

Besprechungsaufsatz

Zur Methodik des Sprachvergleichs *

1. Wenn auch in der Linguistik immer häufiger mit quantitativen Daten gearbeitet wird, so verbirgt sich dahinter der Wunsch, den speziellen Gegenstandsbereich besser zu erfassen, um die Erkenntnis über ihn zu vertiefen. Untersuchungen, die auf dieses Ziel gerichtet sind, müssen somit begrüßt werden, denn sie zeugen davon, daß die betreffenden Autoren sich nicht mit klassifikatorischen (kategorischen) Daten zufrieden geben wollen. Gerade weil das Arbeiten mit quantitativen Daten ein wichtiges Verfahren der Linguistik ist, muß man darauf achten, den aus diesem Unterfangen erwachsenden Forderungen Genüge zu tun. Vor allem diesen allgemeinen methodologischen Aspekt wollen wir bei der Behandlung der Arbeit von Zita Hess zugrunde legen.

Diese Arbeit stellt sich zum Ziel, «die heutigen romanischen Sprachen auf rein *synchronischer* Basis zu vergleichen» [12]. Es wäre gut gewesen, die Ausgangshypothese stärker zu explizieren: daß nämlich die Einteilung der romanischen Sprachen auf synchroner Grundlage nicht der entsprechenden Einteilung auf diachroner Grundlage entspricht. Um einen Vergleich zu ermöglichen, der nicht willkürlich einzelne – heterogene – Merkmale auswählt und insgesamt unvollständig bleibt, beschränkt sich Zita Hess auf die phonologische Ebene. Dabei sollen die phonologischen Teilsysteme möglichst weitgehend erfaßt werden [12]. Damit meint die Autorin vor allem den Wunsch, nicht nur qualitative Aspekte – etwa die Zusammensetzung des Phoneminventars –, sondern auch quantitative Aspekte in den Sprachvergleich einzubeziehen. Dies erfordert umfangreiches empirisches Arbeiten, doch sind hier zwei Arten von Daten zu unterscheiden: (a) solche Daten, die direkt aus der Stichprobe, die für eine Sprache erhoben wird, gewonnen werden, wie z. B. die Phonemfrequenzen, und (b) solche Zahlen, die aus den erstgenannten Daten berechnet werden, Zahlen, die die jeweilige Sprache kennzeichnen sollen. Der Anspruch der Autorin, die phonologischen Teilsysteme möglichst vollständig zu erfassen, kann nur für die unter (a) genannten Daten gelten.

* Anlässlich von Zita Hess, *Typologischer Vergleich der romanischen Sprachen auf phonologischer Basis* (Heidelberger Beiträge zur Romanistik, 8), Bern–Frankfurt (Lang) 1975, 412 S.

Eine umfassende quantitative Charakterisierung der romanischen Sprachen unterbleibt – offensichtlich aus Unkenntnis der entsprechenden Methoden und Verfahren, und dies heißt auch, in Unkenntnis der hier relevanten linguistischen Literatur. Eben weil keine entsprechende Methode herangezogen wurde, beschränkt sich Zita Hess auf einen synchronen Sprachvergleich und verzichtet auf die Nennung einer *Klassifikation* als Ziel, die aber von der Ausgangshypothese her eigentlich anzustreben wäre. Auch einen *globalen* Vergleich der romanischen Sprachen (auf phonologischer Ebene) führt sie nicht durch. Angeboten werden dem Leser eine ganze Reihe von Vergleichen, die aber jeweils nur 1 Merkmal berücksichtigen. Durch sie werden aber keine Informationen zusammengefaßt, wie man es von einer guten Klassifikation erwarten kann, sondern es wird im Gegenteil eine Vielzahl von Detailinformationen neu geschaffen, die eher dazu geeignet sind zu verwirren, als dazu, Überblick zu verschaffen (man schaue sich etwa die Seiten 174 ff., 178 ff., 183 ff., 232 ff., 272 ff. und 297 ff. an).

Gleich zu Beginn setzt sich die Autorin kritisch mit dem Klassifikationsversuch von Muljačić (1967) auseinander, der bekanntlich an 12 romanische Sprachen 40 'Fragen' stellt, auf die mit „ja“ oder „nein“ zu antworten war. Zita Hess stellt die Repräsentativität der von Muljačić ausgewählten Merkmale in Frage und kritisiert die Vermischung synchroner und diachroner Merkmale [12]. Tatsächlich kann a priori nicht entschieden werden, ob und welche (dieser 40) Merkmale für eine Klassifikation (der romanischen Sprachen) wichtig, relevant sind (vgl. Sneath/Sokal, 1973, 109 ff.). Auch hat Muljačić seine Merkmale schon insofern gewichtet, als er nur solche aufgenommen hat, die mindestens eine Sprache von den übrigen unterscheiden. Trotz möglicher Einwände im einzelnen ist Muljačić in seinem Bemühen um einen Vergleich und eine Klassifikation einen bedeutenden Schritt weiter als die Autorin des vorliegenden Buches gegangen, wenn er auch den letzten Schritt zu einer eigentlichen Klassifikation noch nicht getan hat. Davon wird noch die Rede sein.

Der Hauptteil der Arbeit umfaßt zwei Teile, einen «Vergleich der Phoneminventare der romanischen Sprachen» [13–186] und einen «phonematischen Vergleich der romanischen Sprachen auf Grund von statistischen Untersuchungen» [187–299]. Im ersten Teil werden nach einigen Vorüberlegungen [13–34] die Phoneminventare der sieben hier untersuchten romanischen Sprachen (Französisch, Italienisch, Katalanisch, Okzitanisch, Portugiesisch, Rumänisch und Spanisch) dargestellt und erläutert. Die erfreuliche Ausführlichkeit dieser Darlegungen zeugt davon, daß es unmöglich ist, ein 'richtiges' Phoneminventar für eine Sprache anzugeben, daß für die verschiedenen Alternativen vielmehr andere, außersprachliche Kriterien zur Beurteilung herangezogen werden müssen. Zu begrüßen ist auch die einheitliche Darstellungsweise, die das ganze Buch durchzieht; auf diese Weise wird ein Nachschlagen einzelner Daten sehr

erleichtert. Der eigentliche Vergleich der Phonemsysteme umfaßt die Seiten 172–186. Hier wird ausgezählt – einmal bezüglich der silbischen Phoneme, bezüglich der asilbischen Phoneme (= Konsonanten) und dann hinsichtlich der Prosodeme, schließlich hinsichtlich aller Aspekte gleichzeitig – wie viele Phoneme zwei Sprachen gemeinsam sind und hinsichtlich wie vieler Phoneme sie sich unterscheiden. Die Problematik eines solchen Vorgehens wird nicht thematisiert (vgl. hierzu Altmann/Lehfeldt, 1973, 69f.). Für die Vergleichbarkeit zweier Phoneme aus verschiedenen Sprachen reicht es ja nicht aus, daß beide Phoneme durch gleiche Grapheme repräsentiert werden, wie dies die Zusammenstellung auf S. 172 suggerieren kann. Als eine Art Distanzmaß zwischen zwei Sprachen wird ein Quotient aus der Zahl der gemeinsamen, geteilt durch die Zahl der differierenden Phoneme gebildet. Weist dieser Index Werte > 1 auf, so überwiegen die gemeinsamen Phoneme, Zahlen < 1 deuten auf ein Überwiegen differierender Phoneme. Die Werte dieses Quotienten bewegen sich in dem Intervall $< 0; \infty >$, was nie günstig ist. Wenn man dagegen feststellt, wie viele Phoneme zwei Sprachen gemeinsam sind, und dann die nächstliegende Frage stellt, nämlich „von wie vielen?“, bekommt man automatisch einen Index, dessen Werte sich in dem Einheitsintervall $< 0; 1 >$ bewegen und folglich überhaupt erst vergleichend interpretiert werden können. Die hier und im folgenden von der Autorin gewählte Darstellungsweise ist denkbar schlecht, da sie platzraubend, schreibintensiv und redundant ist. So werden für jede Sprache die Distanzen zu jeder anderen Sprache umständlich aufgelistet, wo man doch mit einer 7×7 -Matrix ausgekommen wäre. Die Redundanz der Darstellung wird deutlich, wenn man sich überlegt, daß die Informationen, die die Angaben zur letzten Sprache (dem Spanischen) liefern, sämtlich Wiederholungen sind – die Distanz des Spanischen zum Französischen findet sich schon beim Französischen, die Distanz zum Italienischen schon beim Italienischen usw. Die Auswertung der Werte des genannten Quotienten geschieht dergestalt, daß zu jeder Sprache die übrigen Sprachen in eine Rangordnung gebracht werden. Die Ergebnisse werden dann noch kommentierend interpretiert. Im Prinzip werden alle Sprachvergleiche in diesem Buch auf die angedeutete Art und Weise durchgeführt (krasses Beispiel: «Beim Vergleich der einzelnen Sprachen miteinander überwiegt immer das, was die Sprachen voneinander trennt, vor dem, was sie verbindet», 299).

Kommen wir zu dem eigentlich quantitativen Teil der Arbeit, zur Phonemstatistik. Hier werden in einem ersten Abschnitt die *Phonemfrequenzen* behandelt. Zita Hess bestimmt diese auf der Grundlage einer pragmatischen Stichprobe von jeweils ca. 20 000 Phonemen. Dabei ist sie sich der Problematik des 'zufälligen' Auswählens der zugrundezulegenden Texte bewußt. Auch den Einfluß, den die Tatsache haben kann, daß die gewählten Texte jeweils nur von *einem* native speaker gelesen wurden, bedenkt die Autorin. So findet man dankenswerterweise im Anschluß an

alle Angaben 'Sonderauswertungen', aus denen die Veränderungen ersichtlich werden, die bei einer veränderten phonologischen Interpretation vorgenommen werden müßten. Störend wirken in den einleitenden Bemerkungen Ausdrücke wie «tables de 'nombres au hasard'» [189], «standard error», «error type», «sample», «statistical universe», «standard error of difference» [alle 200]. Wird im deutschsprachigen Raum keine Statistik betrieben?

Vor allem unter drei Gesichtspunkten will Zita Hess die Frequenzwerte interpretieren: (1) Der Anteil der Konsonanten am Inventar (p_i) soll in Beziehung gesetzt werden zu dem entsprechenden Anteil an der Stichprobe (p_t). Dazu übernimmt sie den Quotienten p_i/p_t von Krámský (1959). Dies kann – schon bei Krámský – nicht als besonders glückliche Lösung bezeichnet werden, denn jeder Zahlenwert muß interpretierbar sein (vgl. auch Altmann/Lehfeldt, 1973, 81 ff.). Dazu gehört zunächst, daß das Intervall bekannt ist, in das die Werte des betreffenden Quotienten fallen können. Für vergleichende Zwecke, dies sei noch einmal gesagt, bildet man am besten immer einen solchen Index, dessen Werte in dem Einheitsintervall $< 0; 1 >$ liegen. Der Wertebereich des genannten Indexes ist aber nach oben hin offen. Ist der Wert größer als 1, so interpretiert Zita Hess dies im Anschluß an Krámský als eine Unterbelastung der Konsonanten (bzw. als eine Überbelastung der Vokale). Hier aber drängt sich sofort die Frage auf, um wie viel größer als 1 der Wert denn sein müsse, damit man das Ergebnis für bedeutsam, die Abweichung für 'signifikant' halten könne. Diese Frage ist theoretisch wichtig, denn jeder Zahlenwert sollte linguistisch-qualitativ interpretiert werden können; sie ist aber auch in diesem konkreten Falle von praktischer Wichtigkeit: Ist eine relevante Abweichung schon bei 1.005 (im Französischen) oder erst bei 1.08 (im Portugiesischen) gegeben? Ganz sicher aber doch bei 1.31 (im Katalanischen)? Es ist klar, daß nur ein statistischer Test diese Frage befriedigend, d. h. nichtsubjektiv, beantworten kann; er hat hier die folgende Form: Die Annahme (Nullhypothese), daß zwischen dem Anteil \hat{p} der Konsonanten an der Stichprobe des Umfangs N und ihrem Anteil p am Inventar *kein* signifikanter Unterschied besteht, wird bei zweiseitiger Fragestellung auf dem $\alpha = 0.05$ - oder 5%-Signifikanzniveau *verworfen*, wenn

$$z = \frac{|\hat{p} - p|}{\sqrt{\frac{p \cdot q}{N}}} \geq 1.96$$

(vgl. Kriz, 1973, 162 ff.; Siegel, 1976, 40). So erhält man z. B. im Französischen

$$z = \frac{\left| \frac{11639}{20610} - \frac{21}{37} \right|}{\sqrt{\frac{\frac{21}{37} \left(1 - \frac{21}{37}\right)}{20610}}} = 0.82347$$

Der Anteil der Konsonanten an der Stichprobe des Französischen kann also als nicht signifikant verschieden von dem Anteil der Konsonanten am Inventar betrachtet werden. In allen anderen Sprachen ist die Differenz signifikant. Zu beachten ist bei diesem Test, für den der Quotient p_1/p_t gar nicht benötigt wird, daß die 'erlaubte' Schwankung des Anteils an der Stichprobe im Vergleich zum Anteil am Phoneminventar *kleiner* wird, wenn der Stichprobenumfang erhöht wird. Es steht zu vermuten, daß in jeder Sprache eine signifikante Differenz ermittelt werden kann, wenn nur die Stichprobe groß genug gewählt wird – was aus anderen Gründen wünschenswert ist. So fragt es sich denn, ob diese Fragestellung überhaupt ergiebig ist, da sie es selten gestattet, die Sprachen zu differenzieren.

(2) Um die 'Kompaktheit' des jeweiligen Phonemsystems zu kennzeichnen, ermittelt Zita Hess die Spannweite (Frequenzdifferenz) zwischen dem häufigsten und dem seltensten Phonem. Dies ist eine von vielen Möglichkeiten, eine Frequenzverteilung durch eine Maßzahl zu kennzeichnen. Generell sollte man aber nie mit Prozentzahlen arbeiten, sondern mit relativen Häufigkeiten, denn die Zahl der Stellen, die bei Prozentzahlen üblicherweise nach dem Komma noch angegeben werden (2 oder 3), ist für eine genaue Auswertung einfach zu niedrig (vgl. unten).

(3) Problematischer ist der nächste Gesichtspunkt. «Ein phonologisches System wäre vollkommen 'harmonisch', wenn alle Phoneme gleichmäßig stark belastet wären» [201], was die Verf. durchgängig auch den 'Idealzustand' nennt. Bevor man eine solche Wertung vornimmt, sollte klargestellt werden, worum es sich rein numerisch-rechnerisch handelt: um die hypothetische Gleichverteilung der Phoneme, die dann gegeben ist, wenn jedes Phonem die relative Häufigkeit $1/K$ aufweist, wobei K für die Zahl der Phoneme einer Sprache steht. 'Ideal' kann ein Wert nur in Zusammenhang mit bestimmten anderen Kriterien genannt werden, es ist dies also ein relationaler Begriff. Zita Hess benutzt offensichtlich ein ästhetisches ('Harmonie'-)Kriterium; nun ist dieses genau so 'richtig' oder 'falsch' wie jedes andere; in einer informationstheoretischen Interpretation ist aber die Gleichverteilung, d. h. die Gleichwahrscheinlichkeit aller Phoneme, durchaus kein Idealzustand. Warum nicht? Vor die Aufgabe gestellt, das nächste Phonem einer Sequenz zu erraten, ohne Kenntnis der Sprachstruktur zu haben, liegt man in einem solchen Falle auf lange Sicht gleich gut oder schlecht, welches Phonem man auch wählen mag. Die a

priori-Unsicherheit ist, mit anderen Worten, im Fall der Gleichwahrscheinlichkeit der Phoneme am größten. Jede Sprache weicht von diesem theoretischen Zustand mehr oder weniger stark ab. Die sogenannte *relative Entropie* H_{rel} eines Phonemsystems ist nichts anderes als ein Maß für den *Grad der Annäherung* der tatsächlichen Verteilung an die theoretische. Sie berechnet sich als

$$H_{\text{rel}} = \frac{- \sum_i \hat{p}_i \text{ld } \hat{p}_i}{\text{ld } K}$$

Mit 0.90616 hat das Französische hier den höchsten Wert, das Katalanische mit 0.84663 den niedrigsten.

Wenn man die Frequenzangaben der Phoneme hat, so ist es naheliegend zu fragen, welche Phoneme sich hinsichtlich ihrer Frequenz signifikant von dem Erwartungswert ihrer Frequenz, der Gleichverteilung, unterscheiden. Der Erwartungswert der absoluten Häufigkeit ist definiert als N/K . Zita Hess arbeitet mit dem «standard error of difference», um die Abweichung des beobachteten Wertes vom 'Idealwert' auf Signifikanz zu testen, und teilt die Phoneme in solche ein, deren Frequenz nicht signifikant kleiner ist als der Erwartungswert, und in solche, für die eben dies gilt. Einfacher zu handhaben als die gewählte Teststatistik ist die folgende Prüfgröße:

$$X^2 = \frac{(n_i - E_i)^2}{E_i} = \frac{(n_i - \frac{N}{K})^2}{\frac{N}{K}}$$

wobei n_i für die absolute Häufigkeit des Phonems i steht. Dieser Test wurde schon früh eben auf diese Fragestellung angewendet (vgl. Altmann, 1967). Die Differenz ist signifikant, wenn $X^2 \geq 3.84$. So ergibt sich z. B. im Französischen, daß das Phonem $/y/$ mit seiner Frequenz von 496 vom Erwartungswert $20610/37 = 557.02703$ signifikant abweicht, da

$$X^2 = \frac{(496 - 557.02703)^2}{557.02703} = 6.68603$$

Insgesamt ergibt sich, daß 14 Phoneme stark belastet sind, 1 Phonem normal belastet ist und 22 schwach belastet sind. Will man diese Information vollständig auswerten, so kann man das Französische durch den folgenden Vektor kennzeichnen, der auch zum Sprachvergleich geeignet ist:

$$FB_{\text{Franz.}} = 1/37 [14, 1, 22]$$

In einem weiteren großen Abschnitt beschäftigt sich Zita Hess mit der *Silbenstruktur* der romanischen Sprachen. Vor der Untersuchung der Silbenstruktur steht natürlich die Frage der Bestimmung der Silbengrenzen; hier vertritt Zita Hess (zu Recht) einen phonologischen, keinen phonetischen Standpunkt. Es ist aber bedauerlich, daß sie die neuere Literatur zu diesem Thema nicht verarbeitet hat; wir meinen hier insbesondere Pulgram (1970) und Lehfeldt (1971). Pulgram hat 5 Regeln ausgearbeitet, nach denen generell die Silbentrennung erfolgen kann, die notwendige methodisch korrekte Berücksichtigung der Frequenzen der Phonemverbindungen, die auch Zita Hess fordert, und eine Präzisierung verschiedener wichtiger Begriffe wurde von Lehfeldt geleistet. So ist zwar der Wille zu begrüßen, auch bei der Silbentrennung statistisch zu arbeiten, doch was die Autorin dazu im einzelnen ausführt, taugt leider nicht viel. Das von ihr verwendete Prinzip der Silbentrennung läßt sich folgendermaßen verdeutlichen: Wenn z. B. im Französischen in einer Stichprobe der Auslauttyp $-VC$ 186mal am Wortende, der Anlauttyp $CV-$ 252mal am Wortanfang vorkommt und eine Sequenz $VCCV$ auf zwei Silben zu verteilen ist, so argumentiert sie: «bei einer Silbentrennung in $VC + CV$ läge die relative Wahrscheinlichkeit bei $186 + 252 = 438 \dots$ » [241] und zieht diese Trennung der Segmentierung $VCC + V$ vor, da sich hier aus der Frequenz des Auslaut- und des Anlauttyps nur eine Summe von 66 ergibt. An anderen Stellen spricht sie im Zusammenhang mit solchen Zahlen auch von «statistischer Wahrscheinlichkeit» [z. B. 241]. Dazu ist zunächst zu sagen, daß Wahrscheinlichkeiten per definitionem immer im Intervall $< 0; 1 >$ liegen. Ein Manipulieren mit absoluten Zahlen hat mit dem Feststellen von Wahrscheinlichkeiten so gut wie nichts zu tun. Wie schon angedeutet, wird diejenige Silbentrennung gewählt, bei der die *Summe* der absoluten Frequenzen des resultierenden Codatyps und des resultierenden Onsettyps, bezogen auf ihr Auftreten an Wortgrenzen, am größten ist. Die gemeinsame Wahrscheinlichkeit unabhängiger Ereignisse wird aber durch Multiplikation berechnet, z. B. die «Wahrscheinlichkeit, beim ersten *und* beim zweiten Wurf eine 6 zu erhalten», als $1/6$ mal $1/6$ gleich $1/36$. So ist die «Verfeinerung», die Zita Hess [248] vorschlägt, keine solche, denn absolute Zahlen besagen so gut wie nichts, da der stets notwendige Bezugspunkt fehlt. Auch ist ein Verfahren, das die Frequenz jeder einzelnen Konsonantenkombination berücksichtigt, *nicht* «zu schwerfällig» [248], wie H. meint, wenn man freilich die entsprechenden Hilfsmittel zur Verfügung hat, ohne die man heute eine solche Untersuchung sinnvoll kaum noch durchführen kann. Alle Untersuchungsergebnisse, auf die die Silbentrennung Einfluß hat, sind deshalb als vorläufig zu betrachten.

Im einzelnen nennt Zita Hess die absolute Frequenz der verschiedenen Silbentypen, woraus sich leicht weitere Kennzahlen berechnen lassen, so z. B. das Verhältnis von offenen zu geschlossenen Silben etc. Auch hier sind

die 'Sonderauswertungen' zu begrüßen. An dieser Stelle sind einige Fehler zu korrigieren. Das Verhältnis von offenen zu geschlossenen Silben beträgt im Rumänischen nicht 6585:2844, wie auf S. 269 angegeben, sondern 6651:2778; die Differenz entspricht genau der Frequenz des Silbentyps CCCV. In der Stichprobe des Rumänischen gibt es auch nicht 7525, sondern 5725 Silben mit 2 Phonemen [269]. Ganz allgemein gilt: Wer in einer Untersuchung Zahlen veröffentlicht, kann sich oft mehrfach der Korrektheit der Ergebnisse vergewissern. Ein Beispiel: Zita Hess gibt für jede Sprache an, mit welcher Frequenz ein bestimmter struktureller Silbentyp, etwa CVC, auftritt. Daraus kann man leicht eine Verteilung berechnen, die angibt, wie oft eine Silbe bestimmter Länge auftritt (bei Zita Hess sind diese Angaben explizit nur unvollständig gegeben). So erhält man beispielsweise für das Katalanische die folgenden Angaben:

Silbenlänge x_1	Häufigkeit f_1	Produkt $f_1 x_1$
1	776	776
2	5024	10048
3	2787	8361
4	458	1832
5	20	100
	9065	21117

Die Zahl 9065 wird als Gesamtzahl der Silben auch auf S. 261 genannt. Wenn man jede Silbenlänge x_1 mit ihrer Häufigkeit f_1 multipliziert und die Summe dieser Produkte bildet, erhält man offensichtlich die Gesamtzahl aller Phoneme der Stichprobe. Diese Zahl von 21117 Phonemen, die dem auf S. 218 angegebenen Stichprobenumfang entspricht, bekommt man aber nicht, wenn man die absoluten Phonemfrequenzen summiert. Hier ergibt sich 21118, darunter 9066 Vokale. Wenn jeder Vokal eine Silbe bildet und jede Silbe genau einen Vokal enthält, sollte also die Zahl der Vokale in der Stichprobe (9066) gleich der Zahl der Silben (9065) sein. Nun wird zwar auf S. 261 die Frequenz von CVCC mit 282 angegeben, auf S. 262 f. aber mit 283 gerechnet, doch erklärt dies nicht die genannte Differenz, denn es gäbe ja dann nicht nur einen Vokal, sondern eben auch 3 Konsonanten mehr. Hier helfen auch die Angaben der prozentualen Häufigkeit nicht weiter, denn es ist nicht eindeutig auszumachen, ob Zita Hess als Gesamtzahl 21117 oder 21118 ihren Berechnungen zugrunde gelegt hat, und außerdem ist nicht ganz ersichtlich, nach welchem Verfahren sie gerundet hat. So ist z.B. die Frequenz des Phonems /a/ im Katalanischen $5082/21117 = 0.2406592$ bzw. 24.066%. Die Verf. schreibt 24.065%. Andererseits gibt sie für /e/ statt $505/21117 = 0.0239144$ bzw. 2.391% hier 2.392% an. Diese Ungenauigkeiten in den Prozentzah-

len sind hier relativ unwichtig, da die absoluten Frequenzen ja angegeben werden; in vielen anderen Arbeiten findet man aber eben nur Prozentzahlen. Ähnliche Abweichungen zwischen Silbenzahl und Summe der Vokalfrequenzen ergeben sich für das Okzitanische (7058:7059) – ohne daß sich hier jedoch die Gesamtzahl der Phonemokkurrenzen ändert – sowie für das Rumänische (9429:9433). Da Zita Hess auch für das Verhältnis von Vokalen zu Konsonanten nur die ungenauen Prozentzahlen angibt, ist nicht auszumachen, welche absoluten Zahlen ihren Berechnungen zugrundeliegen.

Ein letzter großer Abschnitt ist der Untersuchung der *Konsonantenkombination* («asilbische Phonemkombinationen») im Rahmen der Silbe gewidmet. Das Material besteht hier aus Frequenzangaben zu jeder in der Stichprobe vorgefundenen Konsonantenkombination. Die Fragestellungen, die anhand dieses Materials überprüft werden sollen, sind – mit Verlaub – dürftig [vgl. 276]. Bei Kenntnis der entsprechenden Literatur hätte eine bessere Auswertung der Daten vorgenommen werden können. Der Aufsatz von Altmann (1973 b), an den wir hier denken, ist zugegebenermaßen an relativ entlegener Stelle publiziert, doch finden sich Hinweise auf ihn auch an anderer Stelle. Auch die Präsentation der Daten hätte verbessert werden können. So ist eine Wiedergabe der Frequenzen in Form einer Matrix in jedem Falle der einfachen Auflistung vorzuziehen, da sie übersichtlicher ist und zugleich mehr Information vermittelt, indem sie auf einen Blick darüber Auskunft gibt, auf welche Phoneme sich die auftretenden Phonemverbindungen konzentrieren. Weiterhin sollten zwei- und dreigliedrige Cluster stets getrennt aufgeführt und gezählt werden.

Den Abschluß des Buches bildet ein Vergleich der gewonnenen Einzelergebnisse mit traditionellen Klassifikationen der romanischen Sprachen, der ein allgemeines Bild der synchronen Einteilung im Verhältnis zur diachronen nur schlecht vermitteln kann.

Zusammenfassend kann man sagen, daß die vorliegende Arbeit gewissenhaft und umsichtig bei der Begründung der verwendeten Phoneminventare und sorgfältig bei der Darstellung der Ausgangsdaten durchgeführt worden ist, wenn auch die Darstellungsweise nicht in allen Punkten befriedigen kann. Eine bessere Literaturkenntnis hätte Mängel bei der Silbentrennung vermeiden helfen und eine bessere Materialauswertung befördert. So trägt die Auswertung über weite Stellen starken 'ad-hoc-Charakter'. Hier zeigen sich einmal mehr die starken Wechselbeziehungen, die zwischen Methodenkenntnis und linguistischen Fragestellungen bestehen.

2. Der quantitative Vergleich einer Gruppe von Sprachen, so wie wir ihn im Anschluß an die Verfahrensweisen der numerischen Taxonomie und an Altmann, Lehfeldt (1973) demonstrieren wollen, durchläuft folgende allgemeine Schritte:

- (1) Aufstellen linguistischer Hypothesen, die überprüft werden sollen;
- (2) Gewinnung des Datenmaterials (Stichprobenerhebung);
- (3) Bearbeitung des Materials (Berechnung von Indizes);
- (4) Zusammenstellen von Vektoren;
- (5) Berechnung von Distanzen bzw. Ähnlichkeiten;
- (6) Klassifizierung;
- (7) Interpretation der Klassifikation;
- (8) Entscheidung über die linguistischen Hypothesen.

Damit sind die Ziele einer Sprachtypologie nicht erschöpft, insbesondere lassen wir hier die Untersuchung des gegenseitigen Zusammenhanges der Sprachmerkmale beiseite, doch verläuft jedes Vergleichen von Sprachen auf quantitativer Grundlage im Prinzip auf die angegebene Weise. Im folgenden werden wir uns mit den Schritten (2) bis (7) begnügen, da es uns hier lediglich um das Demonstrieren der *Methode* geht. Zuvor zwei Bemerkungen. Für das Unterfangen eines quantitativen Sprachvergleichs, wie er im folgenden durchgeführt wird, benötigt man zumindest einen Taschenrechner nicht der einfachsten Sorte und eine ziemlich große Gleichmut stumpfsinniger Rechnerei gegenüber, oder besser, Zugang zu einer Großrechenanlage (wie er für Zita Hess nicht gegeben war, wie sie selbst anmerkt). In bestimmten Bereichen erscheint der Einsatz eines Großrechners sogar unumgänglich, besonders wenn es um die Manipulation der ursprünglichen Daten geht. Wir behaupten weder, das vorliegende Material vollständig und in der einzig möglichen Weise ausgeschöpft zu haben, noch, daß es Zita Hess möglich gewesen wäre, eine Bearbeitung des Materials auch nur in dem hier vorgeführten Umfang vorzunehmen, einfach deshalb, weil dazu bestimmte technische Voraussetzungen gegeben sein müssen und weil wir uns zu einem geringen Teil auf Quellen stützen, die während der Abfassung der rezensierten Arbeit noch nicht veröffentlicht waren.

Der Deutlichkeit halber werden wir die genannten Schritte anhand unseres Beispiels einzeln besprechen.

(2) Gewinnung des Materials. Rufen wir uns ins Gedächtnis zurück, welche Angaben uns aus der Arbeit von Zita Hess vorgegeben sind. Es sind dies zunächst natürlich der Umfang des Phoneminventars jeder Sprache, dann der Umfang der in jeder Sprache erhobenen Stichprobe. Aus diesen Stichproben haben wir im wesentlichen drei Gruppen von Daten: (a) die absoluten Frequenzen eines jeden Phonems, (b) die absoluten Frequenzen aller belegten Silbenstrukturtypen und (c) die absoluten Frequenzen der (zweigliedrigen) Konsonantenkombinationen im Rahmen der Silbe. Was macht man nun mit diesen Daten?

(3) Bearbeitung des Materials. An dieser Stelle wollen wir auf Einzelheiten der Berechnung nicht eingehen, sondern die verschiedenen Indizes nur verbal beschreiben. Sie werden der Eindeutigkeit halber durchnumeriert. Zunächst kennzeichnen wir jede Sprache durch den relativen Anteil ihrer Vokale am Inventar (1) und an der Stichprobe (2). Sodann

wird für jede Sprache die relative Entropie (Index 3, vgl. oben) sowie die (relativierte) Wiederholungsrate (4) ermittelt, die nichtrelativiert als die Summe der Quadrate der relativen Phonemhäufigkeiten definiert ist. Hinzu nehmen wir die Spannweite zwischen dem häufigsten und dem seltensten Phonem (5); nach Altmann (1969) berechnen wir die durchschnittliche Frequenzdifferenz der Phoneme (6). Mithilfe des Chi-quadrattests (vgl. Altmann, 1967) ermitteln wir, wie oben angedeutet, die relative Zahl der stark (7), der normal (8) und der schwach belasteten Phoneme (9). Soweit die Auswertung der Phonemfrequenzen.

Bei den Silben stellen wir zunächst den relativen Anteil der Silben der Phonemlänge 1 bis 6 an der Stichprobe fest (10 bis 15) und bestimmen die durchschnittliche Silbenlänge (16). Nach Altmann/Lehfeldt (1973, 84 ff.) stellen wir weiterhin den relativen Anteil der mit einem Vokal beginnenden Silben an den Silbentypen (17) und an den Silben der Stichprobe (18) fest. Analog verfahren wir mit dem relativen Anteil der offenen Silben (19 und 20). Die gleiche Messung wiederholen wir auf der Ebene der lexikalischen Einheiten, indem wir den relativen Anteil der mit einem Vokal beginnenden (21) bzw. endenden (22) Wörter ermitteln.

Schließlich wenden wir uns den Konsonantenkombinationen zu. Hier berechnen wir nach Altmann (1973 b) die relative Zahl der bevorzugten (P), der aktualen (A), der marginalen (M), der virtuellen (V) und der 'verbotenen' (I) Cluster. Die zugehörigen Indizes tragen die Nummern 23 bis 27. Für die Berechnung der folgenden Werte abstrahieren wir von der Frequenz der Verbindungen und berücksichtigen nur ihre Existenz (+) bzw. ihr Fehlen (—). Nach Altmann/Lehfeldt (1972) bestimmen wir die relative Anzahl der attraktiven (T), der semiattraktiven (ST) und der nichtattraktiven Konsonanten (NT) sowie die entsprechenden Werte für die Grade der Aggressivität (G, SG, NG); die Indizes haben die Nummern 28 bis 33. Durch Kombination beider Kriterien entsteht eine Kreuzklassifikation der Konsonanten. Entsprechend der Anzahl der Klassen – neun – ermitteln wir neun Indexwerte (34 bis 42). Nach Altmann (1973 a) fügen wir die Anteile der qualitativ-symmetrischen (S), der qualitativ-semisymmetrischen (SS) und der qualitativ-asymmetrischen Konsonanten (AS) sowie die Anteile der quantitativ-symmetrischen (QS) und der quantitativ-asymmetrischen (QAS) Konsonanten hinzu und tragen die Werte unter den Indizes 43 bis 47 ein. Nach Altmann/Lehfeldt (1977) schließlich berechnen wir die Werte von sechs distributionellen Typologismaßen (48 bis 53) und ermitteln fünf Werte der durchschnittlichen distributionellen Ähnlichkeit (54 bis 58). Damit ist die quantitative Beschreibung der sieben romanischen Sprachen abgeschlossen, und wir kommen zu dem

(4) Zusammenstellen der Vektoren. Jede Sprache ist jetzt durch 58 Zahlenwerte charakterisiert; diese Zahlenwerte setzen wir in der genannten Reihenfolge zu einem Vektor aus 58 Elementen zusammen. Hier muß

Tabelle I: Vektorwerte

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
FRANZOES	0.43243	0.43527	0.90616	0.01837	0.07986	0.01255	0.37838	0.02703	0.59459	0.07045	0.61186	0.26954	0.04615	0.0020
KZITANI	0.25000	0.41844	0.89789	0.02477	0.11031	0.01643	0.42857	0.07143	0.50000	0.06163	0.54236	0.34330	0.04959	0.0031
PORTUGIE	0.40000	0.44420	0.90603	0.02058	0.09162	0.01325	0.37143	0.02857	0.60000	0.14740	0.51069	0.28804	0.05109	0.0026
ITALIENI	0.22581	0.43157	0.86672	0.03199	0.11752	0.01740	0.35484	0.06452	0.58065	0.04767	0.62936	0.28304	0.03805	0.0018
ATALANI	0.25000	0.42930	0.84663	0.05808	0.23790	0.01925	0.35714	0.03571	0.60714	0.08560	0.55422	0.30745	0.05052	0.0022
UMAENIS	0.21212	0.43207	0.88167	0.02637	0.10503	0.01548	0.42424	0.00000	0.57576	0.07212	0.60717	0.26175	0.05144	0.0072
PANISCH	0.20833	0.46719	0.86960	0.03615	0.13257	0.02148	0.41667	0.00000	0.58333	0.08724	0.58019	0.28501	0.04624	0.0013
	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(23)	(24)	(25)	(26)	(27)	(28)	(29)
FRANZOES	0.00000	0.69673	0.25000	0.08483	0.33333	0.77561	0.25203	0.57491	0.08390	0.06576	0.02494	0.77531	0.04989	0.0000
KZITANI	0.00000	0.70503	0.20000	0.09323	0.40000	0.63616	0.21771	0.47675	0.08163	0.04762	0.00907	0.80499	0.05669	0.0000
PORTUGIE	0.00011	0.69364	0.30769	0.22908	0.30769	0.62828	0.27526	0.44474	0.08844	0.04535	0.02041	0.77098	0.07483	0.0000
ITALIENI	0.00000	0.69853	0.22222	0.12134	0.44444	0.68289	0.19867	0.97153	0.04861	0.02083	0.01563	0.86979	0.04514	0.0000
ATALANI	0.00000	0.69967	0.30000	0.13999	0.30000	0.63409	0.29058	0.46930	0.07029	0.02041	0.01134	0.84127	0.05669	0.0000
UMAENIS	0.00032	0.69838	0.26667	0.12918	0.33333	0.70538	0.24926	0.58260	0.07101	0.04290	0.01183	0.82396	0.05030	0.0000
PANISCH	0.00000	0.69644	0.33333	0.13545	0.40000	0.69531	0.27574	0.55614	0.01939	0.01385	0.00554	0.94460	0.01662	0.0000
	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)
FRANZOES	0.23810	0.76190	0.00000	0.19048	0.80952	0.00000	0.00000	0.00000	0.00000	0.04762	0.14286	0.00000	0.19048	0.6190
KZITANI	0.23810	0.76190	0.00000	0.14286	0.85714	0.00000	0.00000	0.00000	0.00000	0.04762	0.09524	0.00000	0.19048	0.6666
PORTUGIE	0.23810	0.76190	0.00000	0.23810	0.76190	0.00000	0.00000	0.00000	0.00000	0.04762	0.19048	0.00000	0.19048	0.5714
ITALIENI	0.12500	0.87500	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.12500	0.8750
ATALANI	0.14286	0.85714	0.00000	0.04762	0.95238	0.00000	0.00000	0.00000	0.00000	0.04762	0.00000	0.00000	0.09524	0.8571
UMAENIS	0.11538	0.88462	0.00000	0.15385	0.84615	0.00000	0.00000	0.00000	0.00000	0.03846	0.11538	0.00000	0.07692	0.7692
PANISCH	0.10526	0.89474	0.00000	0.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.10526	0.8947
	(43)	(44)	(45)	(46)	(47)	(48)	(49)	(50)	(51)	(52)	(53)	(54)	(55)	(56)
FRANZOES	0.00000	1.00000	0.00000	0.85714	0.14286	0.17460	0.02721	0.08451	0.15584	0.29478	0.00000	0.16752	0.25837	0.0075
KZITANI	0.04762	0.95238	0.00000	0.90476	0.09524	0.13832	0.04082	0.17308	0.29508	0.19501	0.00000	0.05943	0.18272	0.0147
PORTUGIE	0.00000	1.00000	0.00000	0.90476	0.09524	0.15420	0.04535	0.17241	0.29412	0.21769	0.00000	0.08098	0.13796	0.0306
ITALIENI	0.00000	1.00000	0.00000	0.91667	0.08333	0.08507	0.00347	0.02083	0.04082	0.16319	0.00000	0.05730	0.21454	0.0000
ATALANI	0.04762	0.95238	0.00000	0.90476	0.09524	0.09751	0.02268	0.13158	0.23256	0.14966	0.00000	0.03393	0.14985	0.0134
UMAENIS	0.00000	1.00000	0.00000	0.88462	0.11538	0.12574	0.02071	0.08974	0.16471	0.21006	0.00000	0.09287	0.12439	0.0066
PANISCH	0.00000	1.00000	0.00000	0.94737	0.05263	0.03878	0.00000	0.00000	0.00000	0.07756	0.00000	0.00418	0.09357	0.0000
	(57)	(58)												
FRANZOES	0.25003	0.02476												
KZITANI	0.14559	0.01976												
PORTUGIE	0.12191	0.02675												
ITALIENI	0.15697	0.00000												

man nur darauf achten, daß die einmal gewählte Reihenfolge für alle Sprachen beibehalten wird, so daß vergleichbare Zahlenwerte untereinander stehen. Tabelle I zeigt alle ermittelten Informationen. Um die Darstellung an dieser Stelle nicht zu kompliziert werden zu lassen, berücksichtigen wir zwei Probleme nicht. Erstens dürfen eigentlich nur logisch voneinander unabhängige Merkmale benutzt werden. Da es jedoch in diesem Falle auf die Ergebnisse der Klassifikation (s. u.) praktisch keinen Einfluß hat, wenn die logisch abhängigen und bedeutungslosen Merkmale (Null für alle Sprachen) gestrichen werden, gehen wir hier auf dieses Problem nicht weiter ein. Das zweite Problem betrifft die mathematische Transformation der Meßwerte: nicht alle Werte können in dem Intervall $< 0; 1 >$ gleichermaßen variieren. Dies auszugleichen gibt es verschiedene Methoden. Auch nach einem 'ranging' der ursprünglichen Meßwerte ändert sich die Klassifikation der Sprachen prinzipiell nicht. (Zu beiden Problemen vgl. die entsprechenden Abschnitte bei Sneath/Sokal, 1973).

(5) Berechnung der Distanz. Um die Sprachen hinsichtlich aller Eigenschaftsausprägungen zu vergleichen, wählen wir ein Distanzmaß, in diesem Falle die *Euklidische Distanz*:

$$D(A, B) = \left[\sum_{i=1}^n (a_i - b_i)^2 \right]^{1/2}$$

Ein jeder Vektor stellt die Koordination einer Sprache in einem euklidischen Raum mit (hier) $n = 58$ Dimensionen dar, und die euklidische Distanz gibt an, wie weit zwei Sprachen A und B in diesem abstrakten Raum voneinander 'entfernt' sind. Die praktische Berechnung ist einfach: Von zwei untereinanderstehenden Vektorwerten a_i und b_i bildet man die Differenz, quadriert sie und bildet die Summe aller dieser quadrierten Differenzen für die 58 Zahlenpaare, aus der zum Schluß die Wurzel gezogen wird. Auf diese Weise erhalten wir bei 7 Sprachen $7(6)/2 = 21$ verschiedene Distanzwerte, erreichen also eine weitgehende Reduzierung des ursprünglichen Datenmaterials. Die Matrix der Tabelle II zeigt diese Distanzwerte.

Es sei angemerkt, daß bis zu diesem Schritt schon Muljačić (1963; 1967) gekommen ist, wenn auch mit einfacheren, nämlich binären Daten und einer entsprechend einfacheren Distanzberechnung.

(6) Klassifizierung. Um aus vorliegenden Distanz- bzw. Ähnlichkeitsmatrizes eine (hierarchische) Klassifikation der Objekte zu gewinnen, gibt es eine ganze Reihe von Verfahren; eine deutschsprachige Übersicht und Einschätzung gibt z. B. Späth (1975). Wir wählen hier das Cluster-Verfahren von Johnson (1967) in seiner Maximum-Variante, weil uns dafür ein komfortables Computerprogramm zur Verfügung steht. Das Vorgehen ist im Prinzip einfach: Man sucht die kleinste Zahl der Matrix (wobei natürlich die Diagonale unberücksichtigt bleibt), hier die Distanz

zwischen dem Rumänischen und dem Katalanischen. Beide Sprachen werden zu einer Gruppe vereinigt, und eine Sprache wird als Vertreter der Gruppe beibehalten (beispielsweise das Katalanische). Bevor nun das Rumänische aus der Matrix gestrichen wird, werden die Distanzen beider Sprachen zu den anderen Sprachen verglichen; falls eine Distanz der beibehaltenen Sprache niedriger ist als die entsprechende Distanz der zu streichenden Sprache, so wird der (niedrigere) Wert der beibehaltenen Sprache durch den (höheren) Wert der zu streichenden Sprache ersetzt. Daraufhin beginnt man wieder mit der Suche nach der kleinsten Zahl etc. und fährt solange fort, bis alle Sprachen vereinigt sind. Eine graphische Darstellung des Ergebnisses zeigt Abbildung 1. Es verbleibt die

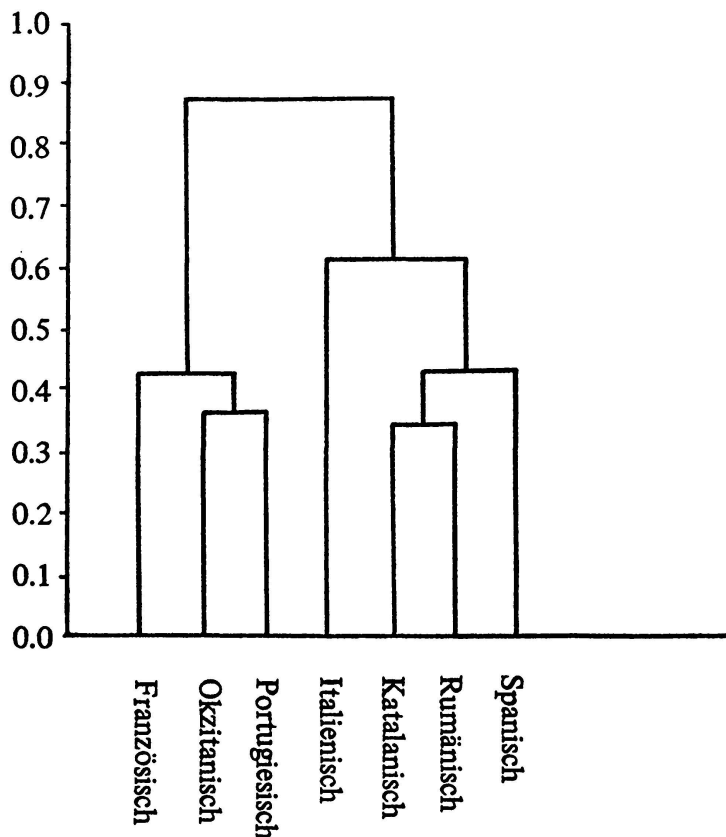


Abbildung 1: Hierarchische Klassifikation 7 romanischer Sprachen

(7) Interpretation der Klassifikation. Sie besteht darin, zu den ursprünglichen Vektor-Werten zurückzugehen und zu schauen, welche Eigenschaftsausprägungen die Sprachen der einzelnen Klassen charakterisieren. Zu diesem Zweck wird ein klassifikatorischer Schlüssel konstruiert, der die Menge der Sprachen fortlaufend durch Abfragen bestimmter Bedingungen zerlegt und somit schließlich jede Sprache singular kennzeichnet. Dieser Algorithmus hat hier folgende Form:

1. Wenn der Anteil der nichtaggressiven Phoneme > 0.8 ist, gehe zu 2., sonst gehe zu 9.
2. Wenn der Anteil der mit einem Vokal endenden lexikalischen Einheiten > 0.9 ist, gehe zu 3., sonst gehe zu 4.
3. Es handelt sich um das Italienische.
4. Wenn die durchschnittliche Frequenzdifferenz > 0.02 ist, gehe zu 5., sonst gehe zu 6.
5. Es handelt sich um das Spanische.
6. Wenn der Anteil der Silben der Länge 6 > 0 ist, gehe zu 7., sonst gehe zu 8.
7. Es handelt sich um das Rumänische.
8. Es handelt sich um das Katalanische.
9. Wenn die durchschnittliche attraktive distributionelle Ähnlichkeit > 0.1 ist, gehe zu 10., sonst gehe zu 11.
10. Es handelt sich um das Französische.
11. Wenn der Anteil der Silben der Länge 1 > 0.1 ist, gehe zu 12., sonst gehe zu 13.
12. Es handelt sich um das Portugiesische.
13. Es handelt sich um das Okzitanische.

Den 'Typ' einer jeden Sprache können wir kennzeichnen, indem wir jeder Sprache die Eigenschaftsausprägungen zuschreiben, die ihr nach diesem klassifikatorischen Schlüssel zukommen (dabei benutzen wir Abkürzungen, die aus dem vorher Gesagten leicht verständlich sein dürften):

Rumänisch:	$ NT > 0.8$	$-V\# < 0.9$	$\bar{D}_{tr} < 0.02$	$SI_6 > 0$.
Katalanisch	$ NT > 0.8$	$-V\# < 0.9$	$\bar{D}_{tr} < 0.02$	$SI_6 = 0$.
Spanisch:	$ NT > 0.8$	$-V\# < 0.9$	$\bar{D}_{tr} > 0.02$.	
Italienisch:	$ NT > 0.8$	$-V\# > 0.9$.		
Portugiesisch:	$ NT < 0.8$	$\bar{S}_{at} < 0.1$	$SI_1 > 0.1$.	
Okzitanisch:	$ NT < 0.8$	$\bar{S}_{at} < 0.1$	$SI_1 < 0.1$.	
Französisch:	$ NT < 0.8$	$\bar{S}_{at} > 0.1$.		

Schließlich soll noch die Klassifikation vorgeführt werden, die sich auf der Grundlage der Daten von Muljačić (1967) ergibt. Dafür benutzen wir jedoch nicht die Distanzwerte, die Muljačić selber berechnet hat, sondern gehen etwas anders vor. Der Grund ist folgender: Wenn man mit binären Daten arbeitet, sollte man es auch konsequent tun oder lassen. An einigen Stellen finden wir in Muljačićs Tabelle jedoch ein „ \pm “ was heißen soll, daß «es unmöglich war, eine kategorische binäre Antwort zu bekommen» [27]. Weil in diesen Fällen also die Möglichkeit gegeben ist, mit der Existenz einer Erscheinung zu rechnen, haben wir in diesen Fällen provisorisch ein „+“ gesetzt und auf dieser Grundlage die Distanzen neu berechnet, wobei folgende Distanzfunktion benutzt wurde:

$$f_i = \begin{cases} 1, & \text{wenn } a_i \neq b_i \\ 0, & \text{wenn } a_i = b_i \end{cases}$$

wobei a_i und b_i für die i -ten Vektorelemente stehen. Die Distanz zwischen zwei Sprachen A und B ist dann definiert als

$$D(A, B) = \frac{\sum_{i=1}^{40} f_i}{40}$$

so daß D Werte des Einheitsintervalls $< 0; 1 >$ annimmt. Das Ergebnis der Klassifikationsprozedur zeigt Abbildung 2. Wir legen auf dem Distanzniveau 0.4 einen Schnitt durch das Dendrogramm und erhalten so vier Sprachgruppen:

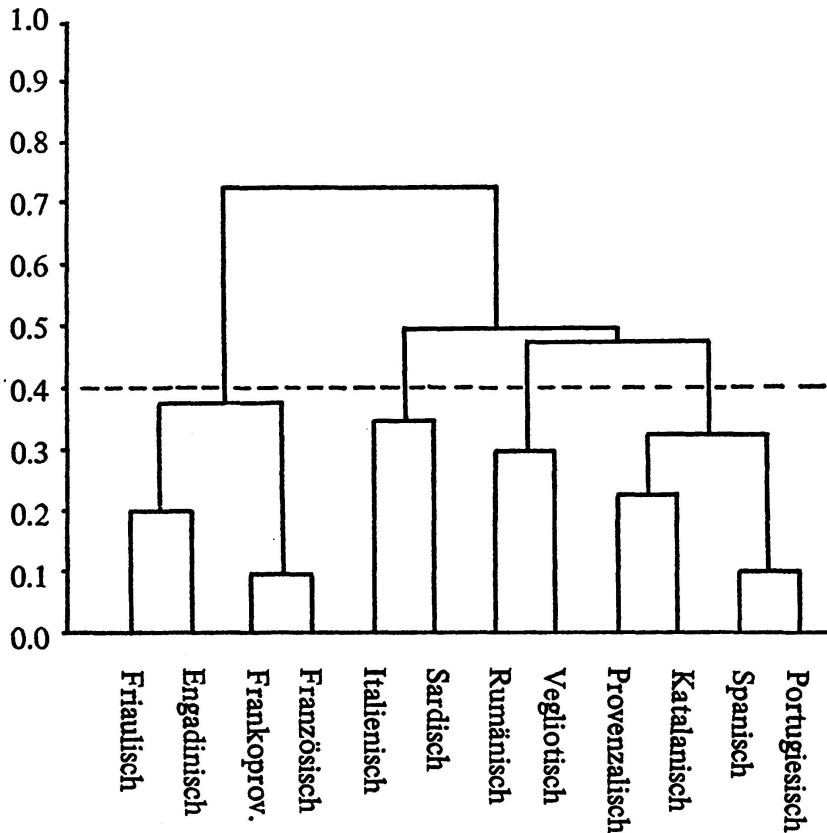


Abbildung 2: Hierarchische Klassifikation 12 romanischer Sprachen

- A: Friaulisch, Engadinisch; Frankoprovenzalisch, Französisch;
 B: Italienisch, Sardisch;
 C: Rumänisch, Vegliotisch;
 D: Provenzalisch, Katalanisch; Spanisch, Portugiesisch.

Auffällig ist die starke Korrelation zwischen der geographischen Lage der Sprachen und ihrer sprachlichen Entwicklung, soweit sie in den 40 Fragen erfaßt wird. In der Klassifikation zeigt sich also eine Eigenschaft, die als Klassifikationskriterium nicht benutzt wurde: Im ersten Schritt werden jeweils die Sprachen zusammengefaßt, die geographisch benachbart sind. Wir verzichten darauf, den vollständigen klassifikatorischen

Schlüssel aufzuführen und nennen nur die Eigenschaften, die die vier Gruppen von Sprachen auszeichnen:

- A: die lat. Phoneme k, g + a werden palatalisiert (Frage 36).
 B: doppelte Konsonanten bestehen (Frage 7).
 C: Futur I nicht des Typs *cantare + habeo* (u. ä.) (Frage 20).
 D: die lat. Phoneme k, g + a werden nicht palatalisiert (Frage 36); die intervokalischen stimmlosen Konsonanten p, t, k, s werden weiterentwickelt (Frage 30).

Damit wurde an zwei Beispielen das allgemeine Verfahren eines Sprachvergleichs, der sich der Methoden der numerischen Taxonomie bedient, demonstriert; wenn hier naturgemäß auch nicht alle Einzelprobleme zur Sprache kommen konnten, so ist doch zu hoffen, daß das generelle Prinzip deutlich geworden ist.

Konstanz

SEBASTIAN KEMPGEN

Tabelle II: Euklidische Distanzen

Französisch	---						
Okzitanisch	0.42718	---					
Portugiesisch	0.41116	0.37394	---				
Italienisch	0.71336	0.71222	0.88095	---			
Katalanisch	0.59445	0.41599	0.55652	0.62257	---		
Rumänisch	0.42909	0.37560	0.47639	0.54419	0.34924	---	
Spanisch	0.72516	0.62783	0.76876	0.50145	0.40526	0.43718	---
	Franz.	Okzit.	Portug.	Italien.	Katalan.	Rumän.	Span.

Literaturverzeichnis

Altmann, G.

1967: *Functional-probabilistic classification of phonemes* (Beiträge zur Linguistik und Informationsverarbeitung, 10), 27–33

1969: *Differences between phonemes* (Phonetica, 19), 118–132

1973a: *Asymmetrie der Distribution linguistischer Einheiten* (Phonetica, 28), 86–96

1973b: *Probabilistische Klassifikation von Konsonantenverbindungen des Indonesischen* (Zeitschrift der Deutschen Morgenländischen Gesellschaft, 123), 98–116

Altmann, G. / Lehfelddt, W.

1972: *Typologie der phonologischen Distributionsprofile* (Beiträge zur Linguistik und Informationsverarbeitung, 22), 8–32

1973: *Allgemeine Sprachtypologie, Prinzipien und Meßverfahren*, München (W. Fink)

1977: *Einführung in die quantitative Phonologie*, (im Druck)

Johnson, S. C.

1967: *Hierarchical clustering schemes* (Psychometrica, 32), 241–254

Krámský, J.

- 1959: *A quantitative typology of languages* (Language and speech, 2), 72–85
- Kriz, J.
1973: *Statistik in den Sozialwissenschaften, Einführung in die kritische Diskussion*, Reinbek bei Hamburg (Rowohlt)
- Lehfeldt, W.
1971: *Ein Algorithmus zur automatischen Silbentrennung* (Phonetica, 24), 212–237
- Muljačić, Ž.
1963: *Nova klasifikacija romanskih jezika i dalmatski*, Naučno društvo SR Bosne i Hercegovine: Radovi, knjiga XX (Odjeljenje istorijsko-filoloških nauka, knjiga 7), Sarajevo, 77–96
1967: *Die Klassifikation der romanischen Sprachen* (Romanistisches Jahrbuch, 18), 23–37
- Pulgram, E.
1970: *Syllable, word, nexus, cursus*, The Hague-Paris (Mouton)
- Siegel, S.
1976: *Nichtparametrische statistische Methoden*, Frankfurt (Fachbuchhandlung für Psychologie)
- Sneath, P.H.A. / Sokal, R.R.
1973: *Numerical Taxonomy*, San Francisco (W. H. Freeman)
- Späth, H.
1975: *Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion*, München-Wien (R. Oldenbourg)