

Peter Hagoort

Die einzigartige, größtenteils aber unbewusste Fähigkeit der Menschen zu sprachlicher Kommunikation¹

Zusammenfassung

Kommunikation mit Hilfe einer natürlichen Sprache gibt es ausschließlich beim Menschen. Sie erfolgt in unterschiedlicher Form: durch Sprachen, Zuhören, Schreiben, Lesen oder Gebärdensprache. In allen diesen Fällen haben wir es mit komplexen kognitiven Abläufen zu tun, die aus einer Kaskade von Verarbeitungsprozessen bestehen. Im Verlauf dieser Verarbeitung werden zugehörige Informationen aus dem Gedächtnis abgerufen. Zu den Informationsquellen gehören dabei begriffliche Kenntnisse, Kenntnisse über die Syntax der Wörter und Sätze sowie Kenntnisse über Lautmuster der Sprache oder über die orthografischen Regeln des Lesens und Schreibens. Die verschiedenen Abruflvorgänge laufen mit hoher Geschwindigkeit ab und werden millisekundengenau koordiniert. Neue bildgebende Verfahren zur Darstellung des Gehirns ermöglichen heute die Untersuchung der neuronalen Grundlagen unserer Sprachfähigkeit. Wie sich dabei herausstellt, erfordert schon die relativ einfache Aufgabe, Einzelwort-Äußerungen hervorzu- bringen, die gemeinsame Tätigkeit mehrerer Gehirnareale, vorwiegend solcher aus der linken Hemisphäre. Im letzten Abschnitt werden kurz die Folgerungen für eine Persönlichkeitstheorie erörtert.

1. Einleitung

Der *Homo sapiens* hat als einzige Spezies im Verlauf seiner Evolution ein Kommunikationssystem entwickelt, in dem eine endliche Zahl von Symbolen in Verbindung mit einer Reihe von Gesetzmäßigkeiten zu ihrer Kombination die Erzeugung unendlich vieler Ausdrücke ermöglicht. Dieses System

¹ Dieser Beitrag ist eine überarbeitete Fassung des Vortrages, den Prof. Dr. Peter Hagoort auf dem Symposium „Das gläserne Gehirn: Moderne Bildgebung – Der Schlüssel zum menschlichen Bewusstsein?“ des Wissenschaftszentrums Nordrhein-Westfalen am 4. Dezember 2002 in Düsseldorf gehalten hat. Übersetzung: Dr. Sebastian Vogel, Köln.

einer natürlichen Sprache versetzt die Angehörigen unserer Spezies in die Lage, Gedanken nach außen zu tragen und innerhalb der sozialen Gruppe, durch die Erfindung der Schrift aber auch innerhalb der Gesamtgesellschaft auszutauschen. Sprechen und Sprache sind sehr wirksame Mittel zur Förderung des sozialen Zusammenhalts in Gesellschaften, in denen das Kraulen – die bevorzugte Methode zur Herstellung von Bindungen bei unseren nächsten genetischeren Verwandten, den Primaten der alten Welt – wegen der Gruppengröße diese Funktion nicht mehr erfüllen kann (Dunbar 1998; Levelt 1999).

Grundlage der Leistungsfähigkeit unserer menschlichen Sprache ist ihr dreiteiliger Aufbau (Jackendoff 1997; 1999). In diesem Aufbau werden sprachrelevante Informationen in mindestens drei unterschiedlichen Darstellungsformaten codiert: einem für die Bedeutung, einem für die Syntax und einem für die Geräuschstrukturen der Wörter und Äußerungen. Durch Überlagerung dieser Darstellungsstrukturen werden die begrifflichen Strukturen, die den Inhalt einer Aussage des Sprechers festlegen, als lineare Abfolge von Sprachlauten (Sprechen) ausgedrückt. Andererseits werden beim Zuhören begriffliche Strukturen aus einer linearen Kette von Sprachlauten abgeleitet. Bei diesem Kartierungsprozess stellt die Kombination der Einheiten zu hierarchischen Satzstrukturen (Syntax) die notwendige Verbindung zwischen begrifflichen Strukturen und Schallstrukturen (phonetischen Strukturen) her.

Entsprechend der Arbeitsteilung zwischen den verschiedenen Wissenschaftsgebieten, die sich mit der Sprachfähigkeit des Menschen befassen, haben Linguisten die Aufgabe, die beteiligten Darstellungsstrukturen zu benennen; die Psycholinguisten befassen sich mit der Frage, wie diese Strukturen beim Zuhören und Sprechen zugänglich gemacht und genutzt werden. Die neurowissenschaftliche Kognitionsforschung schließlich steht vor der Frage, wie das Gehirn die menschliche Sprache möglich macht, und sie will herausfinden, welches räumlich-zeitliche Profil neurophysiologischer Aktivitäten dem Sprechen (Schreiben) und Zuhören (Lesen) zu Grunde liegt.

Wenn wir uns ein genaueres Bild davon machen wollen, welche Kenntnisse über die Sprachfähigkeit der Menschen wir diesen verschiedenen Wissenschaftsgebieten verdanken, müssen wir zunächst klären, was der umfassende Begriff „menschliche Sprachfähigkeit“ eigentlich bedeutet. Er bezeichnet die Gesamtheit folgender komplexer, untereinander verwandter, gleichzeitig aber auch unterschiedlicher Fähigkeiten: Sprechen und Zuhören, Lesen und Schreiben sowie unter Hörbehinderten den Gebrauch der Gebärdensprache. Jede dieser Fähigkeiten setzt voraus, dass andere Darstellungsstrukturen im Gedächtnis in Echtzeit zugänglich gemacht und genutzt

werden. Die kognitiven Strukturen dieser Fähigkeiten bestimmen darüber, welche Darstellungsstrukturen beteiligt sind und wie sie in Echtzeit funktionieren. Die neuronale Architektur legt fest, wie diese Fähigkeiten sich in der physischen Struktur des menschlichen Gehirns im Einzelfall ausprägen. Dabei sollte man aber nicht vergessen, dass die Unterscheidung zwischen kognitiver und neuronaler Architektur eine Idealisierung darstellt. Es ist nützlich, in erster Näherung eine Unterscheidung zwischen Berechnungen in symbolischen Begriffen (kognitiver Architektur) und neurophysiologischen Begriffen (neuronaler Architektur) vorzunehmen, in einer umfassenden neurowissenschaftlichen Erforschung der Sprache sollten diese beiden Ebenen jedoch ineinander fließen.

In diesem Kapitel werde ich zunächst die kognitive Architektur einer derartigen Fähigkeit etwas genauer erörtern und mich dann mit der neuronalen Architektur befassen. Am Ende werde ich daraus einige Schlussfolgerungen für eine Persönlichkeitstheorie ableiten.

2. Die kognitive Architektur

Ein zentraler Bestandteil unserer Sprachfähigkeit ist das mentale Lexikon. Das mentale Lexikon ist ein Teil des deklarativen Gedächtnisses und beinhaltet die Kenntnisse, die ein Sprachnutzer über die Wörter seiner Muttersprache besitzt. Nach Schätzungen haben die Sprecher einer Sprache einen aktiven Wortschatz von mindestens 40.000 Wörtern (eine allgemeine Einführung über das mentale Lexikon findet sich bei Aitchison 1997). Der Sprecher weiß, was diese Wörter bedeuten und wie sie klingen. Außerdem erkennt er ihre syntaktischen Eigenschaften, beispielsweise die Wortart (Substantiv, Verb, usw.). Alle diese lexikalischen Informationen werden aus dem Gedächtnis sehr schnell abgerufen. Ein durchschnittlicher Sprecher bringt zwei bis drei Wörter pro Sekunde hervor. Dies erfordert nicht nur den Abruf von Informationen über diese Wörter aus verschiedenen Quellen, sondern auch die koordinierte Aktivierung zahlreicher Muskeln, die an der Artikulation der gesprochenen Sprache beteiligt sind. Am Sprechen wirken rund 100 Muskeln mit, deren Innervierung millisekundengenau koordiniert werden muss. Trotz der großen Komplexität und Schnelligkeit dieser kognitiven Tätigkeit sind Sprecher sehr genau und machen im Durchschnitt noch nicht einmal einen Fehler je 1000 Wörter. Dennoch liefern solche gelegentlichen Sprechfehler wichtige Aufschlüsse über die Struktur des Sprachprozesses. So werden beispielsweise manchmal Laute zwischen verschiedenen

Wörtern ausgetauscht („heft lemisphere“ an Stelle von „left hemisphere“, Fromkin 1973) oder vorweggenommen und zu früh hervorgebracht („it’s a meal mystery“ an Stelle von „it’s a real mystery“, Fromkin 1973). Solche Sprechfehler machen deutlich, dass die Wörter im Gedächtnis nicht als Einheiten gespeichert werden, sondern jedes Mal, wenn wir ein Wort hervorbringen, aus ihren einzelnen Phonemen neu zusammengesetzt werden müssen. Dieser Vorgang des Zusammensetzens läuft manchmal fehlerhaft ab, so dass Laute im Umfeld des falschen Wortes auftauchen.

Als Beispiel wollen wir betrachten, welche Vorgänge sich abspielen, wenn wir ein bestimmtes Bild auf der Netzhaut als Johannes Paul II. erkannt haben und dann die Lautfolge „Papst“ artikulieren.

Das Sprechen beginnt damit, dass wir den begrifflichen Inhalt der Äußerung festlegen. Dies kann wie in unserem Beispiel durch einen visuellen Input erfolgen, in vielen Fällen ist jedoch ein innerer Input der Ausgangspunkt, beispielsweise die Absicht des Sprechers, einen bestimmten Gedanken mitzuteilen. Unabhängig von dem Auslöser der begrifflichen Festlegung einer Äußerung muss der Sprecher aus den Kenntnissen in seinem Gedächtnis einen bestimmten Begriff oder eine Abfolge von Begriffen auswählen und dann eine Entscheidung über die Art der Äußerung treffen – ein Gedanke kann beispielsweise als Behauptung ausgedrückt werden, aber auch als Frage, mit oder ohne Ironie usw. In unserem Beispiel kann der Sprecher sich beispielsweise entscheiden, ob er „Johannes Paul der Zweite“ oder „der Papst“ sagt, um nur zwei mögliche Alternativen zu nennen.

Dieses Auswählen und Festlegen von Begriffen geht dem eigentlichen Formulieren voraus, bei dem präverbale begriffliche Strukturen den Abruf der erforderlichen linguistischen Strukturen auslösen, mit deren Hilfe der Gedanke als Abfolge von Sprachlauten ausgedrückt wird. Dazu müssen zwei ganz unterschiedliche Arten linguistischer Informationen abgerufen werden: Die eine legt die Eigenschaften des Wortklanges fest, die andere betrifft die grammatikalischen Eigenschaften des Wortes.

Jedes Wort im mentalen Lexikon ist mit der syntaktischen Information über dieses Wort assoziiert (Levelt 1989, 1999; Roelofs 1992, 1993). Dieser zweite Informationstyp wird auch als Lemma-Information bezeichnet. Lemmata legen die syntaktischen Eigenschaften der Wörter fest, beispielsweise die Wortart (Substantiv, Verb, Adjektiv usw.). Bei Sprachen, die das Geschlecht kennzeichnen, wird hier für Substantive auch das grammatikalische Geschlecht festgelegt (*Pferd* ist zum Beispiel im Französischen männlich, im niederländischen dagegen Neutrum). Lemmata von Verben enthalten Informationen über den syntaktischen Rahmen (die Argumentstruktur) und

über die thematische Bedeutung des syntaktischen Arguments (die thematische Struktur). Das Lemma des Wertes *donate* besagt beispielsweise, dass es eine substantivische Phrase (SP) als grammatikalisches Subjekt und eine substantivische Phrase als grammatikalisches Objekt erfordert, wobei zusätzlich die Möglichkeit einer präpositionalen Phrase (PP) als indirektes Objekt besteht (z.B. *John* <Subjekt – SP> *donates a book* <direktes Objekt – SP> *to the library* <PP als optionales indirektes Objekt>). Außerdem wird die Überlagerung dieses syntaktischen Rahmens über die thematischen Rollen festgelegt. Zu *donate* ist der *Handelnde* das Subjekt, das *Thema* ist das direkte Objekt, und das *Ziel* oder der *Nutznießler* der durch das Prädikat ausgedrückten Tätigkeit ist das indirekte Objekt.

In der nächsten Phase des Formulierens sorgt die Auswahl der geeigneten Lemmata für den Abruf des Lautmusters der Äußerung. In dieser Phase werden die Sprachlaute (Phoneme) des Wortes verfügbar. Neben den Phonemen wird auch die metrische Information über das Wort abgerufen, die über die Zahl der Silben und die Betonung bestimmt. In einem Verarbeitungsschritt, der als phonetische Codierung bezeichnet wird, werden die Phoneme in der Reihenfolge von links nach rechts ihren jeweiligen Silben zugeordnet. Das Ergebnis der phonetischen Codierung ist das „phonetische Wort“, das die Information über die Laute des Wortes als Abfolge von Silben mit der richtigen Betonung enthält. Die Silben sind der Code, der die Grundlage für die Artikulationsbewegungen von Stimmbändern, Gaumensegel, Zunge, Kiefern und Lippen bildet. Je nachdem, ob man mit oder ohne Pfeife im Mund spricht, nehmen die Artikulationsbewegungen einen unterschiedlichen Verlauf, sie werden aber stets von demselben abstrakten Silbencode gesteuert. Das Endergebnis dieser Kaskade von Abruf- und Aktivierungsvorgängen ist ein akustisches Signal, aus dem der Zuhörer den beabsichtigten Inhalt ableiten kann.

Dass zwischen dem Abruf der Informationen über Lemma und Wortform ein Unterschied besteht, ist experimentell belegt; uns allen ist aber auch ein Phänomen vertraut, das ebenfalls für eine solche Unterscheidung spricht. Damit meine ich den oftmals peinlichen Zustand des „es liegt mir auf der Zunge“: Wir wissen ganz genau, dass wir das Wort kennen, wir wissen sogar, dass es sich beispielsweise um ein Substantiv mit einem ganz bestimmten grammatikalischen Geschlecht (z.B. Neutrum) handelt, aber aus irgendeinem Grund ist der Abruf der zugehörigen Lautform behindert. Die Tatsache, dass wir Zugang zu einem Aspekt der Information über das Wort haben, andere Aspekte jedoch nicht abrufen können, spricht für die Idee, dass verschiedene Aspekte der Information über ein Wort unterschiedlich gespeichert und

abgerufen werden. Die übermäßig vereinfachte Vorstellung, Wörter seien Einheiten, die irgendwo im Gehirn abgelegt sind, ist schlicht und einfach falsch. „Wort“ ist nicht mehr als ein bequemes Schlagwort für ein ganzes Orchester der Informationen, dessen Mitglieder unterschiedliche Instrumente spielen. Trotz aller Unterschiede in den Einzelheiten verschiedener Sprachmodelle besteht bei den Fachleuten für Sprachverarbeitung allgemein Einigkeit, dass der Vorgang den zeitlich koordinierten Abruf der zuvor erörterten verschiedenartigen Informationen erfordert.

Bisher habe ich den Sprechvorgang als vorwärts gerichteten Ablauf von der Absicht bis zur Artikulation beschrieben (Levelt 1989). Introspektiv haben wir jedoch häufig den Eindruck, dass unsere Art, Gedanken auszudrücken, auch unsere Absichten prägt oder schärfer hervortreten lässt. Mit anderen Worten: das Sprechen beginnt mit einer Absicht, aber Absichten leiten sich auch (teilweise) aus dem Sprechen ab. Dieses Gefühl führte zu Kritik an einer Beschreibung, in der das Sprechen ein Informationsfluss von der Absicht zur Artikulation ist (vgl. Dennett 1994). Nach meiner Überzeugung kann man dieser Kritik aber sehr einfach Rechnung tragen, wenn man erkennt, dass wir als Sprecher gleichzeitig auch Zuhörer sind. Das heißt, wir hören unsere eigenen sprachlichen Äußerungen und analysieren sie mit dem gleichen Apparat, den wir auch auf Äußerungen anderer anwenden. Beim Zuhören leiten wir die Absicht aus den Sprachlauten ab, die in unsere Ohren gelangen. Sprechen ist ein stark durch Zuwachs geprägter Prozess, das heißt, wenn wir mit dem Formulieren beginnen, haben wir noch keineswegs alle Einzelheiten unserer präverbalen Aussage festgelegt. Durch diese Tatsachen eröffnet sich die Möglichkeit, dass das Hören auf unsere eigenen Äußerungen dazu beiträgt, unsere Absichten zu formen und zu prägen, und auf diesem Weg können sie den laufenden Formulierungsprozess beeinflussen. Angesichts der Tatsache, dass unser kognitiver Sprachapparat über Bestandteile für Produktion und Verstehen verfügt, wird auf diese Weise ein ständiger innerer Dialog zwischen „Sprecher“ und „Zuhörer“ in Gang gesetzt, und dieser führt introspektiv zu dem Gefühl, dass Absichten nicht nur die Quelle, sondern auch ein Nebenprodukt des Sprechens sind.

In dem beschriebenen Beispiel habe ich in groben Umrissen die kognitive Struktur des Sprechens skizziert, wobei ich mich vorwiegend auf das Sprechen einzelner Wörter beschränkt habe. Ein vollständiges Modell des Sprechens beschreibt außerdem, wie Wörter zu längeren Äußerungen kombiniert werden, wie die Betonungen solcher Mehrwortäußerungen festgelegt werden, usw. Ähnliche Skizzen kann man auch für das Zuhören, Lesen und Schreiben zeichnen (detaillierte Beispiele bei Brown und Hagoort 1999). In allen

diesen Fällen stützt sich die Beschreibung der kognitiven Struktur für verschiedene Sprachfähigkeiten in ihren Einzelheiten auf eine Kombination aus Begriffsanalyse, Computermodellen und klug angelegten Experimenten. Wenn wir die kognitive Struktur kennen, können wir auf sinnvolle Weise fragen, wie die verschiedenen Sprachfähigkeiten auf neuronaler Ebene in Gang gesetzt werden. Ohne solche ausgefeilten Modelle ist die Untersuchung der neuronalen Sprachmechanismen zum Scheitern verurteilt.

Als Beispiel für diese Art der wissenschaftlichen Arbeit möchte ich ein einfaches Experiment etwas genauer erörtern. Wenn man lexikalische Begriffe wie PAPST oder PRIESTER in einem Netzwerk speichert und wenn die Aktivierung eines bestimmten Begriffs sich zum Teil auf benachbarte Begriffsknoten in dem Netzwerk ausbreitet, kann man für die Verarbeitung von Wörtern, denen semantisch verwandte Wörter vorausgehen, bestimmte Folgen voraussagen. Dies lässt sich folgendermaßen überprüfen. Die Versuchspersonen sehen auf einem Computerbildschirm Wortpaare aufblitzen. Zunächst leuchtet ein Wort eine halbe Sekunde lang auf, dann folgt für einige hundert Millisekunden ein leerer Bildschirm. Anschließend ist das zweite Wort eine halbe Sekunde lang auf dem Schirm zu sehen. Die Teilnehmer erhalten die Anweisung, dieses zweite Wort vorzulesen, sobald es auf dem Bildschirm erscheint. Mit der Präsentation des zweiten Wortes beginnt eine Uhr zu laufen, die durch die verbale Reaktion der Versuchspersonen wieder angehalten wird. Sobald die Versuchsperson mit dem lauten Vorlesen des Wortes auf dem Bildschirm beginnt, bleibt die Uhr stehen. Auf diese Weise kann man die Reaktionszeit der Versuchspersonen messen. Bei einer Form des Experiments sehen die Teilnehmer das fragliche Wort (z.B. „Priester“), nachdem zuvor ein semantisch verwandtes Wort („Papst“) eingeblendet wurde. In der zweiten Form hat das erste Wort in seiner Bedeutung nichts mit dem zweiten Wort zu tun (z.B. „Pferd“). Nach der Vorhersage des Netzwerkmodells führt der Anblick des Wortes „Papst“ über den Zusammenhang zwischen den Begriffen PAPST und PRIESTER zu einer partiellen Aktivierung des Wortes „Priester“. Erscheint nun „Priester“ unmittelbar im Anschluss an „Papst“ auf dem Bildschirm, sollte „Priester“ schneller vorgelesen werden, als wenn es allein auftaucht, weil es zuvor durch das Wort „Papst“ teilweise aktiviert wurde. Das Wort „Pferd“ dagegen verbreitet keinen Teil seiner Aktivierung auf „Priester“, weil die Begriffe PFERD und PRIESTER in dem semantischen Netzwerk so weit voneinander entfernt sind, dass sie sich gegenseitig nicht beeinflussen. Wird „Priester“ also unmittelbar nach „Pferd“ vorgelesen, sollte sich die Lesegeschwindigkeit nicht beschleunigen. Die Ergebnisse solcher Experimente stehen im Einklang mit den Vorhersagen, die

sich aus dem Netzwerkmodell ergeben. Die Versuchspersonen lesen den Begriff „Priester“, dem das Wort „Papst“ vorausgegangen ist, um mehrere zehntel Millisekunden schneller vor, als wenn das Wort „Pferd“ vorausging. Dieser so genannte Vorbereitungseffekt legt die Vermutung nahe, dass die Information über die Bedeutung von Wörtern im Gedächtnis nicht in Form getrennter Pakete mit einzelnen Wortbedeutungen gespeichert wird, sondern als Netzwerk aus untereinander verbundenen Informationsbruchstücken.

3. Die neuronale Architektur

Die neuronale Architektur legt die räumlich-zeitliche Dynamik der Gehirnprozesse fest, durch die beispielsweise das Netzhaubild von Johannes Paul II. in die Sprachlaute [pa:pst] umgewandelt wird. Wir müssen also herausfinden, welche Gehirnareale während der beteiligten Vorgänge aktiviert werden und wie die einschlägigen Aktivierungsvorgänge zeitlich koordiniert sind.

In jüngerer Zeit führte die Aufzeichnung der elektrischen Vorgänge im Gehirn zu einer recht fein aufgelösten Abschätzung für den zeitlichen Verlauf der verschiedenen beteiligten Prozesse. Normalerweise wird die elektrische Aktivität des Gehirns mit Elektroden an der Kopfhaut aufgezeichnet. Führt man eine solche Aufzeichnung im zeitlichen Zusammenhang mit sensorischen, motorischen oder kognitiven Vorgängen durch, erhält man ereignisbezogene Gehirnpotenziale (eventrelated brain potentials, ERPs), in denen sich die Summe der gleichzeitig ablaufenden, postsynaptischen Aktivitäten einer großen Zahl von Neuronen widerspiegelt. Die ERPs zeigen eine hohe zeitliche Auflösung, die in der Größenordnung weniger Millisekunden liegt. Auf der Grundlage dieser Komponenten kann man wichtige Informationen über den zeitlichen Verlauf der zu Grunde liegenden Vorgänge gewinnen.

ERPs wurden zwar aus verschiedenen Gründen (vgl. Hagoort und van Turenout 1997) in der Sprachforschung vorwiegend im Zusammenhang mit Aspekten des Sprachverständnisses aufgezeichnet; in Forschungsarbeiten der jüngsten Zeit wurden die Verfahren aber auch erfolgreich zur Untersuchung der Sprachproduktion eingesetzt (van Turenout, Hagoort und Brown 1997, 1998). Ich möchte hier nicht auf Einzelheiten eingehen, aber auf der Grundlage dieser und anderer Untersuchungen kann man begründete Vermutungen über den zeitlichen Ablauf des Vorganges anstellen, durch den ein visuelles Bild in eine artikulierte Abfolge von Sprachlauten umgewandelt wird (Einzelheiten bei Hagoort und van Turenout 1997). In unserem Beispiel dauert es mindestens 150 Millisekunden, bis das Netzhaubild als Johannes

Paul II. eingeordnet ist (Thorpe, Fize und Marlot 1996). Die nachfolgende Aktivierung des Begriffs PAST nimmt weniger als 200 Millisekunden in Anspruch. Auf sie folgt eine Kaskade von Abrufvorgängen. Die Aktivierung der syntaktischen Eigenschaften von Past (die Information über das Lemma) erfolgt rund 40 Millisekunden früher als der Abruf des ersten Phonems des Wortes [pa:pst] (van Turennout, Hagoort und Brown 1998). Wichtig ist, dass die Information über die phonetische Form eines Wortes nicht sofort zur Verfügung steht, sondern sich erst in der Reihenfolge von rechts nach links summiert. Für Wörter aus drei Phonemen (/p/a:/pst/) dauert es maximal 80 Millisekunden, die restlichen Abschnitte abzurufen, nachdem das erste Phonem des Wortes zur Verfügung steht (van Turennout, Hagoort und Brown 1997). Da rund 600 Millisekunden vergehen, bevor die Artikulation des Wortes [pa:pst] beginnt, wird die restliche Zeit für die Vorbereitung (und auch teilweise für die Ausführung) des motorischen Programms benötigt, das auf der Grundlage der phonologischen Information für die Artikulation sorgt.

Neben der Beantwortung solcher Fragen nach dem zeitlichen Ablauf müssen wir auch herausfinden, welche Gehirnareale an der Kaskade der Verarbeitungsvorgänge beim Sprechen mitwirken. Zu diesem Zweck müssen wir entweder auf Befunde an Patienten mit Gehirnschäden zurückgreifen oder die Gehirnaktivität mit modernen bildgebenden Verfahren wie Positronenemissionstomographie (PET) oder funktioneller Kernspinspektroskopie (fMRI) untersuchen. Gehirnschäden kann man untersuchen, wenn ein Schlaganfall oder ein Gehirntumor zu Sprachbehinderungen geführt hat. Eine genaue Analyse von Ort und Größe des Schadens einerseits sowie der jeweiligen spezifischen Sprachbehinderung andererseits ermöglichen Rückschlüsse darüber, welche Gehirnareale für einen bestimmten Aspekt der Sprachverarbeitung zuständig sind. Schäden im frontalen Kortex, von denen das Broca-Zentrum betroffen ist, führen beispielsweise häufig zu Störungen bei der Produktion der richtigen Lautfolge für die beabsichtigten Wörter; demnach liegt die Vermutung nahe, dass dieser Gehirnbereich normalerweise neben anderen Aufgaben an der Zusammenstellung der Lautfolge eines Wortes beteiligt ist. Stellt man auf diese Weise Zusammenhänge zwischen dem Ort einer Schädigung und den Symptomen der Behinderung her, kann man die Sprachfunktionen einzelnen Gehirnstrukturen zuordnen.

Mit PET und fMRI misst man hämodynamische Signale. Mit ihrer Hilfe kann man lokale Veränderungen der Durchblutung (PET) oder der Sauerstoffsättigung des Blutes (fMRI) nachweisen und sichtbar machen, die durch bestimmte Funktionen entstehen und demnach vermutlich im Zusammenhang mit der Aktivierung des umgebenden Nervengewebes stehen (eine allge-

meine Einführung findet sich bei Raichle 1994). Grob gesagt, weist man auf diese Weise den Ort der Nervenaktivität, die mit kognitiven Prozessen im Zusammenhang steht, auf dem Weg über die Blutgefäße nach. Daraus folgt, dass das zeitliche Auflösungsvermögen solcher Verfahren durch die zeitliche Dynamik der Veränderung von Durchblutung oder Sauerstoffsättigung des Blutes begrenzt wird, und die liegt in der Größenordnung von einigen hundert Millisekunden bis zu mehreren Sekunden. Im Gegensatz dazu stehen die elektrophysiologischen Aufzeichnungen im unmittelbaren Zusammenhang mit der Nerventätigkeit, und deshalb liegt ihre zeitliche Auflösung im Bereich weniger Millisekunden. Dafür leiden die ERP-Messungen aber am so genannten Umkehrproblem; dieses macht es schwierig, den Ort der elektrischen Potenziale zu ermitteln, die an der Kopfhaut gemessen werden. Derzeit liefern nur PET und fMRI Messungen mit der erforderlichen räumlichen Auflösung.

In den letzten Jahren wurde die Sprachverarbeitung in einer zunehmenden Zahl von PET- und fMRI-Studien untersucht. Da es für die Sprachen kein Tiermodell gibt, sind wir stark auf diese neuen bildgebenden Verfahren angewiesen, wenn wir das Gehirn während sprachlicher Aufgaben unmittelbar bei seiner Tätigkeit beobachten wollen. Den meisten derartigen Untersuchungen an der Sprache liegen folgende Überlegungen zu Grunde: Man beobachtet, nach welchen Gesetzmäßigkeiten das Gehirn bei Aufgaben aktiviert wird, die in der Kaskade der Sprachprozesse einen bestimmten Schritt (zum Beispiel den Abruf der Lemma-Information) beinhalten, und vergleicht die Befunde mit Aktivierungsmustern bei Aufgaben, an denen gerade dieser Prozess nicht beteiligt ist. Durch solche Vergleiche kann man herausfinden, welche Gehirnareale bei dem fraglichen Schritt (in unserem Beispiel dem Lemma-Abruf) des Gesamtvorganges stärker aktiviert werden. Man geht davon aus, dass solche stärker aktivierten Regionen an dem betreffenden Aspekt des Sprechens besonders beteiligt sind. Man kann die Versuchspersonen beispielsweise bitten, Wörter und Pseudowörter laut vorzulesen. Pseudowörter sind phonetisch erlaubte Buchstabenketten, die aber in einer Sprache wie zum Beispiel dem Englischen nicht als Wörter existieren. Ein Beispiel ist die Buchstabenkette *floke*. Dieses Wort kann jeder vorlesen, aber niemand weiß, was es bedeutet. Vergleicht man das englische Wort *smoke* mit dem Pseudowort *floke*, so treten zwei Unterschiede auf, wenn man es geschrieben sieht und laut ausspricht. Erstens wird für *smoke* im Laufe des gesamten Vorganges die Bedeutung zugänglich gemacht, für *floke* dagegen gibt es im Gedächtnis keine semantische Repräsentation. Außerdem können wir für das Pseudowort *floke* aus dem Gedächtnis auch keinen phonetischen Code

abrufen. Stattdessen müssen wir einen solchen Code selbst zusammensetzen, indem wir die einzelnen Grapheme in entsprechende Phoneme umsetzen. Misst man das Muster der Gehirnaktivität beim lauten Vorlesen von Wörtern und Pseudowörtern, so sind die Unterschiede zwischen beiden auf den Abruf von Wortbedeutungen und phonetischen Eigenschaften und/oder auf das Zusammensetzen des phonetischen Codes für die Pseudowörter zurückzuführen. Um weiter zwischen der Gehirnaktivierung im Zusammenhang mit Bedeutung und Phonetik zu unterscheiden, muss man andere Aufgaben vergleichen. Auf diese Weise können wir herausfinden, welche Gehirnareale während der verschiedenen Schritte des Sprechprozesses differenziell aktiviert werden. Natürlich sind manche Gehirnfelder für beide Aufgaben von entscheidender Bedeutung, denn neben den Unterschieden gibt es beim Lesen von Wörtern und Pseudowörtern auch Gemeinsamkeiten. So beobachtet man beispielsweise in beiden Fällen eine Aktivierung der primären Sehrinde, denn der ganze Lesevorgang beginnt mit der Analyse des visuellen Musters, das auf die Netzhaut fällt.

In einer neuen Meta-Analyse von mehr als 50 Studien, in denen die Produktion einzelner Wörter mit bildgebenden Verfahren am Gehirn untersucht wurde, fassten Indefrey und Levelt (2000) unsere derzeitigen Kenntnisse über die neuronalen Schaltkreise zusammen, die der zuvor beschriebenen kognitiven Tätigkeit zu Grunde liegen. Alle entscheidenden Schritte des Sprechprozesses werden von Feldern in der linken Gehirnhälfte ausgeführt, die bei der großen Mehrzahl der Menschen die dominante Hemisphäre für die Sprachfähigkeiten ist. An der Auswahl des richtigen Begriffs für das Sprechen (PAPST) ist anscheinend der linke mittlere Gyrus temporalis beteiligt. Von dort breitet sich die Aktivierung auf das Wernicke-Zentrum aus, das entscheidend daran mitwirkt, den phonetischen Code eines im Gedächtnis gespeicherten Wortes abzurufen. Das Wernicke-Zentrum spielt eine unverzichtbare Rolle im gesamten Netzwerk der Sprachverarbeitung, weil es die lexikalischen Aspekte einer Wortform mit den weit verstreuten Assoziationen verknüpft, die seine Bedeutung definieren. Diese Funktion hat das Wernicke-Zentrum sowohl bei der Sprachproduktion als auch beim Sprachverständnis (Mesulam 1998). Die Informationen über die lexikalische Wortform werden an das Broca-Zentrum im linken frontalen Kortex und/oder zum mittleren Teil des oberen Schläfenlappens in der linken Gehirnhälfte weitergegeben. Diese Felder sind von Bedeutung für die Umwandlung des phonetischen Codes aus dem Gedächtnis in phonetische Wörter, aus denen das abstrakte Artikulationsprogramm abgeleitet wird. In der letzten Phase der Vorbereitung auf die Artikulation und ihrer Ausführung werden sensorisch-motorische Felder akti-

viert, wobei als Ergänzung möglicherweise auch weitere motorische Felder und das Kleinhirn einen Beitrag leisten.

Unser begriffliches Wissen über die Wörter unserer Sprache ist offenbar weiter verteilt als die lexikalischen Aspekte von Lemma und Form; es beschränkt sich auch nicht auf die linke Gehirnhälfte. Außerdem weisen sowohl Studien mit bildgebenden Verfahren als auch die Untersuchung von Gehirnschäden im Zusammenhang mit der Semantik konkreter Substantive darauf hin, dass der Zugang zu Wahrnehmungs- und Funktionseigenschaften einer Wortbedeutung über unterschiedliche Gehirnteile erfolgt, wobei die Wahrnehmungseigenschaften den primären sensorischen Feldern näher sind, während die funktionellen Eigenschaften eher in der Nachbarschaft des motorischen Kortex liegen (Überblick bei Saffran und Sholl 1999). Der transmodale Kortex einschließlich des mittleren Schläfenlappens und des Wernicke-Zentrums sind Konvergenzzonen, kritische Übergänge für den Zugang zu einschlägigen Informationen, die dezentral repräsentiert sind (Mesulam 1998). Geht man von der Gesamtstruktur des Kortex aus, so eignen sich diese Felder gut dazu, unser dezentrales begriffliches Wissen in eine einzige Wortform einfließen zu lassen. Damit alles, was wir über Johannes Paul II. wissen, in eine einzige Wortform [pa:pst] einfließen kann, müssen Gehirnareale beteiligt sein, die auf die Verknüpfung der dezentralen Kenntnisbruchstücke zu einem einzigen Ergebnis – in diesem Fall eine einzige Wortform – spezialisiert sind.

Die Kombination unserer Kenntnisse über die Gesamtstruktur des Gehirns mit dem Wissen über spezifische Aktivitätsmuster und zeitliche Dynamik seiner Tätigkeit im Zusammenhang mit Sprachprozessen versetzen uns in die Lage, Erkenntnisse über die neuronale Organisation unserer ausschließlich menschlichen Fähigkeit zur Kommunikation durch eine natürliche Sprache zu gewinnen. Wenn wir dieses höchst komplexe Kommunikationssystem verstehen wollen, sind zahlreiche sorgfältige experimentelle Untersuchungen von Einzelfragen notwendig. Die Kenntnisse, die wir auf diese Weise gewinnen, sind letztlich auch von weit reichender Bedeutung für zentrale Fragen nach der Persönlichkeit des Menschen.

4. Sprache und Persönlichkeitstheorie

Eine vollständige Persönlichkeitstheorie erfordert eine genaue Beschreibung der Wege, auf denen die Signale von sehr unterschiedlichen Systemen wie Sprache, Gedächtnis, Gefühle, motorische Tätigkeit usw., denen jeweils eigene neuronale Schaltkreise gewidmet sind, ein Gefühl für das Ich und die eige-

ne Persönlichkeit entstehen lassen. Wie das im Einzelnen geschieht, ist bisher weitgehend unbekannt. Die Frage, wie das Gehirn das Problem löst, die Signale dieser unterschiedlichen Systeme zu einem einheitlichen Ich-Gefühl zusammenzuführen, das eine Kontinuität von der Vergangenheit in die Zukunft bewahrt, ist für die heutige neurowissenschaftliche Kognitionsforschung fast völlig unbekanntes Terrain. Es gibt zwar nützliche Modelle verschiedener kognitiver Systeme, aber eine umfassende Persönlichkeitstheorie in der neurowissenschaftlichen Kognitionsforschung fehlt bisher. Auch wenn wir im Einzelnen keine solche Theorie kennen, kann man jedoch vernünftigerweise annehmen, dass sie ganz anders aussähe, wenn wir keine Sprache besäßen. Arbib vertritt im Zusammenhang mit seiner Schematheorie die Ansicht, das Ich sei eine Enzyklopädie aus vielen hunderttausend Schemata, mit deren Hilfe ein Mensch neue Informationen interpretiert und seinem Gedächtnis hinzufügt. Die Enzyklopädie der Schemata dient dazu, „eine Geschichte zu erzählen“, die zu den neuen Daten passt. Man beobachtet nicht nur, dass die Metaphern, mit deren Hilfe wir das Gefühl von Ich und Persönlichkeit beschreiben, sehr häufig aus dem Bereich der Sprache stammen (siehe auch Gazzanigas Vorschläge für Dolmetscher, Gazzaniga 1992), sondern klar ist auch, dass Sprache uns in die Lage versetzt, unsere Schemata-Enzyklopädie mit verblüffender Geschwindigkeit zu erweitern. Eine kognitiv-neurowissenschaftliche Erforschung der Sprache erklärt zwar nicht den Inhalt unserer Schemata-Enzyklopädie, sie ist aber unentbehrlich für die Erklärung des Apparats, mit dessen Hilfe wir diese große Enzyklopädie aufbauen. Ohne unsere Sprachfähigkeiten wäre unser Ich- und Persönlichkeitsgefühl zweifellos wesentlich stärker eingeschränkt und viel langweiliger. In dieser Hinsicht ist Sprache also ein entscheidender Bestandteil einer Persönlichkeitstheorie, und deshalb sind die Beiträge einer kognitiv-neurowissenschaftlichen Erforschung der Sprache höchst erwünscht.

Eine ähnliche Geschichte kann man auch über die Beziehung zwischen Sprache und Wahrnehmung erzählen. Zwar ist das Bewusstsein bis heute unter den Gesichtspunkten der neurowissenschaftlichen Kognitionsforschung nicht zufriedenstellend erklärt, aber trotz aller Behauptungen des Gegenteils (ein anregender Bericht findet sich bei Chalmers 1996) sieht die Sache bei der Wahrnehmung besser aus. Sie hat mit unserer Fähigkeit zu tun, Beurteilungen über Phänomene und über unsere Empfindungen sprachlich auszudrücken. Genau wie PET und fMRI durch das Filter der Hämodynamik einen Blick auf die neuronale Aktivität ermöglichen, so ermöglicht auch die Wahrnehmung durch das linguistische Filter einen Blick auf das Bewusstsein. Die Eigenschaften dieses Filters zu kennen, ist in diesem Fall ebenfalls von

entscheidender Bedeutung, wenn man das zentrale, aber immer noch meist sehr schwer fassbare Phänomen des Bewusstseins besser verstehen will.

Prof. Dr. Peter Hagoort ist Direktor des FC Donders Centre for Cognitive Neuroimaging, Nijmegen, Niederlande.

Literatur

- Aitchison, J. (1997), *Wörter im Kopf: eine Einführung in das mentale Lexikon*. Tübingen: Niemeyer.
- Brown, C. M. & Hagoort, P. (1999), *Neurocognition of Language*. Oxford: Oxford University Press.
- Chalmers, D. J. (1996), *The conscious mind: In search of a fundamental theory*. Oxford, UK: Oxford University Press.
- Dennett, D. C. (1991), *Philosophie des menschlichen Bewusstseins*. Hamburg: Hoffmann und Campe.
- Dunbar, R. (1998), *Klatsch und Tratsch – Wie der Mensch zur Sprache fand*. München: C. Bertelsmann.
- Fromkin, V. A. (Hg.) (1973), *Speech errors as linguistic evidence*. Den Haag: Mouton.
- Gazzaniga, M. S. (1992), *Nature's mind: The biological roots of thinking, emotions, sexuality, language, and intelligence*. New York: BasicBooks.
- Hagoort, P., & van Turennout, M. (1997), *The electrophysiology of speaking: Possibilities of event-related potential research on speech production*. In: W. Hulstijn, H. F. M. Peters, & P. H. H. M. Van Lieshout (Hg.), *Speech production: Motor control, brain research and fluency disorders* (S. 351-361). Amsterdam: Elsevier.
- Hillyard, S. A., & Kutas, M. (1983), *Electrophysiology of cognitive processing*. *Annual Review of Psychology*, 34, 33-61.
- Indefrey, P., & Levelt, W. J. M. (2000), *The neural correlates of language production*. In: M. Gazzaniga (Hg.), *The cognitive neurosciences*, 2nd edition. Cambridge, MA: The MIT Press.
- Jackendoff, R. (1997), *The architecture of the language faculty*. Cambridge, MA: The MIT Press.
- Jackendoff, R. (1999), *The representational structures of the language faculty and their interactions*. In: C. M. Brown & P. Hagoort (Hg.), *Neurocognition of Language* (S. 37-81). Oxford: Oxford University Press.
- Levelt, W. J. M. (1989), *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press
- Levelt, W. J. M. (1999), *Producing spoken language: A blueprint of the speaker*. In: C. M. Brown & P. Hagoort (Hg.), *Neurocognition of Language* (S. 83-122). Oxford: Oxford University Press.
- Mesulam, M.-M. (1998), *From sensation to cognition*. *Brain*, 121, 1013-1052.
- Raichle, M. E. (1994), *Bilder der Geistestätigkeit*. Heidelberg: Spektrum der Wissenschaft, Juni 1994, 56-63.
- Roelofs, A. (1992), *A spreading-activation theory of lemma retrieval in speaking*. *Cognition*, 42, 107-142.
- Roelofs, A. (1993), *Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs*. *Cognition*, 47, 59-87.
- Saffran, E. M., & Sholl, A. (1999), *Clues to the functional and neural architecture of word meaning*. In: C. M. Brown & P. Hagoort (Hg.), *Neurocognition of Language* (S. 241-272). Oxford: Oxford University Press.
- Thorpe, S., Fize, D., & Marlot, C. (1996), *Speed of processing in the human visual system*. *Nature*, 381, 520-522.
- Turennout, M. van, Hagoort, P., & Brown, C. M. (1997), *Electrophysiological evidence on the time course of semantic and phonological processes in speech production*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 787-806.
- Turennout, M. van, Hagoort, P., & Brown, C. M. (1998), *Brain activity during speaking: From syntax to phonology in 40 milliseconds*. *Science*, 280, 572-574.

IMPRESSUM:

Wissenschaftszentrum Nordrhein-Westfalen

Jahrbuch 2002/2003

Herausgeber: Gert Kaiser

Redaktion: Carolin Grape, Sabine Wiegand

Wissenschaftszentrum Nordrhein-Westfalen

Reichsstraße 45, 40217 Düsseldorf

Telefon 02 11/38 790-0, Telefax 02 11/37 05 86

<http://www.wz.nrw.de>

E-mail: pr@mail.wz.nrw.de

Layout & Satz:

MedienPark Moers AG, Juliusstraße 9-21, 47053 Duisburg

Druck:

B.O.S.S. Druck & Medien GmbH, Geefacker 63, 47533 Kleve

© Das Jahrbuch 2002/2003 und alle in ihm enthaltenen Beiträge sind urheberrechtlich geschützt. Das Copyright kann jedoch jederzeit beim Wissenschaftszentrum eingeholt werden und wird in der Regel erteilt, wenn die Quelle ausdrücklich genannt wird.

Namentlich gekennzeichnete Beiträge geben nicht unbedingt die Meinung des Herausgebers wieder.

1. Auflage 2003

ISBN 3-929 483-21-1

INHALT

Seite

- 6 **Grußwort**
Hannelore Kraft
- 8 **Editorial**
Gert Kaiser

BIO- UND LEBENSWISSENSCHAFTEN

- 12 **Henrik Walter**
Können wir Gedanken lesen? Prinzipien, Möglichkeiten und Grenzen der funktionellen Bildgebung
- 25 **Ansgar Beckermann**
Vom Gehirn zum Geist? Bilddaten und ihre Aussagekraft
- 33 **Peter Hagoort**
Die einzigartige, größtenteils aber unbewusste Fähigkeit der Menschen zu sprachlicher Kommunikation
- 47 **Thomas F. Münte**
Kognitive Neurowissenschaft der Fehlerverarbeitung
- 56 **Bert Sakmann**
Of men and mice – Mausmodelle von menschlichen Erbkrankheiten des Nervensystems
- 66 **Gereon R. Fink**
Normale und gestörte Aufmerksamkeitsprozesse und ihre therapeutische Beeinflussung:
Von der Grundlagenforschung zur klinischen Anwendung
- 75 **Joachim Groß**
Tomographische Abbildung funktioneller Kopplungen im menschlichen Gehirn mithilfe der
Magnetenzephalographie