III. Computational Linguistics III: Special Methodical Problems
Computerlinguistik III: Besondere methodische Probleme

9. Status und Funktion quantitativer Verfahren in der Computerlinguistik

1. Einführung

Die Begründung eines neuen Wissenschaftszweiges ist stets mit der Entdeckung oder Erahnung von etwas bisher Unbekanntem in einem Forschungsobjekt verbunden. Diese erste elementare Stufe einer Wissenschaft wird bald durch die Phase der Begriffsbildung abgelöst; um das Entdeckte besser fassen zu können, wird es in Begriffe gekleidet. Diese sind zunächst meist qualitativ, oft sehr vage oder sogar nur metaphorisch. Sie erfahren aber im Laufe der Zeit eine Präzisierung durch Definition oder Operationalisierung. — Diese Epoche einer Disziplin erscheint den beteiligten Wissenschaftlern als extrem reich an Entdeckungen; jede Einführung eines neuen Begriffs kann das Gefühl vermitteln, dem Forschungsgegenstand ein Geheimnis entrissen zu haben und ihn besser verstehen zu können. Nun ist diese Phase der Begriffsbildung unabdingbare Voraussetzung für die darauffolgenden Stufen der wissenschaftlichen Erkenntnis; es geschieht aber nicht selten, daß eine Disziplin lange Zeit in ihr verharrt. Gerade in den Sprach- und Textwissenschaften ist noch heute eine Position weit verbreitet, die sich als „magische Auffassung der Sprache“ (vgl. Carnap 1969, 119) charakterisieren läßt, da ihre Vertreter zu glauben scheinen, die Phänomene durch Benennen beherrschen zu können. Jedenfalls wird von ihnen die Begriffsbildung als die höchste Stufe der Forschung betrachtet, die wissenschaftliche Auseinandersetzung mit einem Gegenstand erschöpft sich für sie in Begriffsexplikationen, Beschreibungen und Klassifikationen. — Für eine andere Gruppe von Wissenschaftlern indessen eröffnet dieses Stadium erst die Möglichkeit, zu höheren Stufen der Erkenntnis fortzuschreiten. Sie betrachten die qualitativen Begriffe als erste Annäherungen an die Wirklichkeit, die als Ausgangspunkt zur Schaffung der differenzierteren quantitativen Begriffe dienen können. Ihre Vertreter sind der Ansicht, daß man auch in den Sprach- und Textwissenschaften über Theoriebildung zur Erklärung als Ziel der Wissenschaft gelangen kann und muß. — In den Geisteswissenschaften wird gegen eine solche Zielsetzung immer wieder eingewendet, es gehe (1) nicht um Quantitäten, sondern um Qualitäten und (2) nicht um Erklären, sondern um Verstehen. Der erste Einwand entspringt einem grundsätzlichen Mißverständnis: Die Qualitäten, die den Dingen zugeschrieben werden, sind nicht den Dingen inhärent, sondern Eigenschaften unserer Begriffe — auch dann, wenn sie eine reale empirische Entsprechung besitzen (vgl. Hempel 1974, 110—117; Maxwell 1971). Ebenso dienen quantitative Begriffe dazu, eine Ordnung im Forschungsobjekt zu etablieren (und zwar eine feinere, als es mit qualitativen Mitteln möglich ist); die Dinge selbst bleiben wie sie sind — weder qualitativ noch quantitativ (vgl. a. Esser 1971, 64 ff.). Qualitative Begriffe erscheinen zunächst natürlicher zu sein, da sie uns von der natürlichen Sprache nahegelegt werden. In fortgeschritteneren Stadien einer Wissenschaft kommt man aber ohne Quantifizierung nicht aus (vgl. Stegmül-

ler 1970, 44 ff.; Esser 1971, 64—69), sie erfordert aber bereits eine exaktere und festere Vorstellung des Wissenschaftlers von seinem Untersuchungsgegenstand. — Der zweite Einwand baut einen Scheingegensatz auf: In der Wissenschaft geht es sowohl um Verstehen als auch um die *Erklärung* . Während man aus der wissenschaftstheoretischen Beschäftigung mit dem Begriff der Erklärung genau weiß, worauf sie beruht, welche Beziehung sie zu einer Theorie hat und welche Rolle sie in den empirischen Wissenschaften spielt (vgl. Hempel 1965; Stegmüller 1969; Bunge 1969, II; Popper 1973, 213 ff.), ist vorläufig nicht recht klar, was mit „Verstehen" exakt gemeint sein kann.

Trotz aller Einwände ist heute in allen empirischen Wissenschaften die quantitative Denkweise und damit die Möglichkeit, mathematisch-numerische Verfahren zu verwenden, fest verankert. In diesen Disziplinen hat sich eine ganze Reihe mathematischer Methoden etabliert, die inzwischen zum selbstverständlichen Forschungsinstrumentarium gehören. Im Vordergrund stehen diejenigen Methoden, die zur Bildung von Modellen geeignet sind, d. h. abstrakte Gefüge mit internen Mechanismen. Die Erfassung solcher Mechanismen mit symbolischen Mitteln fördert die Theoriebildung: die Findung von Gesetzessystemen, die der Erklärung und Voraussage dienen können. Zu diesen deduktiven Methoden gehören vor allem Differential- und Differenzengleichungen, stochastische Prozesse und andere Wahrscheinlichkeitsmodelle.

Nicht alle mathematischen Hilfsmittel tragen gleichermaßen zur Theoriebildung bei. Viele von ihnen dienen eher der Beschreibung und Systematisierung des Gegenstandsbereiches. Diese induktiven Methoden helfen aber auch bei der Hypothesenbildung, und einige, insbesondere die statistischen Methoden, sind in den empirischen Wissenschaften das einzige Mittel, um Hypothesen zu testen und die Wahrscheinlichkeit einer falschen Entscheidung über die Annahme einer Hypothese objektiv zu berechnen. — In denjenigen Disziplinen, die in der Modellbildung noch nicht weit fortgeschritten sind, überwiegt aus heuristischen Gründen die Anwendung induktiver Methoden. Von diesen sind in der Computerlinguistik die Taxonomie und die Statistik besonders verbreitet.

2. Taxonomie

Unter diesem Sammelbegriff werden alle Arten von Klassifikation und Typologie zusammengefaßt. Die Anzahl der Methoden, die zur maschinellen Ausführung einer Klassifikation geeignet sind, wächst jährlich um einige Dutzend an (vgl. u. a. Bock 1974). Die Frage nach dem gnoseologischen Wert, ob diese neuen Methoden imstande sind, neue Erkenntnisse zu vermitteln, ist nicht leicht zu beantworten. Kennt man nämlich die Struktur des untersuchten Datenfeldes, so läßt sich erst a posteriori entscheiden, welche Klassifikationsmethode geeignet ist, diese Struktur abzubilden. Auch über die Kriterien der Eignung lassen sich keine allgemeingültigen Aussagen machen. Der heuristische Wert ist in diesem Fall also minimal. Kennt man aber die Struktur des Objekts noch nicht, so können taxonomische Methoden zum Erkenntniszuwachs beitragen. Je nach verwendeter Methode erbringen taxonomische Verfahren unterschiedliche Leistungen, die im folgenden beschrieben werden.

2.1. Schaffung von Ordnung

Die elementare Leistung jeder Klassifikationsmethode ist die Schaffung einer Ordnung in einer Daten- oder Gegenstandsmenge. Diese Ordnung kann auf einem einzigen Kriterium beruhen (etwa beim Ordnen von Texten nach ihrer Länge), das einem ganz bestimmten Zweck gerecht wird; für andere Zwecke ist eine solche Klassifikation dann meist ungeeignet. Eine Mehrzweckklassifikation kann man durch Anwendung einer beliebigen Zahl von Kriterien erhalten; es ist jedoch nicht möglich, eine Allzweckklassifikation aufzustellen. Dementsprechend ist die Anwendung taxonomischer Verfahren ohne genaue vorherige Festlegung ihres Zwecks nicht sinnvoll. — Es ist üblich, zwischen praktischen und wissenschaftlichen Klassifikationen zu unterscheiden. Eine praktische Ordnung kann und soll einfach sein (vgl. Carvell/Svartik 1969), aber in bezug auf eine wissenschaftliche Ordnung ist dieses Desideratum nicht verwendbar, weil Einfachheit fast allen Desideraten der Wissenschaftlichkeit widerspricht (vgl. Bunge 1963). Die größte Vereinfachung erfolgt durch die Reduktion und Anordnung der Taxate auf eine Zahlenachse (vgl. Benzecri 1970), wodurch die in der Linguistik seit Humboldt (1827—29) bekannten Extremtypen entstehen (vgl. Lehfeldt/Altmann 1975).

Wie Hempel (1965, 159) zeigt, gehört diese Taxonomie den frühen Stadien der Entwicklung einer wissenschaftlichen Disziplin an. Sie ist nicht weit entfernt von dem elementaren Bedürfnis des Menschen, in seiner Umwelt Ordnung zu schaffen, aus welchen Gründen auch immer.

2.2. Reduktion der Variabilität

Die Reduktion der Variabilität ist eine taxonomische Leistung, die insbesondere durch die natürliche Sprache erbracht wird. Jedes Ding ist ein von allen anderen Dingen unterschiedenes Individuum. Die natürliche Sprache und eine einfache Taxonomie reduzieren diese unendliche Vielfalt so, daß sie die Dinge in Klassen einordnen. Die natürliche Reduktion ist eine 'ethnolinguistische' Klassifikation aufgrund von Ähnlichkeiten, die wissenschaftlich völlig irrelevant sein können. Eine solche Variabilitätsreduktion ist wichtig für die Minimierung des Gedächtnisaufwandes und für die Orientierung des Menschen in seiner Umwelt. Wissenschaftlich kann die Reduktion der Variabilität dazu führen, daß in scheinbar chaotischen Daten Regularitäten auftauchen, die zur Aufstellung von Hypothesen führen können.

2.3. Begriffsbildung

Taxonomie trägt zur Begriffsbildung bei. Eine Klassifikation läßt sich auf drei Weisen durchführen: (i) Eintragung der Taxate auf eine Zahlengerade (Skala), eine Ebene, einen dreidimensionalen Raum usw. und Trennung der Klassen durch 'geeignete' Kriterien. (ii) Anordnung der Taxate in eine Hierarchie und Trennung der Klassen durch ein Abbruchkriterium. (iii) Bildung von getrennten Klassen durch eine clusterbildende taxonomische Methode. In jedem dieser Fälle werden bestimmte Ganzheiten voneinander getrennt: Dinge, die irgendwelche gemeinsamen Eigenschaften haben, oder Eigenschaften, die miteinander korrelieren. Diese Dinge oder Eigenschaften werden unter einen gemeinsamen Begriff gefaßt, der entweder aus praktischen oder aus wissenschaftlichen Gründen verwendet werden soll. Je nach Methode führt dies zu einer Begriffsbildung, bei der die entstehenden Klassen disjunktiv sind (z. B. bei der Clusteranalyse), oder zu unscharfen Klassen (wie bei der Skalierung, Diskriminationsanalyse oder hierarchischer Taxonomie ohne Abbruchskriterium). — Sehr viele linguistische Untersuchungen geben sich an diesem Punkt mit dem Erreichten zufrieden: Die Einordnung der Taxate in Klassen hat eine Ordnung erzeugt, durch die taxonomische Darstellung ergab sich offenbar eine Struktur, und durch Begriffsprägung wurde ein neues Stück der Realität erfaßt. Obwohl dieser Schritt selbstverständlich ein anerkanntes und sogar unbedingt notwendiges Ziel wissenschaftlicher Arbeit ist, erschöpft sich die Forschung keineswegs in Deskription und Begriffsbildung. Als höchstes Ziel jeder Wissenschaft sieht man heute die Erklärung an (vgl. Popper 1973, 213 ff.). Damit stellt sich die Frage nach dem möglichen Beitrag der Taxonomie zur Explanation.

2.4. Heuristik

Erklärungen selbst sind nur mit Hilfe von Theorien möglich, zu deren Aufbau Begriffe, Konventionen (Desiderata, Operationen, Kriterien, Regeln u. a.) und vor allen Dingen Hypothesen benötigt werden. Die heuristische Bedeutung taxonomischer Verfahren liegt darin, daß sie bei der Aufstellung dieser wichtigsten Bestandteile von Theorien (vgl. Spinner 1974, 117 ff.) helfen können. Bunge (1969, 253 ff.) unterscheidet vier verschiedene Typen von Hypothesen, je nach theoretischer und empirischer Begründetheit: (i) Spekulationen, die weder theoretisch begründet (abgeleitet) noch empirisch überprüft sind, (ii) empirische Generalisierungen, die durch Verallgemeinerung singulärer Aussagen entstanden und durch Erfahrung gut bestätigt sind (zu diesem Typ gehören z. B. Lautgesetze, grammatische Regeln u. a.), (iii) deduktive (bzw. valide, plausible) Hypothesen, die aus einer Theorie, aus Axiomen oder Gesetzen abgeleitet und damit theoretisch begründet sind, aber noch keiner empirischen Überprüfung unterzogen wurden, und (iv) Gesetze, die sowohl theoretisch begründet (abgeleitet) als auch empirisch gut bestätigt sind. — Eine Theorie enthält vor allem Gesetze; wenn auch empirische Generalisierungen zu ihren Bestandteilen zählen, fällt sie in die Klasse der deduktiv-induktiven Theorien. Entsprechend sind Wissenschaften, die als höchsten Hypothesentyp empirische Generalisierungen enthalten, als vortheoretisch zu kennzeichnen. Unter diese Kategorie fallen (noch) fast sämtliche computerlinguistischen Teildisziplinen sowie die Mehrzahl der linguistischen und textwissenschaftlichen Bereiche. In diesem Zusammenhang ist es wichtig sich zu vergegenwärtigen, daß die Anwendung taxonomischer Verfahren eine wichtige

heuristische Funktion erfüllt, andererseits aber keinen Ersatz für Hypothesenbildung darstellt: Klassifikationen können Hinweise auf Phänomene liefern, über die Hypothesen zu bilden sind; sie tragen jedoch zur Gewinnung der Hypothesen wenig bei.

2.5. Inferenz

Schließlich bieten taxonomische Verfahren die Möglichkeit zur Inferenz. Vom theoretischen Gesichtspunkt aus wird eine Klassifikation dann fruchtbar, wenn mit ihrer Hilfe Schlußfolgerungen gezogen werden können. Dies ist dann der Fall, wenn aufgrund der der Klassifikation zugrundeliegenden Merkmale Aussagen über andere Merkmale möglich sind, die in die Klassifikation nicht eingingen. Klassifiziert man beispielsweise Texte nach ihrer durchschnittlichen Satzlänge, so kann man aus der Klasseneinteilung auf die Zugehörigkeit eines Textes zu einer Textsorte oder auf den Autor schließen. — Besteht eine derartige Zusammenhangshypothese bereits vor der Durchführung der Klassifikation, so stellt diese eine Überprüfungsinstanz der Hypothese dar. Ist eine solche Hypothese nicht vorhanden, so kann eine evtl. resultierende Inferenz kaum gedeutet werden (vgl. etwa Caroll 1960, 290 ff.), vor allem ist ein so gewonnener Befund (vgl. Sapir 1921, 134 ff.) ohne (deduktive) Hypothesen theoretisch nicht verwertbar. Unter diesen Umständen sind Schlußfolgerungen nur tentativ möglich, und sie müssen unabhängig von der vorgenommenen Klassifikation überprüft werden. Das Schließen auf genetische Aussagen aus synchron angelegten Klassifikationen wird im allgemeinen vollständig abgelehnt.

Wir können zusammenfassend feststellen, daß die Taxonomie ein induktives Verfahren ist, das eine wichtige und unangefochtene praktische Bedeutung besitzt. Für den theoretischen Bereich spielt sie darüber hinaus eine Hilfsrolle in zweifacher Hinsicht: (a) Vor dem Aufbau einer Theorie hilft sie bei der Begriffsbildung und beim Auffinden von Gebieten, in denen Hypothesen über Zusammenhänge, latente Mechanismen, verborgene und unterschwellige Kräfte und über die Dynamik des Gegenstandsbereiches aufgestellt werden können (vgl. Altmann/Lehfeldt 1973, 15). Sie selbst liefert keine Hypothesen; denn diese gehen nicht aus den Daten hervor, sondern sind Schöpfung des analysierenden Wissenschaftlers. (b) Folgt der Klassifikationsansatz aus einer Theorie, handelt es sich also um eine „ideale Typologie“ im Sinne von Hempel (1965, 166 ff.), so kann man sie als Überprüfungsinstanz für die Theorie betrachten. Bei Nichtübereinstimmung von Theorie und Klassifikationsresultat sollte jedoch das Primat der Theorie über jedes induktive Verfahren beachtet werden, das insofern den theoriebildenden Wert aller Methoden der numerischen Taxonomie relativiert. (Einen anderen, empiristischen Begriff von „idealer Typologie“ verwenden Sokal und Sneath 1963.) — Die an eine Klassifikation zu stellenden Desiderata (vgl. Carvell/Svartvik 1969, 33 ff.; Floodgate 1962; auch Altmann/Lehfeldt 1973, 52 ff.) beziehen sich auf den induktiven Ansatz (a).

3. Statistik

Während sich Klassifikationsverfahren in den Sprach- und Textwissenschaften schon seit langer Zeit (meist allerdings in nichtnumerischer Form) großer Beliebtheit erfreuen, ist die Anwendung statistischer Methoden auf sprachliche Daten nicht so verbreitet, sie wird sogar oft mit Skepsis betrachtet. Auf der anderen Seite wird die Statistik, wo sie zum Einsatz kommt, sehr häufig ohne ausreichenden Sachverstand schematisch benutzt, so daß die Resultate völlig wertlos sein können, ohne daß der Anwender dies bemerkt. Beides beruht auf der ungenügenden Einbeziehung numerischer und statistischer Methodik in die (computer-)linguistische Ausbildung. Der Computerlinguist kann jedoch nicht erwarten, daß ihm linguistisch geschulte Statistiker die Übersetzung seiner Hypothesen in die Sprache der Statistik und die Rückübersetzung der Resultate in eine sprachwissenschaftlich verwertbare Form abnehmen. Er muß sich vielmehr die Methoden und ihre Voraussetzungen so weit zu eigen machen, daß er mit ihnen sicher und idealerweise kreativ umgehen kann. — In der Computerlinguistik eignen sich statistische Verfahren zu folgenden Zwecken:

3.1. Beschreibende Statistik

Die Beschreibung von linguistischen oder textwissenschaftlichen Gegenständen mit Hilfe statistischer Mittel ist u. a. aus zwei Gründen notwendig: (1) In den Fällen, wo es um die Erfassung eines Trends, einer Tendenz, einer stochastischen Abhängigkeit, einer Korrelation usw. geht, sind qualitative Begriffe zu ungenau und teilweise völlig ungeeignet. (2) Die Charakterisierung der Dateneigenschaften sehr vieler (insbesondere

unendlich vieler) Objekte ist oft nur mit quantitativen Begriffen möglich und stets exakter als die mit qualitativen Begriffen. Verteilungsparameter und andere statistische Kenngrößen erlauben es, mit wenigen (meist zwei oder drei) Zahlen das Verhalten von beliebig vielen Objekten zu beschreiben.

Bei der linguistischen Deskription greift man im allgemeinen auf die bekannten Kenngrößen wie Mittelwert, Varianz, Modus, Proportion oder eine Häufigkeitsverteilung zurück, deren Eigenschaften gut bekannt sind (vgl. Kendall/Stuart 1963), nicht selten jedoch ist es nötig, eigene Indices zu entwerfen, um die Ausprägungen einer Eigenschaft linguistischer Daten quantitativ zu erfassen.

Die wichtige Prozedur der Indexbildung in der Computerlinguistik soll hier nur kurz besprochen werden; ausführlich behandelt wird sie von Galtung (1967) und Scheuch/Zehnpfennig (1974). Zugrunde gelegt wird bei der Bildung eines Index ein theoretischer quantitativer Begriff, der einen latenten n-dimensionalen Eigenschaftsraum aufspannt. So spannt etwa der Begriff „Vokalhaltigkeit einer Sprache“ den zweidimensionalen Raum „Vokalverwendung im Text und Vokalverwendung im Lexikon“ auf. Diese theoretischen Begriffe entsprechen dann den Beobachtungsbegriffen „Proportion der Vokale im Text“ und „Proportion der Vokale im Lexikon“, die auf verschiedene Weisen operationalisiert werden (vgl. Altmann/Lehfeldt 1973, 78 ff.). Darunter versteht man die Angabe einer Vorschrift, nach der die gegebenen Begriffe auf ein Zahlenintervall abgebildet werden. Der Index entsteht schließlich durch Reduktion der numerischen Werte für alle beteiligten Dimensionen auf eine einzige Zahl. Diese resultierende Zahl repräsentiert die Ausprägung der untersuchten Eigenschaft. — Obwohl diese Prozedur recht einfach erscheint, kann sie in der Forschungspraxis mit zahlreichen Problemen verbunden sein. Einigen von ihnen begegnet man durch bestimmte Anforderungen an die mathematischen Eigenschaften der Indices. In den Sprachwissenschaften ist es jedoch leider üblich, Zahlen ohne entsprechende Prüfung aufeinander zu beziehen und so die bekannten Quotienten zu bilden. Da auf die so entstandene Deskription fast nie eine Hypothesenüberprüfung vorgenommen wird, treten die Schwächen dieser Indices selten offen zutage (vgl. Altmann 1978 b). —

Ein Index sollte jedenfalls nicht ohne zugrundeliegende Hypothese verwendet oder konstruiert werden. Messungen dienen entweder zur Überprüfung von Hypothesen, also von Sätzen einer Theorie (vgl. Galtung 1967, 242; Kuhn 1961; Koopmans 1947), oder zur Beschreibung von Objekten auf der Grundlage von solchen Sätzen: Ohne eine solche Grundlage sind Indices kaum interpretierbar. Die Stichprobenverteilung verwendeter Indices sollte zumindest asymptotisch bekannt sein, da dies statistische Vergleichbarkeit schafft und eine Beurteilung der beobachteten Werte ermöglicht (vgl. Galtung 1973, 241; Altmann/Lehfeldt 1980, 29). Darüber hinaus soll ein Index maximale Information erbringen, auf den Stichprobenumfang bezogen und stabil, d. h. für kleinste Veränderungen unempfindlich (Galtung 1967, 241) sein. Oft wird auch Einfachheit des Index gefordert; dies ist eine weder notwendige noch hinreichende Bedingung (vgl. Bunge 1961), sie erleichtert jedoch in der Praxis die ersten Schritte beim Aufbau einer Theorie.

3.2. Heuristik

Ähnlich wie taxonomischen Verfahren kommt auch der Statistik eine heuristische Bedeutung zu, nämlich bei der Auffindung von Korrelationen zwischen Variablen. So erhält man etwa mit Hilfe der Faktorenanalyse die Dimension der miteinander korrelierten Eigenschaften eines Gegenstandsbereiches; weder über Art noch über Richtung der (möglicherweise auch gegenseitigen) Abhängigkeit der Variablen ist dagegen mit einem solchen Analyseergebnis etwas ausgesagt. Dennoch bietet die Statistik auch hier explorative Verfahren zur weitergehenden Analyse. Die Verwendung pfadanalytischer Methoden ermöglicht es, komplexe Eigenschaftssysteme aufzustellen und beliebige Konstellationen von Variablen und Abhängigkeitsrichtungen als Hypothesen (zunächst linear) zu überprüfen. Da die Hypothesenbildung in diesem Falle mechanisch (induktiv) erfolgt, ergibt sich allerdings bereits bei nur wenigen Variablen eine unendliche Zahl möglicher Hypothesen, wenn man berücksichtigt, daß außer den rekursiven auch nichtrekursive und blockrekursive Systeme berechnet werden müssen, und vor allem nicht nur lineare Modelle in Betracht kommen. Letztendlich müssen die Hypothesen doch auf deduktive Weise entwickelt werden, selbst wenn man mit Hilfe des maschinellen Verfahrens solche Hypothesen aussondern

könnte, die mit den Daten verträglich sind. Bis zu diesem Punkt jedoch sind induktive statistische Methoden wie die Pfadanalyse erfolgreiche Mittel, um Hypothesenbereiche abzustecken, und sie leisten gute Dienste für Voraussagen und Analysen anhand komplexer Systeme, die ohne sie nicht überschaubar wären.

3.3. Überprüfung von Hypothesen

Die wichtigste Rolle spielt die Statistik in ihrem Einsatz bei der Überprüfung von Hypothesen, wenn man von der statistischen Erklärung absieht, deren Status problematisch und ungeklärt ist (vgl. Stegmüller 1969; 1973; Salmon 1971; Hempel 1965). Zu den wichtigsten Typen von Hypothesenüberprüfungen in der Computerlinguistik zählen (i) der Test, ob eine Stichprobe zu einer Grundgesamtheit mit vorgegebenen Parametern gehört; (ii) der Schluß auf die Parameter einer Stichprobe gegebener Größe aus einer vorgegebenen Grundgesamtheit; (iii) der Test, ob beobachtete Daten durch eine vorgegebene theoretische Funktion oder Verteilung erfaßt werden oder signifikant von ihr abweichen. Dabei werden die Parameter aus den Daten abgeschätzt, und der Test erlaubt eine Aussage darüber, ob die allgemeine Form der theoretischen Funktion oder Verteilung adäquat ist; (iv) der Test auf Gleichheit zweier Stichproben, bei dem geprüft wird, ob sie aus derselben Grundgesamtheit stammen; (v) die Anpassung einer theoretischen Funktion oder Verteilung an beobachtete Daten und Voraussagen für nicht beobachtete Werte durch Interpolation oder Extrapolation; (vi) die Untersuchung zweier Klassifikationen auf Unabhängigkeiten; (vii) die Überprüfung einer Stichprobe auf Erfüllung von Qualitätskriterien durch sukzessives Herausziehen von Elementen.

Obwohl in der Linguistik — wie in anderen empirischen Wissenschaften auch — eigene, spezielle Fragestellungen, Maße und Methoden entwickelt wurden, verläuft die Anwendung einer statistischen Testprozedur grundsätzlich nach einem allgemeingültigen Schema in fünf Schritten (vgl. Altmann 1973): (1) Das Aufstellen einer linguistischen Hypothese. Eine solche Hypothese liegt im allgemeinen in qualitativer Form vor: der Linguist formuliert eine Frage, eine Behauptung, einen Wenn-dann-Satz usw.; seltener hat die zu testende Hypothese selbst quantitativen Charakter (alle allgemeinen wissenschaftlichen Aussagen sind Hypothesen, auch wenn sie schon gut bestätigt wurden). Erfahrene Forscher formulieren ihre Hypothesen bereits im Hinblick auf bestimmte mathematische Modelle. — Alle linguistischen Hypothesen sollten zumindest prinzipiell überprüfbar sein, auch wenn die tatsächliche Durchführung einer empirischen Untersuchung oft enorme praktische Schwierigkeiten mit sich bringt oder die Einbeziehung anderer Disziplinen erfordert, in denen ein Linguist meist nicht kompetent ist. Diese Forderung ist in anderen Wissenschaften selbstverständlich, in den Sprach- und Textwissenschaften sind unüberprüfbare Aussagen eher die Regel: einige sind von vornherein untestabel, andere werden durch die Einführung begrifflicher Vagheiten oder durch Bildung zusätzlicher ad-hoc-Hypothesen dazu gemacht (vgl. dazu Bunge 1969, 261 ff.). Aus dem umfangreichen Problemkreis der linguistischen Hypothesenbildung (s. Stepanov 1980) soll im folgenden nur das eine Rolle spielen, was mit der statistischen Überprüfung von Hypothesen zusammenhängt. (2) Nach der Formulierung einer linguistischen Hypothese muß diese in die Sprache der Statistik übersetzt werden. Zunächst werden die in der Hypothese enthaltenen qualitativen Begriffe metrisiert, d. h. in quantitative Begriffe umgeformt, so daß die statistischen Modelle auf sie angewendet werden können. Darauf wird die Hypothese selbst umformuliert und in eine statistische Form (z. B. eine Nullhypothese) gebracht. Schließlich wird ein mathematisches Modell gewählt, das es ermöglicht, durch geeignete Operationen die Wahrscheinlichkeit zu berechnen, mit der die Hypothese in bezug auf die untersuchten Daten zutreffend ist. — Bemerkenswert ist, daß die Arbeitsweise der Statistik den Linguisten nicht nur zur disziplinierten Begriffs- und Hypothesenbildung zwingt, sondern auch zur Hypothesenbildung anregt. Die Kenntnis mathematischer Modelle hat schon oft zur Entdeckung von Strukturen in den bearbeiteten Daten geführt, nach denen zunächst nicht einmal gesucht worden war. — Bei der Wahl des mathematischen Modells muß darauf geachtet werden, daß die Daten, die zur Überprüfung herangezogen werden, die Bedingungen für die Anwendbarkeit des Modells erfüllen. Wird etwa zur Durchführung eines Tests die Normalverteilung als statistisches Modell in Betracht gezogen, so muß die betroffene Variable unbedingt auch in den Daten normalverteilt bzw. entsprechend transformierbar sein. (3) Der dritte Schritt besteht

aus dem Sammeln und Auswerten von Daten und den Berechnungen, die aus der Wahl des Modells folgen. Die Benutzung von Rechenanlagen zwingt oft zu einer bestimmten Art von Metrisierung und zu einer Bevorzugung bestimmter Modelle sowie Zähl-, Meß- und Testverfahren. — Dieser Schritt wird durch Computereinsatz erleichtert und in vielen Fällen, besonders bei der Verarbeitung von Textkorpora, erst ermöglicht. Andererseits verführen die Möglichkeiten der maschinellen Datengewinnung auch dazu, Datensammlungen ohne vorausgehende Hypothesenbildung anzulegen ('Datengräber'; vgl. Schmitz 1983), die leider allzuoft keinem bekannten Zweck dienlich sind. (4) Das Resultat des dritten Schrittes wird durch eine oder mehrere Zahlen, eine Funktion oder eine Klassifikation dargestellt; es muß nun von einem Statistiker beurteilt werden. Er kann feststellen, ob die Hypothese aufrechtzuerhalten oder abzulehnen ist, ob ein Testresultat als signifikant zu gelten hat, ob eine theoretische Funktion oder Verteilung die Daten adäquat repräsentiert usw. Diese Entscheidung trifft er aufgrund bestimmter Kriterien, die jedoch nicht in den Daten liegen, sondern vom Linguisten vorgegeben sind: es sind Konventionen. Wie solche Konventionen festgesetzt werden, hängt sowohl von der wissenschaftlichen Disziplin als auch von der Art der Fragestellung, der Wichtigkeit des Problems, der Einstellung des Forschers und der Menge der Daten ab (so sind möglicherweise an medizinische Untersuchungen strengere Kriterien anzusetzen als an eine Textanalyse zur Autorenbestimmung). — Ein Testresultat ist in keinem Fall ein Beweis, es führt lediglich zu einer Entscheidung darüber, ob die getestete Hypothese vorläufig beibehalten werden kann, oder ob sie, eventuell mit der Aufforderung, weitere Daten zu sammeln, abgelehnt wird. Statistische Methoden bieten dabei den Vorteil, daß sie das Risiko einer Fehlentscheidung numerisch exakt auszudrücken erlauben. Grundsätzlich wird keine Hypothese endgültig bestätigt, sondern nur entweder verworfen oder als mehr oder weniger gestützt betrachtet. (5) Der letzte Schritt des Verfahrens besteht in der linguistischen Interpretation des Resultats, also in der Rückübersetzung in die Sprache der Linguistik und im Ziehen von Schlüssen, die aus der nun bekräftigten oder abgelehnten linguistischen Hypothese folgen. Für diesen Schritt gelten allein linguistische Maßstäbe. Es muß jedoch betont werden, daß aus Indizes (z. B. Proportionen) ohne statistischen Test keine Schlüsse gezogen werden dürfen. Das Konstatieren eines intuitiv noch so großen Unterschiedes zwischen zwei Indizes etwa hat ohne Test keinen Wert. Andererseits lassen sich aufgrund eines durchgeführten Tests auch nur die Schlußfolgerungen ableiten, die aus der überprüften Hypothese folgen; hat diese selbst weitere Konsequenzen, so müssen auch diese überprüft werden.

Betrachtet man diese fünf Phasen der statistischen Hypothesenüberprüfung im Zusammenhang, so sieht man, daß am Anfang und am Ende der Linguist steht, in dessen Auftrag die dazwischenliegenden Schritte durchgeführt werden. Wenn es sich nicht um einen statistisch geschulten Computerlinguisten handelt, muß er mit einem Statistiker, der eine Hilfsrolle beim Aufstellen der mathematischen Modelle übernimmt, und evtl. einem Programmierer für die Datengewinnung und die Durchführung der Berechnungen zusammenarbeiten. Zwar ist das Ideal denkbar, daß alle nötigen Kenntnisse in der Person des linguistischen Forschers vereinigt sind, doch kommt das in der Praxis kaum vor. Zudem scheinen heute gerade linguistische Probleme eine interdisziplinäre Zusammenarbeit zu fordern.

4. Literatur (in Auswahl)

G. Altmann 1973 · G. Altmann 1978 b · G. Altmann/W. Lehfeldt 1973 · J. P. Benzecri 1970 · H. H. Bock 1974 · M. Bunge 1961 · M. Bunge 1963 · M. Bunge 1969 · R. Carnap 1969 · J. B. Carroll 1960 · W. K. Essler 1971 · G. D. Floodgate 1962 · J. Galtung 1967 · C. G. Hempel 1965 · C. G. Hempel 1974 · W. v. Humboldt 1963 · M. G. Kendall/A. Stuart 1963 · T. Koopmans 1947 · T. S. Kuhn 1961 · W. Lehfeldt/G. Altmann 1975 · G. Maxwell 1971 · K. R. Popper 1973 · W. C. Salmon (Hrsg.) 1971 · E. Sapir 1921 · E. K. Scheuch/H. Zehnpfennig 1974 · U. Schmitz 1983 · R. R. Sokal/P. H. A. Sneath 1963 · H. Spinner 1974 · W. Stegmüller 1973 · W. Stegmüller 1969 · W. Stegmüller 1970 · H. S. Stepanov (Hrsg.) 1980.

Reinhard Köhler/Gabriel Altmann,
Bochum (Bundesrepublik Deutschland)