

Genetische Klassifikation von Sprachen

Teil 2: statistische Methoden

Zeittiefe der historischen Rekonstruktion

- drei mögliche Positionen:
 - generelle Grenze für Zeittiefe, die mit historischer Linguistik erreicht werden kann
traditionelle Position
 - Grenzen existieren, hängen aber von verfügbarer Information und Methoden ab
 - keine grundsätzliche Grenze (abgesehen von der biologischen Evolution der Sprachfähigkeit)
Greenberg, Ruhlen

Argumente für pessimistische Position

- solider Beweis für Verwandtschaft zweier Sprachen:
 - Rekonstruktion der gemeinsamen Ursprache
 - Rekonstruktion des Sprachwandels zu den von der Ursprache zu den fraglichen Sprachen
- verlangt Identifikation von *cognates*
- Sprachwandel verwischt Ähnlichkeiten
- außerdem könne Cognates auch durch Entlehnung entstehen
- nach maximal 10 000 Jahren keine Verwandtschaft mehr nachweisbar

Argumente für pessimistische Position

Hock & Joseph (1996):

Let us pursue this issue a little further by taking a closer look at the relationship between Modern Hindi and English – pretending that we do not yet know that they are related, and trying to establish their relationship by vocabulary comparison. This is actually more difficult than it appears. It is all too easy to be influenced by one's knowledge of the historical relationship between the two languages and therefore to notice the genuine cognates, or even to underestimate the effects of linguistic change on the recognizability of genuine cognates.

Argumente für pessimistische Position

Hock & Joseph (1996):

Clearly, one correspondence is not enough; nor are twenty. And just as clearly, a thousand correspondences with systematic recurrences of phonetic similarities and differences would be fairly persuasive. Are 500 enough, then? And if not, are 501 sufficient? Nobody can give a satisfactory answer to these questions. And this is no doubt the reason that linguists may disagree over whether a particular proposed genetic relationship is sufficiently supported or not.

Wortlisten

- Methoden der klassischen komparativen historischen Linguistik sind vermutlich tatsächlich weitgehend ausgereizt
- alternative Herangehensweise: Gebrauch von **Wortlisten**
- Identifikation von lautlichen Ähnlichkeiten im Wortschatz verschiedener Sprachen => Maß für Verwandtschaft der Sprachen
- kein Anspruch auf Rekonstruktion der Ursprache

Wortlisten

- Erstellung von *Konzeptlisten* – universeller Wortschatz, der in alle Sprachen übertragen werden kann
- Übersetzung dieser Liste in alle fraglichen Sprachen
- für jedes Übersetzungspaar stellt man fest, ob Verwandtschaft bzw. Ähnlichkeit besteht
- Verwandtschaft entspricht Prozentsatz an ähnlichen Wörtern

Wortlisten

- Morris Swadesh (1909-1967)
 - amerikanischer Linguist
 - untersuchte u.a. genetische Klassifikation von Indianersprachen
 - Pionier der **Lexikostatistik** und er **Glottochronologie**
 - erstellte s.g. **Swadesh-Liste** von 207 Konzepten, die in allen Sprachen/Kulturen vorkommen:
http://www.uni-erfurt.de/sprachwissenschaft/personal/lehmann/Fundus/Swadesh_list.html

Lexikostatistik

- Verwandtschaft oder Zufall?
(Quelle: Maiwald & Willeke)

Deutsch

Latein

Herz

cord-

Horn

cornu

Hund

canis

hundert

centum

Hirsch

cervus

Lexikostatistik

- Verwandtschaft oder Zufall?

Englisch

Hawaiisch

Begriff

sew	humu	nähen
smell	honi	riechen
snow	hau kea	der Schnee
stab	hou	niederstechen
star	hoku	der Stern
swell	ho'opehu	(an)schwellen

Glottochronologie

- Grundannahme:
 - jede Sprache erneuert im Laufe der Zeit ihr Vokabular
 - Prozess verläuft häufig stetig
 - Wortschatz einer Sprache hat also „Halbwertszeit“
 - diese Halbwertszeit ist für alle Sprachen ungefähr gleich
 - Wortlistenvergleich lässt Rückschlüsse zu, wann Sprachen sich getrennt haben

Glottochronologie

- Eichung anhand Vgl. Englisch vs. Spanisch
- Schätzung: nach 1000 Jahren sind noch 81% des 200-Swadesh-Liste erhalten

- Formel:

$$t = \frac{\ln c}{\ln r}$$

- *t*: Zeittiefe (Zeit seit Ursprache)
- *t/2*: zwischen zwei Schwestersprachen aufgeteilte Zeittiefe (Abstand von dem nächsten ihnen gemeinsamen Knoten)
- *c*: Prozentsatz des gemeinsamen Basiswortschatzes ($0 < c < 1$)
- *r*: glottochronologische Konstante. (81%)

Glottochronologie

- Vorteile:
 - auf beliebige Sprachpaare anwendbar
 - geringer Analyseaufwand
 - gibt über beliebig entfernte Verwandtschaftsverhältnisse Auskunft
 - gibt Auskunft über Zeittiefe, nicht nur über Verwandtschaft

Glottochronologie

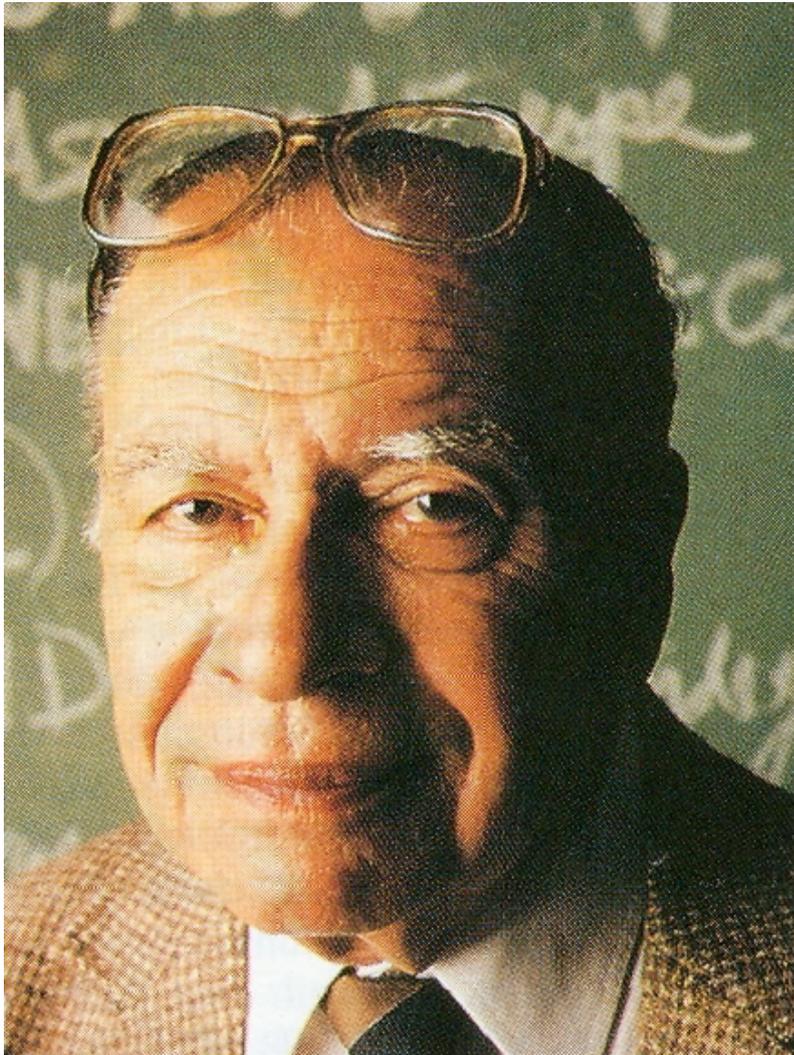
- Nachteile

- linguistische Verwandtschaft schlägt sich auch (oder sogar stärker) in Grammatik nieder als in Wortschatz
- Swadesh-Liste ist nicht universal (manche Sprachen haben keine Wörter für *wenn*, *fünf*, *riechen* usw.)
- Identifikation von Cognates fraglich, wenn Geschichte von Sprache nicht bekannt ist
- Halbwertszeit ist nicht konstant, sondern hängt von einer Reihe von Faktoren ab, darunter Sprachkontakt, Tabu, schriftliche Tradition, Nationalstolz.

Fazit

- „Klassische“ Lexikostatistik & Glottochronologie
 - interessanter Ansatz
 - basiert jedoch auf z.T. fraglichen Hintergrundannahmen
 - unzureichende mathematische Unterfütterung, um Rolle von Zufall korrekt einzuschätzen => Statistik

Joseph Greenberg (1915 - 2001)



- einer der wichtigsten Linguisten des 20. Jhd.
- Begründer der linguistischen Typologie
- bahnbrechende Untersuchungen zu linguistischen Universalien
- Klassifikation der afrikanischen und amerikanischen Sprachen (stark umstritten)

Mass lexical comparison

- Greenberg:
 - statistische Zuverlässigkeit der Lexikostatistik kann erhöht werden, wenn man Anzahl der Datenpunkte erhöht
 - statt Vergleich zweier Sprachen Vergleich der Grundwortschätze ganzer **Sprachgruppen**

Mass lexical comparison

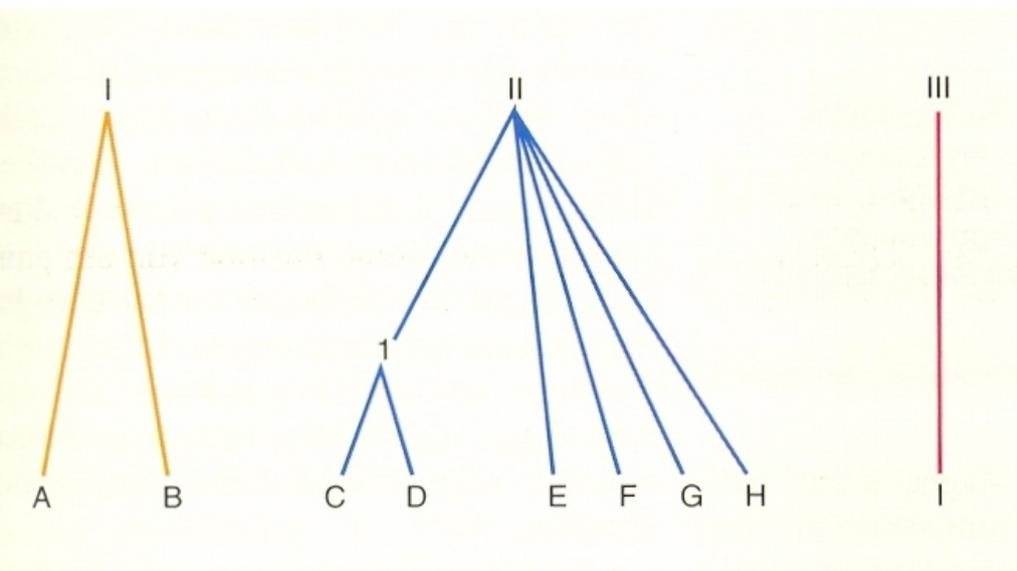
Eine Klassifikationsübung nach der vergleichenden Methode à la Merritt Ruhlen:

Sprache	zwei	drei	ich	du	wer?	nicht	Mutter	Vater	Zahn	Herz	Fuß	Maus	er trägt
A	ʔiθn-	θalāθ-	-ni	-ka	man	lā	ʔumm-	abū	sinn	lubb	rijl-	fār	yaḥmil-
B	ʃn-	šaloš	-ni	-ka	mi	lo	ʔem	aβ	šen	leβ	regel	ʃaḵbər	nošeh
C	duvā	tráyas	mām	tuvám	kás	ná	mātár	pitár-	dant-	hṛd-	pád	muš-	bháрати
D	duva	θrāyō	mām	tuvəm	čiš	naē-	mātar-	pitar-	dantan-	zərəd	paiḏya		baraiti
E	duo	treîs	eme	sú	tís	ou(k)	māter	pater	odón	kardiā	pod-	mûs	phérei
F	duo	trēs	mē	tū	kwis	ne-	māter	pater	dent-	kord-	ped-	mūs	fert
G	twai	θreis	mik	θu	hwas	ni	aiθei	faðar	tunθus	haírtō	fōt		baíriθ
H	dó	trí	-m	tú	kía	ní-	máθir	aθir	dēt	kride	traig	lux	berid
I	iki	üč	ben-i	sen	kim	deyil	anne	baba	diš	kalp	ayak	sičan	tašiyor

Mass lexical comparison

Eine Klassifikationsübung nach der vergleichenden Methode à la Merritt Ruhlen:

Sprache	zwei	drei	ich	du	wer?	nicht	Mutter	Vater	Zahn	Herz	Fuß	Maus	er trägt
A	ʔiθn-	θalāθ-	-ni	-ka	man	lā	ʔumm-	abū	sinn	lubb	rijl-	fār	yaḥmil-
B	ʃn-	šaloš	-ni	-ka	mi	lo	ʔem	aβ	šen	leβ	regel	ʃaḳbər	nošeh
C	duvā	tráyas	mām	tuvám	kás	ná	mātár	pitár-	dant-	ḥrd-	pád	muṣ-	bhárati
D	duva	θrāyō	mām	tuvəm	čiš	naē-	mātar-	pitar-	dantan-	zərəd	paiḏya		baraiti
E	duo	treis	eme	sú	tís	ou(k)	māter	pater	odón	kardiā	pod-	mūs	phérei
F	duo	trēs	mē	tū	kwis	ne-	māter	pater	dent-	kord-	ped-	mūs	fert
G	twai	θreis	mik	θu	hwas	ni	aiθei	faðar	tunθus	haírtō	fōt		baíriθ
H	dó	trí	-m	tú	kía	ní-	máθir	aθir	dēt	kride	traig	lux	berid
I	iki	üč	ben-i	sen	kim	deyil	anne	baba	diš	kalp	ayak	sičan	tašiyor



Klassifizieren Sie die angegebenen neun Sprachen (von A bis I) in Familien und Unterfamilien und vergleichen Sie den Wortschatz für die 13 Wörter, die hier in phonetischer Umschrift geboten werden. Lösung: Sprache A und B (Arabisch und Hebräisch) gehören zur Familie der semitischen Sprachen. Die sechs Sprachen C bis H (Sanskrit, Awestisch, Altgrie-

chisch, Latein, Gotisch und Altirisch) sind indogermanische Sprachen. I (Türkisch) läßt sich keiner Familie zuordnen. Mit einer längeren Wortliste kann man nach demselben Verfahren die Familien wieder in Überfamilien einteilen usw. Der Stammbaum, den man so erhält, würde dann beweisen, daß alle Sprachen von einer Muttersprache abstammen.

Mass lexical comparison

Multilateraler Sprachenvergleich

Schlichtes Vergleichen einiger Allerweltswörter erhellt bereits die Verwandtschaftsverhältnisse unter den Sprachfamilien Indoeuropäisch (mit den Zweigen Germanisch, Romanisch und Slawisch) sowie Uralisch-Jukagirisch und Baskisch.

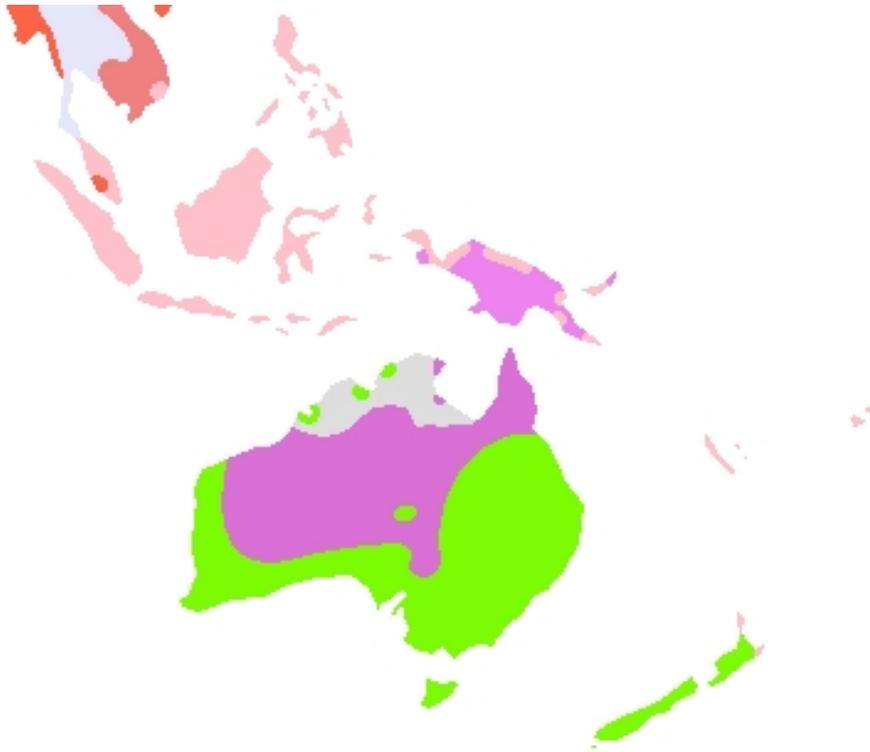
Sprachfamilie	Sprache	eins	zwei	drei	Kopf	Auge	Nase	Mund
<i>Germanisch</i>	Schwedisch	en	tvo	tre	hyvud	øga	næsa	mun
	Niederländisch	ēn	tvē	dri	hōft	ōx	nōs	mont
	Englisch	wæn	tū	θrī	hæd	ai	nouz	mauθ
	Deutsch	ains	tsvai	drai	kopf	auge	nāzə	munt
<i>Romanisch</i>	Französisch	œ/yn	dø	trwa	tət	œj	ne	buš
	Italienisch	uno	due	tre	təsta	okjo	naso	boka
	Spanisch	uno	dos	tres	kabesa	oxo	naso	boka
	Rumänisch	un	doi	trei	kap	oki	nas	gure
<i>Slawisch</i>	Polnisch	jeden	dwa	trzy	gwowa	oko	nos	usta
	Russisch	odin	dva	tri	galava	oko	nos	rot
	Bulgarisch	edin	dva	tri	glava	oko	nos	usta
<i>Uralisch-Jukagirisch</i>	Finnisch	yksi	kaksi	kolme	pæ	silmæ	nenæ	sū
	Estnisch	yks	kaks	kolm	pea	silm	nina	sū
<i>Baskisch</i>	Baskisch	bat	bi	hiryr	byry	begi	sydyr	aho

Afrika



- Greenberg 1963 „The languages of Africa“
 - nur vier Sprachfamilien in Afrika
 - Afroasiatisch (ersetzt das trad. „Hamito-Semitisch“)
 - Niger-Kongo
 - Nilosaharanisch
 - Khoisan
 - nicht unumstritten, aber weitgehend anerkannt ²¹

Pazifik/Ozeanien



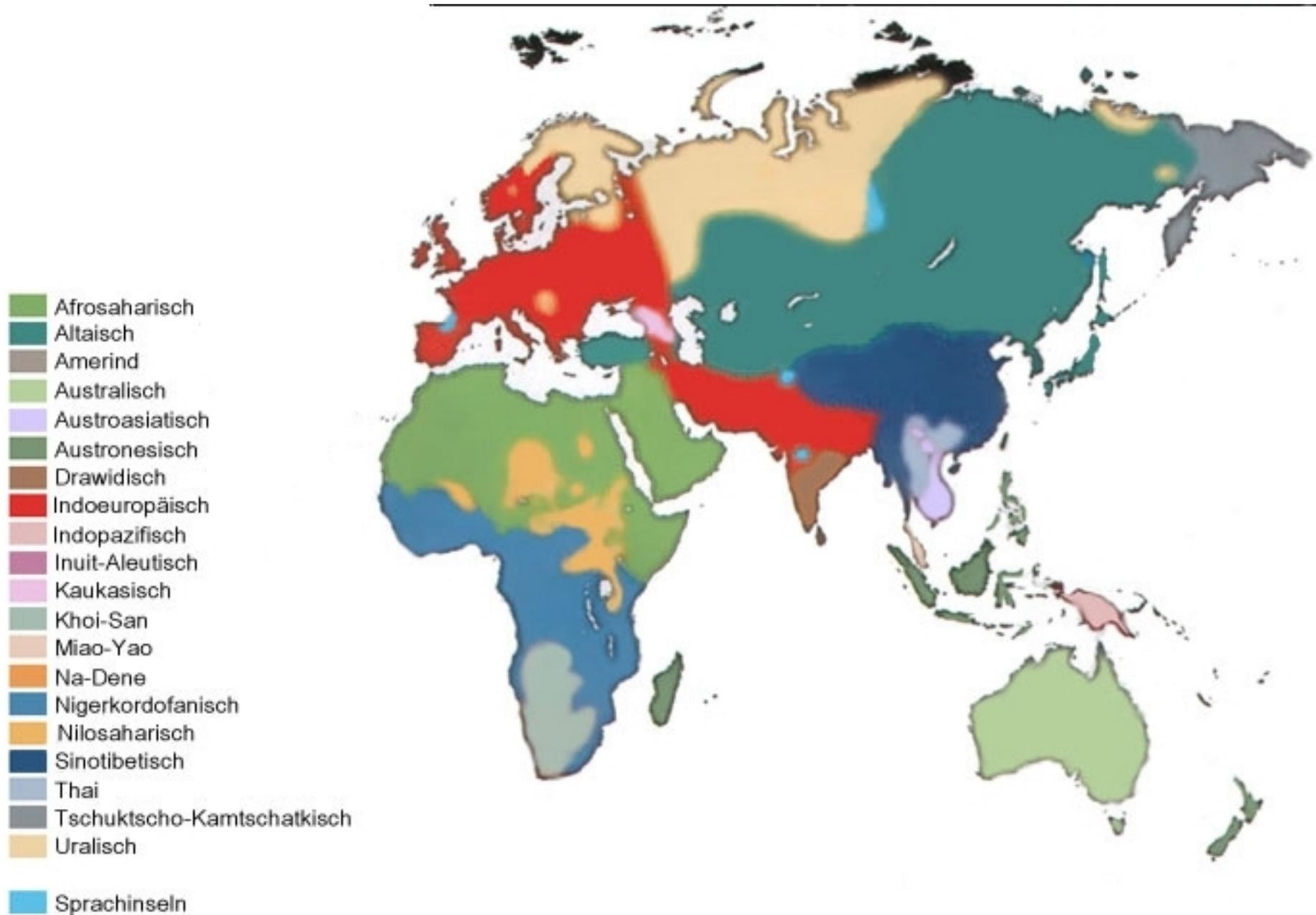
- 1971: „The Indo-Pacific Hypothesis“
 - Indo-pazifische Makrofamilie
 - soll Papuasprachen, andamanische und tasmansische Sprachen umfassen
- heute übereinstimmend verworfen

Amerika



- 1987: „Language in the Americas“
- drei Makrofamilien:
 - Eskimo-Aleutisch
 - Na-Dené
 - Amerindisch
- letztere Klasse stark umstritten
- Standard-Lehrmeinung: Greenbergs Amerindisch besteht aus ca. 200 Familien und isolierten Sprachen

Eurasien



Eurasien

- 2000/2002: „Indo-European and Its Closest Relatives“
 - Makrofamilie „Eurasiatisch“
 - Unterfamilien:
 - Indoeuropäisch
 - uralische Sprachen
 - Altaisch (Turksprachen, Mongolisch, Tungusisch, Koreanisch, Japanisch)
 - Eskimo-Aleutisch
 - diverse isolierte Sprachen (z.B. Etruskisch)
 - ebenfalls umstritten

Ruhlen

- Merritt Ruhlen
 - Schüler von Greenberg
 - noch radikalere Anwendung von Greenbergs Methoden der genetischen Klassifikation
 - Hypothese, dass partielle Rekonstruktion von „Proto-World“, also der Ursprache der Menschheit möglich sei
 - von der Fachwelt fast einhellig abgelehnt

Kritik an Greenberg/Ruhlen

- Grundsätzliche Einwände gegen Verwendung von Wortlisten
 - nicht kulturunabhängig
 - Annahme von „universalen“ Konzepten fraglich
- konkrete Einwände gegen Greenbergs Methode
 - sehr sehr weitherzige Auslegung von „semantischer Entsprechung“
 - keine solide statistische Auswertung
 - Boë et al. 2003: Ruhlens Rekonstruktion von Proto-World basiert auf statistisch nicht signifikanten Daten (Ähnlichkeiten könnten auch Zufall sein)
 - Greenberg ist z.T. zu brauchbaren Ergebnissen gekommen, aber bislang konnte niemand anderes seine Methode überzeugend anwenden => Erfolg mehr aufgrund Greenbergs extrem guter Intuition als aufgrund einer validen Methode

Der genetische Stammbaum des Menschen

- Luigi Luca Cavalli-Sforza (1922*)
 - Evolutionbiologe (Kollege von Greenberg in Stanford)
 - Human Genome Diversity Project
 - Versuch, biologischen Stammbaum der modernen Menschheit mit Hilfe von Gen-Analysen zu erstellen

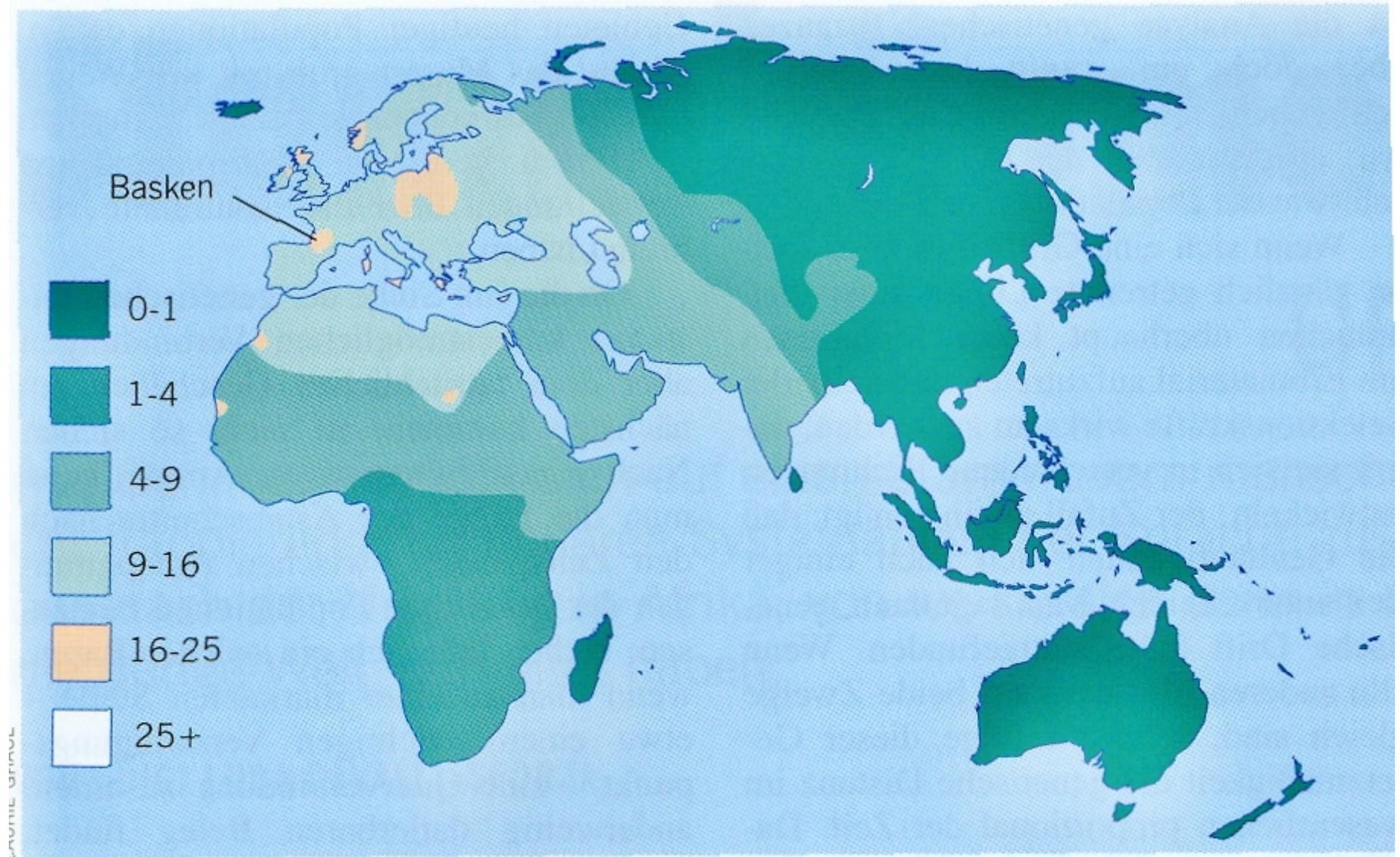


Cavalli-Sforza

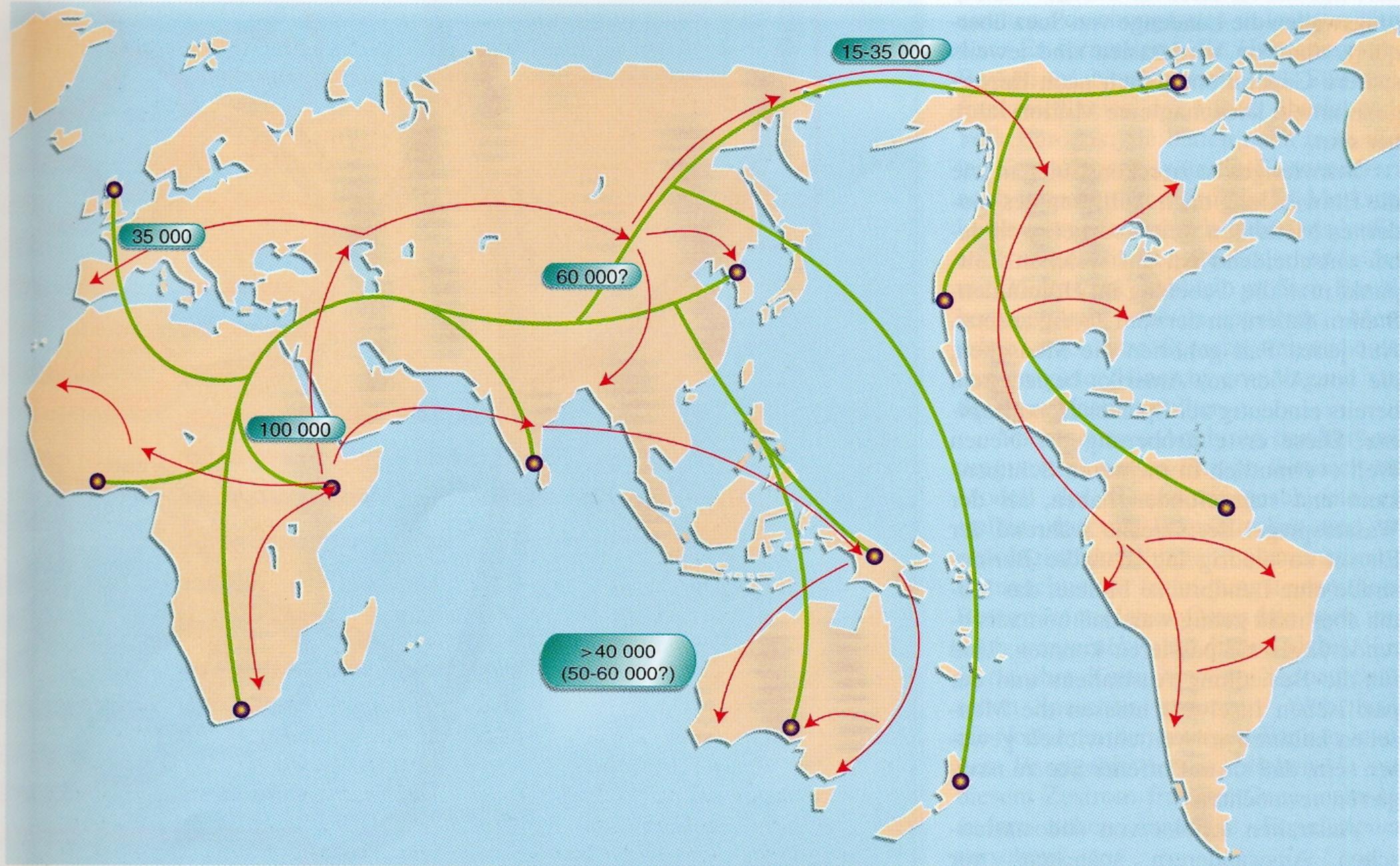
- Grundidee:
 - Evolution basiert auf **natürlicher Auslese**: Allele, die Fitness (d.h. Überlebens- und Fortpflanzungsfähigkeit) erhöhen, setzen sich allmählich in Population durch
 - Variation entsteht durch genetische **Mutation**
 - viele Mutationen haben aber keinen phänotypischen Effekt
 - in diesen Fällen findet keine Auslese statt
 - ob sich derartige neutrale Mutationen durchsetzen, ist purer Zufall => sog. **genetischer Drift**

Cavalli-Sforza

- Beispiel:
 - Rhesusfaktor des Bluts
 - hat keinen Einfluss auf Fitness
 - ist erblich
 - zusammenhängende Population hat spezifischen Prozentsatz von Rh-
 - wenn Population getrennt wird, triffen diese spezifischen Werte allmählich auseinander
 - Differenz von Prozentsatz von Rh- somit grober Gradmesser der Verwandtschaft von Populationen
 - Verwendung mehrerer Hunder derartiger neutraler genetischer Marker



Anteile der Menschen mit dem Blutfaktor rhesus-negativ in der Alten Welt und in Australien. Bei der baskischen Bevölkerung kommt er am häufigsten vor; nach Osten und Süden zu wird er immer seltener. Die Basken scheinen eine sehr alte Bevölkerung zu sein, die erst spät Kontakt zu Einwanderern aus dem asiatischen Raum bekam und deshalb viele ihrer ursprünglichen Merkmale bewahrt hat, so auch ihre Sprache. Die Zahlen geben den Prozentsatz rhesus-negativer Menschen an.



Rekonstruktion der Ausbreitung des Menschen in vorgeschichtlicher Zeit. Ein erster genetischer Stammbaum (grün) wurde derart auf eine Weltkarte projiziert, daß die Endpunkte der Zweige in den heutigen Regionen der einzelnen Populationen liegen. Das Ergebnis paßt recht gut zu einer Rekonstruktion nach archäologi-

schen und fossilen Funden (die Zahlen bezeichnen das Auftauchen des anatomisch modernen Menschen in Jahren vor der Gegenwart). Neuere genetische Untersuchungen (rot) lassen vermuten, daß der *Homo sapiens* auf zwei Routen nach Asien gelangte; die Details der Wege beruhen aber auf Spekulation.

Cavalli-Sforza

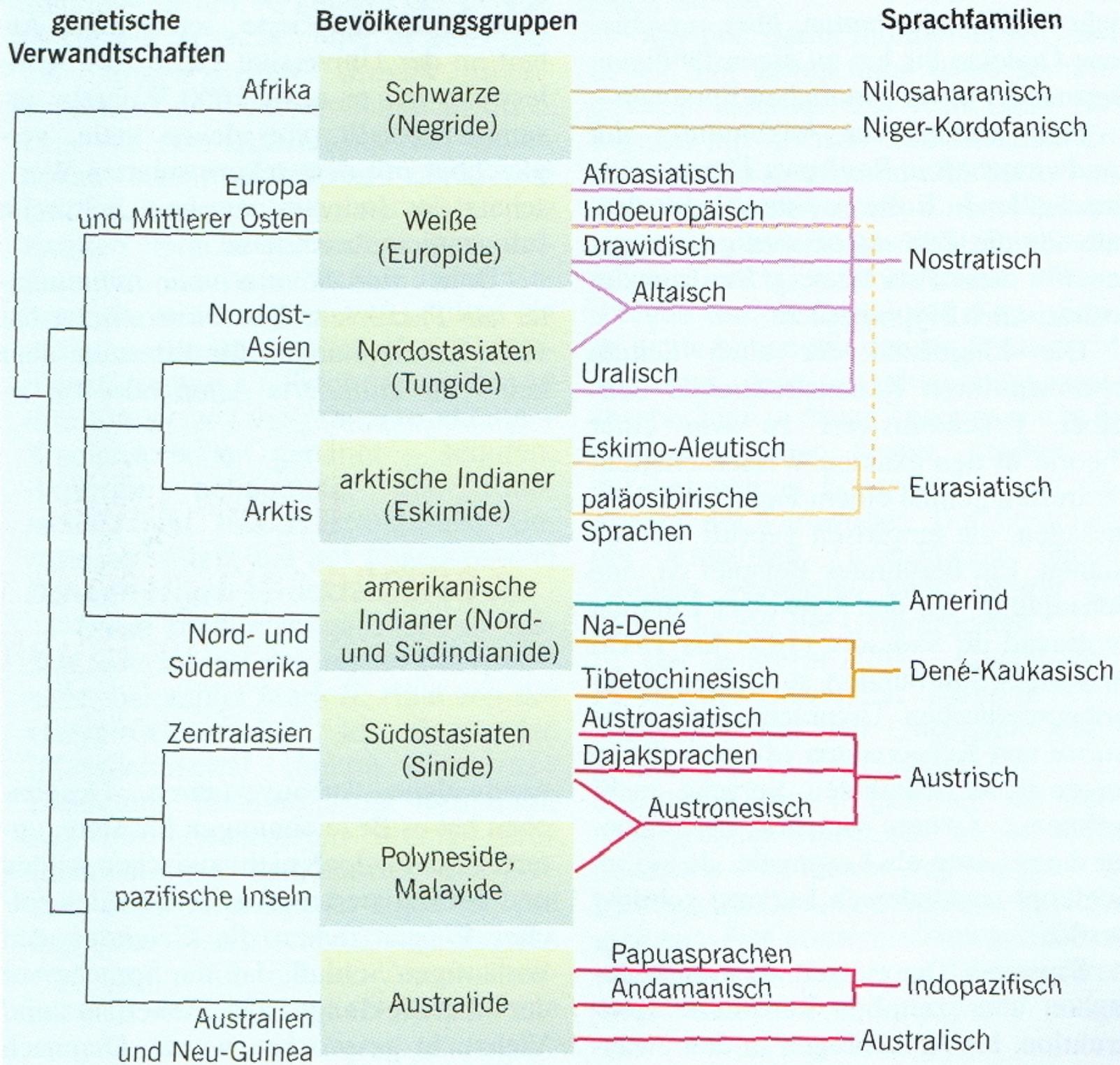
- Projekt z.T. kritisiert, weil es angeblich Konzept von Rasse wiederbelebt
- nicht stichhaltig; im Gegenteil:
 - Ergebnisse zeigen, dass es keine menschl. Rassen im biolog. Sinne gibt (i.S. von völliger reproduktiver Isolation)
 - Aufspaltung der menschl. Teilpopulationen geschah vor – evolutionär gesehen – sehr kurzer Zeit (ca. 60 000 Jahre)
 - genet. Variation innerhalb einer Population häufig größer als zwischen Populationen
 - äußere Merkmale wie Hautfarbe, Haarkonsistenz etc. betreffen nur kleine Anzahl von Genen

Cavalli-Sforza & Greenberg

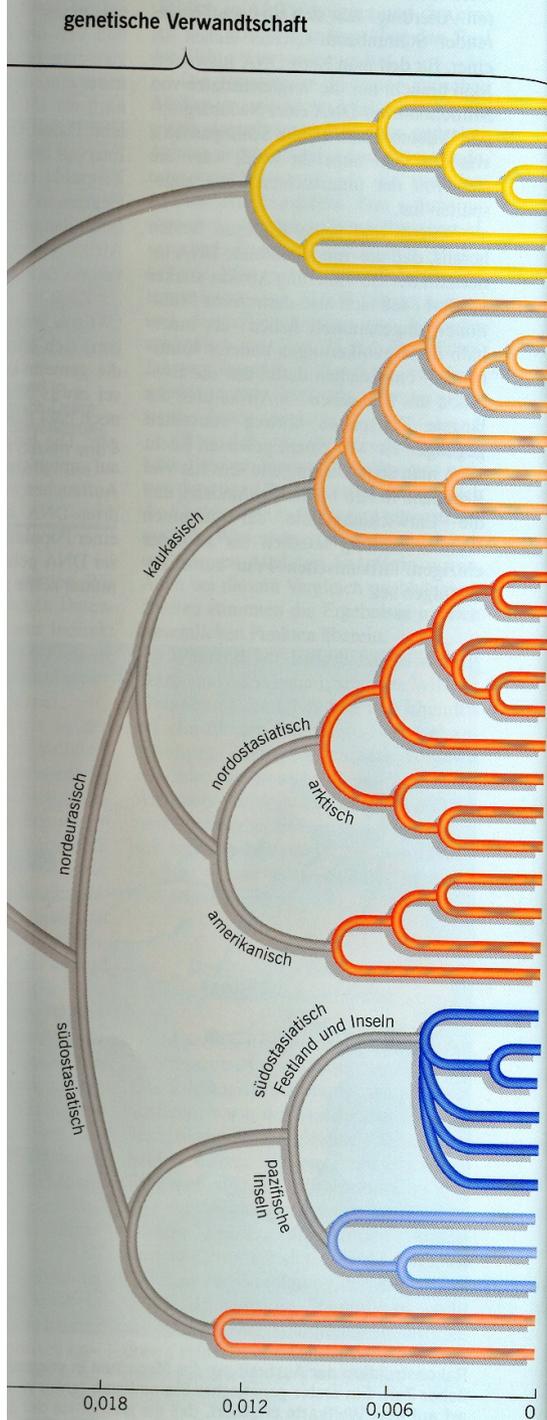
- verblüffend gute Übereinstimmung zwischen Cavalli-Sforzas und Greenbergs Klassifikation der menschlichen Populationen
- zusätzliches Argument für die Korrektheit von Greenbergs Untersuchungen

Cavalli-Sforza & Greenberg

- einzelne Divergenzen:
 - Ungarn sprechen uralische Sprache, unterscheiden sich aber genetisch nicht von Nachbarvölkern
 - Lappen (Samen) sprechen auch uralische Sprache, stehen aber genetisch den anderen Skandinaviern und mongoloiden Sibiriern näher
 - Äthiopier sprechen afro-asiatische Sprache, stehen aber genetisch schwarzafrikanischen Populationen näher als nordafrikanischen



Der biologische und der linguistische Stammbaum



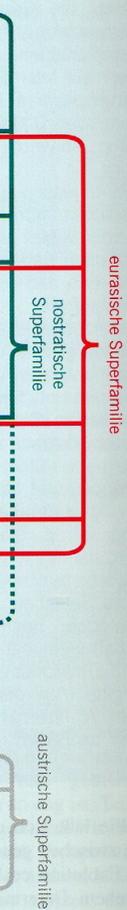
Genetischer Stammbaum (links) und der linguistisch ermittelte

Populationen

- Mbuti-Pygmäen
- Westafrikaner
- Bantu
- Nilbevölkerung
- San (Buschleute)
- Äthiopier
- Berber, Nordafrikaner
- Südwestasiaten
- Iraner
- Europäer
- Sarden
- Inder
- Südostinder
- Lappen
- Samojeden
- Mongolen
- Tibeter
- Koreaner
- Japaner
- Ainu
- Sibirier
- Eskimo
- Tschuktschen
- Indianer Südamerikas
- Indianer Mittelamerikas
- Indianer Nordamerikas
- Ind. Nordwestamerikas
- Südchinesen
- Mon-Khmer
- Thai
- Indonesier
- Malaiien
- Filipinos
- Polynesier
- Mikronesier
- Melanesier
- Neuguineer
- Australier

Sprachfamilien

- (Ursprache nicht bekannt)
- Niger-Kordofanisch
- Nilosaharisch
- Khoi-san
- Afroasiatisch
- Indoeuropäisch
- Drawidisch
- Uralisch
- Tibetochinesisch
- Altaiisch
- Eskimo-Aleutisch
- Tschuktschisch-Kamtschatkisch
- Amerind
- Na-Dené
- Tibetochinesisch
- Austroasiatisch
- Thai
- Austronesisch
- Indopazifisch
- Australisch



Die Sprachenklassifikation stammt aus „A Guide to the World of Languages“ von Merritt Ruhlen.

Sprachenstammbaum (rechts) heute lebender Völker in der Gegenüberstellung

Quellen

- http://www.uni-erfurt.de/sprachwissenschaft/personal/lehmann/CL_Lehr/Wandel/Wandel_Glottochronologie.html
- <http://lms.cms.hu-berlin.de/moodle/course/view.php?id=1011>
- <http://lms.cms.hu-berlin.de/moodle/mod/resource/view.php?id=40966>
- Die Evolution der Sprache, Spektrum der Wissenschaft -- Dossier 1/2000
- Ruhlen, M., On the Origin of Languages: Studies in Linguistic Taxonomy. Stanford University Press, 1996.
- Ruhlen, M., The Origin of Language: Tracing the Evolution of the Mother Tongue. Wiley, 1996.