

Hochleistung in Millisekunden – Sprechen und Sprache verstehen

Professor WILLEM J. M. LEVELT

Max-Planck-Institut für Psycholinguistik, Nijmegen

Sprache fängt mit dem ersten Schreien des Neugeborenen an; es ist seine Eintrittskarte in die Welt menschlicher Kommunikation. Die Fähigkeit, zu sprechen und Sprache zu verstehen, ist dem Menschen angeboren. Das langgestreckte [ae] – [ae] ist das erste Anzeichen dafür, daß der Neuankömmling mit diesem artspezifischen Kommunikationsvermögen ausgerüstet ist (Abb. 1).

Aber ist das schon Sprache? Wohl kaum. Beim ersten Schreien werden zwar zum ersten Mal Lunge und Stimmbänder zur Erzeugung von Lauten verwendet, aber man kann noch nicht von Artikulation sprechen. Erst die Artikulation, das Modulieren des Klangs der Stimme durch koordinierte Bewegungen von Lippen, Zunge, Gaumen und Kehlkopf, macht wirkliches Sprechen möglich.

Wenn ein Erwachsener spricht, werden viele Dutzende von Muskeln auf die Millisekunde genau gesteuert, um sprachliche Laute zu erzeugen und zu Wörtern und sinnvollen Äußerungen zu verbinden. Sprechen gehört zu den komplexesten psychomotorischen Leistungen, zu denen der Mensch fähig ist. Das Verstehen von Sprache kann man gleichfalls eine kognitive Hochleistung nennen. Was sich im Bereich von Millisekunden abspielt, wenn ein Mensch Sprache hört, ist verblüffend. Und häufig ist sich der Mensch dieser Hochleistung nicht einmal bewußt. Die Prozesse verlaufen größtenteils automatisch, ohne uns zu ermüden. Die Sprache ist unser treuer Vasall, der erst etwas von seinen Geheimnissen preisgibt, wenn er noch nicht oder nicht mehr gut funktioniert. Das Plappern des kleinen Kindes, das zufällige Sichversprechen des Erwachsenen und das schreckliche Experiment der Natur, die Aphasie: Sie sind von jeher wichtige Informationsquellen für den Psycholinguisten gewesen, der die Prozesse des Sprechens und des Verstehens von Sprache untersucht.

Diese Informationsquellen bleiben für unsere Arbeit am Max-Planck-Institut für Psycholinguistik, einem Forschungsinstitut, das sich ausschließlich

Überarbeitete Fassung des Festvortrags bei der Hauptversammlung der Max-Planck-Gesellschaft am 12. Juni 1987 in Hamburg.



Abb. 1: Mit dem Schrei des Neugeborenen fängt Sprache an (Foto: Rothar Reinbader, Kempten).

mit dem Erwerb, der Produktion und der Wahrnehmung von Sprache (language and speech) beschäftigt, eminent wichtig. Ich möchte einige der Prozesse, die sich beim Sprechen und beim Verstehen von Sprache abspielen, erläutern. Vieles davon stützt sich auf die experimentelle Forschung, die in den letzten Jahren an unserem Institut stattgefunden hat, häufig in Zusammenarbeit mit verschiedenen Fachbereichen der Universität Nijmegen.

Der Säugling und die Sprache

Wenden wir uns wieder unserem Neugeborenen zu. Wenn es 6 bis 7 Monate alt ist, beginnt es, die Artikulatoren intensiv zu gebrauchen, ja man kann sagen: zu *üben*. Und zu Beginn des zweiten Lebensjahres steht ihm fast das vollständige artikulatorische Repertoire zur Verfügung.

Wenn ein Baby mit etwa 6 bis 7 Monaten zu artikulieren beginnt, welches Repertoire entwickelt es dann? Imitiert es die Laute der Muttersprache, oder umfaßt das Repertoire auch andere Laute? Die Antwort auf diese Frage ist bekannt. Das Plapper-Repertoire ist anfänglich universell. Alle Babies auf der Welt machen dann ungefähr dasselbe. Erst etwa mit dem ersten Geburtstag beginnt das Kind, sich auf diejenigen Laute zu konzentrieren, die für die jeweilige Muttersprache spezifisch sind. Andere Sprachlaute verschwinden dann langsam aber sicher aus dem Repertoire, und diese können später nur noch mit großer Mühe erneut erworben werden.

Das Artikulieren ist dem Menschen angeboren. Aber wie steht es mit dem Wahrnehmen, mit dem perzeptuellen Unterscheiden zwischen Sprachlauten? Wenn die Muttersprache beim Artikulieren anfänglich nahezu keine Rolle spielt, ist dies dann vielleicht so, weil das Baby überhaupt noch nicht zwischen Sprachlauten unterscheiden kann? Das sollte nicht verwunderlich sein, denn das Unterscheiden zwischen Sprachlauten verlangt vom Gehör eine Hochleistung in Millisekunden. Dies möchte ich am Beispiel der Sprachlaute [ede] und [ete] erläutern. Bei diesen Lauten wird die Vokalisation, das Schwingen der Stimmbänder, kurz unterbrochen. Dies ist gut zu sehen, wenn wir uns den Verlauf der Amplitude bei den Lautfolgen [ede] und [ete] ansehen (Abb. 2).

Abb. 2 zeigt links bei 1 [ede] und rechts bei 10 [ete]. Der einzige Unterschied zwischen [ede] und [ete], wie hier gezeigt, ist die Dauer des stillen Intervalls, der sogenannten Stimmlatenz. Wenn dieses stille Intervall sehr kurz ist, hören wir [ede], wenn es länger ist, hören wir [ete]. Die Grenze

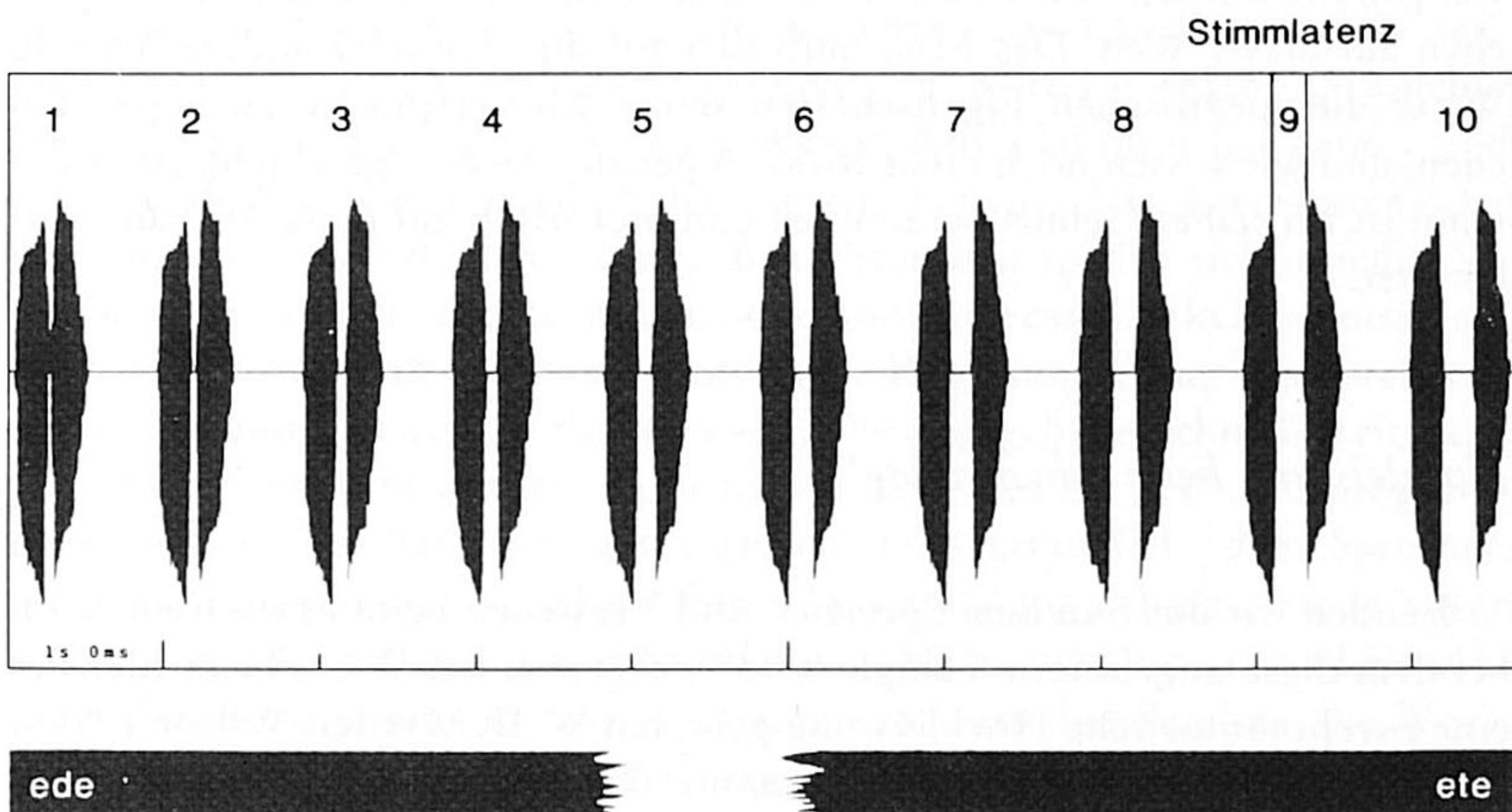


Abb. 2: Kategorischer Übergang von Sprachlaut [ede] zu Sprachlaut [ete]. Die Abb. zeigt den Verlauf der Amplitude über die Zeit in zehn Sprachlauten mit zunehmender Stimmlatenz. Der Umschlagpunkt liegt bei 5 oder 6, bei einer Stimmlatenz von etwa 80 Millisekunden.

zwischen diesen beiden Lauten liegt bei einer Stimmlatenz von etwa 80 Millisekunden; das ist ungefähr zwischen 5 und 6 auf Abb. 2. Wenn wir nun diese 10 Laute, deren Stimmlatenz von links nach rechts stets um 10 Millisekunden zunimmt, nacheinander hören, so hören wir nicht einen allmählichen Übergang von [ede] zu [ete]. Nein, der Übergang ist ziemlich abrupt, wenn wir bei 5 oder 6 ankommen. Alle kürzeren Latenzen hören wir als [d], alle längeren als [t]. Dieses Phänomen heißt *kategorische Wahrnehmung*.

Diese Grenze von etwa 80 Millisekunden bestimmt für den Erwachsenen auch den Unterschied zwischen [p] und [b] und zwischen [k] und [g], d. h., den wichtigen Unterschied zwischen stimmlosen und stimmhaften Plosiven.

Wann lernt nun das kleine Kind diesen genauen Millisekunden-Unterschied? Die Antwort auf diese Frage ist einfach: Das Kind lernt ihn überhaupt nicht. Aus den bahnbrechenden Arbeiten von P. D. Eimas und seinen Kollegen (z. B. Eimas, 1975) zeigt sich, daß diese Differenzierungsfähigkeit schon direkt nach der Geburt vorhanden ist. Das Neugeborene unterscheidet sofort zwischen Laut 5 und Laut 6, die wir als [d] und [t] hören. Es ist jedoch absolut unmöglich für das Baby, die Laute 1 und 2 zu unterscheiden, die wir beide als [d] hören und die den gleichen Unterschied in der Stimmlatenz beinhalten wie die Laute 5 und 6. Entsprechende Forschung ist auch für eine Menge anderer Sprachlaute durchgeführt worden, und die Schlußfolgerung ist immer dieselbe: Sprachrelevante Unterschiede werden durch das Kind sofort erkannt, selbst wenn der Unterschied in der Muttersprache keine Rolle spielt.

Zusammenfassend können wir sagen, daß das Kind beim Entwickeln seiner sprachlichen Fähigkeiten auf ein reichhaltiges und genetisch vorprogrammiertes Repertoire zurückgreifen kann. Sicherlich, es muß auch sehr viel gelernt werden, sonst hätten wir nicht beinahe 6000 verschiedene Sprachen auf dieser Welt. Das Kind muß sich auf die eine oder andere Art und Weise die spezifischen Eigenschaften seiner Muttersprache zu eigen machen, und wir wissen noch nicht wirklich genau, wie dies geschieht. Aber das Kind ist bei seiner Geburt perzeptuell und motorisch auf diese Aufgabe vorbereitet.

Hochleistung beim „shadowing“

Wenden wir uns nun dem Sprechen und Verstehen beim Erwachsenen zu, bei dem diese angeborene Fähigkeit voll ausgereift ist. Wie sehr es hier um eine psychomotorische Hochleistung geht, hat W. D. Marslen-Wilson (1985) mit seiner Untersuchung über sogenannte *close shadowers* gezeigt. Der folgende Auftrag wird „shadowing“ genannt: Eine Versuchsperson hört einen gesprochenen Text, und ihre Aufgabe ist es, diesen Text mitzusprechen. Die Versuchsperson wird gebeten, dies mit so geringer Verzögerung wie möglich zu tun. Diese Verzögerung kann sehr klein sein. Es gibt sogar Versuchs-

personen, die einen Text mit einer mittleren Verzögerung von nicht mehr als 250 Millisekunden nachsprechen können. Sie werden „close shadowers“ genannt. 250 Millisekunden entsprechen der mittleren Dauer einer gesprochenen Silbe. Was alles tut so eine Versuchsperson in diesen 250 Millisekunden? Sie erstellt eine auditive Analyse des fortlaufenden akustischen Signals, sie identifiziert die Sprachlaute und entscheidet vermutlich, welches Wort sie gerade hört (Fehler in dem vorgesprochenen Text werden häufig korrigiert). Dies geschieht, bevor das Wort vollständig ausgesprochen ist, und dazu verwendet sie den Satzzusammenhang. Aufgrund des identifizierten Wortes ruft sie das Programm zum Aussprechen dieses Wortes auf und gibt ihm ein zum Satzzusammenhang passendes Intonationsmuster. Schließlich wird die Artikulationsmotorik ausgelöst. Dies alles in der Zeitspanne einer einzigen Silbe. Lassen sie uns jetzt diese Hochgeschwindigkeitsprozesse des Sprechens und Verstehens etwas näher unter die Lupe nehmen.

Sprechen

Beginnen wir mit den Prozessen beim Sprechen. Ein erwachsener Sprecher mit normaler Ausbildung in unserer Kultur verfügt über einen Wortschatz von etwa 30 000 und manchmal noch viel mehr verschiedenen Wörtern. Während des Sprechens ruft er im Mittel pro Sekunde zwei bis drei Wörter aus seinem Lexikon ab. Es gibt sogar Situationen, in denen ein Sprecher bis zu 7 Wörter pro Sekunde produziert. Diese Wörter geben fast immer die Absicht des Sprechers richtig wieder, und außerdem erscheinen sie fast immer in der richtigen Reihenfolge, im grammatikalisch richtigen Augenblick. Jedes dieser Wörter besteht wiederum aus artikulatorischen Segmenten wie Vokalen und Konsonanten. Im Mittel werden 15 solcher Sprachlaute pro Sekunde produziert. An der Artikulation jeden Lautes sind etwa hundert Muskeln beteiligt, die sich zusammenziehen, entspannen oder ihre Spannung beibehalten müssen. Beim Sprechen spielen sich in jeder Sekunde viele hundert genau miteinander koordinierte Muskelereignisse ab.

Es ist deutlich, daß die psychomotorische Programmierung des Sprechens nicht vollständig bewußt verlaufen kann. Die angegebene Schnelligkeit kann nur dadurch erreicht werden, daß wichtige Teile des Prozesses vorprogrammiert sind und vollautomatisch ablaufen. Dies ermöglicht dem Sprecher, seine Aufmerksamkeit beinahe vollständig darauf zu richten, *was* er sagen will, auf den Inhalt, den er ausdrücken will. Er braucht sich fast überhaupt nicht darum zu kümmern, *wie* er es sagen soll, um den Satzbau, die Wortwahl, und schon gar nicht um die Frage, welchen Vokal oder Konsonant er im nächsten Moment produzieren muß. *Wie* der Inhalt in Sprechen umgesetzt wird, kann der Sprecher fast vollständig an Automatismen überlassen, die wir *Module* nennen. In Abb. 3 sehen Sie, wie unsere Vorstellung vom Aufbau dieses Systems ist:

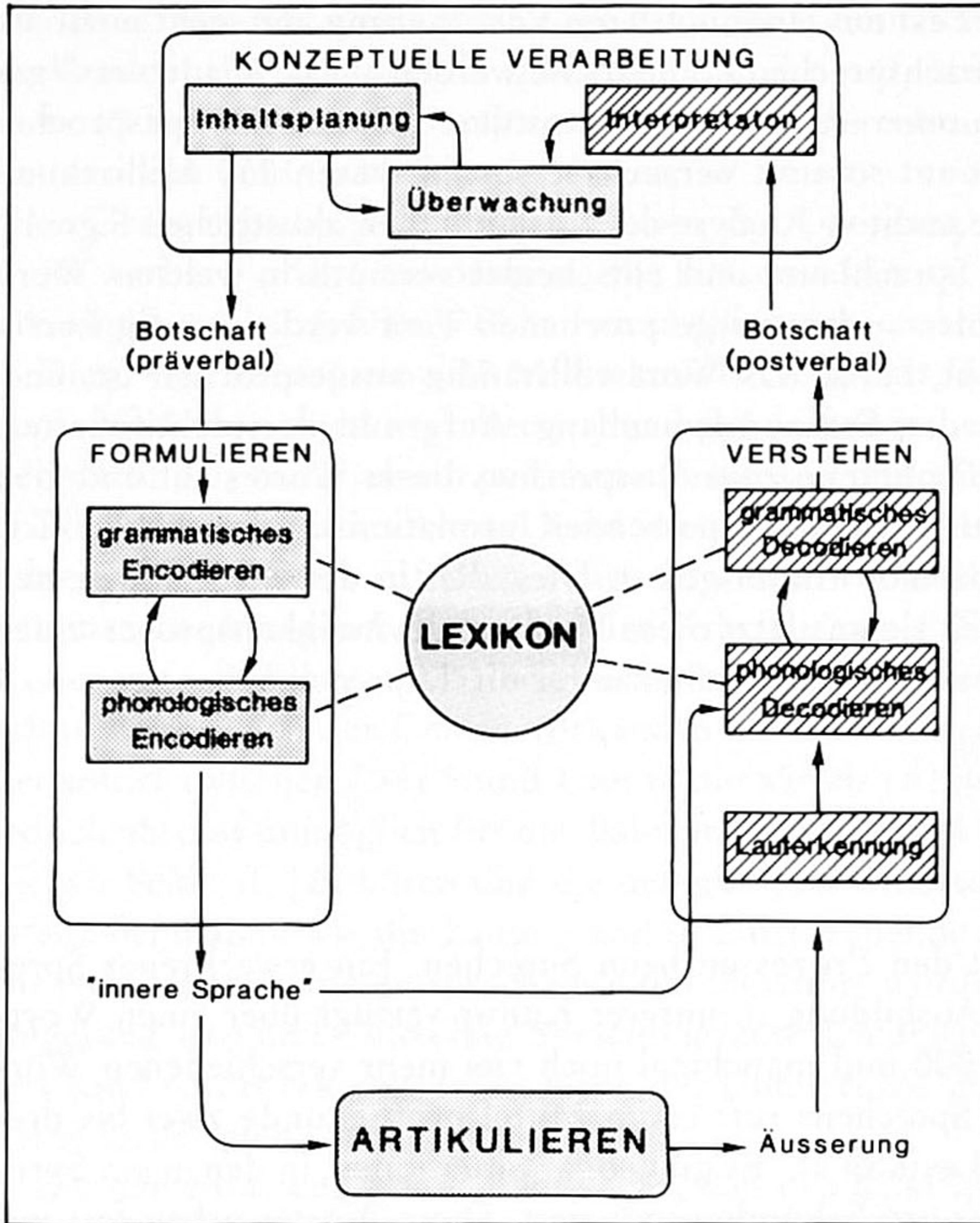


Abb. 3:
Prozeßmodell
des
Sprachbenutzers.

Wie schon gesagt, richtet der Sprecher seine Aufmerksamkeit beinahe ausschließlich auf die Sinnggebung, auf die Planung dessen, was er zum Ausdruck bringen will. Dieser intentionale Prozeß ist in Abb. 3 oben als *Inhaltsplanung* wiedergegeben. *Was* wir zu sagen haben, hängt von unseren Absichten ab, von der Situation, in der wir uns befinden, davon, was unser Gesprächspartner und was wir selbst schon gesagt haben, und von noch vielen anderen Faktoren. Nennen wir das, was der Sprecher zum Ausdruck bringen will, die *Botschaft*. Diese Botschaft muß nun in Sprache umgesetzt werden, und dafür benützt der Sprecher mehrere Module. Zum ersten einen *Formulator* (Abb. 3 links), der aus zwei Teilen besteht. Der erste Teil übernimmt die *grammatikalische Enkodierung*: Er sucht im Lexikon die passenden Wörter auf und fügt sie zu grammatikalisch korrekten Satzstrukturen zusammen. Der zweite Teil übernimmt die *phonologische Enkodierung*. Er entfaltet für jedes Wort und für die gesamte Äußerung einen phonetischen Plan, den wir informell auch manchmal als *inneres Sprechen* bezeichnen. Das nächste Modul, der Artikulator (Abb. 3 unten), übernimmt die Vorbereitung und das Ingangsetzen der Sprechmotorik. Das Ergebnis hören wir als tatsächliche Äußerung. Aber das ist noch nicht alles. Der Sprecher ist auch sein eigener Hörer. Er kann seinem eigenen Sprechen zuhören, so wie er auch ei-

ner anderen Person zuhört. Dafür ist das Modul rechts in Abb.3 verantwortlich. Er kann seine Aufmerksamkeit auch auf sein eigenes inneres Sprechen richten. So kann er das, was er gerade sagt, einigermaßen auf Form und Inhalt hin kontrollieren. Dieses Überwachen ist in Abb.3 oben angegeben.

Die hohe Geschwindigkeit, mit der dieses System Sprache produziert, beruht auf sogenannter *paralleler Verarbeitung*. Die verschiedenen Module arbeiten nicht nacheinander, sondern alle gleichzeitig. Während der Sprecher eine Äußerung artikuliert, denkt er schon darüber nach, was er als nächstes sagen will. Mehr noch: Häufig beginnt das Aussprechen eines Satzes schon, bevor der gesamte Inhalt, geschweige die vollständige Satzform, geplant ist. Wir können uns das etwa vorstellen wie in Abb. 4.

Sobald ein erster Teil des Inhalts geplant ist, beginnt sofort die grammatische Enkodierung. Das heißt, die dem Inhalt entsprechenden Wörter werden gesucht, und es wird mit der Konstruktion des Satzes begonnen. Sobald ein Wort aus dem Lexikon abgerufen ist, beginnt dessen phonologische Enkodierung – ein phonetisches Programm für dieses Wort wird erstellt. Und das Artikulieren kann schon beginnen, sobald das phonetische Programm für die erste Silbe des Wortes erstellt ist. Das Überwachen schließlich folgt dem phonologischen Enkodieren auf dem Fuße. Sobald das Stadium des inneren Sprechens erreicht ist, können Korrekturen angebracht werden.

Statt anhand von experimentellen Resultaten kann jeder dieser Prozesse anhand von Beispielen erläutert werden. Jedes Beispiel zeigt ein Entgleisen oder eine Störung eines dieser Prozesse. Aus solchen Entgleisungen können wir wichtige Schlußfolgerungen über den normalen und meistens vollautomatischen Verlauf dieser Prozesse ziehen.

Beginnen wir mit der Inhaltsplanung. Sie wird normalerweise durch bewußte Aufmerksamkeit gesteuert, aber manchmal entgeht der Aufmerksam-

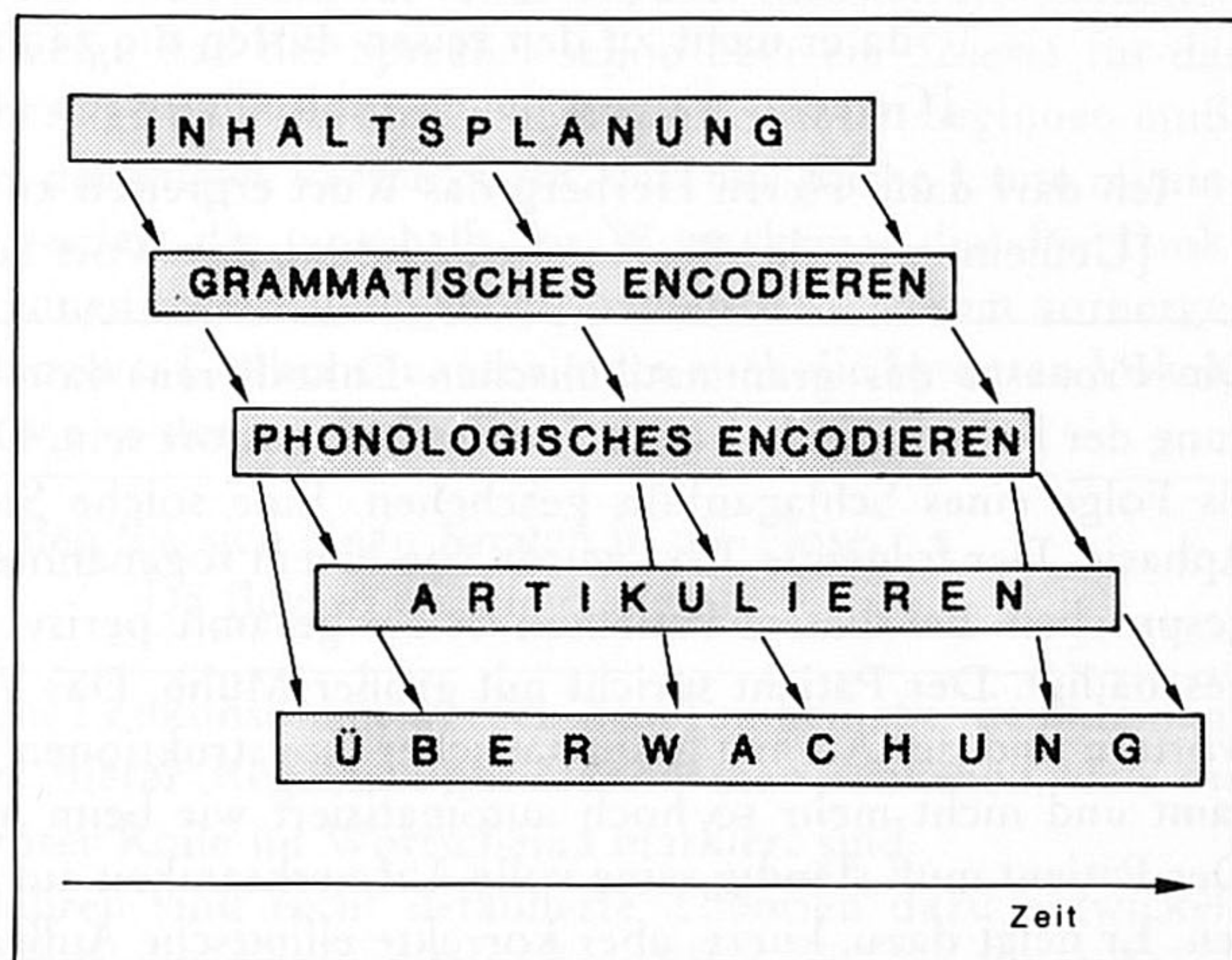


Abb. 4:
Inkrementelle
parallele
Verarbeitung in der
Sprachproduktion.

keit etwas, und der Sprecher läßt einen Gedanken in der Botschaft zu, der eigentlich nicht beabsichtigt ist. Hier ist ein klassisches Beispiel, das Sigmund Freud berichtet hat. Ein junger Mann spricht eine Dame auf der Straße mit den Worten an:

Wenn Sie gestatten, mein Fräulein, möchte ich Sie begleitdigen.

Er dachte offenbar, schreibt Freud, er möchte sie gern begleiten, fürchtete aber, sie mit dem Antrag zu beleidigen. Solche Versprecher, bei denen der Sprecher zwei Intentionen gleichzeitig ausdrückt, kommen auch heutzutage häufig vor. Viktorianische Sexualmoral ist keine Voraussetzung für ihr Auftreten. Sehen Sie sich folgenden Versprecher an:

Was hält die Koalition noch zustande?

Man kann hier eine Freudsche Interpretation zwar nicht ausschließen, aber sie liegt doch nicht so nahe. Auf jeden Fall handelt es sich hier um die Vermischung von zwei Gedanken: zustandebringen und zusammenhalten.

Auch in der nächsten Phase – dem grammatikalischen Enkodieren – kann es zu Entgleisungen und Fehlern kommen. Es kann passieren, daß wir mitten im Satz plötzlich das Wort, das wir benötigen, nicht finden können. Es kommt auch vor, daß ein falsches Wort aus dem Lexikon abgerufen wird, ein Wort, das wir sicher nicht meinen. Hier sind zwei Beispiele:

Sie sitzt also zwischen zwei Türen [Gemeint: Stühlen].

Dann muß man auch dem anderen eine Waffe in die Hand nehmen [Gemeint: geben].

Auch bei der Konstruktion der syntaktischen Struktur eines Satzes kann es zu Fehlern kommen, so wie in diesen Beispielen:

da er nicht zu den reisen dürfen die zählt, ...
[Gemeint war: zu denen zählt, die reisen dürfen]

Ich darf dann Herrn Herberg das Wort ergreifen zu seinem Vortrag
[Gemeint war: das Wort geben / bitten, das Wort zu ergreifen, ...]

Die Prozesse des grammatikalischen Enkodierens können durch Schädigung der linken Hirnhälfte auch dauerhaft gestört sein. Dies kann vor allem als Folge eines Schlaganfalls geschehen. Eine solche Störung nennen wir Aphasie. Der folgende Text wurde von einem sogenannten Broca-Patienten gesprochen. Bei diesem Patienten ist das gesamte perisylvische Gebiet links geschädigt. Der Patient spricht mit großer Mühe. Das Finden von Inhaltswörtern und der Aufbau grammatischer Konstruktionen sind stark verlangsamt und nicht mehr so hoch automatisiert wie beim normalen Sprecher. Der Patient muß ständig seine volle Aufmerksamkeit auf das Sprechen richten. Er neigt dazu, kurze, aber korrekte elliptische Äußerungen zu verwenden.

den, die es ermöglichen, syntaktische Bindewörter wie Artikel, Präposition und Hilfsverben auszulassen.

Ja gestern Fußball angeguckt, a Deutschland gewonnen ... nja hm Holland verloren, aber Olland gut gespielt ... ja Schnee ja ... hm Niesel hm hm hm Regen, sch- hm hm -eschneit.

Es besteht kein Zweifel, daß dieser Patient genau weiß, *was* er ausdrücken will, ungeachtet seines mühseligen Formulierens.

Die Niederländische Forschungsgemeinschaft ZWO (Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek) stellt unserem Institut für die Dauer von 10 Jahren die Mittel für ein Forschungsprojekt zur Verfügung, in dem wir – zusammen mit Neurologen und Sprachpsychologen der Universität Nijmegen – derartige Aphasien erforschen. Über die außerordentlich interessanten Ergebnisse dieses Projektes wurde u. a. von C. Heesch (1985) und A. Friederici (1985) berichtet.

Wenden wir uns nun der phonologischen Enkodierung zu, also dem Aufbau des phonetischen Programms einer Äußerung. Dabei erstellt der Sprecher für jedes Wort zunächst einen phonologischen Rahmen, in den dann die richtigen Sprachlaute eingesetzt werden. Dieses Einsetzen der Sprachlaute mißlingt allerdings gelegentlich, wie z. B. im folgenden Versprecher:

Eine Sorte von Tacher.

Der Sprecher wollte hier eigentlich eine *Torte von Sacher* sagen, aber dabei gerieten die Anfangskonsonanten von *Torte* und *Sacher* durcheinander. Das Bemerkenswerte an diesem Versprecher ist, daß der Sprecher, nachdem er anfangs versehentlich „Sorte“ gesagt hat, später nicht „Acher“ sagt, obwohl er das [s] ja schon vorher gebraucht hat. Anscheinend muß das Wort auf jeden Fall mit einem Konsonanten beginnen und nicht mit einem Vokal. Und dazu wählt er dann den noch nicht verwendeten, aber falschen Konsonanten [t] von *Torte*. Dies zeigt, daß der Sprecher schon über ein *Schema* für das Wort verfügt, demzufolge das Wort mit einem Konsonanten beginnen muß.

Allgemein gilt bei derartigen Versprechern, daß nur solche Laute miteinander verwechselt werden, die innerhalb des Wortschemas dieselbe Funktion haben: Dies können zwei Anfangskonsonanten sein, wie im vorhergehenden Beispiel, oder zwei Endkonsonanten, oder auch die betonten Vokale von zwei Wörtern, wie in den nächsten Beispielen:

Binden Sie sich einen Knaten in die Nase.
Da flug so ein Flogzeug.

Ein Anfangs- und ein Endkonsonant werden dagegen beinahe nie miteinander vertauscht. Aus dieser Regelmäßigkeit kann man schließen, daß die Laute hinsichtlich ihrer Rolle im Wortschema markiert sind.

In den letzten Jahren sind recht detaillierte Theorien dazu entwickelt worden, wie der Sprecher seine phonologischen Programme aufbaut. Diese

Theorien basieren unter anderem auf der statistischen Analyse umfangreicher Sammlungen von Versprechern. Die meisten Versprecher, die ich Ihnen hier vorgeführt habe, stammen zum Beispiel aus dem Corpus von Prof. Manfred Bierwisch von der Akademie der Wissenschaften in Berlin, der unserem Institut als Auswärtiges Wissenschaftliches Mitglied angehört. Die Theorien werden aber auch zunehmend im Laborversuch überprüft. So sind z. B. Methoden entwickelt worden, Versprecher im Labor gezielt hervorzurufen und auf Band aufzunehmen, was natürlich eine viel genauere Analyse zuläßt. Ein von uns im Labor ausgelöster Versprecher ist in Abb. 5 wiedergegeben.

Die Sprecherin beschreibt hier ein Bild mit farbigen Punkten und macht dabei nacheinander zwei Versprecher. Der erste ist ein Fehler bei der Auswahl des richtigen Wortes: Beinahe hätte sie *grüner* statt *roter* gesagt. Aber offenbar hat sie ihre Äußerung gut genug überwacht, um diesen Fehler noch zu entdecken und wählt dann das neue Wort *roter*. Aber nun läuft sie leider beim phonologischen Enkodieren arglos in die Falle: Sie sagt *Prunkt* statt *Punkt*. Der zweite Konsonant von dem abgebrochenen *grün* ist offenbar noch verfügbar und drängt sich als zusätzlicher Konsonant in das Wort *Punkt*.

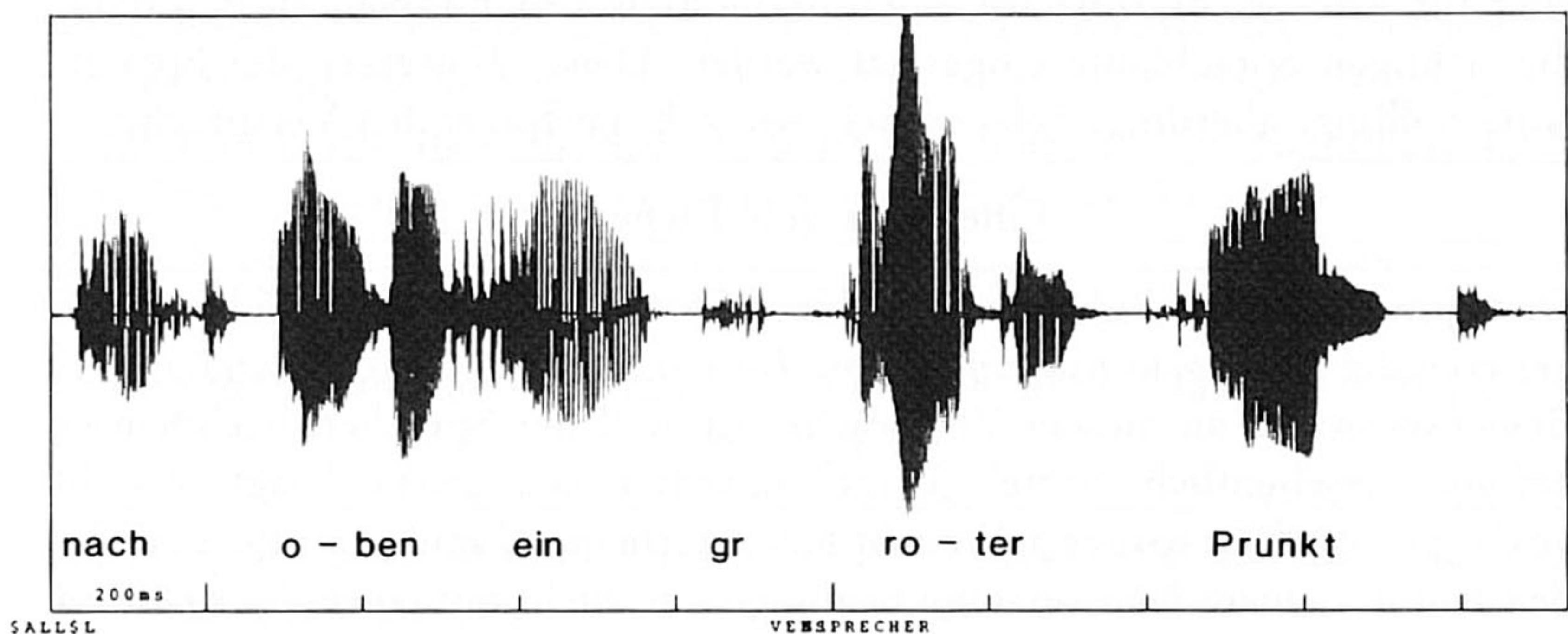


Abb. 5: Entgleisung des phonologischen Enkodierens: Der Versprecher enthält den Sprachlaut „Prunkt“ anstatt „Punkt“.

Gleich kommen wir noch einmal auf derartige Selbstkorrekturen zurück, aber sehen wir uns zunächst kurz das folgende Modul an, das für die Artikulation verantwortlich ist. Auch wenn ein Wort phonologisch richtig geplant ist, kann immer noch die Artikulation mißglücken. Es gibt einige Störungen, die ganz spezifisch dieses Modul betreffen. Die wohl häufigste und bekannteste Störung dieser Art bei vermutlich intaktem Gehirn ist das Stottern. Das Wort ist vollständig geplant, aber der Sprecher kann außer dem Anfangskonsonanten nichts davon über die Lippen bringen. Das Artikulationsprogramm für das Wort muß mehrmals von neuem aufgestartet werden, bis das Wort endlich flüssig gesagt werden kann. Es gibt auch neurologische Veränderungen, sog. *Disarthrien*, die zu spezifischen Störungen der Artikulation führen.

Selbstüberwachung und Selbstkorrektur

Kehren wir noch einmal zu der Vorstellung von Verarbeitungsebenen in der Sprachproduktion (Abb. 3) zurück. Wir haben gesehen, daß der Sprecher sich vor allem auf die Planung des Äußerungsinhaltes konzentriert. Die anderen Prozesse laufen mehr oder weniger automatisch ab. Die zuständigen Module sind alle gleichzeitig am Werk und erstellen verschiedene Teile der Äußerung. Jedes Modul kann dabei auf seine eigene Art entgleisen, und dabei entstehen dann Fehler in der Äußerung, die charakteristisch sind für das jeweilige Modul.

Oft bemerkt der Sprecher, daß er einen Fehler gemacht hat, und verbessert sich. Häufig gelingt das sogar schon, bevor das falsche Wort überhaupt artikuliert worden ist. Dann hört man allenfalls noch ein kurzes Stocken im Sprachfluß. Aber oft hat der Sprecher das kritische Wort auch schon teilweise oder vollständig ausgesprochen, bevor er den Fehler bemerkt. Wann ein Fehler bemerkt wird, hängt unter anderem davon ab, wieviel seiner Aufmerksamkeit der Sprecher auf die Selbstüberwachung verwenden kann. Wir haben an unserem Institut untersucht, wie ein Sprecher sein eigenes Sprechen überwacht und gegebenenfalls korrigiert (W. J. M. Levelt, 1983; W. J. M. Levelt und A. Cutler, 1983). Offenbar wird ungefähr die Hälfte der auftretenden Fehler vom Sprecher bemerkt. Die Entdeckung eines Fehlers geschieht blitzschnell. In 70% der Fälle unterbricht der Sprecher sich schon innerhalb oder unmittelbar nach Ende des kritischen Wortes. Abb. 6 und 7 sind Beispiele dafür. Im ersten Fall fand die Unterbrechung am Ende des falschen Wortes statt (Abb. 6). Im zweiten Fall (Abb. 7 b), war die Sprecherin noch schneller: Hier unterbricht sie das fehlende Wort *gelber* schon in der ersten Silbe.

Aus der Art und Weise, wie die Korrektur durchgeführt wird, können wir allerlei über die Struktur des Sprechprogramms ableiten. Die Sprecherin wiederholt bei der Korrektur in Abb. 7 b – ein roter Punkt – den Artikel *ein*,

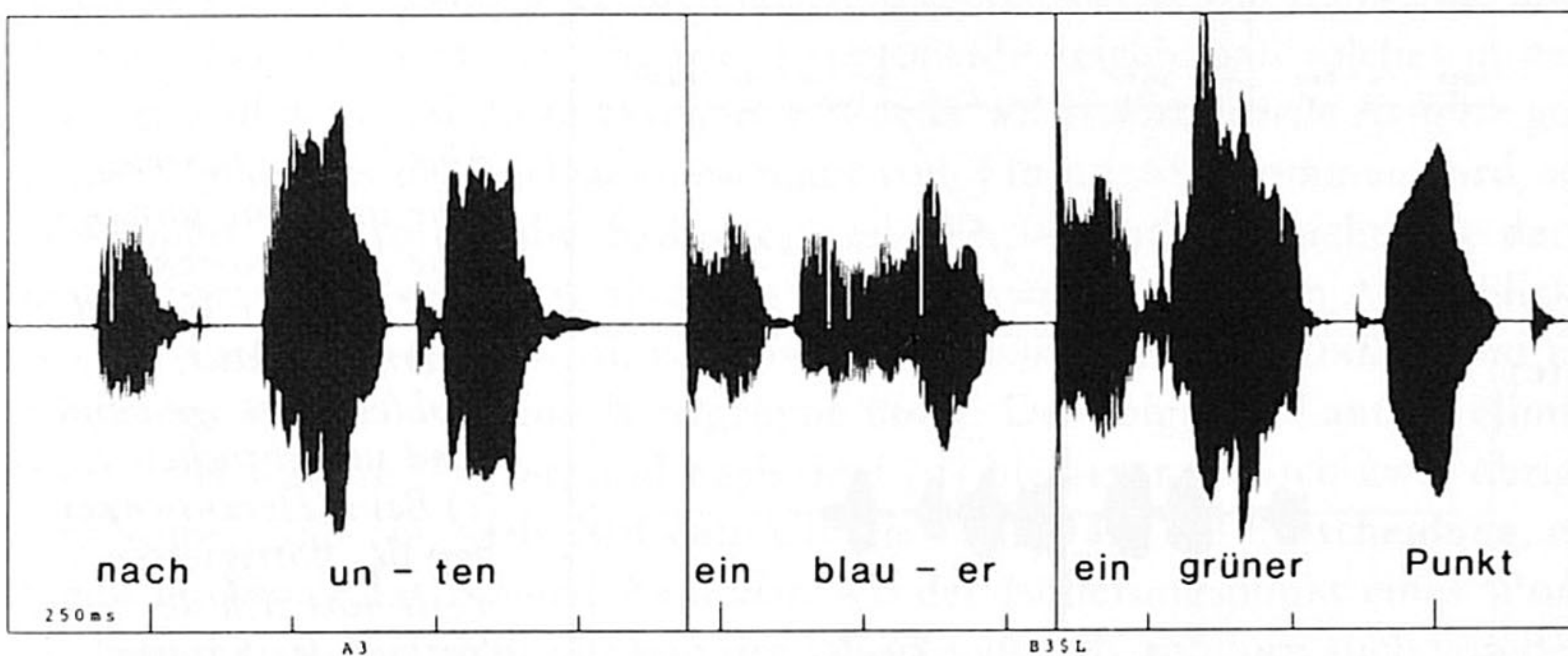


Abb. 6: Selbstkorrektur: „ein blauer“ wird ersetzt durch „ein grüner“.

obwohl er ja richtig war. Wir finden bei Verbesserungen häufig, daß der Sprecher etwas weiter zurückgeht, als eigentlich nötig ist. Reculer pour mieux sauter. Den Schritt zurück kann der Sprecher allerdings nicht beliebig wählen. Die von uns entwickelte Theorie (Levelt, 1983) sagt vorher, welche Schritte möglich sind und welche nicht. Von den vier Fällen, die in Abb. 7 a stehen, wird einer durch die Theorie ausgeschlossen. Dieser Fall ist kursiv geschrieben. Die Sprecherin könnte sagen: *nach oben ein ge - roter Punkt*, oder auch *nach oben ein ge - ein roter Punkt*. Hingegen kann sie nach unserer Theorie nicht sagen: *nach oben ein ge - oben ein roter Punkt*. Und in der Tat, derartige Fälle treten in unseren Daten nicht auf. Der letzte Fall in Abb. 7 a, *nach oben ein ge - nach oben ein roter Punkt*, ist dann wieder möglich. Die Theorie sagt, daß das Programm des Formulators bei der Selbstunterbrechung nicht verlorenggeht. Es bleibt verfügbar, und die Verbesserung wird genau darauf abgestellt. Deshalb kann man das Stück, das der Sprecher verbessert und das in Abb. 7 b mit Raster unterlegt ist, einfach wegschneiden. In Abb. 7 c ist dieses Stück aus der ursprünglichen Äußerung herausgeschnitten worden. Das Resultat hört sich beim Abspielen wieder wie ganz normale Sprache an (Levelt, 1984).

Mit dem Thema der Selbstüberwachung sind wir bereits bei den Prozessen des Sprachverstehens angekommen. Der Sprecher kann seinen eigenen

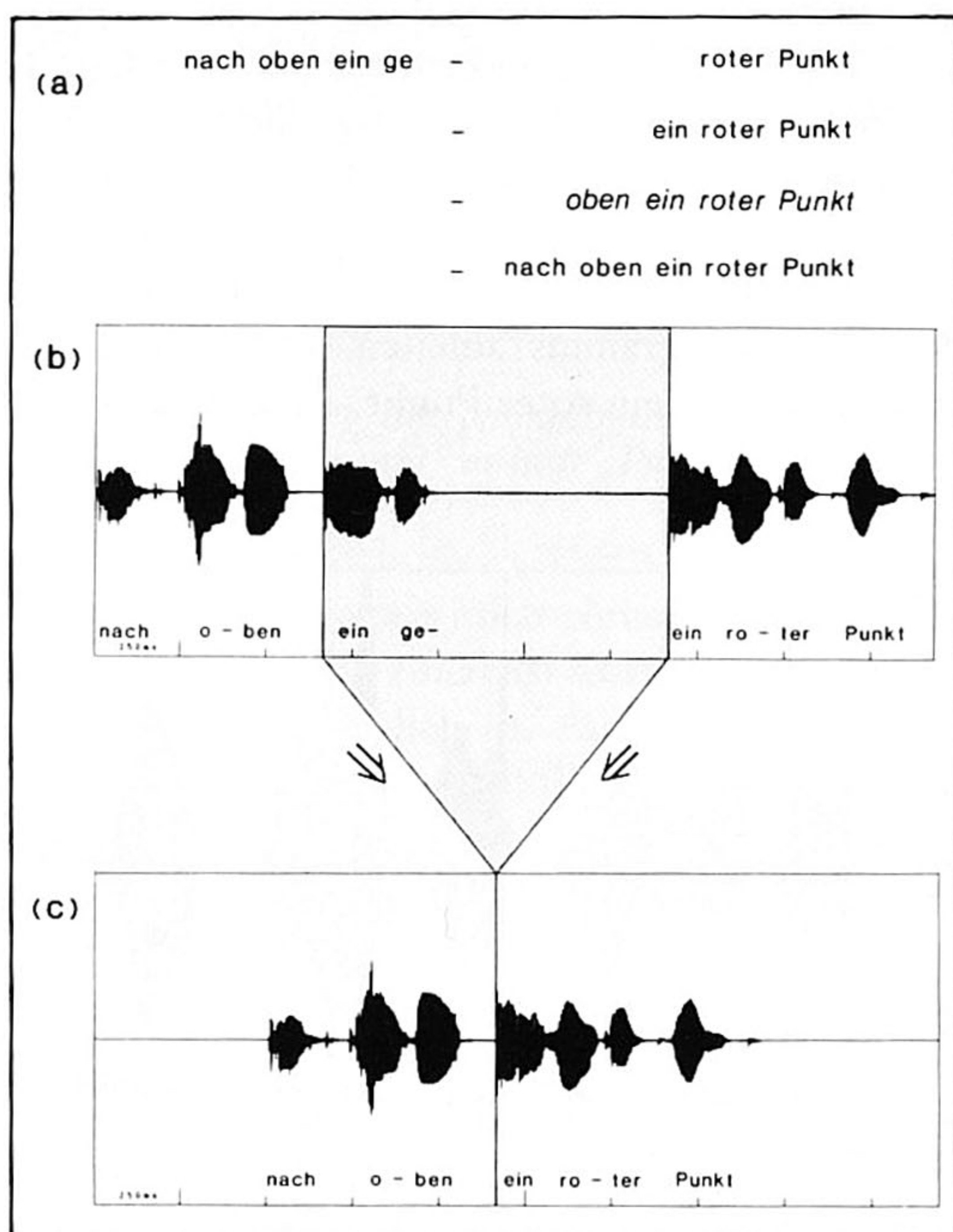


Abb. 7: (a) Drei mögliche und eine unmögliche Selbstkorrektur. (b) Zeitlicher Verlauf einer Korrektur; das Problemwort „gelber“ wird unterbrochen. (c) Beim Herausschneiden des korrigierten Teils hört sich der Verlauf der Sprache wieder normal an.

Äußerungen zuhören, genauso wie er jemandem anderen zuhören kann. Wir haben gesehen, daß die Selbstkontrolle blitzschnell vonstatten geht. Mit ebenso beeindruckender Geschwindigkeit können wir andere verstehen.

Sprache verstehen

Wie schnell erkennt ein Hörer ein gesprochenes Wort? Der Fall des „shadowing“ hat schon gezeigt, daß dies offenbar sehr schnell gehen muß. Ein Wort wird meistens schon erkannt, bevor es überhaupt vollständig ausgesprochen ist. Das gilt besonders dann, wenn das Wort Teil eines gesprochenen Satzes ist. Der Prozeß des Worterkennens wird an unserem Institut intensiv untersucht. Lassen Sie uns zuerst das Erkennen isolierter Wörter betrachten. Eine unserer Methoden, die Worterkennung zu untersuchen, ist das sogenannte *gating*-Verfahren. Dabei hört die Versuchsperson nur ein kurzes Stück des Wortanfangs, zum Beispiel [k], und sie muß raten, um welches Wort es sich handelt. Bei so einem kurzen Stück ist das natürlich unmöglich, aber wir bieten stets längere Teile an, und die Versuchsperson versucht jedesmal wieder, das Wort zu erraten. So entsteht mehr und mehr Gewißheit über das Wort, bis hin zum gelungenen Erraten. Der Zeitpunkt, zu dem das Wort richtig erkannt wird, liegt in diesen Experimenten meistens deutlich vor dem Wortende. So eine Serie immer längerer Wortteile ist in Abb. 8 dargestellt. Sie betrifft das Wort *Kandidat*. Es wird von Versuchspersonen meistens zu Beginn des zweiten [a] identifiziert. Warum? Die an unserem Institut entwickelte Theorie der Worterkennung, die sogenannte Kohortentheorie (W. D. Marslen-Wilson, 1987), erklärt dies folgendermaßen:

Ein Hörer erkennt ein Wort, sobald es sich von allen anderen Wörtern unterscheidet. Die ersten drei oder vier Sprachlaute am Wortanfang aktivieren alle Wörter im mentalen Lexikon des Hörers, die mit diesen Lauten beginnen. In der linken Spalte von Abb. 7 stehen die Wörter, die in unserem Gehirn durch den Wortanfang [kan] aktiviert werden. Darunter befindet sich das Wort *Kandidat*, aber auch allerlei andere Wörter wie *Kannibale* und *Kantate*. Das ist nicht nur Theorie; Experimente zeigen, daß solches in der Tat der Fall ist. Diese Liste aktivierter Wörter wird *wortinitiale Kohorte* genannt. Wenn nun der nächste Sprachlaut vom Hörer aufgenommen wird, so verkleinert sich sofort die Kohorte, weil alle Wörter, die nicht mit dem neuen Sprachlaut vereinbar sind, aus ihr verschwinden. In dem Augenblick, wo der Laut [d] gehört wird, verschwinden *Kannibale* und *Kantate*, und es bleibt nur noch eine kleine Wortgruppe übrig. Der folgende Laut [i] eliminiert zwei weitere Wörter, und nach dem [d] bleiben nur noch zwei übrig, ein Nomen und ein Verb. Mit dem letzten Vokal fällt die Entscheidung, es bleibt nur noch das Nomen *Kandidat*. Wo der Isolierungspunkt eines Wortes liegt, hängt also nicht nur von dem Wort selbst ab, sondern auch von der Menge anderer Wörter, die mit denselben Sprachlauten beginnen.

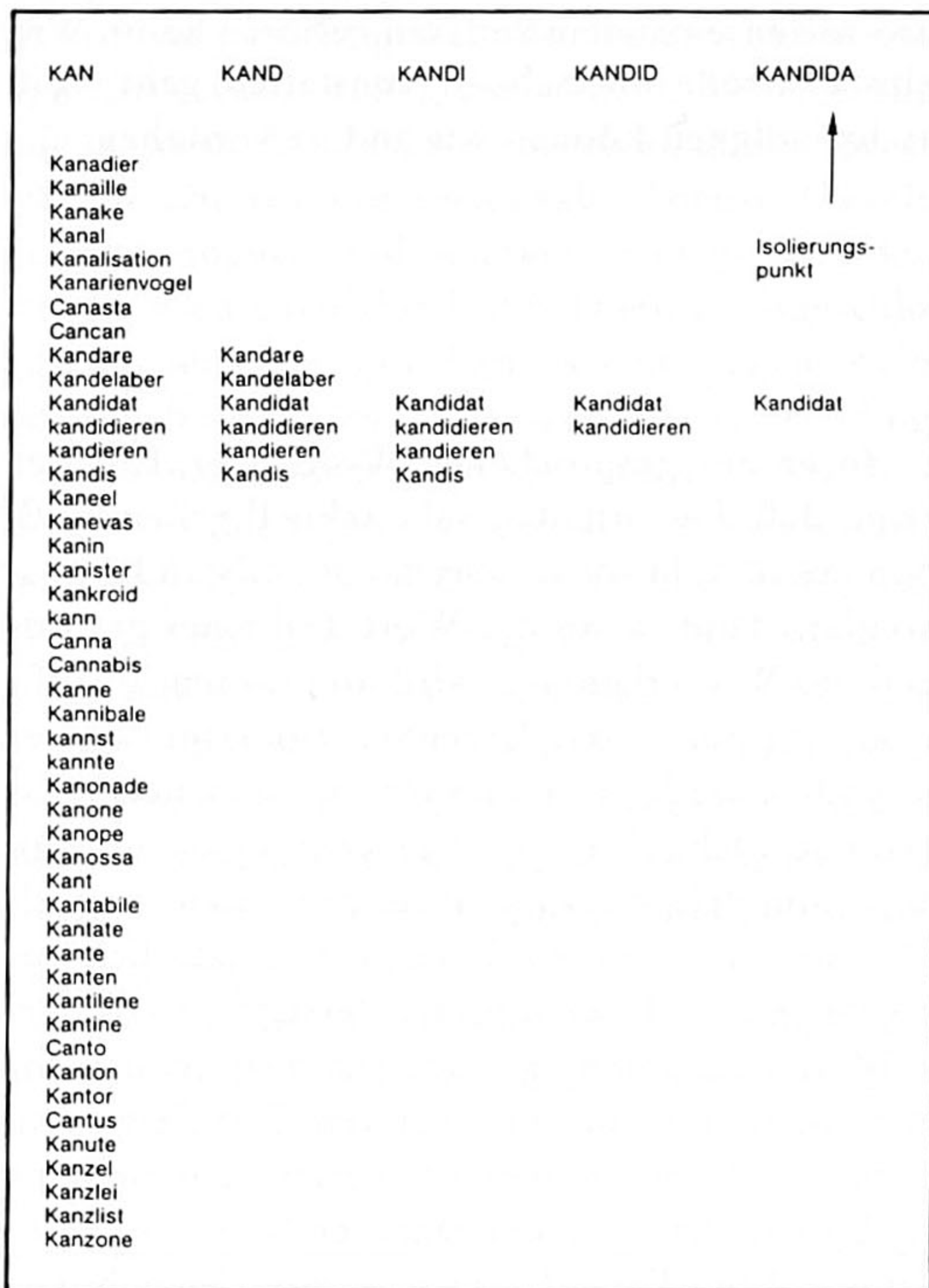


Abb. 8: Kohortenentwicklung für „Kandidat“. Links die wortinitiale Kohorte, alle Wörter die mit dem Anfangslaut [kan] kompatibel sind. Nach rechts: Verkleinerung der Kohorte durch neu hinzukommende Sprachlaute.

Der Isolierungspunkt ist der früheste Zeitpunkt, zu dem ein allein stehendes Wort überhaupt erkannt werden *kann*. Es ist aber noch die Frage, ob es da auch *tatsächlich* erkannt wird. Aus der Analyse von Erkennungszeiten ergibt sich, daß der Isolierungspunkt eine sehr gute Annäherung an den tatsächlichen Erkennungspunkt darstellt. Das bedeutet also, daß das Worterkennungssystem in zeitlicher Hinsicht optimal funktioniert – arbeitete es noch schneller, würde es die Wörter regelmäßig falsch identifizieren. Das System arbeitet also genauso schnell, wie es die einlaufende Information zuläßt (W. D. Marslen-Wilson und L. K. Tyler, 1981).

Soviel zum Erkennen alleinstehender Wörter. Wenn Wörter im Satzzusammenhang auftreten, geht der Erkennungsprozeß noch etwas schneller vonstatten. Mit Hilfe des Kontextes kann die wortinitiale Kohorte schneller verkleinert werden. Wenn es sich aus dem Satzzusammenhang ergibt, daß als nächstes Wort ein Nomen kommen muß, können alle Verben sofort aus der Kohorte ausgeschlossen werden. In unserem Beispiel kann der Hörer sich dann also schon entscheiden, wenn er [Kandi] hört; *Kandidat* ist dann ja das einzige Nomen, das noch in der Kohorte verbleibt. Daß derartige Kontexteffekte tatsächlich auftreten, ist in verschiedenen Experimenten gezeigt worden (z. B. L. K. Tyler und W. D. Marslen-Wilson, 1986). Sowohl

der syntaktische als auch der semantische Zusammenhang können die Eingrenzung der Kohorte beschleunigen. Allerdings scheint der Satzzusammenhang keinen Einfluß auf den Umfang der wortinitiellen Kohorte zu haben, so wie sie links in Abb.7 steht. Diese wird ausschließlich durch die akustische Information des Wortanfangs bestimmt.

Akustisch kann man derartige Effekte leicht vorführen. Insbesondere der sog. „Warren-Effekt“ ist dazu sehr geeignet (R. M. Warren, 1970). Eine einigermaßen äquivalente visuelle Darstellung solcher Kontexteffekte ist aber auch möglich. In Abb.9 sind zwei Sätze geschrieben, allerdings ist die Abbildung leider von Tintenflecken beschmutzt. Die Sätze sind aber noch gut lesbar: *Eine Abteilung zu leiten macht viel Spass* und *Kriege bringen entsetzlich viel Leiden in die Welt*.

Das Wort *leiten* im ersten Satz ist wegen der Beschmutzung dem Wort *Leiden* im zweiten Satz visuell identisch. Trotzdem werden sie als unterschiedliche Wörter identifiziert, und das kann nur durch den Satzzusammenhang veranlaßt sein¹.

Auch akustischer Kontext erlaubt es uns, ein Wort frühzeitig zu identifizieren, und zwar auch dann, wenn die akustische Information an sich nicht ausreichend ist. Dieser Kontexteffekte sind wir uns gewöhnlich nicht bewußt. Wie das Sprechen ist das Sprachverstehen nämlich ein Prozeß, der sich größtenteils automatisch vollzieht und nicht unserer bewußten Kontrolle unterliegt. Man kann fast sagen, daß es mehr *für* uns getan wird als durch uns. Erst wenn diese automatischen Prozesse gestört werden, merken wir, wie unabhängig wir von ihnen sind.

Gerade weil diese Automatismen der Sprachwahrnehmung in ihrem Verlauf der bewußten Kontrolle entzogen sind, gibt es feste Rahmenbedingungen für ihr Funktionieren. Eine dieser Rahmenbedingungen betrifft die Intonation. Um Wörter im Satzkontext zu erkennen, sollte der Intonationsverlauf eine gewisse Kontinuität aufweisen. Mit der Satzmelodie erzeugt der Sprecher eine bestimmte Gruppierung der Wörter zu sinnvollen syntaktischen Einheiten. Der Hörer benutzt die Satzmelodie, um diese sinnvolle



Abb. 9: Effekt des Bedeutungskontextes bei der Worterkennung. Beim Lesen dieser Sätze werden die beschmutzten Wörter „leiten“ und „leiden“ trotz physischer Identität korrekt unterschieden.

¹ Während des Festvortrags wurde mit diesen Sätzen der Warren-Effekt vorgeführt.

Gruppierung der Wörter zu erkennen und zu interpretieren. Die Satzmelodie ist darüber hinaus so eine Art Trägerwelle für die normalen Worterkennungsprozesse. Das automatische Erkennen der Wörter kann vollständig mißlingen, wenn die Intonation bei jedem neuen Wort unterbrochen wird und neu anfängt. Was in der schriftlichen Sprache gerade eine notwendige Bedingung für das Sprachverständnis ist – die Isolierung der Wörter –, führt in der gesprochenen Sprache das Verständnis in die Irre. Wenn man einen Satz mit Hilfe eines Spracheditierungsprogrammes aus einzelnen klar ausgesprochenen Wörtern zusammenreihet, so ist er äußerst schwierig oder gar nicht zu verstehen. Es fehlt dann die notwendige „Durchverbindung“ der Intonation. Ein gleichartiges Phänomen kann man im schriftlichen Sprachverständnis erzeugen, wenn man die Wortgrenzinformation zerstört. Das kann der Leser beim nächsten Satz mitvollziehen:

Wenne in evors tel lungver worrena usge drück twirdsof olgt dersch lus
snochgarn ich tdass siea uchver wor renged achtwor densei.

So kann unser treuer Vasall seine genaue Millisekundenarbeit nicht vollziehen. Nur das Denken kann uns noch behilflich sein². Dieser Satz wurde 1809 von Heinrich von Kleist geschrieben – und er stimmt offenbar noch immer.

Literatur

Eimas, P. D.: Speech perception in early infancy. In: L. B. Cohen und P. Salapatek (Eds.), *Infant perception: From sensation to cognition*, Vol. II. Academic Press, New York 1975, 193–231.

Friederici, A.: Levels of processing and vocabulary types: Evidence from on-line comprehension in normals and agrammatics. *Cognition*, 19, 133–166 (1985).

Heeschen, C.: Agrammatism versus paragrammatism: a fictitious opposition. In: M.-L. Kean (Ed.), *Agrammatism*. Academic Press, New York 1985, 207–248.

Kleist, H. von: Über die allmähliche Verfertigung der Gedanken beim Reden. In: Heinrich von Kleist. *Sämtliche Werke und Briefe*, Bd. 2. Carl Hanser Verlag, München, 2. Aufl. 1961, 1044–1051.

Levelt, W. J. M.: Monitoring and self-repair in speech. *Cognition* 14, 41–104 (1983).

Levelt, W. J. M.: Spontaneous self-repairs in speech: Processes and representations. In: M. P. R. van den Broeke and A. Cohen (Eds.), *Proc. Tenth Int. Congress of Phonetics*. Foris, Dordrecht 1984, 105–117.

Levelt, W. J. M. and *A. Cutler*: Prosodic marking in speech repair. *Journal of Semantics* 2, 205–217 (1983).

2 Oder eine Fußnote. Der unzerstörte Satz lautet: *Wenn eine Vorstellung verworren ausgedrückt wird, so folgt der Schluß noch gar nicht dass sie auch verworren gedacht worden sei*. Mit diesem Satz wurde während des Vortrages das akustische Phänomen demonstriert.

Marslen-Wilson, W. D.: Speech shadowing and speech comprehension. *Speech Communication* 4, 55-73 (1985).

Marslen-Wilson, W. D.: Functional parallelism in spoken word-recognition. *Cognition* 25, 71-102 (1987).

Marslen-Wilson, W. D. and L. Tyler: Central processes in speech understanding. *Philosophical Transactions of the Royal Society, London, Series B* 295, 317-322 (1981).

Tyler, L. K. and W. D. Marslen-Wilson: The effects of context on the recognition of polymorphemic words. *Journal of Memory and Language* 25, 741-752 (1986).

Warren, R. M.: Perceptual restorations of missing speech sounds. *Science* 167, 392-393 (1970).