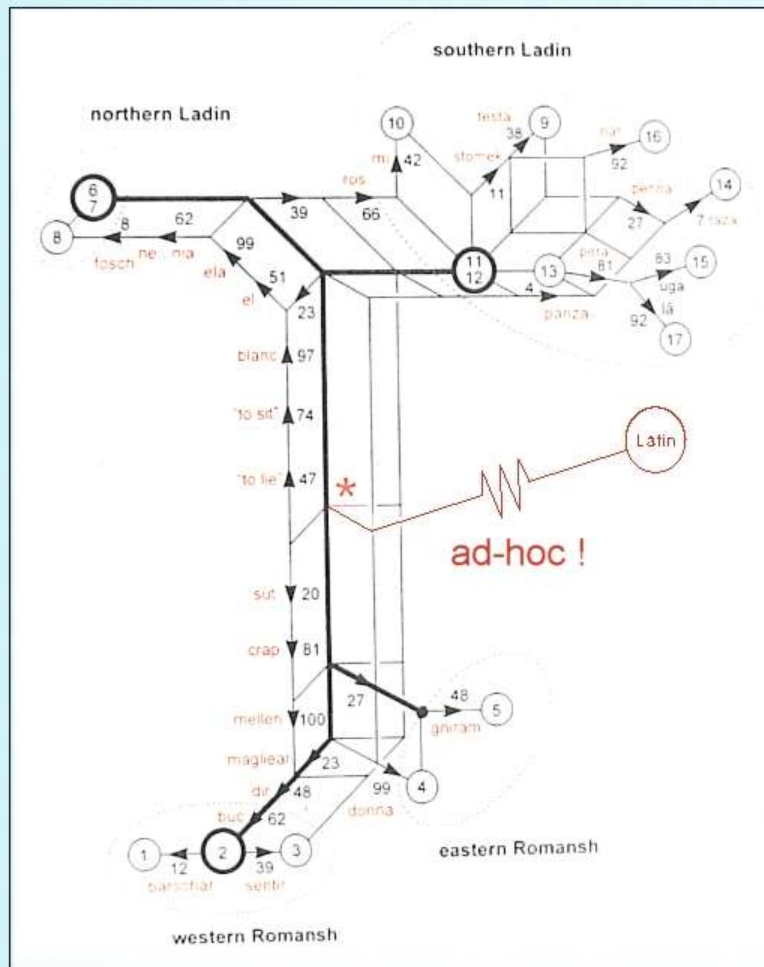
*und ...*



# 3.5 zum Beispiel



rezente Oberflächen-Struktur.



## Rumantsch . aus

(Forster P, Toth A, Bandelt H-J:  
Evolutionary Network Analysis of Word Lists: Visualising the Relationships between Alpine Romance Languages. In JQL 5-3/ 1998:174-87).

Unsinn

- Latein als punktuelle "Wurzel"

*Da bleibt nur ...*

## 3.5 Aphorismen

Schon Raimo Anttila bemerkte spöttisch\*:

"Polarization into opposite camps is quite common in the sciences.

Two general principles are mainly responsible for this,

(1) the drunkard's search:

**It is easier to look for a lost key under a street lamp because "it is lighter there" than where it was actually lost.**

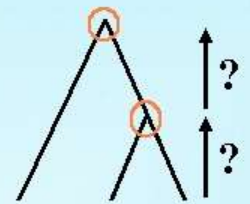
(2) The law of the instrument:

**Give a small boy a hammer, and he will find that everything he encounters needs punding."**

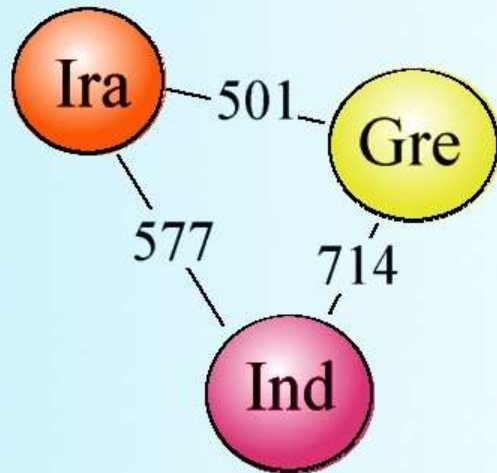
*Nun eine reale Aufgabe ...*

\* Abraham Kaplan, The conduct of inquiry. Methodology for behavior science. Chandler: 1964. (Cited from Anttila 1989:324)

## 4.1 Erweiterung auf 3 Töchter



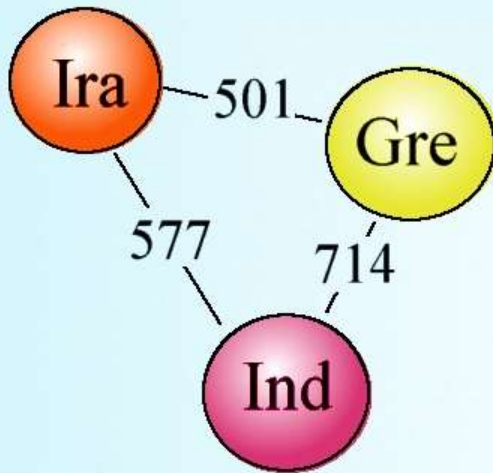
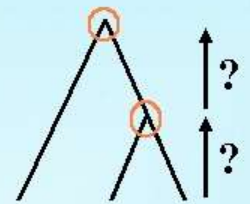
Nun zu realen Daten:



Pokorny liefert für die links dargestellten Sprachen die angegebenen Gemeinsamkeiten.

*Fallen Sie noch darauf herein ?*

## 4.1 Erweiterung auf 3 Töchter



Nach naivem Verständnis  
wäre damit Ind näher mit  
Gre als mit Ira verwandt...

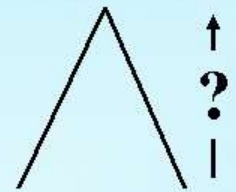
Davon

würden bei diesen Daten alle anderen quantitativen  
Methoden ausgehen;

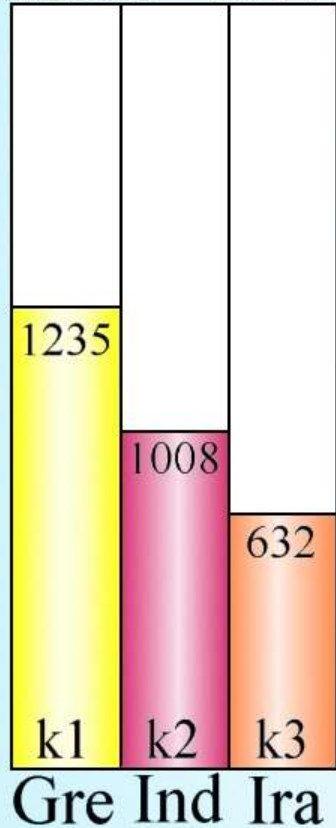
*Wir nicht ...*

*Dafür werden wir die  
Ausgangsmenge berechnen ...*

## 4.2 Ausgangsmengen-Berechnung



$N_{\text{Pok}} = 2044$  = Gesamtbestand Rekonstruktionen Pokorny

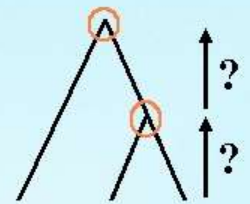


Hier die weiteren bekannten Daten

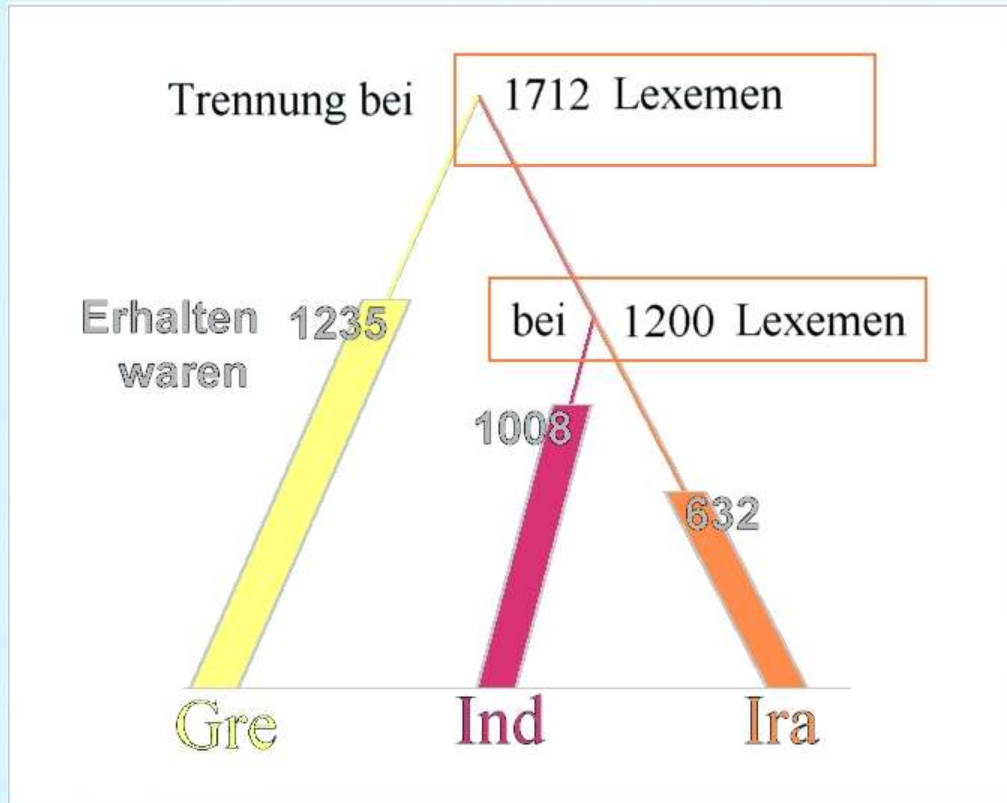
= Erhaltene Kognaten in den drei Sprachen

***Gesucht: die unbekanntes Trennungsmengen ...***

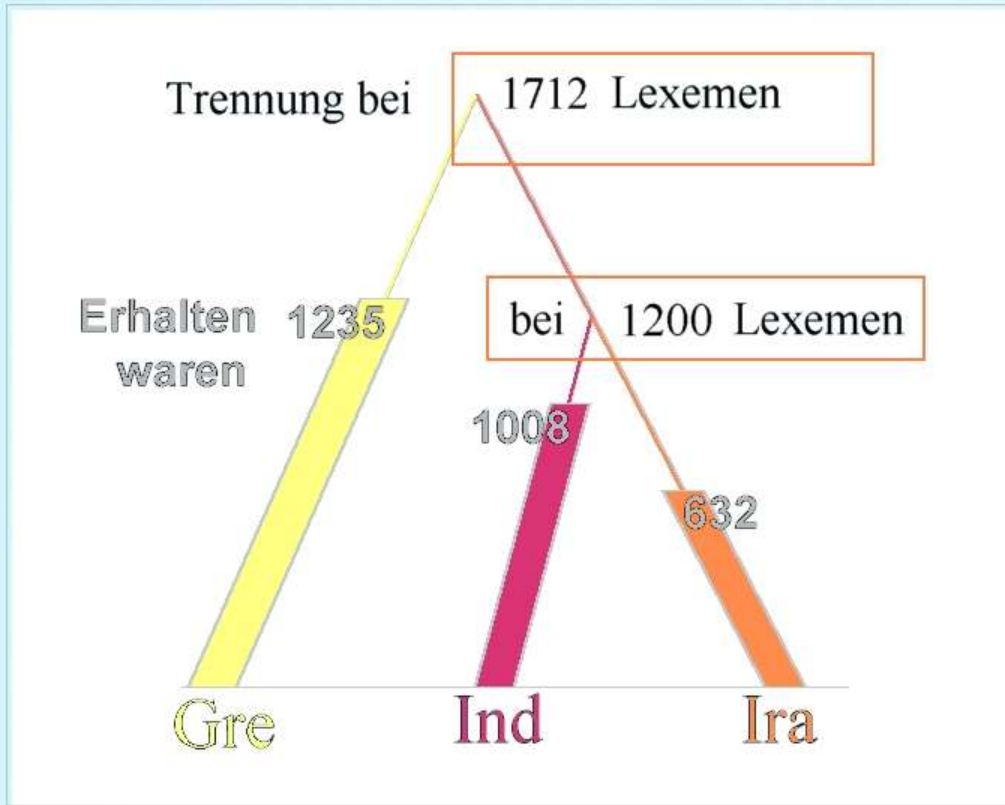
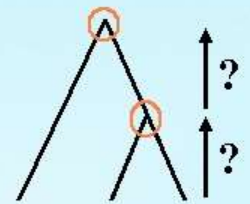
## 4.3 Ausgangsmengen-Berechnung



Wir rechnen kurz einmal nach und erhalten:



## 4.3 Ausgangsmengen-Berechnung

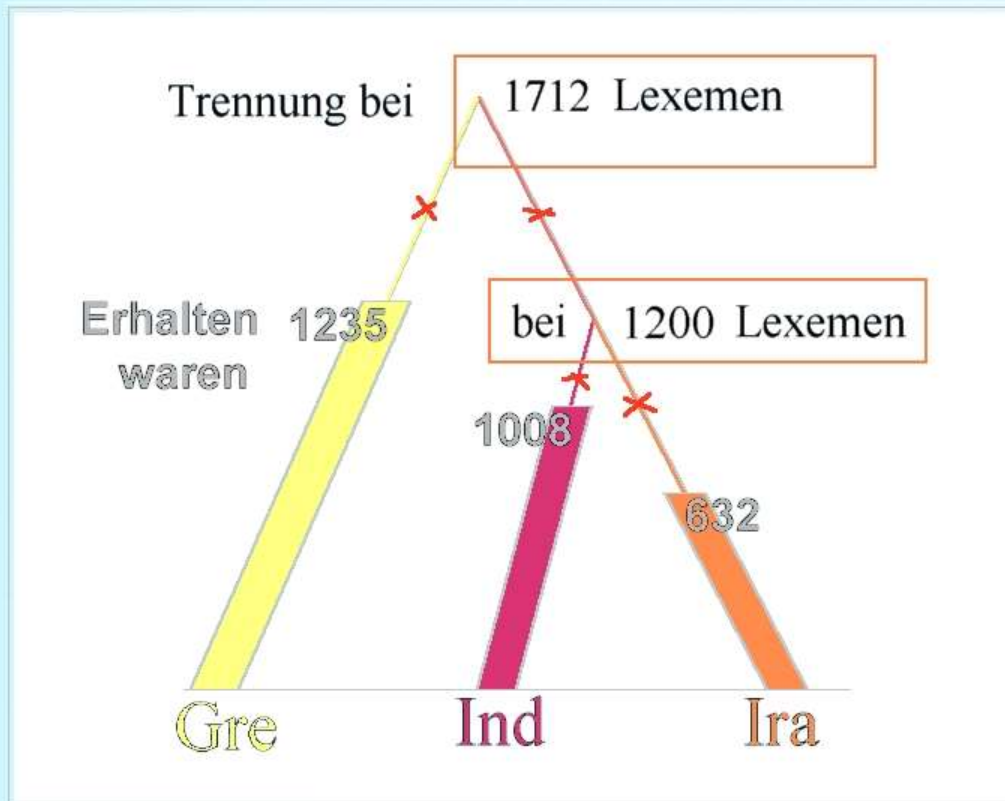
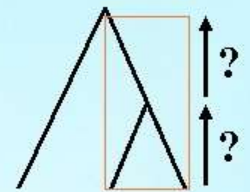


... was der Realität  
doch erheblich näher  
kommt.

*Aber wo bleiben die Neuerungen ?*

*Dies wollen wir an einem konkreten Modell simulieren:*

# 5.1 Modellentwurf



## Modell:

Wir simulieren zufällige  
✗ Neuerungen. Vereinfachend  
setzen wir statt  
1712  $\rightarrow$  100 Erste Trennung

es sollen verbleiben:

Ari. 1200  $\rightarrow$  70 Merkmale

Gri. 1235  $\rightarrow$  72 "

*Für die Ergebnisse  
brauchen wir folgende Tabelle ...*

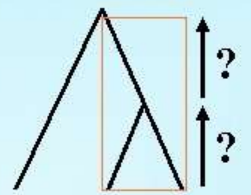


# 5.2. Unabhängig neuern:

Gre auf 72%,

Zufallsergebnisse bezogen auf 1715 = 100 %

Merkmale	In Sprache		Übereinstimmungen				Merkmale	In Sprache		Übereinstimmungen			
	Gri	Arisch	1n2	1n3	2n3	alle		Gri	Arisch	1n2	1n3	2n3	alle
1							51						
2							52						
3							53						
4							54	●					
5							55						
6							56	●					
7							57						
8							58						
9	●						59						
10	●						60						
11							61						
12	●						62						
13							63						
14							64						
15	●						65						
16							66						
17	●						67						
18							68						
19	●						69						
20							70	●					
21	●						71	●					
22	●						72						
23							73						
24	●						74						
25							75						
26							76						
27							77						
28							78						
29							79						
30							80						
31							81						
32							82	●					
33	●						83						
34							84	●					
35							85						
36							86						
37							87	●					
38							88						
39							89						
40	●						90	●					
41							91						
42							92						
43							93						
44							94						
45							95	●					
46	●						96	●					
47							97						
48	●						98	●					
49	●						99	●					
50	●						100	●					
Σ 'k' originaler Merkmale (weiß)							72						
scheinbare individuelle Neuerungen							28						
Rest gemeinsamer Neuerungen													



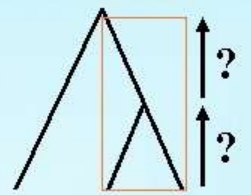
und ...

## 5.2. Unabhängig neuern:

Arisch noch gemeinsam auf 70%

Zufallsergebnisse bezogen auf 1715 = 100 %

Merkmale	In Sprache		Übereinstimmungen				Merkmale	In Sprache		Übereinstimmungen				
	Gri	Arisch	1n2	1n3	2n3	alle		Gri	Arisch	1n2	1n3	2n3	alle	
1							51							
2							52							
3		■					53							
4		■					54	●	■					
5		■					55		■					
6		■					56	●	■					
7							57							
8							58							
9	●						59							
10	●						60		■					
11		■					61							
12	●	■					62							
13		■					63							
14		■					64							
15	●	■					65							
16							66							
17	●						67		■					
18							68							
19	●	■					69							
20		■					70	●	■					
21	●						71	●						
22	●						72		■					
23							73		■					
24	●						74		■					
25							75							
26		■					76							
27							77							
28							78							
29		■					79							
30							80							
31							81		■					
32							82	●	■					
33	●						83							
34		■					84	●	■					
35		■					85		■					
36		■					86							
37							87	●						
38							88							
39							89		■					
40	●	■					90	●	■					
41							91		■					
42		■					92							
43							93							
44							94							
45							95	●						
46	●	■					96	●						
47		■					97							
48	●	■					98	●						
49	●						99	●	■					
50	●						100	●						
Σ 'k' originaler Merkmale (weiß)							:	70						
scheinbare individuelle Neuerungen							:	30						
Rest gemeinsamer Neuerungen							:							



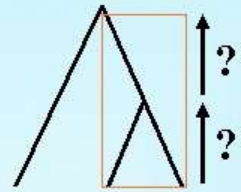
■ = gemeinsame Neuerungen

Im nächsten Schritt trennt sich dann das Arische ...

# 5.2 Es trennen sich und neuern selbständig ...

- Ira auf 40 %,
- Ind auf 59 %.

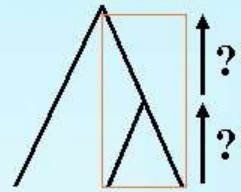
Merkmale	In Sprache			Übereinstimmungen				Merkmale	In Sprache			Übereinstimmungen				
	Gri	Ira	Ind	1n2	1n3	2n3	alle		Gri	Ira	Ind	1n2	1n3	2n3	alle	
1								51								
2								52								
3								53								
4								54								
5								55								
6								56								
7								57								
8								58								
9								59								
10								60								
11								61								
12								62								
13								63								
14								64								
15								65								
16								66								
17								67								
18								68								
19								69								
20								70								
21								71								
22								72								
23								73								
24								74								
25								75								
26								76								
27								77								
28								78								
29								79								
30								80								
31								81								
32								82								
33								83								
34								84								
35								85								
36								86								
37								87								
38								88								
39								89								
40								90								
41								91								
42								92								
43								93								
44								94								
45								95								
46								96								
47								97								
48								98								
49								99								
50								100								
Σ 'k' originaler Merkmale (weiß)									70							
scheinbare individuelle Neuerungen									30							
Rest gemeinsamer Neuerungen																



Was fällt Ihnen auf?

# 5.2 Es trennen sich und neuern selbständig ...

Merkmale	In Sprache			Übereinstimmungen				Merkmale	In Sprache			Übereinstimmungen			
	Gri	Ira	Ind	1n2	1n3	2n3	alle		Gri	Ira	Ind	1n2	1n3	2n3	alle
1								51							
2								52							
3								53							
4								54							
5								55							
6								56							
7								57							
8								58							
9								59							
10								60							
11								61							
12								62							
13								63							
14								64							
15								65							
16								66							
17								67							
18								68							
19								69							
20								70							
21								71							
22								72							
23								73							
24								74							
25								75							
26								76							
27								77							
28								78							
29								79							
30								80							
31								81							
32								82							
33								83							
34								84							
35								85							
36								86							
37								87							
38								88							
39								89							
40								90							
41								91							
42								92							
43								93							
44								94							
45								95							
46								96							
47								97							
48								98							
49								99							
50								100							
Σ 'k' originaler Merkmale (weiß)									70						
scheinbare individuelle Neuerungen									30						
Rest gemeinsamer Neuerungen															



Beobachtung:  
 Die Ersetzungen nach der Trennung überdecken einen Teil der alten gemeinsamen arischen Neuerungen!

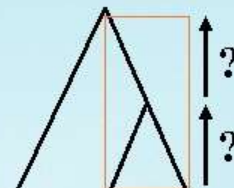
*genauer ...*

# 5.2 Auszählung

Ergebnis:

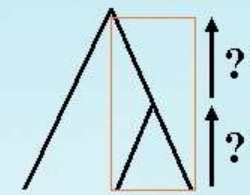
Es wurden  
20 der zunächst  
30 arischen  
Neuerungen  
**unkennlich!**

Merkmale	In Sprache			Übereinstimmungen				Merkmale	In Sprache			Übereinstimmungen				
	Gri	Ira	Ind	1n2	1n3	2n3	alle		Gri	Ira	Ind	1n2	1n3	2n3	alle	
1								01								
2								02								
3		*	●					03	●	●	●					
4								04	●	●	●	*				
5			●					05	●	●	●					
6								06	●	●	●	*				
7			●					07	●	●	●					
8								08								
9	●							09								
10	●	●	●					10		*	●					
11								11								
12	●	●	*					12								
13		*	●					13	●	●	●					
14								14								
15	●	*	●					15	●	●	●					
16		●						16	●	●	●					
17	●	●	●					17	●	●	●	*				
18		●						18								
19	●	●	●					19	●	●	●					
20		●						20	●	●	●					
21	●	●	●					21	●	●	●					
22	●	●						22	●	●	●					
23		●						23	●	●	●					
24	●	●	●					24	●	●	●	*				
25		●						25	●	●	●					
26		*	●					26								
27		●						27								
28								28								
29		●	*					29	●	●	●					
30		●						30								
31								31								
32								32	●	●	●					
33	●	●	●					33	●	●	●					
34		●	*					34	●	*	●					
35		●	*					35	●	●	●					
36		●	*					36	●	●	●					
37		●						37	●	●	●					
38		●	●					38	●	●	●					
39		●						39	●	●	●	*	●			
40	●	●	●					40	●	●	●					
41		●	●					41	●	●	●	*				
42		*	●					42								
43			●					43	●	●	●					
44								44								
45								45	●	●	●					
46	●	●	●					46	●	●	●					
47		●	*					47								
48	●	●	●					48	●	●	●					
49	●	●	●					49	●	●	●					
50	●	●	●					50	●	●	●					
Σ 'k' originaler Merkmale (weiß)								:	72	40	59	4	17	6	28	
scheinbare individuelle Neuerungen								:	28	50	31					
Reste gemeinsamer Neuerungen								:	10 von 30							



# 5.2 Auszählung

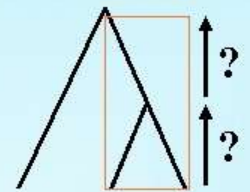
Merkmale	In Sprache			Merkmale	In Sprache		
	Gri	Ira	Ind		Gri	Ira	Ind
1				51			
2				52		●	●
3		*	●	53		●	●
4				54	●	●	*
5			●	55		●	●
6				56	●	●	*
7			●	57		●	●
8				58			
9	●			59			
10	●	●		60		*	●
11				61			
12	●	●	*				
13		*	●				
14							
15	●	*	●				
16		●					
17	●						
18		●					
19	●	●					
20		●					
21	●	●					
22	●	●					
23		●					
24	●	●					
25		●					
26		*	●				
27		●					
28							
29		●	*				
30		●					
31							
32							
33	●		●				
34		●	*				
35		●	*				
36		●	*				
37		●					
38		●	●				
39		●					
40	●	●					
41			●				
42		*	●				
43			●				
44							
45							
46	●	●					
47		●	*				
48	●	●					
49	●	●					
50	●	●					
Σ 'k' originaler Merkmale (weiß)	72	40	59	4	17	6	28
scheinbare individuelle Neuerungen	28	50	31				
Reste gemeinsamer Neuerungen		10	von 30				



Dadurch entstanden

**\*** falsche, scheinbar individuelle Neuerungen!

# 5.3 Ergebnisse des Experiments

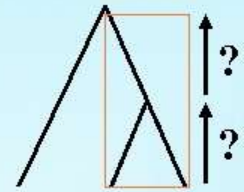


Merk- male	In Sprache			Übereinstimmungen				Merk- male	In Sprache			Übereinstimmungen				
	Gri	Ira	Ind	1n2	1n3	2n3	alle		Gri	Ira	Ind	1n2	1n3	2n3	alle	
Σ 'k' originaler Merkmale (weiß)									72	40	59	4	17	6	28	
scheinbare individuelle Neuerungen									28	50	31					
Reste gemeinsamer Neuerungen										10 von 30						

## 1. Erkenntnis:

Gemeinsame Neuerungen werden mit hoher Wahrscheinlichkeit durch individuelle Neuerungen z.T. zerstört. Dies kann bei einer geringen Anzahl, z. B. morphologischer Merkmale, zu einem Totalverlust führen.

# 5.3 Ergebnisse des Experiments



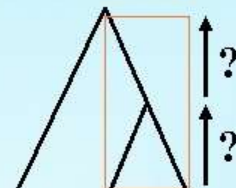
Merk- male	In Sprache			Übereinstimmungen				Merk- male	In Sprache			Übereinstimmungen			
	Gri	Ira	Ind	1n2	1n3	2n3	alle		Gri	Ira	Ind	1n2	1n3	2n3	alle
$\Sigma$ 'k' originaler Merkmale (weiß)									72	40	59	4	17	6	28
scheinbare individuelle Neuerungen									28	50	31				
Reste gemeinsamer Neuerungen											10 von 30				

## 2. Erkenntnis:

*Wir gewinnen praktische Erfahrung über die Genauigkeit der Simulation ...*



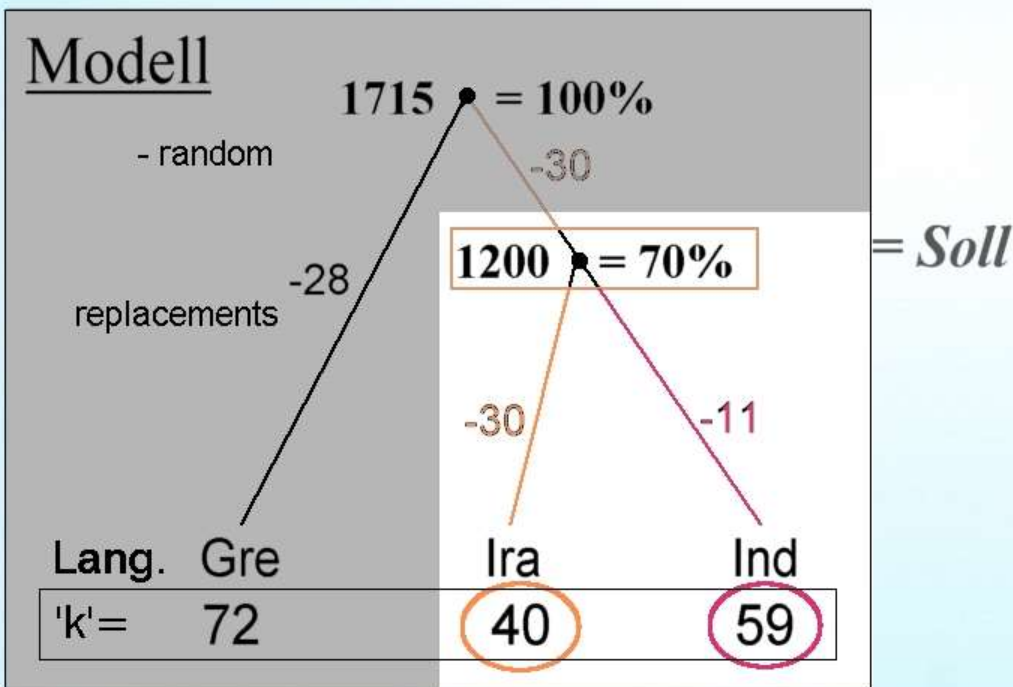
# 5.3 Ergebnisse des Experiments



Merk-male	In Sprache			Übereinstimmungen				Merk-male	In Sprache			Übereinstimmungen			
	Gre	Ira	Ind	1n2	1n3	2n3	alle		Gre	Ira	Ind	1n2	1n3	2n3	alle
Σ 'k' originaler Merkmale (weiß)									72	40	59	4	17	6	28
scheinbare individuelle Neuerungen									28	60	41				
Reste gemeinsamer Neuerungen									10 von 30						

$$\frac{k_2 \cdot k_3}{a} = \hat{N} \rightarrow \frac{40 \cdot 59}{34} = 69,4$$

Wir setzen die erhaltenen Zufallswerte ein

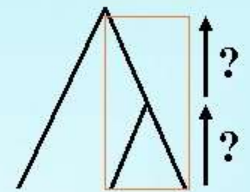


'a'	Gre	Ira	Ind
Ind	17	6	
Ira	4		69,4 = Ist
Gre			
	computed $\hat{N}$		

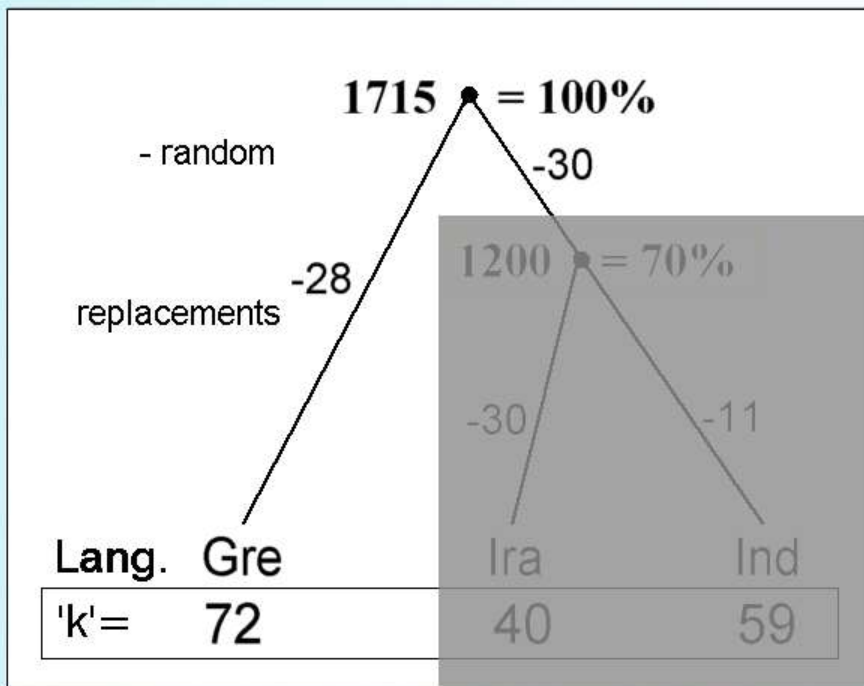
und erhalten für die Aufspaltung Ira : Ind ein sehr gutes Ergebnis (69,4 ~ 70).

beim Griechischen jedoch ...

# 5.3 Ergebnisse des Experiments



ergibt die Rechnung ...



= Soll

'a'	Gre	Ira	Ind
Ind	17	6	
Ira	4		69,4
Gre		90	94,4
	computed N		

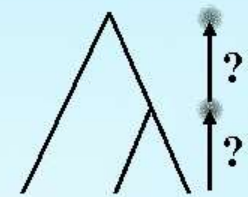
= Ist

eine erhebliche Streuung

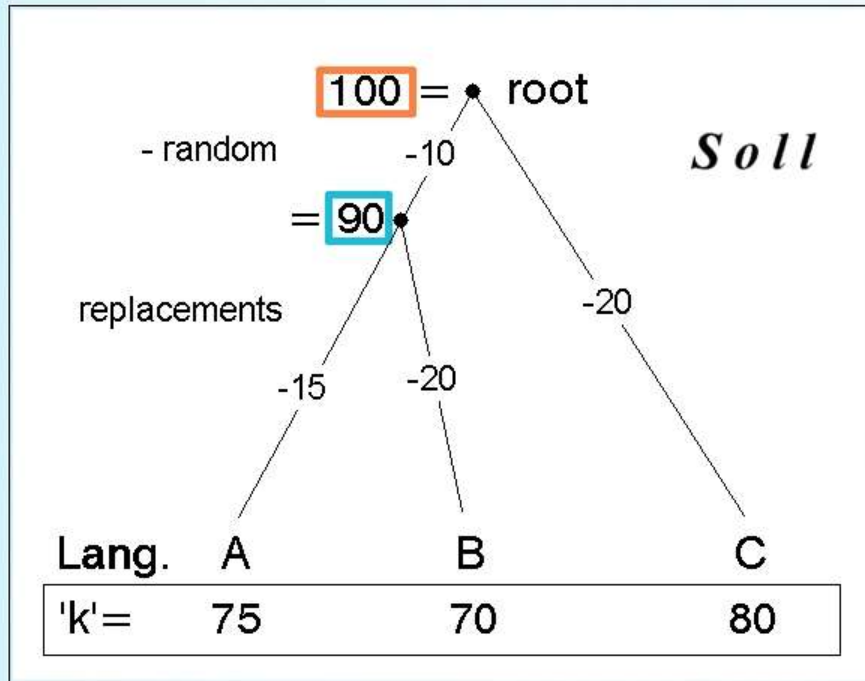
*Zur Klärung*

*kurz ein paar weitere Experimente ...*

# 6.1 Zwei Aufspaltungen - 3 Töchter



## Modell



## Ergebnis der diachronen Simulation

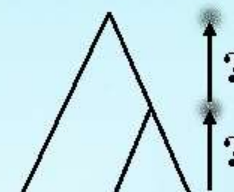
'a'	A	B	C
C	61	58	
B	58		96.6
A		90.5	98.4
	computed $\hat{N}$		

*Ist*

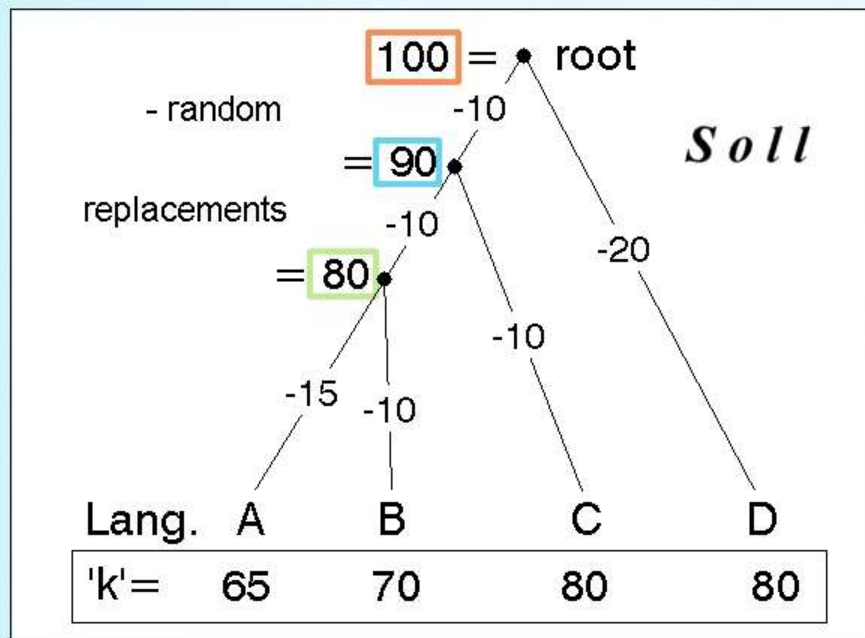
Quelle:

Holm, Hans J. (2007a): Requirements and Limits of the Separation Level Recovery Method in Language Subgrouping, An Update. In: Köhler, R., & P. Grzybek (Eds) Exact Methods in the Study of Language and Text. Dedicated to Professor Gabriel Altmann on the occasion of his 75th birthday. [Quantitative Linguistics 62]. Berlin: de-Gruyter: 223-234.

# 6.2 Drei Aufspaltungen - 4 Töchter



## Modell



## Ergebnis der diachronen Simulation

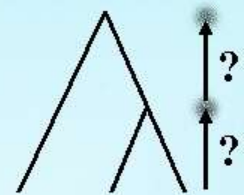
'a'	A	B	C	D
D	54	55	63	
C	58	64		101.6
B	57		87.5	101.8
A		79.8	89.7	96.3
	computed $\hat{N}$			

*Ist*

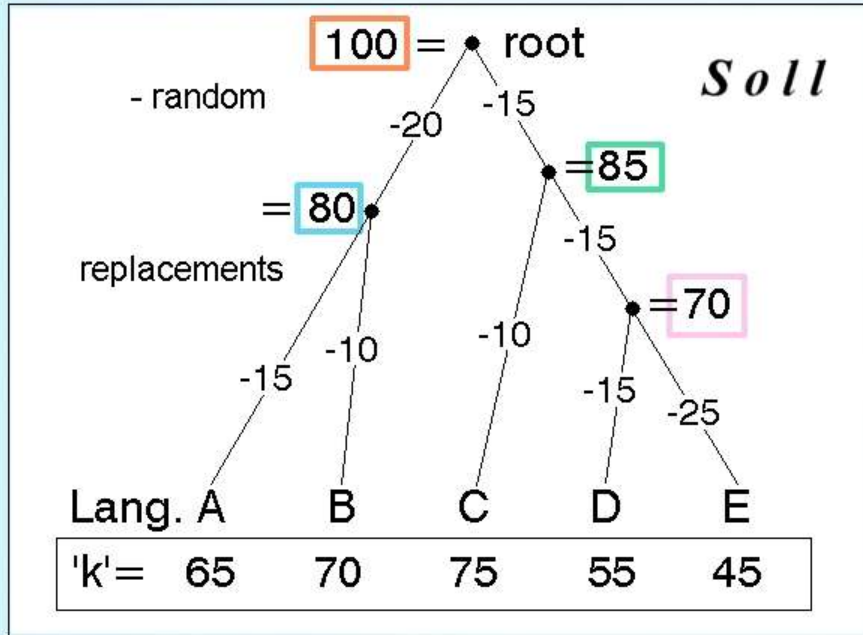
Quelle:

Holm, Hans J. (2007a): Requirements and Limits of the Separation Level Recovery Method in Language Subgrouping, An Update. In: Köhler, R., & P. Grzybek (Eds) Exact Methods in the Study of Language and Text. Dedicated to Professor Gabriel Altmann on the occasion of his 75th birthday. [Quantitative Linguistics 62]. Berlin: de-Gruyter: 223-234.

# 6.3 Vier Aufspaltungen > 5 Töchter



## Modell



## Ergebnis der diachronen Simulation

'a'	A	B	C	D	E
E	33	34	40	35	
D	36	42	48		70.7
C	47	52		85.9	84.4
B	57		101	91.7	92.6
A		79.8	103.7	99.3	88.6
	computed $\hat{N}$				

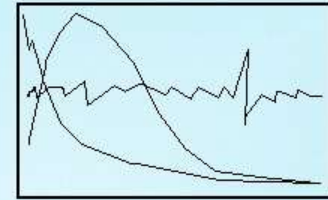
*Ist*

*Es gibt eine zweite Fehlerquelle ...*

Quelle:

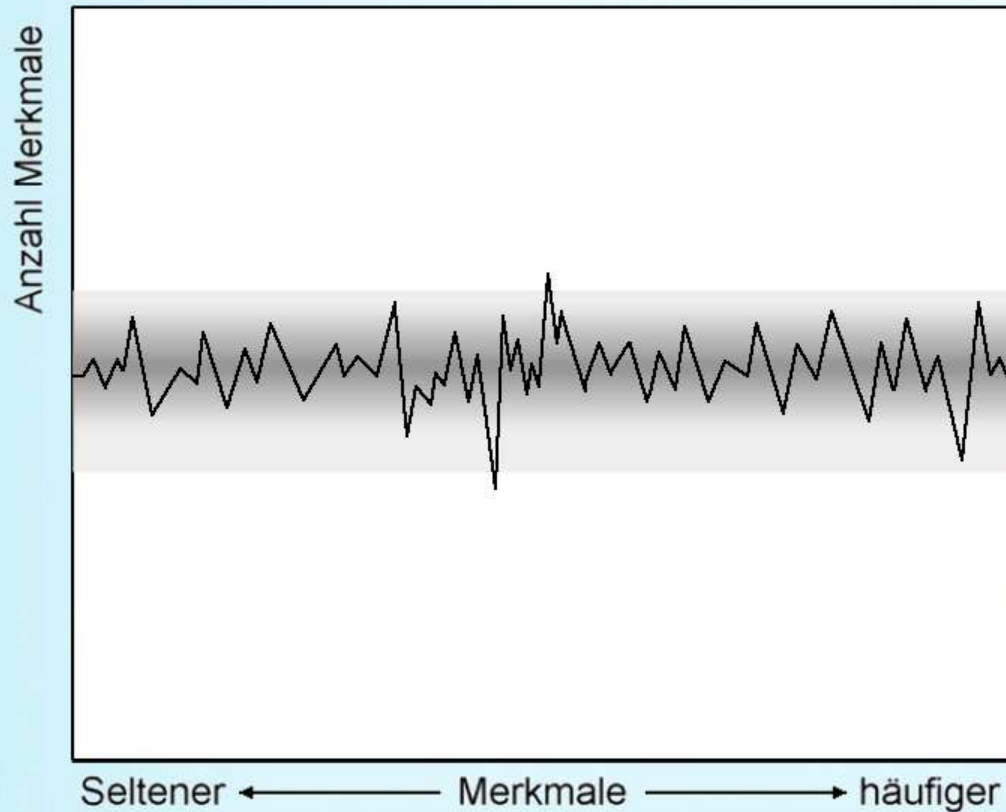
Holm, Hans J. (2007a): Requirements and Limits of the Separation Level Recovery Method in Language Subgrouping, An Update. In: Köhler, R., & P. Grzybek (Eds) Exact Methods in the Study of Language and Text. Dedicated to Professor Gabriel Altmann on the occasion of his 75th birthday. [Quantitative Linguistics 62]. Berlin: de-Gruyter: 223-234.

# 7.1 Bisher 'ideale' Verteilung



Bisher waren die Daten noch "ideal verteilt", d.h.

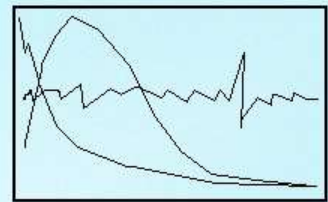
alle Merkmale



hatten die gleiche  
Ersetzungswahrscheinlichkeit.

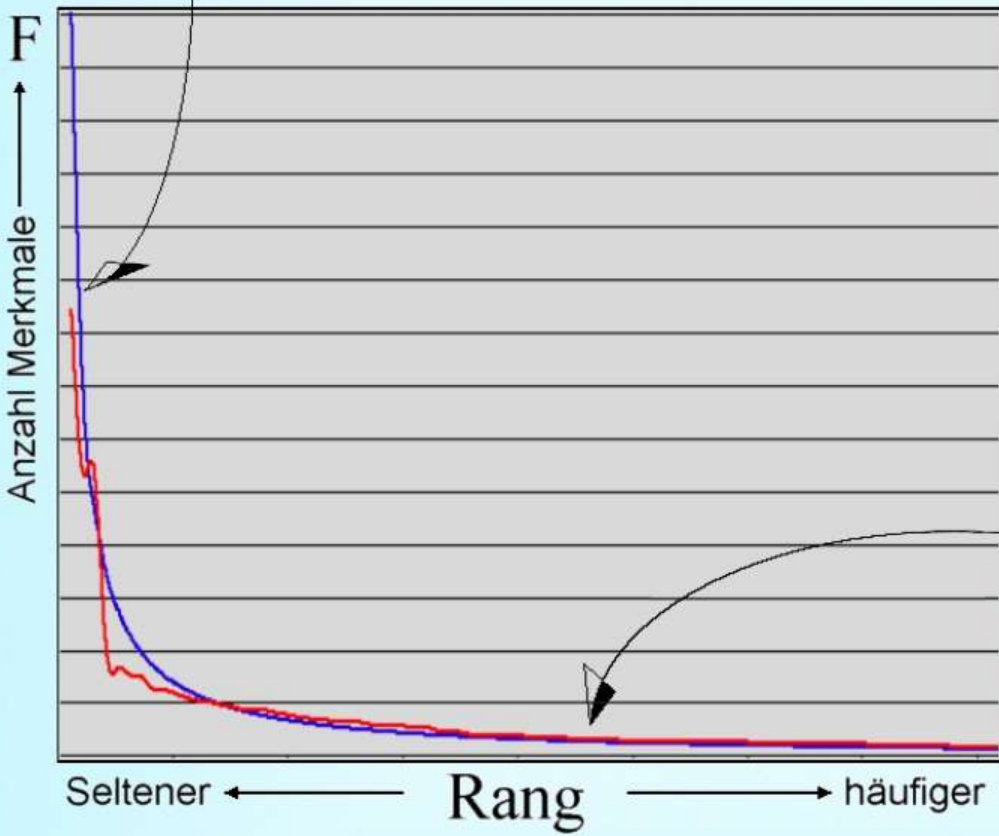
*Leider ist das in natürlichen Sprachen nicht der Fall ...*

# 7.2 Reale Daten ~ "Zipf-verteilt", d.h.



$$R \cdot F = \text{konst.}$$

Viele Wörter, die selten vorkommen,

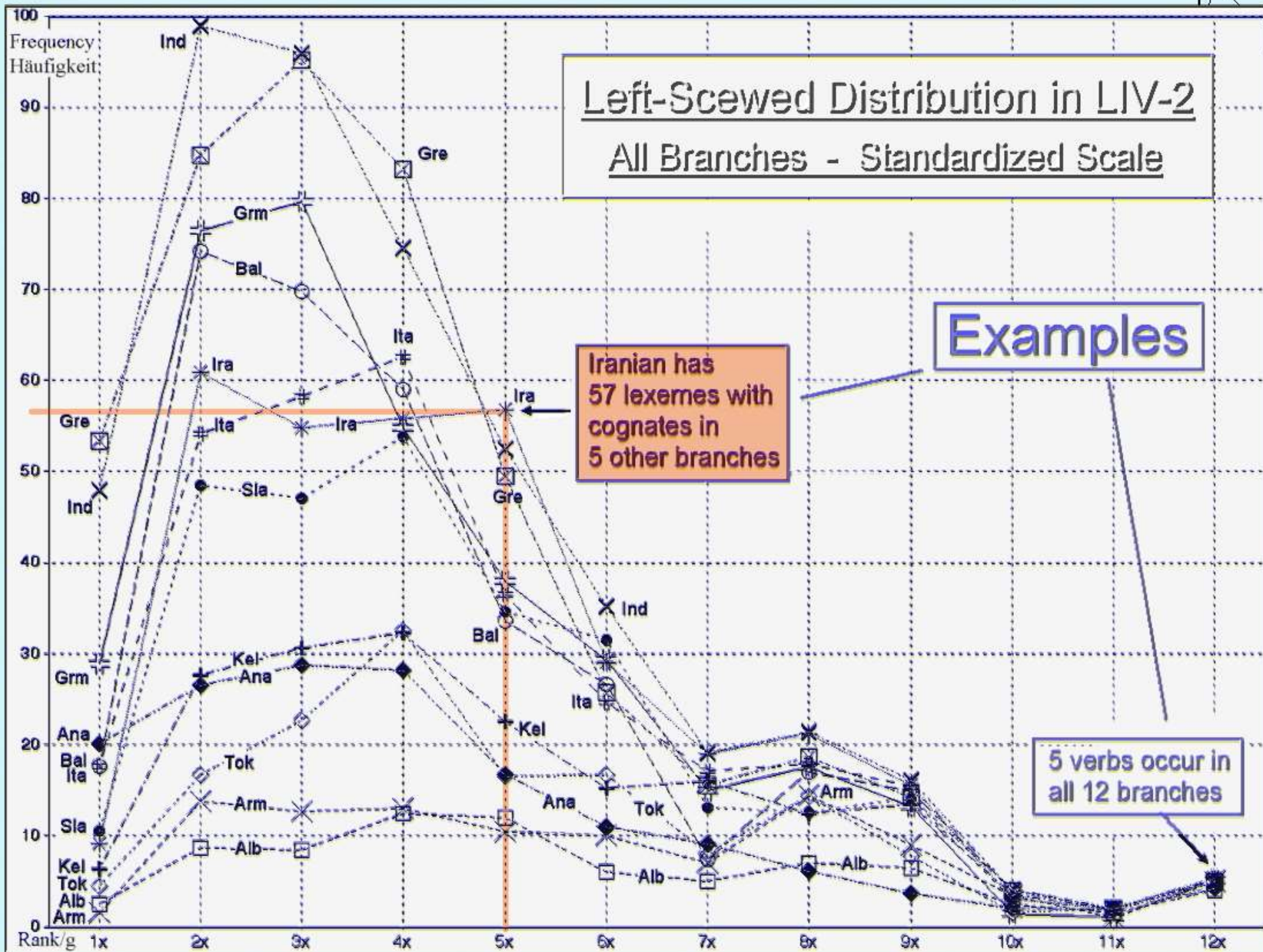


Wenige Wörter, die häufig  
gebraucht werden

*Also erstellen wir eine solche Rang- : Frequenz- Tabelle für das LIV ...*

\* Zipf, George K. (1965) The psychobiology of language...

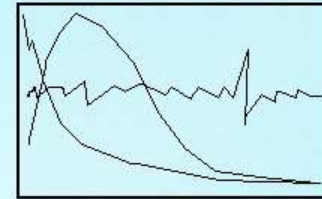
# 7.3 Verteilung im LIV-Korpus



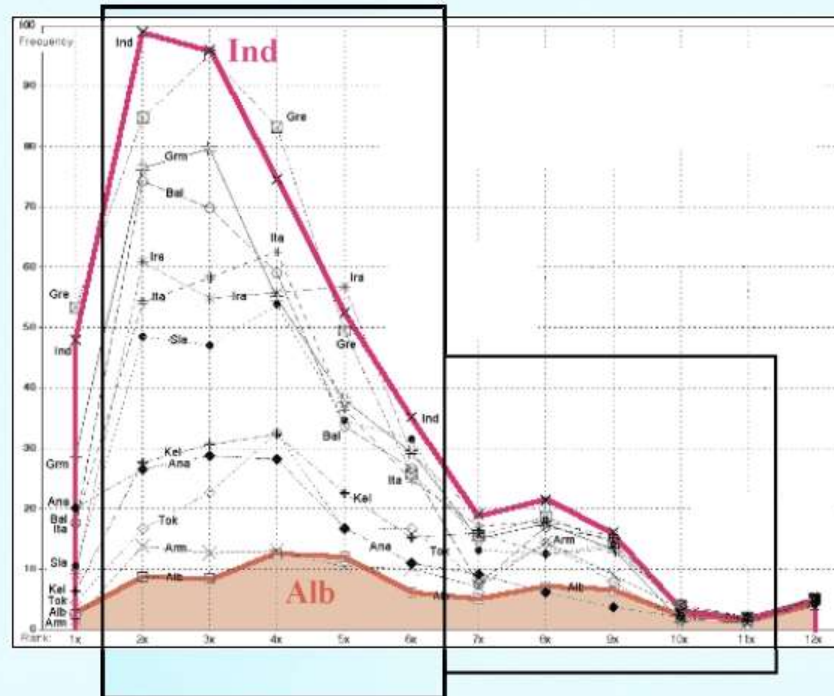
?



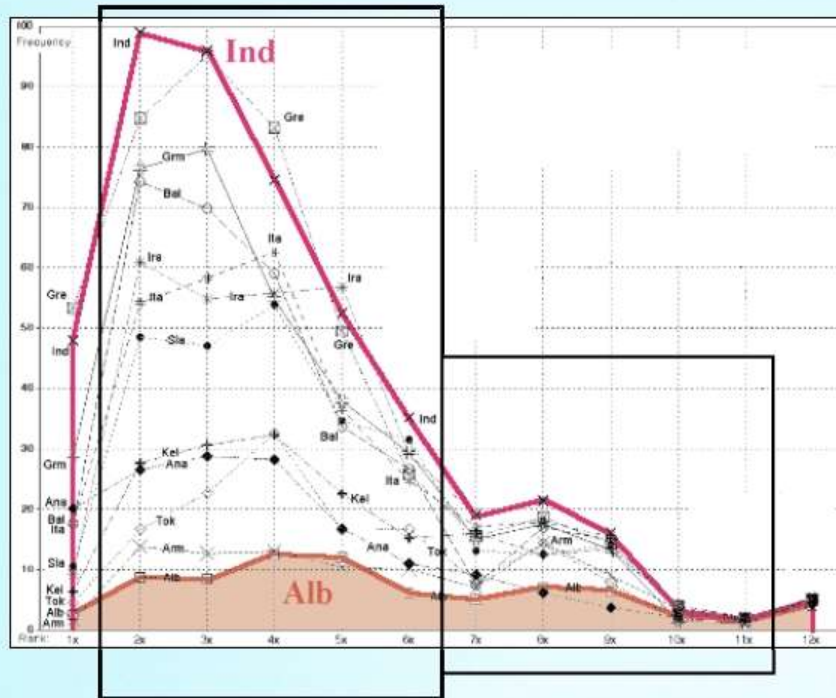
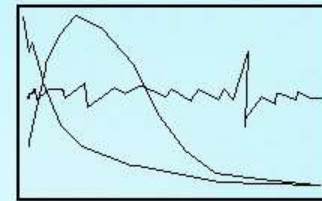
# 7.4 Verfälschende Auswirkung



Vgl.  
linker < > rechtem Kasten!



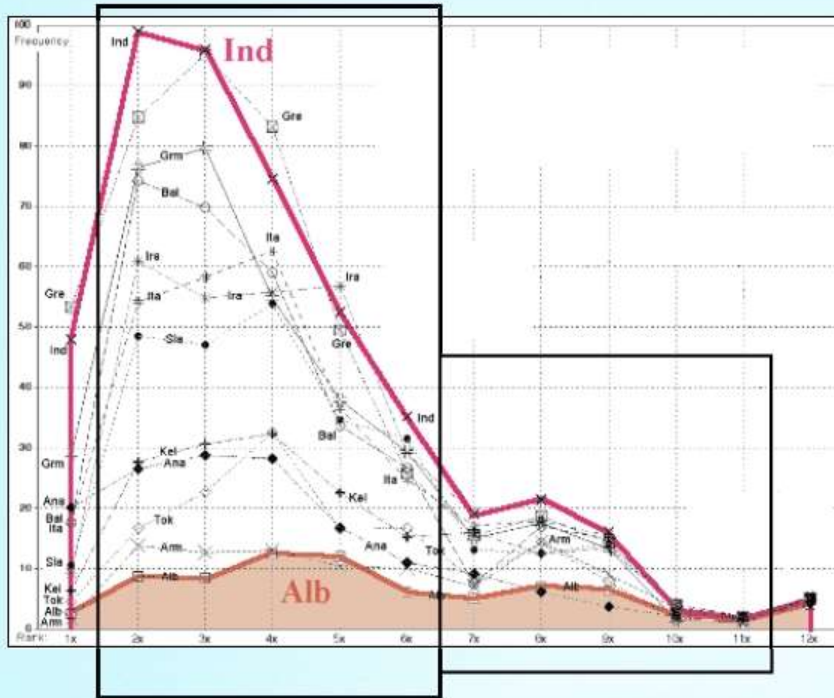
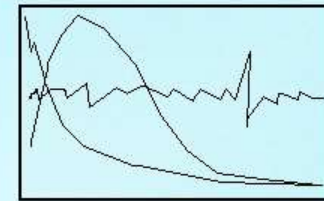
# 7.4 Verfälschende Auswirkung



Im Vergleich rechts zu links zeigt Albanisch gegenüber Indisch einen unverhältnismäßig höheren Anteil an "a (greements)".

$$\frac{k_1 \cdot k_2}{a} = \hat{N}$$

# 7.4 Verfälschende Auswirkung



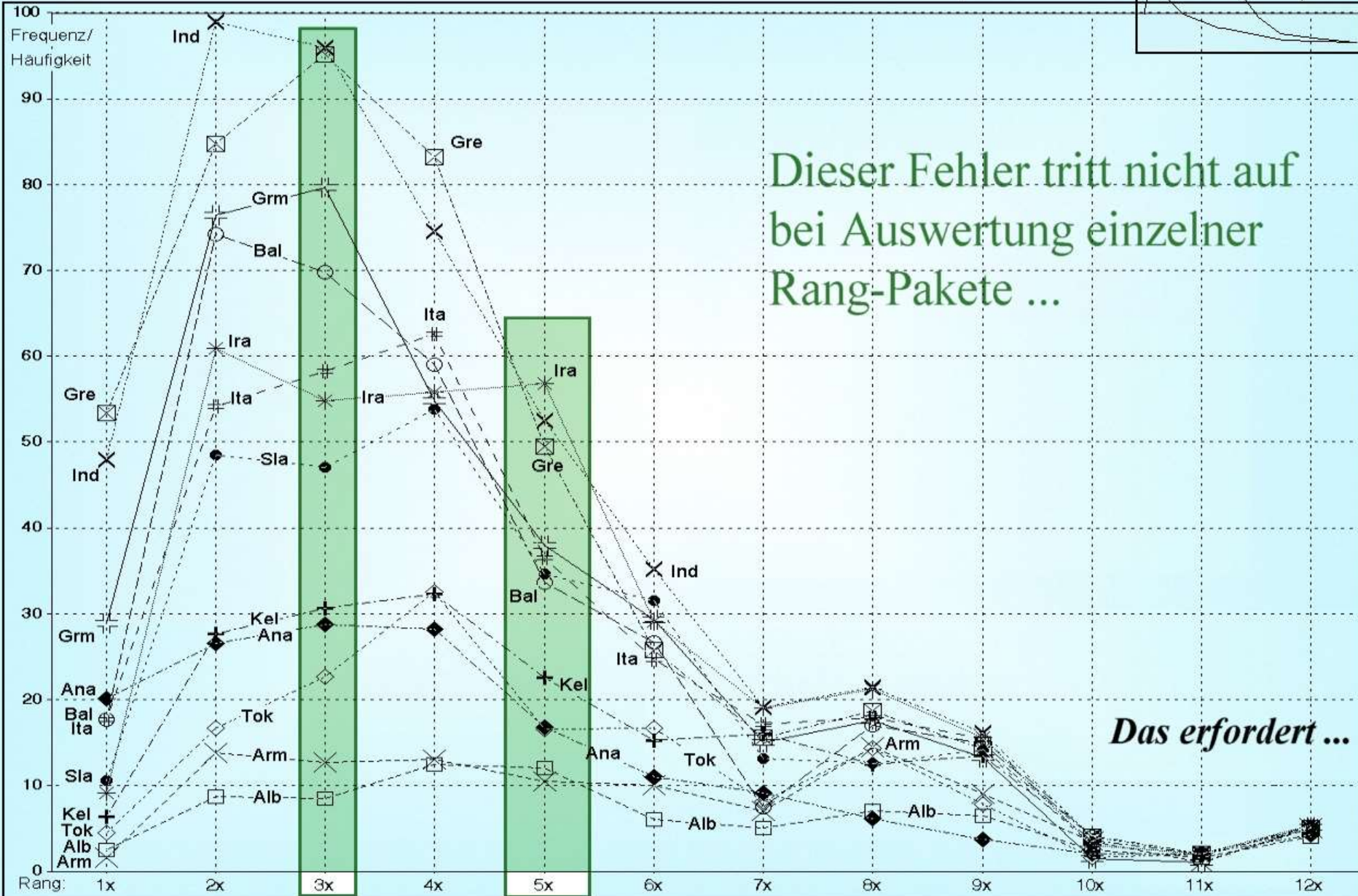
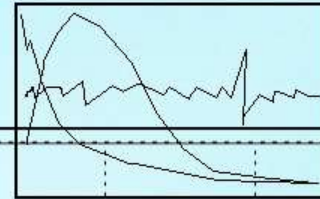
höheren Anteil an  
"a (greements)".

$$\frac{k_1 \cdot k_2}{a \uparrow} = \hat{N} \downarrow$$

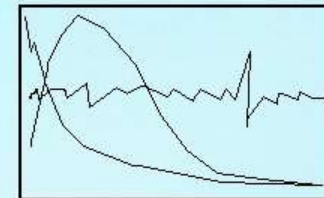
Dies führte bei Kleinkorpussprachen  
zu falsch-niedrigen Trennungsmengen,  
also falsch-späten Aufspaltungen.

*Was tun?*

# 7.4 Verfälschende Auswirkung



# 7.5 Neues SLR-D-Verfahren



(Separation Level Recovery regarding Distribution)

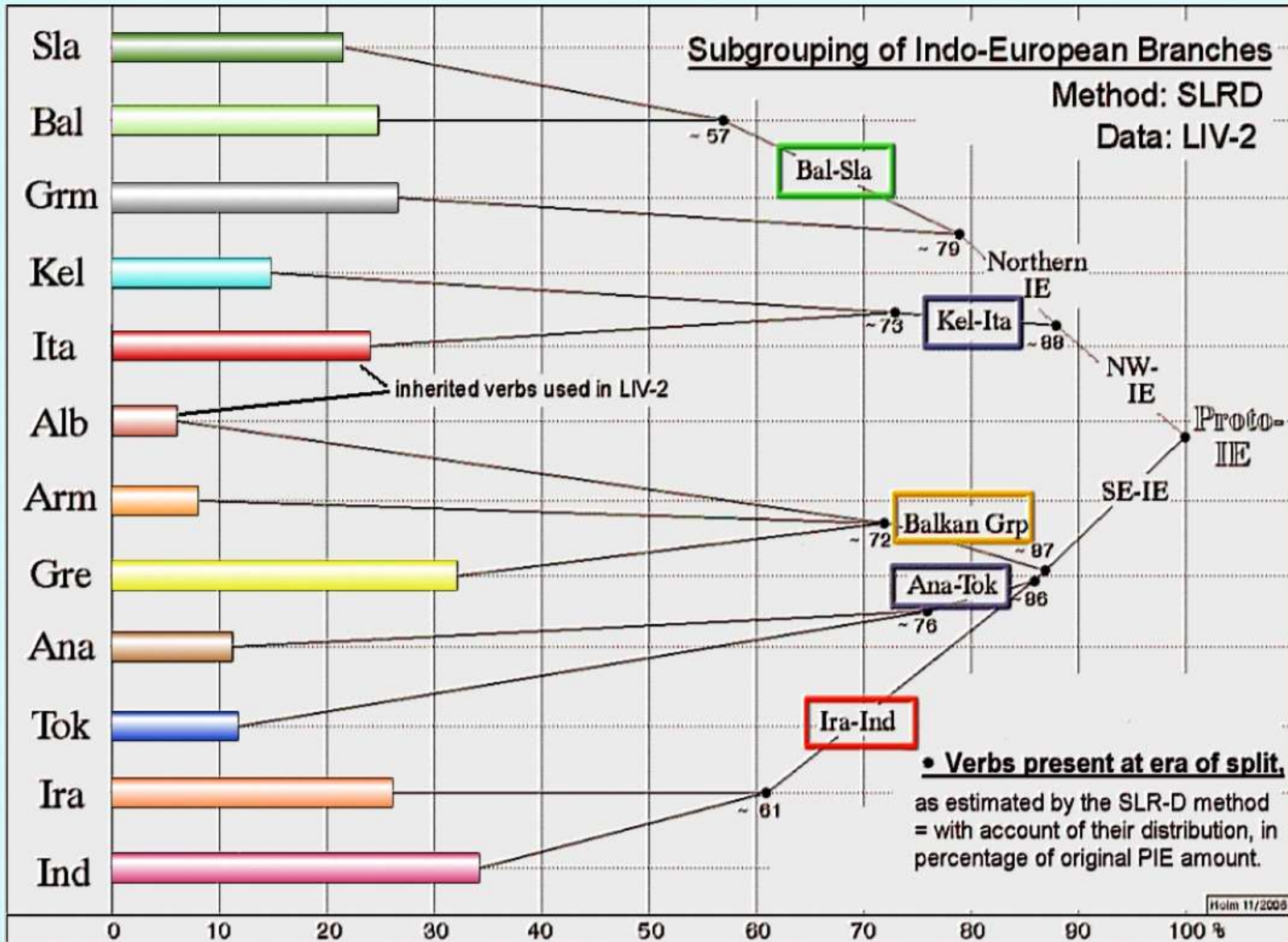
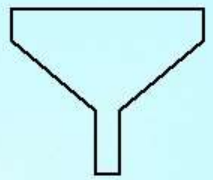
Erfordert getrennte Berechnung von 10 Ersetzungswahrscheinlichkeiten, plus 10·12 Streuungsberechnungen, mal 66 Sprachzweigpaare.

Roh-Ergebnis:

Lang	Sla	Bal	Grm	Kel	Ita	Alb	Arm	Gre	Ana	Tok	Ira	Ind
Sla	Sla 74		104	120	157	110	140	135	135	121	131	163
Bal		Bal	105	115	120	103	141	143	185	140	141	143
Grm			Grm	110	113	111	124	115	141	132	137	135
Kel				Kel 93		126	115	138	122	136	130	142
Ita					Ita	103	118	106	129	117	124	122
Alb						Alb 93	94		98	124	115	113
Arm							Arm 96		106	119	116	113
Gre								Gre	116	114	134	108
Ana									Ana 100		124	108
Tok										Tok	112	109
Ira											Ira 80	
Ind												Ind
K:	265.4	308.1	332.1	181.5	299.1	74.8	100	398.7	139.6	145.4	323.8	424.9

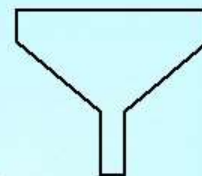
*Dies ergibt folgende Ausgliederungen der Indogermania ...*

# 8.1 Ausgliederung basierend auf LIV-2



*Keine  
Zeiten!!*

## 8.2 Diskussion



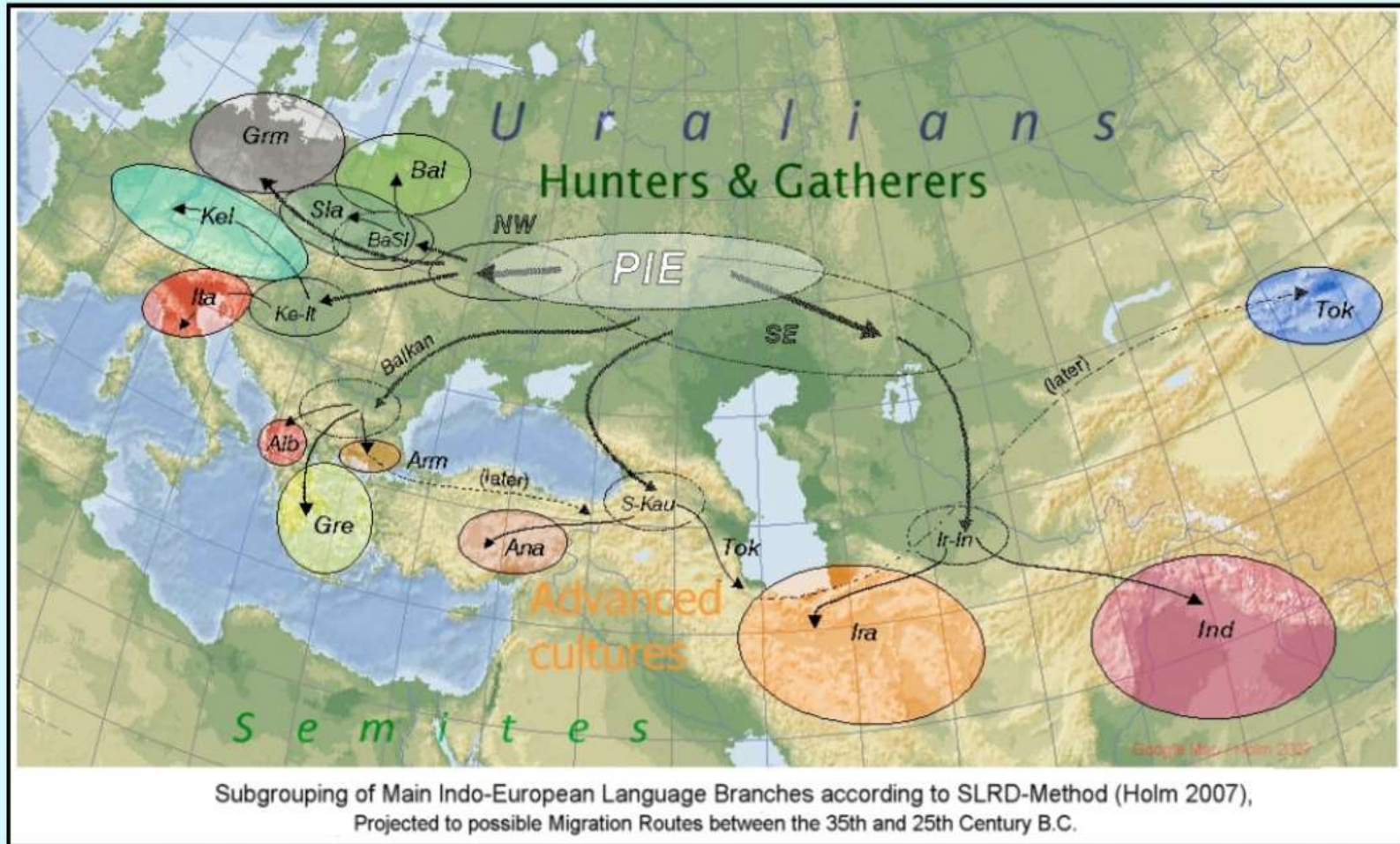
(M)ein Ausgliederungsmodell als "perfekt" zu bezeichnen, wie z.B. Ringe es tut, wäre vermessen; denn es bleiben noch einige "unhandliche" Fehlerquellen :

- Kulturell extrem unterschiedliche Substrate, z.B.
  - Jäger- und Sammler-Kulturen im Norden, vs.
  - Hochkulturen in Anatolien bzw. "Fruchtbarem Halbmond"!
- Intensitätsunterschiede und Fehler in der Forschung selbst;
- (Innovative) Zentrallagen vs. konservative Saumlagen verschiedener Sprachzweige. Irrige Auffassungen gehen von der umgekehrten Annahme aus (Nicols, MDS)!

*Ein wichtiger Test ist die Projektion in die reale Geographie ...*

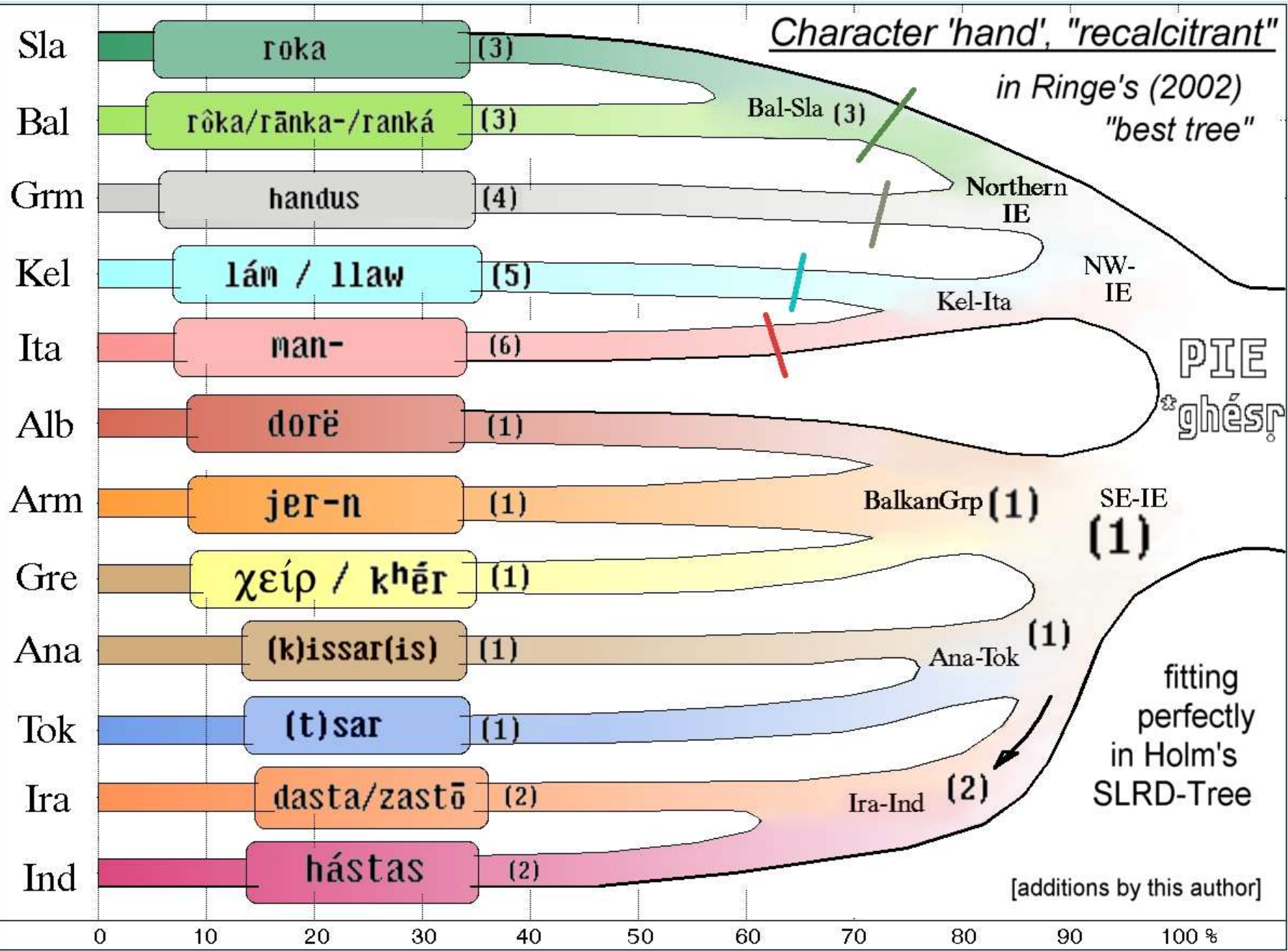
## 8.3 Teste: a) Geografie

Jedes Ausgliederungsergebnis muss in die Realität projizierbar sein!!!



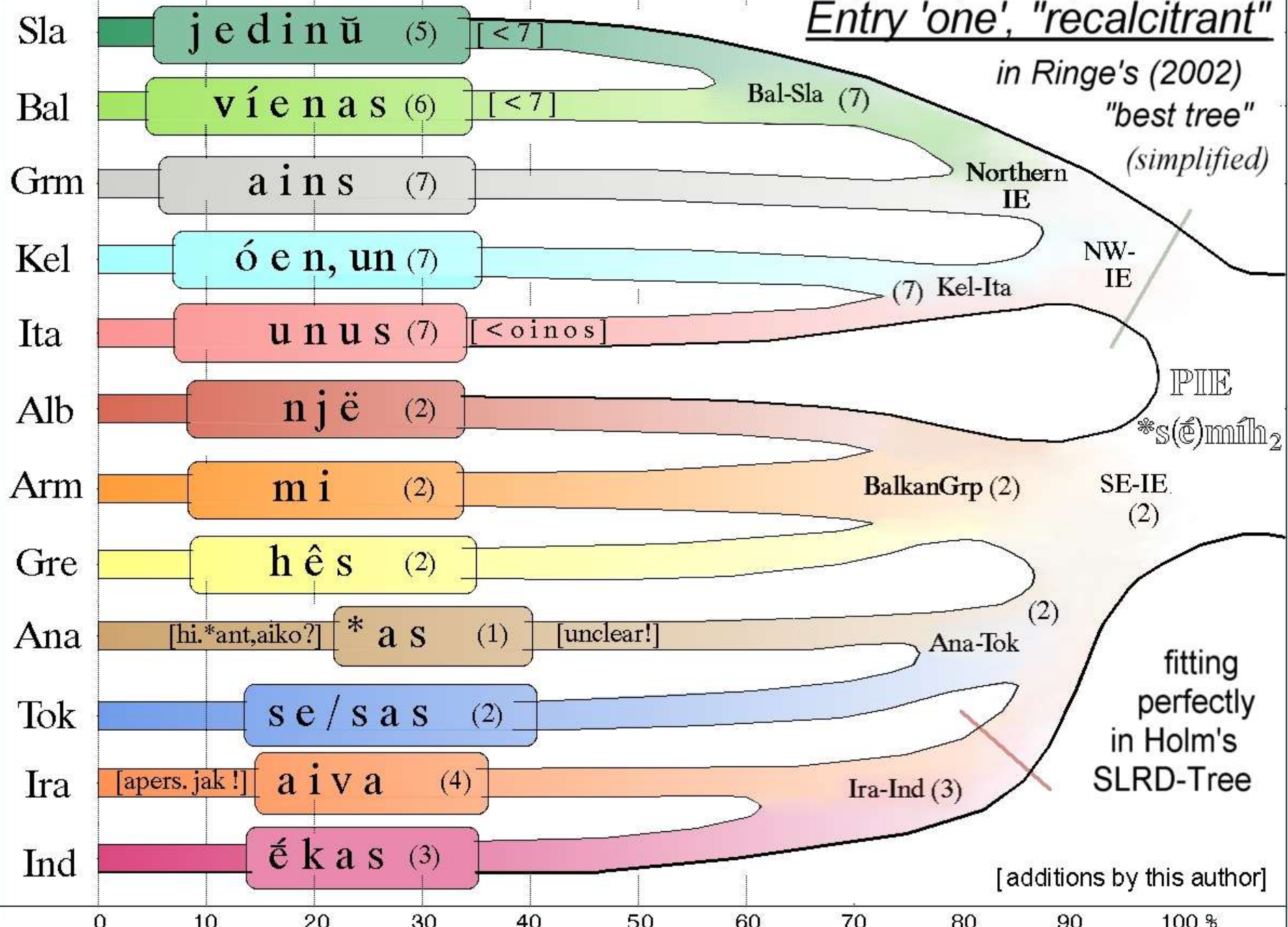
*Versuchen Sie das einmal mit anderen Modellen!*



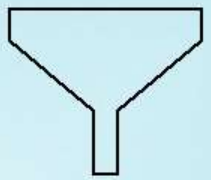


Entry 'one', "recalcitrant"

*in Ringe's (2002)  
"best tree"  
(simplified)*



## 8.4 a) Schluss - Rückblick

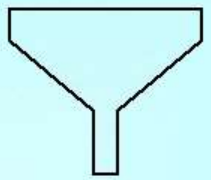


Wir haben erreicht ...

### Stochastisch

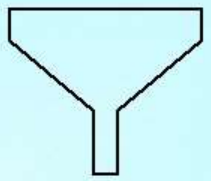
- Die Wiederherstellung des (quantitativen) Trennungszustandes verwandter Sprachen unter Berücksichtigung
  - aller vier beteiligten Parameter, sowie
  - ihrer Verteilungsunterschiede.

## 8.4 b) Schluss - Rückblick



Wir haben erreicht ...

- Linguistisch die
  - primäre Trennung in eine NW- gegen eine SO-Gruppe;
  - Sekundärgruppen
    - Bal-Sla,
    - Kel-Ita,
    - Balkan,
    - Ana-Tok,
    - Ira-Ind;
  - eindeutige Widerlegung einer frühen Abspaltung des Anatolischen von einem "Rest-Idg".



Absicht war, Sie zu befähigen, selbständig lexikostatistische Versuche zu beurteilen.

Achten Sie bitte dabei besonders auf

- die Quantität und Qualität der verwendeten Daten;
- wurden alle 4 Parameter berücksichtigt?
- wurde die Verteilung analysiert?
- versteckte Glottochronologie (G&A).

Lassen Sie sich niemals von trivialen Allerweltsergebnissen täuschen.  
Die großen Sprachzweige wurden noch von der schlechtesten Methode erkannt.

**Конец Фильма**