

Joachim Boy

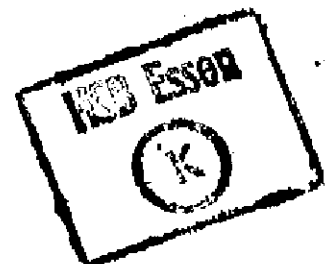
DECHIFFRIERUNGSLGORITHMEN
ZUR PHONETISCHEN IDENTIFIKATION
VON BUCHSTABEN

600

E 101057266 *Mie*

1/37 N 96
Studienverlag Dr. N. Brockmeyer

Bochum 1977



INHALT

Einleitung.....	5
1. Die Ziele der Arbeit.....	5
1.2. Die Klassifikationen Suchotina.....	7
1.3. Die Klassifikation Ševoroškina.....	9
1.4. Verwendetes Sprachmaterial.....	10
1.5. Methodenübersicht.....	11
2. Metrikgebundene Klassifikation.....	13
2.1. Einführung.....	13
2.2. Ordnungsverfahren.....	13
2.2.1. Distanzfunktion $d_0(a_i, a_j)$	13
2.2.2. Mittelwert \bar{a}_i^0	14
2.2.3. Gewichtungsfaktor $g_r(a_i, a_j)$ sowie verallgemeinerte Distanzen $d_r(a_i, a_j)$ und Mittelwerte \bar{a}_i^r	16
2.2.4. Mittelwerte als Ordnungskriterium.....	17
2.2.5. Stabilitätskoeffizient als Abbruchkriterium.....	20
2.2.6. Stabilität der Folge der \bar{a}_i^r für höhere Werte von r	22
2.2.6.1. Prinzip der Kurvenanpassung.....	23
2.2.6.2. Berechnung der Näherungskurven.....	29
2.2.6.3. Bestimmung der Schnittpunkte.....	29
2.2.7. Nachweis der Metrikeigenschaften für die Distanzen $d_0(a_i, a_j)$ und $d_r(a_i, a_j)$	31
2.3. Separierungsverfahren.....	34
2.3.1. Separierung durch Ranggewichtung der Differenzen der verallgemeinerten Mittelwerte.....	34
2.3.1.1. Definition des Grenzwertes.....	37
2.3.2. Weitere Abbruchkriterien.....	38
2.4. Zusammenfassung des metrikgebundenen Baserverfahrens.....	39
2.5. Modifikationen zu Gewichtungsfaktoren im Ordnungsverfahren.....	40
2.5.1. Neudefinition des Gewichtungsfaktors $g_r(a_i, a_j)$ durch vorgezogenen Mittelwertbildung.....	40
2.5.2. Klassifikation der Buchstaben des englischen Alphabets unter Verwendung der Neudefinition von $g_r(a_i, a_j)$	40

2.5.3. Weitere Eigenschaften des Mittelwerts \bar{a}_1^T	44
2.5.4. Ideale Eigenschaften eines Gewichtungsfaktors.....	44
2.5.5. Herleitung des Homogenitätskoeffizienten.....	45
2.5.5.1. Vektorbeträge und Komponentensummen.....	45
2.5.5.2. Inhomogenitätsquotient.....	47
2.5.5.3. Wertebereich des Inhomogenitätsquotienten.....	47
2.5.5.4. Frequenzabhängigkeit des Inhomogenitätsquotienten....	47
2.5.5.5. Definition des Homogenitätsmaßes $H(a_1)$	50
2.5.6. Anwendung von $H(a_1)$ als Gewichtungsfaktor.....	52
2.6. Vorläufige und endgültige Buchstabenklassen; Korrektur der vorläufigen Einteilung.....	55
2.6.1. Endgültige Klassifikation des englischen Alphabets.....	56
2.7. Modifikation des Separierungsverfahrens: die gewichtete Homogenitätsdifferenz als grenzbestimmendes Merkmal.....	57
2.8. Anwendung des gesamten metrikgebundenen Klassifikations- verfahrens auf das französische Alphabet.....	58
3. Metrikfreie Klassifikation:	
Klassifikation durch Teilsummenvergleich.....	63
3.1. Einführung.....	63
3.1.1. Grundsätze einer methodischen Vereinfachung: Zwei Ausgangshypothesen weiterer Klassifikationen.....	64
3.1.2. Das Prinzip der sukzessiven Zerlegung.....	66
3.1.3. Möglichkeiten zur Einschränkung der Objektmenge: Beschreibung des "Zentrums".....	66
3.1.4. Begrenzung des Zentrums: Anteil der Zentrums- buchstaben an Frequenzen aller Buchstaben des Alphabets.....	67
3.2. Klassifikation durch Teilsummenvergleich: Anwendung auf das türkische Alphabet.....	69
3.3. Anwendung auf das deutsche Alphabet.....	78
3.4. Anwendung auf das armenische Alphabet.....	83
3.4.1. Phonetische Transkription der vokalischen Buchstaben des armenischen Alphabets.....	86

4. Metrikfreie Klassifikation: Steigungskongruenzen.....	87
4.1. Einführung.....	87
4.1.1. Möglichkeiten einer Transformation der Distri- butionsangaben auf qualitative Daten.....	87
4.1.2. Definition des Steigungsvektors.....	88
4.2. Definition der Steigungskongruenz.....	89
4.3. Klassifikation durch Vergleich der Steigungskongruenzen; Anwendung auf das deutsche Alphabet.....	91
4.4. Modifikation des Verfahrens: Beschränkung auf das Steigungsprofil in Z.....	102
4.4.1. Frequenzanordnung der Zentrumsbuchstaben.....	102
4.5. Anwendung des modifizierten Verfahrens auf das türkische Alphabet.....	103
5. Metrikfreie Klassifikation: Kombiniertes Verfahren.....	112
5.1. Einführung.....	112
5.1.1. Übersicht über die angewandten Methoden.....	112
5.2. Das kombinierte Klassifikationsverfahren; Anwendung auf das türkische Alphabet.....	113
5.3. Anwendung auf das englische Alphabet.....	122
5.4. Weitere Anwendungen.....	127
6. Schlußbemerkungen.....	136
6.1. Erzielte Resultate.....	136
6.2. Verbesserungsmöglichkeiten.....	136
6.3. Perspektiven weiterer Untersuchungen.....	137
Distributionstabellen: Armenisch.....	139
Deutsch.....	140
Englisch.....	141
Französisch.....	142
Russisch.....	143
Türkisch.....	144
Verzeichnis der Abkürzungen und Symbole.....	145
Literaturangaben.....	147

1. EINLEITUNG

Alle sprachlichen Einheiten lassen sich durch die Feststellung ihrer verschiedenen Eigenschaften und deren Einordnung in entsprechende Bezugssysteme beschreiben. Die Entwicklung geeigneter Meßverfahren ermöglicht die Erfassung ihrer Ähnlichkeiten und Unterschiede und, darauf aufbauend, ihre Klassifikation auch mit dem Ziel eines intra- oder intersprachlichen Vergleiche. Insbesondere in der Phonologie wurden zahlreiche Meßverfahren entwickelt (vgl. Altmann 1969, Harary/Paper 1957, Lohfeldt 1973, Peterson/Harary 1961) und zu typologischen Untersuchungen verwendet (Altmann 1971a, Altmann/Lohfeldt 1973).

1.1. Es kann als Tatsache angenommen werden, daß der Sprachbau, also die Kombination von untergeordneten zu höheren Spracheinheiten, nicht vollkommen willkürlich erfolgt, sondern bestimmten stochastischen Gesetzmäßigkeiten unterworfen ist. Unter diesen Voraussetzungen wirft die Untersuchung inner-sprachlicher Funktionsmechanismen die Frage auf, ob zwischen verschiedenen Eigenschaften einer sprachlichen Einheit, etwa eines Phonems, ein Zusammenhang besteht, wie stark er ist und ob er ein charakteristisches Merkmal aller Elemente seiner Klasse darstellt oder nur diesem einen Individuum eigen ist. Bejaht man grundsätzlich das Bestehen eines solchen Zusammenhangs, so erhebt sich das Problem seines Nachweises. Gerade im Fall typologischer Untersuchungen kann der Nachweis eines Merkmalezusammenhangs von Bedeutung sein: die strikte wechselseitige Abhängigkeit zweier oder mehrerer Bezugsgrößen in einer polythetischen Klassifikation erbringt zwar nicht unmittelbar falsche oder ungenaue Ergebnisse, doch wird in der Untersuchung automatisch, ähnlich wie bei der Verwendung eines nicht-minimalen Axiomensystems in der Mathematik, ein gewisses Maß an Redundanz aufgebaut. Die ausschließliche Verwendung unabhängiger Eigenschaften vergrößert daher den Informationsgehalt einer Klassifikation.

Die vorliegende Arbeit unternimmt den Versuch, den Zusammenhang von verschiedenen Merkmalen einer sprachlichen Einheit zu entschlüsseln. Mit den Buchstaben der Schrift jeder natürlichen Sprache sind phonetische Eigenschaften der jeweils zugehörigen Laute mehr oder weniger eng verknüpft. Diese phonetischen Korrelate der Buchstaben sollen in der folgenden Arbeit der Kürze wegen als "phonetische Eigenschaften der

Buchstaben" bezeichnet werden. Eine solche Bezeichnung erscheint zwar zunächst als mißverständlich, doch wird sie hier nur stets im o.a. Zusammenhang verwendet, so daß sie als zulässig gelten mag. Die Ausgangshypothese bildet die Annahme, daß zwischen den phonetischen Eigenschaften vokalisiert/konsonantisch von Buchstaben und deren Distribution eine so enge Verbindung besteht, daß die exakte Unterteilung aller Buchstaben eines Alphabets in die Klassen der Vokale und der Konsonanten ausschließlich aufgrund ihrer distributionellen Eigenschaften durch eine geeignete mathematische Prozedur möglich ist. (Unter "distributionellen Eigenschaften" wird hier mit Juillard die Gesamtheit der durch die syntagmatischen Beziehungen der Spracheinheiten erzeugten Eigenschaften verstanden (siehe Juillard 1961, 24, 26 ff.).) Da grundsätzlich eine Verknüpfung verschiedener Merkmale auf allen Ebenen der Sprache angenommen werden kann, erscheint es sinnvoll, mit dem allgemeinen Versuch einer Dechiffrierung auf der untersten Ebene, d.h. in der Menge der Buchstaben, zu beginnen.

Bei den hier vorgenommenen Klassifikationsversuchen werden die untersuchten Buchstaben allein anhand ihrer distributionellen Eigenschaften im o.g. Sinn gruppiert, es wird ihnen also vor dem Beginn der jeweiligen Prozeduren keine andere als graphemische Qualität zugemessen. Insbesondere findet keine Vorinterpretation etwa in Bezug auf ihre phonetischen Eigenschaften statt. Ein solches Vorgehen beruht natürlich auf der Vermutung, daß die Verteilung der Buchstaben in den Texten einer natürlichen Sprache nicht willkürlich ist, sondern daß ihr Erscheinen in den realen Positionen durch eine mehr oder weniger invariante Zurechnung phonetischer Eigenschaften beeinflußt oder gar determiniert wird (vgl. auch die Eingangsbeobachtungen bei Suchotin 1962). Eben diese Eigenschaften sollen durch die Klassifikation zum Ausdruck kommen.

Das Problem der Klassifikation der Buchstaben eines Alphabets nach den hier verwendeten Kriterien wurde in der Vergangenheit in der linguistischen Literatur nur relativ selten behandelt. Juillard (1961, Kap. 2.51) schlug eine Verwendung von statistischen Daten für die Kategorienbildung auf der phonologischen Ebene zur Trennung von Konsonanten und Vokalen vor. Er hält ein solches Vorgehen jedoch nur für "certain languages" und "exceptionally" für angebracht und geht daher nicht auf nähere Einzelheiten einer solchen Untersuchung ein. (Juillard 1961, 40).

Intensiver haben sich insbesondere Suchotin (1962, 1963) und Savoroškin (1962) mit diesem Problem auseinandergesetzt. Ihre Arbeiten und die von ihnen verwendeten Methoden zur Buchstabenklassifikation sollen daher hier kurz dargestellt werden.

1.2. Suchotins Grundannahme bei der Formulierung seines Verfahrens besteht darin, daß, statistisch gesehen, die Aufeinanderfolge von Vokal und Konsonant häufiger ist als die von gleichartigen Buchstaben, also jeweils zwei Konsonanten oder Vokalen. Der Klassifikationsalgorithmus arbeitet folgendermaßen: Die im Text vorkommenden Buchstaben werden in gleicher Reihenfolge an den Rändern einer quadratischen Tabelle angeordnet. Stellt man sich diese Anordnung so vor, daß in der linken oberen Ecke der Tabelle mit den Vokalen begonnen wird und daß die Konsonanten sich anschließen, und geht man weiter davon aus, daß die o.g. Annahme der relativ größeren Häufigkeit ungleichartiger Buchstabenverbindungen zutrifft, so erhält man, wenn man an den Kreuzungspunkten der Zeilen und Spalten der Matrix die Häufigkeiten der entsprechenden Buchstabenverbindungen einträgt, eine vierfache Unterteilung der Matrix. In den Feldern links oben (1) und rechts unten (3), die die Frequenzen der Verbindungen gleichartiger Buchstaben beschreiben, wäre unter den zugrundeliegenden Annahmen die Gesamtzahl der erfaßten Verbindungen geringer als in den Feldern 2 und 4, die die Verbindungen ungleichartiger Buchstaben bezeichnen. Da Suchotin bei seiner Untersuchung lediglich die Tatsache des Aufeinandertreffens von Buchstaben feststellt, nicht aber deren Reihenfolge unterscheidet (ebensowenig werden Textlücken durch Wort- oder Satzgrenzen bei der Zählung berücksichtigt), ist die so erhaltene Matrix entlang ihrer Hauptdiagonalen symmetrisch. Beginnt man die Untersuchung mit einer ungeordneten Matrix, bei der die Reihenfolge der Randeintragungen willkürlich ist, jedoch am waagerechten und am senkrechten Rand übereinstimmt, so gestaltet sich das Klassifikationsverfahren unter den genannten Aspekten als der Versuch, die Trennlinien zwischen den Feldern der Matrix so zu plazieren, daß die Summe der Eintragungen in den Feldern 1 und 3 minimal, in den Feldern 2 und 4 dagegen maximal ist. Zu Beginn werden innerhalb der Matrix provisorisch je eine durchgehende horizontale und vertikale Trennlinie zwischen den Feldern 1 bis 4 gezogen. Anschließend werden die Summen der Häufigkeitseintragungen für jeden am Rand vermerkten Buchstaben links und rechts (bzw. oberhalb und unterhalb)

der Trennungslinie berechnet und die linke von der rechten (bzw. die obere von der unteren) Summe subtrahiert. Derjenige Buchstabe, für den sich bei dieser Operation die größte positive Differenz ergibt, wird über die Trennungslinie hinaus in Richtung der linken oberen Ecke der Matrix übertragen. Dieses Verfahren wird so lange wiederholt, bis alle Buchstaben mit positiver Differenz der Summen ihren endgültigen Platz innerhalb der Matrix durch Übertragung gefunden haben. Die Lage der Trennungslinien soll dann auf dem Rand der Matrix die Klassen der Vokale und der Konsonanten abteilen. Die Buchstaben innerhalb der Klassen können mit Hilfe desselben Verfahrens prinzipiell noch weiter geordnet werden (natürlich unter Einführung neuer Trennlinien). Suchotin erwartet bei fortgesetzter Anwendung seiner Methode auf die Klasse der Konsonanten beispielsweise eine Aufteilung in die Gruppen der Verschlusslaute, der Reibelauten, der Sonore usw. Für den Fall von offensichtlichen Fehlentscheidungen innerhalb des Ordnungsablaufs, wenn sich etwa eine Asymmetrie der Matrix ergeben sollte, gibt Suchotin einige Hinweise für gezielte Eingriffe in den Verfahrensablauf. Die gezeigten Anwendungen des Verfahrens auf russische, französische und englische Texte mit einem Umfang von jeweils etwa zehntausend Buchstaben ergeben Resultate, deren Fehlerquote sich als bemerkenswert gering erweist oder bei denen sich charakteristische Ungenauigkeiten zumindest mit hinreichender Plausibilität erklären lassen. In allen durchgeführten Fällen werden die Konsonanten einwandfrei separiert, d.h. in der Klasse der Konsonanten befinden sich am Ende der Prozedur keine Vokale. Umgekehrt sind fehlerhafte Einordnungen nicht immer zu vermeiden, treten jedoch auch nur selten auf.

An späterer Stelle (1963) schlägt Suchotin die folgende Modifikation seines Verfahrens vor: Am senkrechten Rand der Matrix sind die Buchstaben nunmehr in beliebiger Reihenfolge notiert, während die Reihenfolge der Buchstaben am oberen waagerechten Rand dem Resultat der oben beschriebenen Klassifikation entspricht. Das Feld der Matrix wird nunmehr jedoch lediglich, im Gegensatz zur ersten Prozedur, von einer einzigen, vertikalen Trennungslinie unterteilt. Die Untersuchung des Textes erfolgt schrittweise. Es werden am Beginn des Textes fünf Buchstaben abgetrennt. In derjenigen Zeile der Matrix, die mit dem mittleren Buchstaben der Fünfergruppe bezeichnet ist und die, wie alle Stellen der Matrix, bisher nur Nullen enthält, wird an den Kreuzungspunkten mit den Spalten derjenigen Buchstaben, die außer dem mittleren in der abgetrennten Gruppe

enthalten sind, jeweils eine Eins eingetragen. Anschließend wird die - selbe Prozedur wiederholt, wobei der Anfang der Fünfergruppe im Text um einen Buchstaben nach rechts versetzt wird. Dieser Vorgang wird nun fortgesetzt, bis der gesamte Text durchlaufen ist. Die Summen- und Differen - zenbildung erfolgt anschließend nach der Verfahrensweise der zuerst be - schriebenen Prozedur entlang der Trennungslinie. Allerdings ergibt sich trotz der Einbeziehung eines, wenn auch dem Umfang nach nur geringen, Kontextes in die Untersuchung nicht die erhoffte wesentliche Resultats - verbesserung. Wie die ausgedruckten Tabellen zeigen, ist wiederum ins - besondere die Aussonderung der Vokale mit Ungenauigkeiten behaftet.

1.3. Das von Ševorođkin (1963) entwickelte Klassifikationsverfahren ar - beitet nicht mit Tabellen über Buchstabenverbindungen, sondern untersucht Oppositionen innerhalb von fortlaufend aus dem Text abgetrennten Buch - stabengruppen. Die Grundannahme ist, daß in jeder Sprache das zahlen - mäßige Verhältnis von Vokalen und Konsonanten bestimmte Gesetzmäßigkeiten der Lautfolge, also der Struktur von Lautketten, bedingt. Keine Sprache erlaubt eine unbegrenzte Häufung von Lauten einer Klasse, sondern läßt nur bestimmte Maximalverbindungen zu, wobei sich eine Konzentration von einzelnen, von Ševorođkin als charakteristisch angesehenen Typen zeigt. Zur Erstellung seiner Prozedur ist für den Autor speziell die Merkmale - ausprägung bei dreigliedrigen Lautverbindungen von Bedeutung. Es läßt sich empirisch belegen, daß bei Sprachen mit nicht extrem stark ausge - prägtem Vokalismus die weitaus häufigste Form der dreigliedrigen Laut - ketten die Reihenfolge CVC (also Konsonant - Vokal - Konsonant) oder, bei zentralem Konsonant, VCV aufweist, wobei im letzteren Fall der Vokalismus von größerem Einfluß ist. Unter diesen Voraussetzungen stellt sich die Untersuchung dar als die Ermittlung von Gruppen zentraler Elemente sol - cher Dreierketten. Konkurrierende Paare von Umgebungslauten bei Überein - stimmendem zentralem Element ergeben die Einteilung der Laute (oder der Buchstaben) in die Klassen der Konsonanten und der Vokale.

Der zu untersuchende Text wird vom Beginn her schrittweise, jeweils um einen Buchstaben fortschreitend, in Dreiergruppen aufgeteilt. Die Dreiergruppen mit identischem zentralem Buchstaben werden in Spalten untereinander angeordnet. Nun wird untersucht, in welcher Spalte peri - phere Buchstaben mit auffällender Regelmäßigkeit mit zentralen Buchsta - ben kombiniert sind. Wegen der höheren Wahrscheinlichkeit einer

CVC- oder VCV - Verbindung gegenüber einer des Type CCC oder VVV werden in einem solchen Fall die peripheren und die zentralen Buchstaben als kontrastierend angesehen und daher verschiedenen Lautklassen zugeordnet. Auf diese Weise lassen sich weiter fortschreitend durch geeignete Vergleiche auch solche Verbindungen analysieren, die nur einmal im Text auftauchen, so daß sich am Ende der Prozedur zwei kontrastierende Lautgruppen als die Klassen der Vokale und der Konsonanten ergeben. Die Gruppe mit dem zahlenmäßig geringeren Inventar wird als die Klasse der Vokale angesehen. Sein Verfahren demonstriert Ševoroškin an jeweils einem lateinischen und russischen Beispielsatz sowie an einem längeren karischen Text, wobei einwandfreie Resultate erzielt werden. Allerdings stellt sich die Frage nach der Anwendbarkeit des Verfahrens bei Sprachen, in denen aufgrund ihrer speziellen Ausprägung des Vokalismus bzw. des Konsonantismus mit einem Überwiegenden Auftreten von Dreierverbindungen der Typen CVC oder VCV nicht gerechnet werden kann.

Im zweiten Abschnitt seiner Arbeit untersucht Ševoroškin die Möglichkeit, mit Hilfe geeigneter Prozeduren zu einer feineren Unterteilung der Lautklassen zu gelangen. Da eine solche Fragestellung jedoch über den Rahmen des hier zu behandelnden Themas hinausgeht, soll hier auf eine nähere Beschreibung der entsprechenden Überlegungen verzichtet werden.

1.4. Die Ergebnisse der oben beschriebenen Untersuchungen haben gezeigt, daß es durchaus grundsätzlich möglich ist, mit Hilfe von geeigneten mathematischen Verfahren eine Klassifikation von Buchstaben zu erreichen, in der ihre phonetischen Charakteristika zum Ausdruck kommen. Dabei scheinen die von Suchotin angewandten Verfahren prinzipiell besser geeignet zu sein, das angestrebte Ziel zu erreichen. Die in der folgenden Arbeit beschriebenen Prozeduren sollen nach Möglichkeit auf alle natürlichen Sprachen anwendbar, also nicht von der Ausprägung bestimmter Sprachmerkmale abhängig sein, so daß sich eine strukturelle Vorinterpretation des zu untersuchenden Sprachmaterials verbietet. Die in den folgenden Kapiteln beschriebenen Klassifikationsverfahren basieren daher, ebenso wie Suchotins Algorithmen, auf Tabellen von den in einer Sprache vorkommenden Buchstabenverbindungen. Dabei wurden die Tabellen verwendet, die Suchotin als Ergebnis seiner Klassifikationsversuche erhielt (Suchotin 1962). Um eine Beeinflussung des Ergebnisses zu verhindern, wurden die

Randeintragungen und damit die Zahlen der erfaßten Buchstabenverbindungen im Inneren der Matrix auf alphabetische Reihenfolge transformiert. Die Verwendung dieser Tabellen erwies sich als grundsätzlich möglich. Bei der Auszählung der Buchstabenverbindungen des Englischen, des Französischen und des Russischen, die diesen Tabellen zugrundelag, wurden jedoch, wie oben erwähnt, keine Textlücken durch Wort- oder Satzgrenzen berücksichtigt. Da bei einem solchen Vorgehen wort- und satzübergreifende Buchstabenverbindungen ebenso mitgezählt werden wie jene innerhalb eines Wortes, ihr Zustandekommen daher implizit als durch dieselben Spracheigenschaften determiniert angesehen und in der Zählung bewertet wird, scheint hier ein gewisses Risiko für einen fehlerfreien Verfahrensablauf zu liegen. Die Anwendung von Verfahren, die in dieser Hinsicht "empfindlicher" reagieren als die in der folgenden Arbeit beschriebenen, könnte aufgrund dieser Eigenschaften des von Suchotin zusammengestellten Zahlenmaterials zu Fehlentscheidungen führen. In den hier zusätzlich vorgenommenen Zählungen der Buchstabenverbindungen des Armenischen (genauer des Neu-Ostarmenischen), des Deutschen und des Türkischen wurden daher nur Buchstabenverbindungen innerhalb der Wortgrenzen berücksichtigt. (Die Zählungen wurden anhand von modernerer Prosaliteratur vorgenommen.) Alle Tabellen der Buchstabenverbindungen sind der Arbeit am Schluß als "Distributionstabellen I bis VI" angefügt.

1.5. Die hier entwickelten Dechiffrierungsalgorithmen beruhen nicht auf den oben beschriebenen Prozeduren. Um einen allgemeinen Zusammenhang zwischen den distributionellen und den phonetischen Eigenschaften vokalisiert/konsonantisch der Buchstaben nachzuweisen, wurde auf die Anwendung mehrerer voneinander unabhängiger mathematischer Verfahren Wert gelegt. Aus der unterschiedlichen Konzeption der Prozeduren ergibt sich die Aufteilung der Arbeit: Der im zweiten Kapitel beschriebene Algorithmus basiert auf der Definition einer Distanzfunktion zur Bestimmung des Unterschiedes zwischen jeweils zwei Buchstaben; das Verfahren arbeitet also metrikgebunden (die Metrikeigenschaften der Distanzfunktion werden in Abschnitt 2.2.7. nachgewiesen). Demgegenüber verzichten die weiteren Prozeduren auf die Verwendung einer solchen Distanzfunktion. Von diesen in den Kapiteln drei, vier und fünf beschriebenen metrikfreien Verfahren sind die beiden ersten, also das Teilsommenverfahren und das Verfahren unter Verwendung der Steigungskongruenzen, voneinander unabhängig; das dritte, kombinierte Verfahren stellt eine Zusammenfassung der beiden

vorangegangenen Algorithmen dar, kann jedoch wegen verschiedener Veränderungen als selbständig gelten. Alle Verfahren werden ausführlich anhand einer Beispielsprache erläutert und anschließend an mindestens einer weiteren Sprache demonstriert. Das kombinierte Verfahren wird am Ende der Arbeit auf alle hier untersuchten Sprachen angewendet.

Bei der Formulierung der Algorithmen wurde berücksichtigt, daß eine programmierte Anwendung durch elektronische Datenverarbeitungsanlagen möglich sein sollte. Die Programmierung selbst wurde hier nicht vorgenommen, doch ist sie unter Berücksichtigung der einzelnen Verfahrensschritte unschwer möglich. (Eine Programmierung erscheint jedoch nicht als zwingend notwendig; für die Untersuchungen stand hier lediglich ein elektronischer Taschenrechner des Type "Texas Instruments BR 51-A" zur Verfügung, mit dem alle Operationen, sofern sie nicht trivialer Natur waren, durchgeführt wurden.) Neben der Handlichkeit der Prozeduren stand insbesondere ihre Leistungsfähigkeit in Bezug auf die zu bewältigende Aufgabe im Vordergrund. Es können hier nicht alle mit den verschiedensten Methoden vorgenommenen Versuche beschrieben werden, auch wenn sie zum Teil günstige Ergebnisse erbrachten oder zu weiteren interessanten Untersuchungen anregten.

Die Aufteilung der Buchstaben der untersuchten Sprachen in die Gruppen der Vokale und der Konsonanten ist uns bekannt; somit steht eine empirische Kontrollinstanz für die erzielten Resultate zur Verfügung. Diejenigen Dachiffrierungsalgorithmen, die diesbezüglich identische Ergebnisse erbringen, werden als für eine Klassifikation geeignet angesehen.

Der Nachweis eines Zusammenhanges zwischen den phonetischen und den distributionellen Eigenschaften von Buchstaben ist nicht das alleinige Ziel dieser Arbeit. Die hier bearbeitete Aufgabe stellt nur eines von sehr vielen Problemen aus dem Bereich der Entschlüsselung von Merkmalezusammenhängen sprachlicher Einheiten dar. Vielleicht kann diese Arbeit eine, wenn auch nur methodische, Anregung zu weiteren Forschungen in dieser Richtung geben und so einen kleinen Beitrag zur Erarbeitung einer umfassenden Theorie innersprachlicher Funktionsmechanismen leisten.