

## Die 'distinctive features'-Theorie und Probleme der automatischen Erkennung gesprochener Sprache

Von GEORG HEIKE, Marburg a. d. Lahn

Wir möchten auf einige grundsätzliche Probleme eingehen, die sich unter anderem bei unserer Beschäftigung mit der sogenannten automatischen Erkennung gesprochener Sprache stellen. Diese Bezeichnung wie auch der noch ungenauere Terminus 'Spracherkennung' sind offenbar entsprechend zu 'speech recognition' gewählt. G. Ungeheuer schlägt korrekterweise die dem Stand der Forschung angemessenere Bezeichnung 'Identifizierung von Sprachschallmerkmalen' vor. Das Ziel dieser Forschung läßt sich formulieren als die apparative Feststellung von invarianten Sprachschallmerkmalen, die die Signifikanten zu identifizieren und zu unterscheiden gestatten. Insofern gilt es bis zu einem gewissen Grade, nämlich unter Ausschluß von Kontexthilfen, den menschlichen Identifizierungsprozeß durch ein Modell nachzubilden. Das Auffinden und Definieren solcher invarianter Merkmale ist, wie es sich in der Praxis zeigt, weitaus schwieriger, als es die bereits vorliegenden akustisch-phonetischen Untersuchungen erwarten lassen. Zunächst scheint sich die 'distinctive features'-Theorie und die darin formulierten Binärmerkmale in akustischer Definition als weitgehende Näherung an die Lösung des Invarianzproblems anzubieten. Es sind bereits Versuche zur 'speech recognition' zu verzeichnen, die auf diesen Definitionen aufbauen. An dieser Stelle ist es leider notwendig, ausdrücklich darauf hinzuweisen, daß es sich bei der von uns formulierten Aufgabe nicht darum handeln kann, eine apparative Phonemanalyse anzustreben. Noch viel weniger kann beispielsweise der Sonagraph als ein automatischer Phonemdetektor angesehen werden, wie Isaenko in seiner Arbeit über den velaren Nasal im Deutschen ja sehr richtig sagt. Insofern wird der Ansatz von 'distinctive features' zur Beschreibung und weiteren Durchleuchtung eines be-

reits elaborierten Phonemsystems nicht von dieser Feststellung betroffen, nur weil man bei diesem Vorgehen unter anderem auch mit dem Sonagraphen arbeitet. Augenscheinlich basieren solche und andere Mißverständnisse auf der scheinbaren Unversöhnlichkeit von abstrahierender Phonemdefinition und der Untersuchung von Phonemunterscheidungsmitteln im Kommunikationsakt. Der vermeintlich strenge Strukturalist fühlt sich zu einer Ablehnung phonetischer Ansätze verpflichtet und der substanzorientierte Phonetiker und Physiker glaubt, eine apparative Phonemerkennung oder gar Phonemanalyse erreichen zu können. Vielleicht ist auch die übersehene Zweideutigkeit des Terminus 'Phonemanalyse' der Grund für solche Mißverständnisse, da einerseits darunter die Bestimmung des Phoneminventars einer Sprache verstanden wird und andererseits die Zerlegung eines konkreten Textes in phonematische Einheiten, wie sie beispielsweise in einer phonemischen Transkription zum Ausdruck kommt. Selbstverständlich ist der zweite Vorgang nicht möglich ohne den ersten und beide nicht möglich ohne die Basis phonetischer Daten. Ebenso einleuchtend ist es, daß das weitgesteckte Ziel eines sogenannten Phonemerkenners oder -detektors nur durch eine Bewertung apparativ extrahierter Signalmerkmale mit den für die betreffende Sprache maßgebenden distributionellen Regeln möglich sein könnte. Es sind bereits in einigen Arbeiten Ansätze hinsichtlich der Einbeziehung von – wie man sagt – linguistischer Information zu verzeichnen. Für einen Automaten, der in der Lage sein soll, ein einlaufendes, optimal produziertes Sprachsignal in eine Folge von Phonemsymbolen zu codieren, möchten wir den Terminus 'Phonemcoder' vorschlagen. Die Termini 'Erkennung' oder 'Detektor' sind nicht angemessen, da es sich nicht nur um eine Feststellung objektiv vorhandener Signaleigenschaften handeln kann, sondern letztlich um das Ergebnis einer Bewertung solcher Signaleigenschaften nach Maßgabe eines vorgegebenen Codes. Wenn wir die folgende Frage positiv beantworten können, so lassen sich einige Bedingungen für das Modell eines Phonemcoders formulieren. Die Frage lautet: «Ist ein phonetischer Experte in der Lage, ein optimal produziertes Sprachsignal einer ihm unbekanntem Sprache phonemisch zu transkribieren, wenn ihm das Phonemsystem in expliziter Formulierung des Phoneminventars, der phonetischen Beschreibung der zugehörigen Varianten und von Distributionsregeln vorgegeben ist?» Bejahen wir diese Frage – wie wir es gemäß der beanspruchten Geltung phonologischer Analysierergebnisse tun müßten – so könn-

ten folgende Stufen eines Phonemcoders grob formuliert werden: 1. Zerlegung eines einlaufenden Schallsignals in simultane Schichten observabler Merkmale, die für die Beschreibung aller Signale der betreffenden Sprache notwendig sind, 2. Codierung von Merkmalskombinationen oder -komplexen in ein begrenztes Allophoninventar, das in akustischer Definition vorliegt, 3. Codierung so definierter Allophone in ein kleineres Repertoire von Phonemen auf Grund geltender distributioneller Regeln. Die Verwirklichung der Stufe 2 setzt eine Lösung des Invarianzproblems voraus und stellt die Hauptaufgabe der gegenwärtigen Forschung dar. Die Stufe 3 wäre im Gegensatz zur Analyse in 'distinctive features' notwendig, da Fälle möglich sind, wo *ein* Merkmalsbündel nicht für alle positionellen Allophone eines Phonems gilt, zum Beispiel die Allophone des /r/-Phonems im Deutschen und weiter die Codierung des velaren Nasals vor dem Plosiv [k] in das Phonem /n/ und in allen anderen Positionen in die Phonemfolge /ng/, falls man die entsprechende Analyse von *Isačenko* akzeptiert. Der Ansatz von 'distinctive features' ist von großer Bedeutung als Bindeglied zwischen der Phonemdefinition als distinktiver Einheit und den Voraussetzungen, unter denen die Distinktion im Sprechakt garantiert wird. Insofern stellt diese Theorie eine wichtige Basis für die Stufen 1 und 2 unseres Modells dar. Mehrere Probleme, von denen wir im folgenden einige nennen wollen, bedürfen jedoch einer Klärung. Unter anderem ist zu nennen: 1. die relationelle Definition einiger Merkmale, die wohl mit gewissen Bedenken zur Beschreibung eines Phonemsystems ausreicht, wobei alle Phoneme gleichzeitig betrachtet werden, jedoch nicht zur Identifizierung einer einzelnen Phonemrealisation geeignet ist. Ohne Berücksichtigung dieses Problems ergibt sich ein weiteres daraus, daß bei der analytischen Transkription nach erfolgter auditiver Segmentation eine Merkmalsbeschreibung der Segmente vorgenommen wird, während in unserem Fall die Segmentation gerade aus der Merkmalsbeschreibung hervorgehen müßte. Die dichotomische 'distinctive feature'-Definition auf paradigmatischer Ebene findet im kontinuierlichen Schallstrom keine eindeutige Entsprechung etwa in Form eines abrupten Merkmalsprunges. 2. das Invarianzproblem kontextueller und interindividueller Varianten; daraus ergibt sich unter anderem die Unmöglichkeit, die Invarianz von Qualitäten verschiedener individueller Bezugssysteme durch fixe, absolut definierte Merkmale zu erfassen; 3. die nicht ausreichende Entsprechung der Merkmals-

definitionen zu Wahrnehmungsqualitäten. Die Notwendigkeit einer solchen Entsprechung ist von *R. Jakobson* und *M. Halle* mehrmals ausgesprochen worden. Fassen wir die 'distinctive features' mit *G. Ungeheuer* primär als Wahrnehmungsqualitäten auf, so bleibt dennoch die Notwendigkeit, sie nur in ihren genetischen und akustischen Korrelaten quantitativ bestimmen zu können. Aus dem Gesagten ergibt sich die Forderung nach einer umfassenden psychophonetischen Grundlagenuntersuchung, die eine Zuordnung von Signalen zu Wahrnehmungsqualitäten im Sinne auditiv-essentieller Sprachschallmerkmale zum Inhalt hat.

Adresse des Autors: Dr. Georg Heike, Forschungsinstitut für deutsche Sprache, Kaffweg 3, Marburg a. d. Lahn (Deutschland).